WHEN LEARNING CAN SAFELY FORGET: TRANSITION-FAITHFUL CERTIFICATES FOR ADAPTER STATE

Anonymous authors
Paper under double-blind review

ABSTRACT

An inactive direction can disappear from predictions while still determining the next learning update. We construct state release that preserves an unfinished learner's update rule. Future-inactivity permission fixes omitted gate outputs; summaries closed under the remaining optimizer reads preserve their residual effects. Common radial dynamics retain one energy, complete linear-channel queries are characterized by a Gram matrix, and distinguishable quadratic modes retain separate energies. For the radial gated reference, a future cross-entropy budget yields an optimal adverse-probability relaxation; transporting a reliable dual cost reduces repeated checking. The reduction preserves declared learning events through adaptive release, rejection, rollback and corresponding-point restoration. On fixed Letter and CIFAR executions, compilation removes over 96% of dual evaluations while retaining over 90% of fresh-check release integrals. Paired trials, same-candidate interventions and fixed-mode dimension tests validate the information that must be retained. Same-checkpoint permission comparisons identify the additional positive authorizations from loss information, while complete-record measurements quantify storage gains and their acquisition costs.

1 INTRODUCTION

An unfinished learner stores not just predictions but the decisions governing projection, acceptance, rollback and stopping. We ask *when sparse directions can be physically released while preserving this computation*. "Forgetting" here releases directions, not training-data influence.

Two obstacles are independent. Equally energetic zero-gate atoms can give different next gate outputs; permanently inactive energy can still change another parameter's update. Sparse adaptation, screening and checkpoint compression address rank, variable exclusion and storage/performance (Ding et al., 2023; Ndiaye et al., 2017; Li et al., 2024). We preserve the original learner's finite decisions after task inactivity.

Our construction joins *future permission and remaining-read closure*. Permission fixes omitted gate outputs; a closed summary computes every other declared hidden read and its own next state. Radial, channel and spectral dynamics determine the distinguishable information. CE supplies radial permission; the other classes use separate proofs.

Fixed updates separate representation savings from optimizer redesign: parameter-efficient learning gains a release criterion, optimization a finite-execution fidelity standard, and systems resumable checkpoints based on future computation.

Characterize releasable information. We expose both obstacles and characterize closed summaries and indistinguishability for the declared read classes.

Make release executable without changing learning. Future CE and reliable dual transport supply permission; adaptive release, rollback and restoration preserve original events.

Test the joint construction and resource regimes. Read interventions, same-checkpoint bounds, natural releases and dimension tests distinguish the two obligations; complete records and replay identify their costs.

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.

2 THE LEARNER AND THE OBSTRUCTION TO DELETION

Minimal example. An acceptance residual measures a trial's deviation from its local objective model; at zero tolerance it accepts iff nonpositive. Let x be retained, y hidden, and $u \in [0,1]$ its gate:

$$F(x,y,u) = 2x^2 + \tfrac{1}{2}y^2 + \tfrac{1}{2}u^2y^2 + mu, \qquad m > 0. \tag{1}$$

At $u = 0$, the gate forcing is $m > 0$, so a zero-centered proximal gate step returns zero. Take $(x,y) = (1,10)$, an inactive ball constraint, inverse step $M = 4$ (step size $1/M$), and acceptance parameter $\nu = 1/2$. The factor displacement is $(-1,-5/2)$. With zero acceptance tolerance, the full and predictor-only residuals are

$$\mathcal{A}_F = \tfrac{1}{2}\{(4-2) + (1-2)25/4\} = -17/8, \qquad \mathcal{A}_{\text{delete}} = 1. \tag{2}$$

Full accepts $x^+ = 0$; deletion rejects and doubles M to 8, giving $x^+ = 1/2$. Thus hidden energy changes acceptance with an inactive ball; Section 5 separately tests projection.

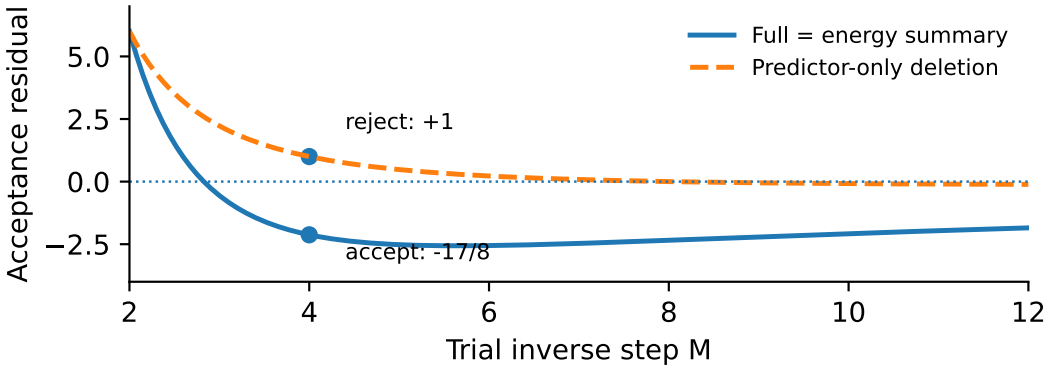

Figure 1: Prediction equality does not preserve acceptance. Full and its energy summary agree; predictor-only deletion reverses the decision at $M = 4$ with an inactive ball. The separate empirical intervention changes 42/1,864 decisions (Section 5).

Gated adaptation. For C classes and fixed features $x_i \in \mathbb{R}^d$, use the LoRA-shaped predictor (Hu et al., 2022)

$$W = W_0 + sB\,\mathrm{Diag}(\gamma)A, \quad B \in \mathbb{R}^{C\times r_0},\ A \in \mathbb{R}^{r_0\times d}, \quad 0 \le \gamma_j \le 1. \tag{3}$$

Column b_j and row $a_j^\top$ form one atom. A *zero episode* is a consecutive interval with literal $\gamma_j = 0$; its directions stay stored until future inactivity is certified. The original scale $s = \alpha/r_0$ stays fixed after release. Write $\Theta = (B,A)$ and its anchored displacement $d_\Theta = \Theta - \Theta_w$; use the Euclidean norm of all factor entries. Let $\mathcal{L} = n^{-1}\sum_i[\log\sum_c e^{(Wx_i)_c} - (Wx_i)_{y_i}]$. The reference minimizes

$$F = \mathcal{L} + \tfrac{\mu}{2}\|d_\Theta\|^2 + \lambda^\top\gamma - P_2(\gamma) + \beta c^\top\gamma, \qquad \|d_\Theta\| \le R, \quad f = F - \lambda^\top\gamma. \tag{4}$$

Here $\mu, \lambda_j > 0$, $\beta, c_j \ge 0$, and P_2 is separable with $P'_{2,j}(0) = 0$ and $0 \le P_{2,j}(u) \le \lambda_j u$ on $[0,1]$. MCP is the experimental choice (Zhang, 2010). Thus $F \ge \mathcal{L} \ge 0$ and $-\Lambda \le f \le F$, where $\Lambda = \sum_j \lambda_j$. Appendix A defines both gate kernels and the penalty.

One specified execution. A factor trial projects $\Theta - M^{-1}\nabla_\Theta f$ onto the original joint ball. At its accepted candidate, the gate block computes

$$\gamma^+ = \arg\min_{u\in[0,1]^{r_0}} \{q^\top u + M_\gamma D_{\mathcal{H}}(u,\gamma)\}, \qquad q = \nabla_\gamma f + \lambda. \tag{5}$$

The separable kernel $\mathcal{H}$ is strongly convex with finite endpoint derivatives; $D_h(a,b) = h(a) - h(b) - h'(b)(a-b)$. In particular, $q_j \ge 0$ keeps a zero center at zero. For $z = (\Theta,\gamma)$ and block i, the Bregman alternating reference (Ahookhosh et al., 2021) accepts a trial when

$$\mathcal{A}_i = f(z^+) - f(z) - \langle\nabla_i f, z_i^+ - z_i\rangle - (1-\nu_i)M_i D_{h_i}(z_i^+, z_i) \le \epsilon\max(1,|f(z)|), \tag{6}$$

with $0 < \nu_i < 1$ and $D_{h_\Theta} = \|\Theta^+ - \Theta\|^2/2$. Rejections double the inverse step; resets, caps, and stopping rules are fixed parts of the learner. A gate-cap failure restores the entire pre-sweep state, including a provisional accepted factor step. A *completed checkpoint* is after a committed sweep and before the next factor trial.

Reference class. The scalar closure fixes data, base, scale, anchor, joint ball, isotropic ridge and functional controls, under future-inactivity permission and Table 1's read set. Its CE permission additionally uses zero anchors, accepted factor inputs with $M \geq \mu$, and the original finite horizon. Channel and spectral dynamics use separately proved permission: "independent" here means a separate proof, not statistical independence.

3 PRESERVE WHAT THE OPTIMIZER CAN OBSERVE

Let D be atoms whose zero gates are certified inactive through the full reference's original remaining horizon; also use D for their factor coordinates. Write d_D for their displacements and d_P for the retained complement. The orientation can be discarded, but its use of the joint ball and its ridge contribution must remain correct. In the *original* objective, $f = f_P + \mu\|d_D\|^2/2$; f_P omits only hidden ridge. Store $v = \|d_D\|^2$. Each factor trial uses

$$t = 1 - \mu/M, \qquad u_P = d_P - \nabla_P f/M, \qquad \kappa = \min\{1, R/\sqrt{\|u_P\|^2 + t^2 v}\}, \quad \psi = \kappa t, \\ d_P^+ = \kappa u_P, \qquad v^+ = \psi^2 v. \tag{7}$$

At a zero denominator use $\kappa = 1$. The hidden candidate is $d_D^+ = \psi d_D$, so the same scalar reconstructs every declared hidden read in Table 1.

Table 1: Read closure after permission. f_P, F_P omit the hidden ridge only. Raw gate gradients are removed by permission, not recovered from energy.

Optimizer operation	Full hidden read	Compact replacement
Joint projection	$\|td_D\|^2$	$t^2 v$
Objective	$\mu\|d_D\|^2/2$	$\mu v/2$
Gradient pairing	$\langle \mu d_D, d_D^+ - d_D \rangle$	$\mu(\psi - 1)v$
Squared step	$\|d_D^+ - d_D\|^2$	$(\psi - 1)^2 v$
Tolerance scale	$\max(1, \|f\|)$	$\max(1, \|f_P + \mu v/2\|)$
Rollback / stop	Objective, steps, patience, budget	Corresponding controls and snapshots
Raw zero-gate query	Direction-dependent forcing	Output fixed at zero by permission

The complete acceptance correction. At the *same retained candidate*, let $\mathcal{A}_P$ contain the retained objective, pairing, and step terms. Then

$$\boxed{\mathcal{A}_F - \mathcal{A}_P = \tfrac{1}{2}[\mu - (1 - \nu_\Theta)M](\psi - 1)^2 v =: D_v.} \tag{8}$$

The hidden objective change, gradient pairing and step penalty are $\mu(\psi^2 - 1)v/2$, $-\mu(\psi - 1)v$, and $-(1 - \nu_\Theta)M(\psi - 1)^2 v/2$; summing proves (8). When $M > \mu/(1 - \nu_\Theta)$, $D_v \leq 0$, so omitting it can reject an accepted trial. The *complete signed guard* is

$$g_F = \mathcal{A}_P + D_v - \epsilon \max(1, |f_P + \mu v/2|), \qquad \text{accept iff } g_F \leq 0. \tag{9}$$

Energy also enters the tolerance on the right of (6); Appendix C isolates its effect with a rational counterexample.

Learning events versus representation work. Compact state is $(d_P, S_D, \gamma_P, \chi, \eta_{\text{cert}})$: χ contains original step, budget and stopping controls; η_{cert} contains permission caches and schedules. Preserved events $\mathcal{O}$ are retained candidates, accept/reject, commit, objective/predictor, rollback and halt. Certificate work is internal and unchecked full needs no matching cache. Functional work counters belong to χ; the budget counts reference events, not elapsed time.

Theorem 1 (Adaptive release preserves the same learner)**.** *Fix Section 2's reference and original scope. Release only at completed checkpoints, after permission valid on full has fixed the omitted gate outputs at zero. Suppose every other functional hidden read and snapshot is closed by Table 1, and no controller reads the omitted raw gate-gradient magnitudes. If internal checking always returns in finite work, any such adaptive release schedule preserves the $\mathcal{O}$ event sequence in exact arithmetic. At fixed D, the learning-event maps satisfy*

$$\Pi_D T_F = T_C^{\mathcal{O}} \Pi_D, \tag{10}$$

where Π_D projects functional state and $T_C^{\mathcal{O}}$ excludes internal certificate work. Disjoint releases merge by $v \leftarrow v + \|d_E\|^2$. Any finite number of lossless compact restores at corresponding interruption points in the same context preserves this sequence.

Proof. Equal controls select the same trial; closed reads give the same candidate and guard. Acceptance updates corresponding summaries; rejection commits neither candidate. Permission fixes omitted gate outputs. Failure restores corresponding snapshots; original controls determine halt. Summary addition preserves the relation at each release. Lossless decoding inserts the identity at interruption. Induction proves the claim; Appendix C supplies the complete argument. □

Restoration recovers the reduced state, not the discarded vectors. Corresponding predictors give $Q(h_C) = Q(h_F)$ for every fixed output-based metric; deterministic selection using preserved values also agrees (Appendix C.2).

3.1 Summary size follows distinguishable dynamics

For channels $Z \in \mathbb{R}^{m \times p}$, let $Z^+ = ZA_t$ and allow channel-inner-product queries whose vectors and updates depend on prior allowed reads and retained state. Then

$$G = Z^\top Z, \qquad \langle Zu, Zw \rangle = u^\top G w, \qquad G^+ = A_t^\top G A_t. \tag{11}$$

Theorem 2 (Exact state for the complete channel-query class)**.** *Under independent permission, G is sufficient. For the class allowing every probe $\|Zu\|^2$, two hidden states are indistinguishable if and only if their Gram matrices agree.*

Equal Grams yield identical adaptive read histories and updates. A differing diagonal is exposed by $u = e_i$; a differing off-diagonal with equal diagonals is exposed by $e_i + e_j$. This proves both directions. Under the open permitted-section and continuous-complete-encoding conditions of Appendix C.3, $p(p+1)/2$ real coordinates are necessary and sufficient.

Table 2: Hidden summary size follows the allowed queries. Necessity concerns continuous complete representations on the permitted domains in Appendices C.3–D, rather than every individual trajectory.

Dynamics	Closed information	Permission source
Common radial update	One energy v	Future CE (Section 4)
p linear channels	Gram G determines every allowed read	Separate, e.g. global ball
K quadratic modes	K distinguishable energies	Separate global-ball bound

Projected heavy-ball needs two norms and their cross term ($p = 2$). Distinct spectral decays need separate energies: identical current norm and ridge can give different next norms. Within-mode dimension does not change that information (Table 3); Appendices C.3–D prove closure, necessity and permission.

4 From a future loss budget to executable permission

We now construct permission from training loss. A finite loss budget prevents all true-class probabilities from approaching zero together, restricting the adverse forcing that could reactivate a gate.

4.1 CROSS-ENTROPY RULES OUT INCOMPATIBLE WORST CASES

For one zero gate, put $w = \lambda_j + \beta c_j$. Under zero anchors and accepted factor steps with $M \geq \mu$, any already-zero prefix has $b = \rho b_0$, $a = \rho a_0$, $0 \leq \rho \leq 1$. Define the samplewise adverse contrasts

$$c_{ic} = s(b_{0,c} - b_{0,y_i})a_0^\top x_i, \qquad a_i^{\mathrm{adv}} = -\min_c c_{ic} \geq 0, \qquad \bar{a} = n^{-1}\sum_i a_i^{\mathrm{adv}}. \tag{12}$$

The true-label contrast is zero. If p_i is its future probability, then

$$q_j = w + \frac{\rho^2}{n}\sum_{i,c}\pi_{ic}c_{ic} \geq w - \frac{\rho^2}{n}\sum_i a_i^{\mathrm{adv}}(1-p_i). \tag{13}$$

The simplex-only bound is $w - \bar{a}$. A future mean-CE bound E, covering accepted intermediate candidates, couples probabilities through $\sum_i -\log p_i \leq nE$. For two adverse magnitudes equal to one, $w = 3/4$ and $E = \log 2$ force $p_1 p_2 \geq 1/4$, hence $p_1 + p_2 \geq 1$: the margin improves from $-1/4$ to $+1/4$. This licenses an already-zero gate. For general $a = (a_i^{\mathrm{adv}})_{i=1}^n$, define

$$V(a,E) = \min_{\substack{0<p_i\leq 1 \\ \sum_i -\log p_i \leq nE}} \sum_i a_i p_i, \qquad \mathcal{C}(a,E) = \bar{a} - V(a,E)/n. \tag{14}$$

This is the *exact worst-case probability relaxation* given a and E; shared-network reachability is relaxed. Binary contrasts $0, -a_i$ attain the relaxed bound (Appendix B).

Theorem 3 (Loss-informed future inactivity)**.** *Let E bound the CE at every future gate evaluation in the original scope. If $E > 0$ and $a \neq 0$, the unique $\theta \in (0, \max_i a_i)$ satisfying*

$$\sum_{a_i > \theta} \log(a_i/\theta) = nE \tag{15}$$

solves (14)*: $V = \sum_i \min(a_i, \theta)$ and $\mathcal{C} = n^{-1}\sum_i (a_i - \theta)_+$. A positive reliable lower bound on $w - \mathcal{C}(a,E)$ certifies that the gate remains zero. At $E = 0$ or $a = 0$, $\mathcal{C} = 0$. For finite E and $a \neq 0$, $w - \mathcal{C}(a,E) > w - \bar{a}$.*

Proof. Minimizing $a_i p - \theta \log p$ on $(0,1]$ gives $p_i^* = \min(1, \theta/a_i)$, with $p_i^* = 1$ for $a_i = 0$. Equation (15) exhausts the CE budget and proves primal–dual optimality. Since $\mathcal{C} \geq 0$, (13) gives $q_j \geq w - \rho^2\mathcal{C} \geq w - \mathcal{C}$. Positive forcing makes every positive coordinate worse than zero in the zero-centered proximal subproblem. This extends the zero prefix by one gate step, then by induction through the scope. Finite E implies $p_i \geq e^{-nE}$, so $V > 0$ when $a \neq 0$. □

Where the future budget comes from. Set $B(F) = \max(1, \Lambda, F)$. Projection and gate optimality combined with (6) imply $F^+ \leq F + \epsilon B(F)$ at every accepted block. If N bounds accepted-block ancestry, then

$$\mathcal{L} \leq F \leq E_N(F_0) := F_0 + B(F_0)[(1+\epsilon)^N - 1]. \tag{16}$$

At a completed checkpoint with H original sweeps left, $N = 2H$ includes a factor candidate whose subsequent gate block fails. Rollback returns to a covered ancestor and restores its envelope; it does not accumulate discarded branches. The nesting $E_{N-1}(F^+) \leq E_N(F)$ permits tighter valid envelopes. Appendix B proves these statements and their outward-bound versions. A failed check retains the direction until a tighter bound or further contraction within that scope permits release.

4.2 COMPILE A DUAL COST, THEN TRANSPORT IT

A numerical water-level solver only proposes a multiplier. For *any* $\theta > 0$, weak duality gives the upper cost

$$\mathcal{C}_\theta(a,E) = \theta E + \frac{1}{n}\sum_{a_i > \theta}\left[a_i - \theta\left(1 + \log\frac{a_i}{\theta}\right)\right] \geq \mathcal{C}(a,E), \qquad \mathcal{C}_{\theta=0} = \bar{a}. \tag{17}$$

Use upper contrasts and a reliable upper E, evaluate $\mathcal{C}_\theta$ upward, and evaluate $w - \mathcal{C}^U$ downward. Only a strictly positive lower endpoint authorizes release. A feasible primal objective has the opposite inequality direction and cannot authorize it.

One scale for transport and comparison. The feasible set is independent of a, so $\mathcal{C}(\alpha a, E) = \alpha\mathcal{C}(a, E)$ for $\alpha \geq 0$. At an episode's first checked zero, call its adverse vector a_{base} and compile $\mathcal{C}_0^U$ (index 0 denotes this episode baseline). After committed contraction ρ_k, current contrasts are $a_k = \rho_k^2 a_{\text{base}}$. Thus

$$\widetilde{\mathcal{C}}_k = \mathcal{C}(a_{\text{base}}, E_k), \qquad \mathcal{C}(a_k, E_k) = \rho_k^2 \widetilde{\mathcal{C}}_k, \qquad q_j \geq w - \rho_k^2 \mathcal{C}_0^U. \tag{18}$$

The fresh and cached bounds are compared on this common base scale; current cost already includes ρ_k^2. Further zero-prefix contraction improves the bound. Until authorization, directions remain stored and the original gate query is executed. Reactivation invalidates the episode; rejected trials do not advance its multiplier.

Budget tightening lowers base cost by $\int_{E_k}^{E_0} \theta(a_{\text{base}}, u)\, du$. A cached cost needs additional squared-contraction clock $\log(\mathcal{C}_0^U / \widetilde{\mathcal{C}}_k)$ to match a positive fresh base cost, conditional on those updates occurring within scope. Appendix B.4 separates this threshold wait from scan detection and fresh authorization. The measured policy scans every ten sweeps; affine-bank and shared-clock variants remain theoretical options.

Algorithm 1 Permission-carrying continuation (measured policy)

Require: Complete checkpoint; fixed reference, context, and original horizon
while the original stopping rule permits another sweep **do**
  Snapshot learner state χ, reduced arrays/summary, and checking state η_{cert}.
  Run original factor/gate trials using the closed reads for released atoms.
  if a block exhausts its trial cap **then**
    Restore the entire sweep snapshot; retain work counts; halt with failure.
  Invalidate reactivated episodes; register zeros; commit contraction products.
  if this completed sweep is a checking boundary **then**
    Compile missing costs for eligible zero episodes.
    for each eligible episode **do**
      if its floating hint is not positive **then**
        Retain directions and the valid cached cost.
      else
        Evaluate a fresh outward margin.
        if its lower endpoint is positive **then**
          Release directions and merge their summary.
        else
          Retain directions and invalidate this cached cost.
  Commit original controls χ; persist valid checking state η_{cert} separately.

Numerical contract. The theorem preserves exact-real learning events; outward arithmetic certifies permission for stored inputs; FP64 pairs test implemented continuation; lossless records reproduce their own compact suffixes. All-future machine branch preservation additionally requires propagated errors and separated functional guards (Appendix E). Algorithm 1 uses floating hints only to request reliable authorization.

5 Experiments: necessary information and measurable release

Every representation comparison fixes data, objective, initialization, updates and original budget. Natural releases test permission and continuation; constructed controls isolate read closure and dimension. Zero-release jobs remain in every aggregate. Trials and overlapping replay suffixes are coverage, not independent samples.

Fixed references. Letter expands 13 classes to 26 with replay, 153 fitting-only quadratic features and width eight (Slate, 1991). Three 1,300/520 fitting/validation splits and two paired kernels give six jobs. CIFAR expands five classes to ten on DINOv2 features (Krizhevsky, 2009; Oquab et al., 2024): three 400/100 resplits of one 500-image pool, again with both kernels. Both use 1,200

sweeps. Additional Letter uses three archive-relative unused, feature-disjoint 650/130 groups sharing a 390-example holdout. Appendix F specifies selection and exposure.

5.1 WHICH HIDDEN INFORMATION CHANGES THE NEXT UPDATE?

On four releasing Letter jobs, fixing the retained candidate and omitting only (8)'s correction flips 42 of 1,864 factor decisions. The smallest flipping guard distance is 2.14×10^{-9}, versus identity error 2.61×10^{-16}. Removing projection energy is a different intervention: for $U = \|u_P\|^2$, $V = t^2 v$, the candidate difference is

$$(\kappa_N - \kappa_F)\sqrt{U}, \qquad \kappa_N = \min(1, R/\sqrt{U}), \quad \kappa_F = \min(1, R/\sqrt{U+V}). \tag{19}$$

It reaches 0.15545; all 1,864 trials have active projection. At $U = 0$ both candidates are zero. Appendix C's rational tolerance-only witness isolates the remaining right-hand read, separately from this empirical denominator.

Decision sensitivity of hidden reads. On the stored natural traces, subtracting D_v from each recorded guard at the *same candidate and full tolerance* flips 100/2,780 nonzero-energy factor trials on Letter and 0/5,200 on CIFAR. For $D_v < 0$, erasure changes acceptance exactly when $D_v < g \leq 0$ (Appendix C.1). All these corrections are negative. Among original acceptances, the minimum margin/correction ratio $(-g)/(-D_v)$ is 0.005419 on Letter but 3.067 on CIFAR: 100 Letter trials enter the flip band, whereas every CIFAR acceptance stays outside. This stored-trace algebraic diagnostic is distinct from the executed 42/1,864 intervention; counts are not pooled. Appendix F.3 gives all denominators and zero-release controls.

Independent-permission controls further show that omitting any heavy-ball Gram component changes the first candidate; the full Gram and spectral summaries preserve decisions (Appendices F–G).

Mode count, not coordinate count. Six matched synthetic tasks ($K = 1, 4$; three seeds) vary coordinates per mode over 18, 180 and 1,800 while preserving initial mode energies. All 2,316 factor trials have nonzero permitted energy, including 2,169 active projections and 1,164 rejections. Summary and direction-retaining lazy both match full decisions. With K fixed, a hundredfold coordinate increase leaves the hidden summary at K energies and the complete records nearly constant (Table 3). Lazy avoids direction updates but still pays to store the directions.

Table 3: Fixed-mode dimension scaling. Complete zlib-6 bytes sum three seeds and two pauses per row. The summary has K energies; all three representations use one codec and identity-bound external context. Intermediate dimension 180 and the protocol summary are in Appendix G.2.

K	Coordinates/mode	Full	Lazy	Summary
1	18	11,120	11,708	10,128
1	1,800	94,893	95,499	10,144
4	18	14,115	14,917	10,305
4	1,800	346,737	344,538	10,362

5.2 HOW MUCH RELEASE DOES CERTIFICATION WORK BUY?

Compare unchecked full, geometric/simplex permission, fresh CE every ten sweeps, and compiled CE with fresh authorization. At the *same recorded release checkpoints*, CE gives a positive lower bound while the implemented simplex lower bound is nonpositive for **6/6 Letter and 12/14 CIFAR releases**; both bounds are positive at the other two. Thus loss information provides additional positive authorization at 18 of 20 recorded events (Appendix F.2). Fresh CE also advances all six Letter releases by 120–140 sweeps over geometric permission.

Relative to geometric permission, compilation increases release integrals by **55.20% on Letter** at 0.641% more time, and by **39.17% on CIFAR while using 33.23% less time**. Relative to fresh CE, dual evaluations fall 430→14 and 1,637→38, retaining 95.57%/90.80% of release integrals. Letter retains six releases with some 10–40 sweeps later; CIFAR retains 14 of 16. Table 4 includes overhead against unchecked full and complete-record costs.

Table 4: Same learner, different permission strategies. Direction integral divides released byte–sweeps by full-direction byte–sweeps. Record saving is signed complete zlib-6 saving at all six pauses, including caches/controls. T sums job medians; T_F, T_E denote unchecked full and fresh CE. Additional Letter has no timing measurement.

Cohort	Permission	Duals	Direction integral	Record saving	Time/full T/T_F	Fresh/time T_E/T
Letter	Geometric	–	2.17%	+0.31%	1.26	3.83
	Fresh CE	430	3.52%	−0.19%	4.84	1.00
	Compiled CE	14	3.37%	−0.66%	1.27	3.81
CIFAR	Geometric	–	13.78%	+6.72%	2.14	3.41
	Fresh CE	1,637	21.13%	+11.68%	7.31	1.00
	Compiled CE	38	19.18%	+9.49%	1.43	5.11
Additional Letter	Fresh CE	327	2.85%	−0.18%	–	–
	Compiled CE	10	2.67%	−0.585%	–	–

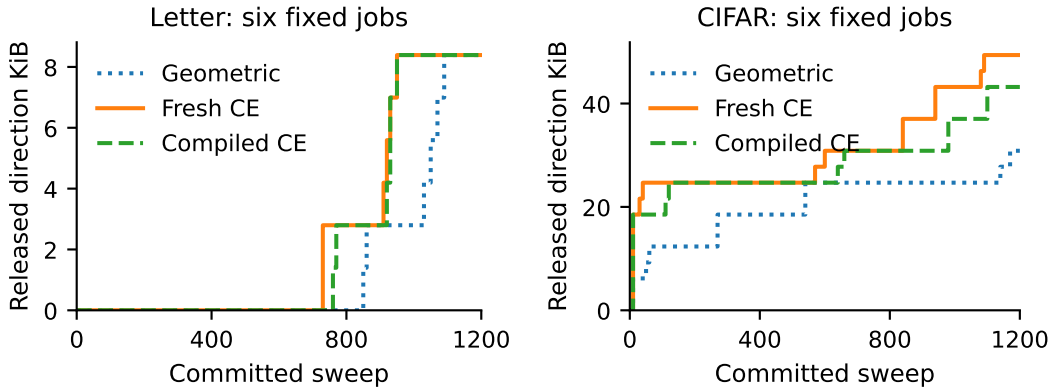

Figure 2: Cumulative released-direction bytes from every archived event across six jobs per dataset. Area is a byte–sweep integral under the common endpoint convention; Table 4 separately measures complete storage and time.

Additional Letter retains 93.90% of fresh release area with 10 rather than 327 checks. CIFAR's lost integral is 79.46% common-event delay and 20.54% missed releases. Split 402 retains 51.61%: its four releases lag by 140/160 sweeps, but only 26 of 600 atom–sweeps (4.33%) are scan waiting. First positive hints precede detection and outward authorization at 980/1100. Most delay therefore precedes the hint, rather than arising from ten-step scans (Appendix F.2).

5.3 Does the released state preserve continuation and restoration?

Compiled continuation matches all 48,539 reference decisions across twelve jobs. Of these, **20,297 trials use nonzero released energy**, including 7,201 active factor projections and 7,117 factor/gate rejections. Retained-candidate errors are at most 9.49×10^{-13}/4.98×10^{-14} on Letter/CIFAR. Appendix F.3 separates blocks, policies and statistical units.

For fixed-mode trials, the maximum same-trial ratio $|g_{C,t} - g_{F,t}|/|g_{F,t}|$ is 1.560×10^{-3} for summary and 2.008×10^{-12} for lazy. No full guard is zero: every observed error stays below its branch-boundary distance (Appendix E).

All 72 natural compiled records and 36 additional-Letter records reproduce their own uninterrupted suffixes in fresh workers. Fixed-mode compact/lazy each pass 36 restores; the billed compact decoder rejects 96 reference/context mutations. Table S6 gives the actual loaders and coverage without pooling overlapping suffixes. Additional Letter matches 28,759 trials per compact arm.

5.4 PRESERVED QUALITY AND COMPLETE RESOURCE REGIMES

Theorem 1 gives $Q(h_C) = Q(h_F)$ for every fixed output-based metric: the representation adds no quality shift. Measured FP64 validation-logit differences are at most 7.38×10^{-14}/1.60×10^{-14} and labels agree. A per-sample logit error δ bounds ideal CE change by 2δ (Appendix C.2); evaluation rounding is separate. These representation effects differ from the external quality gaps in Table 5.

Table 5: Validation CE / accuracy of the fixed reference and strong comparators. Full-to-compact predicted-label changes are zero. The original strict joint quality rules fail on both cohorts; additional Letter passes its separate same-width rule (Appendix F).

Cohort	Fixed tail	Strong same-width	Dense
CIFAR	.142841 / 95.67%	.111070 / 96.00%	.086044 / 97.00%
Additional Letter	1.191097 / 70.00%	1.197864 / 68.46%	1.074874 / 71.54%

Times in Table 4 sum job medians after one warmup and three serial repeats. **Each of the six CIFAR jobs improves all three coordinates over geometric permission**: more release integral, fewer complete six-pause bytes and less continuation time (Appendix F.2). This does not remove the overhead relative to unchecked full.

Complete CIFAR records save 141,706 bytes versus full. Split 403 contributes 76.77%, yet removing it leaves **32,915 saved bytes (3.31%)**; the remaining jobs retain all three improvements over geometric permission. Original/additional Letter instead cost 6,800/6,072 extra bytes. Late Letter pauses save space despite a negative uniform aggregate. Appendix G retains every pause/codec sign and full-process memory cost.

6 RELATED WORK AND IMPLICATIONS

A different correctness object. Safe screening excludes variables from sparse optima (Ndiaye et al., 2017; Wang et al., 2014). SoRA learns proximal gates and removes zero-rank blocks for inference; AdaLoRA allocates adaptation rank during training (Ding et al., 2023; Zhang et al., 2023). We instead preserve a specified learner's finite candidates, acceptance and rollback, including effects that remain after directional inactivity.

ExCP compresses weight/momentum checkpoints (Li et al., 2024); LDAdam and LoRA-Pre design low-dimensional statistics and updates (Robert et al., 2025; Wang et al., 2026). Our guarantee holds the original update rule fixed, separating state representation from update design. Lazy regularization (Bottou, 2012) propagates a multiplier while retaining a base direction: Table 3 separates avoiding updates from releasing direction storage. Abstract interpretation and proof-carrying code provide the program-abstraction viewpoint (Cousot and Cousot, 1977; Necula, 1997); the learning-specific advance is computable future permission coupled to a closed state and its structural necessity.

When release pays off. For declared pause weights p_t and residence times h_t, complete-record saving is $S = \sum_{j,t} p_t h_t (B^F_{jt} - B^C_{jt})$. With explicit unit weights $a_{\mathrm{mem}}, b_{\mathrm{time}} > 0$, net cost is $-a_{\mathrm{mem}} S + b_{\mathrm{time}}(T_C - T_F)$. Longer residence can offset extra computation when bytes are saved; it cannot reverse simultaneous storage and time overhead. Complete records and GPU peaks remain distinct resources.

Conclusion. Releasability belongs to the remaining learning computation, not to the predictor alone. Future permission removes direction-sensitive gate queries; closed summaries preserve residual optimizer effects. Their joint construction releases directions while retaining the original learning events. Same-checkpoint bounds establish the permission increment, structural queries identify indispensable information, and complete records locate the measured resource gains.

REPRODUCIBILITY STATEMENT

The appendix specifies the reference, proofs, data roles and frozen evaluation protocols. The artifact provides immutable traces/records, read-erasure and resource analyses, actual-loader replay, identity checks, and paired manuscript sources. `REPRODUCE.md` separates analysis dependencies from the NumPy-bound environment for archived structural replay; a preflight validates the latter before execution. Original natural-data training and restoration additionally require the indexed historical inputs and worker dependencies. `analysis/reproduce.py` generates the paired permission/resource tables; `code/replay_delivery.py` checks the billed structural records. Missing natural guard pairs and episode contrasts remain explicitly identified.

DATA AND ETHICS STATEMENT

The study uses public Letter Recognition and image datasets, frozen feature pipelines, and constructed controls. Protocols specify fitting, validation and test roles, shared samples and prior exposure. No human-participant recruitment or clinical data collection is involved. Ordinary-head pipeline diagnostics retain their separate roles.

AI USE STATEMENT

Generative AI tools were used to aid manuscript restructuring, language editing, and documentation. We have reviewed all AI-assisted work. The mathematical results, proofs, computational checks, and reported numbers were produced and verified by the authors, who take full responsibility for the final content of the paper.

REFERENCES

Masoud Ahookhosh, Le Thi Khanh Hien, Nicolas Gillis, and Panagiotis Patrinos. Multi-block Bregman proximal alternating linearized minimization and its application to orthogonal nonnegative matrix factorization. *Computational Optimization and Applications*, 79:681–715, 2021. https://doi.org/10.1007/s10589-021-00286-3.

Léon Bottou. Stochastic gradient descent tricks. In Grégoire Montavon, Geneviève B. Orr, and Klaus-Robert Müller (eds.), *Neural Networks: Tricks of the Trade*, volume 7700 of Lecture Notes in Computer Science, pp. 421–436. Springer, 2012. https://doi.org/10.1007/978-3-642-35289-8_25.

Patrick Cousot and Radhia Cousot. Abstract interpretation: A unified lattice model for static analysis of programs by construction or approximation of fixpoints. In *POPL*, pp. 238–252, 1977. https://www.di.ens.fr/~cousot/COUSOTpapers/POPL77.shtml.

Ning Ding, Xingtai Lv, Qiaosen Wang, Yulin Chen, Bowen Zhou, Zhiyuan Liu, and Maosong Sun. Sparse low-rank adaptation of pre-trained language models. In *EMNLP*, pp. 4133–4145, 2023. https://doi.org/10.18653/v1/2023.emnlp-main.252.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *ICLR*, 2022. https://openreview.net/forum?id=nZeVKeeFYf9.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009. https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

Wenshuo Li, Xinghao Chen, Han Shu, Yehui Tang, and Yunhe Wang. ExCP: Extreme LLM checkpoint compression via weight-momentum joint shrinking. In *ICML*, PMLR 235:27575–27588, 2024. https://proceedings.mlr.press/v235/li24m.html.

Eugene Ndiaye, Olivier Fercoq, Alexandre Gramfort, and Joseph Salmon. Gap Safe screening rules for sparsity enforcing penalties. *Journal of Machine Learning Research*, 18(128):1–33, 2017. https://jmlr.org/papers/v18/16-577.html.

George C. Necula. Proof-carrying code. In *POPL*, pp. 106–119, 1997. https://people.eecs.berkeley.edu/~necula/papers.html.

Maxime Oquab et al. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024. https://openreview.net/forum?id=a68SUt6zFt.

Thomas Robert, Mher Safaryan, Ionut-Vlad Modoranu, and Dan Alistarh. LDAdam: Adaptive optimization from low-dimensional gradient statistics. In *ICLR*, 2025. https://proceedings.iclr.cc/paper_files/paper/2025/hash/0e6eb6f6001748b66348b9b53ad434db-Abstract-Conference.html.

David Slate. Letter Recognition [dataset]. UCI Machine Learning Repository, 1991. https://doi.org/10.24432/C5ZP40.

Jie Wang, Jiayu Zhou, Jun Liu, Peter Wonka, and Jieping Ye. A safe screening rule for sparse logistic regression. In *Advances in Neural Information Processing Systems*, 27, 2014. https://papers.nips.cc/paper_files/paper/2014/hash/65b9807722a2a0b28aaebd00b94d6521-Abstract.html.

Zhengbo Wang, Jian Liang, Ran He, Zilei Wang, and Tieniu Tan. Taming momentum: Rethinking optimizer states through low-rank approximation. In *ICLR*, 2026. https://proceedings.iclr.cc/paper_files/paper/2026/hash/2f4b6febe0b70805c3be75e5d6a66918-Abstract-Conference.html.

Qingru Zhang, Minshuo Chen, Alexander Bukharin, Nikos Karampatziakis, Pengcheng He, Yu Cheng, Weizhu Chen, and Tuo Zhao. Adaptive budget allocation for parameter-efficient fine-tuning. In *ICLR*, 2023. https://openreview.net/forum?id=lq62uWRJjiY.

Cun-Hui Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894–942, 2010. https://doi.org/10.1214/09-AOS729.

A Complete reference and information boundary

A.1 Objective, kernels, and literal zero outputs

For the minimax concave penalty used by the experiments, let $a_{\mathrm{MCP}} > 0$ and

$$P_{2,j}(u) = \int_0^u \min(t/a_{\mathrm{MCP}}, \lambda_j)\, dt, \qquad \lambda_j u - P_{2,j}(u) = \begin{cases} \lambda_j u - u^2/(2a_{\mathrm{MCP}}), & u \le a_{\mathrm{MCP}}\lambda_j, \\ a_{\mathrm{MCP}}\lambda_j^2/2, & u \ge a_{\mathrm{MCP}}\lambda_j. \end{cases} \tag{20}$$

Both branches are nonnegative and $P'_{2,j}(0) = 0$. This proves $F \ge \mathcal{L} \ge 0$ and $-\Lambda \le f \le F$ for (4). The representation theorem also covers $P_2 = 0$.

The Euclidean gate kernel is $h(u) = u^2/2$. For an entropy-core kernel, set $h_0(u) = u\log u + (1-u)\log(1-u)$ on $[\delta, 1-\delta]$, with $0 < \delta < 1/2$, and extend it by its second-order Taylor polynomial at each endpoint of that core interval. This gives finite endpoint derivatives and $h'' > 0$. The unique gate solution satisfies

$$\mathcal{P}_h(x; q, M) = \arg\min_{u\in[0,1]} \{qu + MD_h(u, x)\}, \quad 0 \in q + M[h'(u) - h'(x)] + N_{[0,1]}(u). \tag{21}$$

Put $b = h'(x) - q/M$. The oracle returns literal 0 for $b \le h'(0)$, 1 for $b \ge h'(1)$, and $(h')^{-1}(b)$ otherwise. Thus $x = 0, q \ge 0$ returns exactly zero, not a small positive gate.

A.2 State and transaction boundaries

Context fixes data, base, scale, atom identities, objective, rules and absolute budget. Encode χ, η_{cert} and snapshots; only χ corresponds to unchecked full. Factor acceptance is provisional until gate success; failure restores the pre-sweep state and prescribed counters. Release/restore use completed checkpoints and an event budget. Radial closure allows nonzero anchors and real t; CE transport requires zero anchors and $M \ge \mu$. No other control reads omitted raw gate gradients.

B Future cross-entropy, reliable dual bounds, and transport

B.1 A loss envelope on accepted ancestry

Proposition B.1 (Nested finite-budget envelope)**.** *With $B(F) = \max(1, \Lambda, F)$, a state with at most N accepted-block ancestors from F_0 satisfies* (16)*. An accepted block gives $E_{N-1}(F^+) \le E_N(F)$ for $N \ge 1$. Upper $F_0^U \ge F_0$ may replace F_0.*

Proof. Projection optimality gives $\langle \nabla_\Theta f, \Delta\rangle \le -M\|\Delta\|^2$; gate optimality gives $\langle \nabla_\gamma f + \lambda, \Delta\gamma\rangle \le -M_\gamma(D^+_{\mathcal{H}} + D^-_{\mathcal{H}})$, using forward/reverse divergences. Adding (6) yields

$$\begin{aligned} F^+ - F &\le -\tfrac{1}{2}(1+\nu_\Theta)M\|\Delta\|^2 + \epsilon\max(1,|f|), \quad \text{factor}, \\ F^+ - F &\le -\nu_\gamma M_\gamma D^+_{\mathcal{H}} - M_\gamma D^-_{\mathcal{H}} + \epsilon\max(1,|f|), \quad \text{gate}. \end{aligned} \tag{22}$$

Since $-\Lambda \le f \le F$, both imply $F^+ \le F + \epsilon B(F)$ and $B(F^+) \le (1+\epsilon)B(F)$. Iteration gives $F_m \le F_0 + \epsilon B(F_0)\sum_{t<m}(1+\epsilon)^t = E_m(F_0) \le E_N(F_0)$; at $\epsilon = 0$ use $F_m \le F_0$. Substitution gives $E_{N-1}(F^+) \le F + \epsilon B(F) + (1+\epsilon)B(F)[(1+\epsilon)^{N-1} - 1] = E_N(F)$. Monotonicity permits F_0^U. A provisional accepted factor is covered before querying gates; rollback restores an ancestor. Thus $N = 2H$ covers H remaining sweeps, without accumulating discarded branches. □

A verified decrease $d \ge 0$ sharpens $U \ge F$ to $U - d + \epsilon B(U)$. Envelope minima require the same future scope: a discarded branch's bound need not cover its sibling. At $\epsilon = 0$ the real envelope is unlimited, while records retain their declared scope.

B.2 Optimal probability relaxation and safe evaluation

For $E > 0, a \ne 0$, the continuous $S(\theta) = \sum_{a_i > \theta} \log(a_i/\theta)$ decreases strictly from infinity to zero, giving a unique water level. Set $p_i^* = \min(1, \theta/a_i)$ (1 at $a_i = 0$). It uses nE loss; for feasible p,

$$\sum_i a_i p_i \ge \sum_i (a_i p_i - \theta\log p_i) - \theta nE \ge \sum_i (a_i p_i^* - \theta\log p_i^*) - \theta nE = \sum_i a_i p_i^*. \tag{23}$$

Coordinate minimization proves the second inequality. Actual probabilities satisfy (13); positive forcing extends zero prefixes through provisional candidates and rollback. Binary contrasts $0, -a_i$ attain the relaxation at p_i^* (limits at boundaries), while shared-network reachability can be tighter. Finite E gives $V \geq e^{-nE} \sum_i a_i > 0$.

For any $\theta > 0$, the minimum of $a_i p - \theta \log p$ is $\phi = a_i$ when $a_i \leq \theta$, and $\theta[1 + \log(a_i/\theta)]$ otherwise. Weak duality gives $\sum_i \phi - \theta n E \leq V$, proving (17); at zero multiplier, $\mathcal{C}_{\theta=0} = \bar{a}$. The support form $\mathcal{C} = \max_{p \in \mathcal{P}_E} n^{-1} a^\top (1-p)$ proves monotonicity in a, E. Upper inputs and cost, with lower w, give a valid lower margin. Unresolved/nonfinite checks retain directions.

A primal value has the wrong direction: for $n = 1, a = 1, E = \log 2, w = .4$, the true margin is $-.1$; treating feasible $p = 1$ as a lower objective bound falsely gives $.4$. An enclosed dual stays safe without solver convergence.

B.3 Compiled cost and scheduling

Proposition B.2 (Transport and affine reuse)**.** *At a checked zero, let a_{base} be the adverse vector and E_0 cover the remaining scope. A compiled $\mathcal{C}_0^U \geq \mathcal{C}(a_{\text{base}}, E_0)$ gives $q_j \geq w - \rho_k^2 \mathcal{C}_0^U$ on the zero episode. Reliable positivity proves future inactivity. At a tighter E_k^U, a valid upper cost is*

$$\mathcal{C}_{\text{bank}}^U(E_k^U) = \min\{\bar{a}_{\text{base}}^U, A_1^U + \theta_1 E_k^U, \ldots, A_b^U + \theta_b E_k^U\}, \tag{24}$$

where $\theta_\ell \geq 0$ and A_ℓ^U bounds (17)*'s sum term at a_{base}.*

Proof. Homogeneity and further contraction $0 \leq \sigma \leq 1$ give $q_j \geq w - \sigma^2 \rho_k^2 \mathcal{C}(a_{\text{base}}, E_0) \geq w - \rho_k^2 \mathcal{C}_0^U$. Positivity extends the zero prefix; until then, directions and ordinary queries remain. Every bank multiplier is legal at the tighter budget, so the minimum remains an upper bound. □

For $\psi_t > 0$, the clock $\tau_k = -\sum_{t<k} \log \psi_t^2$ reaches margin $0 < \delta < w$ when $\tau_k - \tau_{t_0} \geq \max\{0, \log[\mathcal{C}_0^U/(w-\delta)]\}$. Zero cost/contraction is handled directly. A shared-clock heap needs $O(T + J\log(1+J))$ scalar work and $O(J)$ storage (one insertion/removal per episode), excluding compilation, gradients and repairs. Algorithm 1 instead uses per-episode products and ten-sweep scans, with invalidation and rollback.

B.4 Budget tightening, common scales and scan delay

On $A = \{i : a_i > \theta\}$, $k = |A|$, solve the water equation and differentiate $\mathcal{C} = n^{-1}(\sum_{i \in A} a_i - k\theta)$:

$$\theta(E) = \Big(\prod_{i \in A} a_i\Big)^{1/k} e^{-nE/k}, \qquad \mathcal{C}_E' = \theta(E), \qquad \mathcal{C}_E'' = -n\theta(E)/k. \tag{25}$$

Water levels match at active-set boundaries, with right limit $\max_i a_i$ at $E = 0$ (zero if $a = 0$). Integrating the continuous decreasing derivative proves

$$\mathcal{C}(a, E_0) - \mathcal{C}(a, E_k) = \int_{E_k}^{E_0} \theta(a, u)\, du, \quad E_0 \geq E_k \geq 0, \tag{26}$$

with bounds $\theta(a, E_0)(E_0 - E_k)$ and $\theta(a, E_k)(E_0 - E_k)$. This is the budget-gap diagnostic. The cached gap additionally includes initial upper-bound slack.

Corollary B.3 (Conditional contraction delay)**.** *Let $\widetilde{\mathcal{C}}_k = \mathcal{C}(a_{\text{base}}, E_k) > 0$ be the fresh* base-scale *cost, at most $\mathcal{C}_0^U$. If $\rho_k^2 \widetilde{\mathcal{C}}_k \leq w - \delta$ with $0 < \delta < w$, an additional squared contraction R_d^2 satisfying*

$$-\log R_d^2 \geq \log(\mathcal{C}_0^U / \widetilde{\mathcal{C}}_k) \tag{27}$$

makes the cached bound reach that margin. If actual successful updates obey $\psi_t^2 \leq r < 1$, then $\lceil \log(\mathcal{C}_0^U/\widetilde{\mathcal{C}}_k)/(-\log r) \rceil$ such updates suffice within the original scope.

Proof. Multiply $R_d^2 \mathcal{C}_0^U \leq \widetilde{\mathcal{C}}_k$ by ρ_k^2; use $R_d^2 \leq r^d$ for the count. Fresh validity supplies inactivity. Current cost is $\rho_k^2 \widetilde{\mathcal{C}}_k$, so it must not be multiplied by ρ_k^2 again. Zero costs/contractions are handled directly. □

Proposition B.4 (Scan waiting). *In an uninvalidated zero episode, the hint first becomes and stays positive at completed sweep k_*. Scans occur at multiples of h; the first positive scan is within scope and its fresh check succeeds. Then $k_{\mathrm{auth}} = h\lceil k_*/h \rceil$ and $0 \le k_{\mathrm{auth}} - k_* \le h - 1$.*

Proof. This is the first scan not before k_*; all earlier hints are nonpositive. Consecutive scans have at most $h - 1$ intervening integer sweeps. Invalidations or failed fresh checks require separate accounting. □

C Closed state and exact continuation

Energy alone cannot fix an unlicensed gate. For binary softmax with $x = a = s = 1$, $y = 1, W_0 = 0, \gamma = 0, w = .1$, choices $b = (\pm 1, 0)^\top$ have energy 2 and identical logits, yet forcing $.1 \mp .5$. Euclidean $M_\gamma = 1$ gives next gates $.4, 0$. Permission cannot be inferred from energy; Section 2 proves the converse obstruction.

C.1 Complete guard and composition across release boundaries

Put $H(z) = \max(1, |z|)$ and $g_P = \mathcal{A}_P - \epsilon H(f_P)$. Table 1 gives, at the same retained candidate,

$$g_F - g_P = \tfrac{1}{2}[\mu - (1 - \nu_\Theta)M](\psi - 1)^2 v - \epsilon\{H(f_P + \mu v/2) - H(f_P)\}. \tag{28}$$

The second term has magnitude at most $\epsilon\mu v/2$, since H is 1-Lipschitz; its sign need not be fixed when $f_P < 0$. Thus residual-only and tolerance-only interventions test distinct reads.

In (1), use $x = 1, y^2 = 30, u = 0, M = 4, \nu = 1/2, \epsilon = 1/100$, with inactive ball. Then $f_P = 2, f_F = 17, \mathcal{A}_P = 1, \mathcal{A}_F = 1 - 30/32 = 1/16$. Full accepts $1/16 \le 17/100$; retained-only tolerance rejects $1/16 > 1/50$.

When a read changes a decision. At fixed candidate and full tolerance, erasing residual read D changes guard g to $g - D$. Acceptance at $g \le 0$ gives the exact flip conditions

$$D < 0 :\ D < g \le 0; \qquad D > 0 :\ 0 < g \le D; \qquad D = 0 :\ \text{no flip}. \tag{29}$$

If $D < 0$, only $g \le 0 < g - D$ can flip; if $D > 0$, only $g - D \le 0 < g$ can flip. These prove the conditions, including endpoints. For accepted g and $D < 0$, the criterion is $(-g)/(-D) < 1$. Projection erasure changes the candidate and is separate.

Proof of adaptive composition. Relate states by equal retained values/χ, $S_D = S(d_D)$, corresponding snapshots and valid permissions. Equal controls select M; zero gates make retained task values/gradients equal. Equations (7)–(9) give equal candidates/guards. Acceptance commits corresponding states; rejection updates the same backtracking control. Permitted gate outputs have zero displacement, pairing and Bregman terms. Other gate reads agree; hidden ridge cancels from their objective difference but remains in the tolerance. No other control reads omitted gradients.

Gate-cap failure restores corresponding pre-sweep snapshots; the same resets, caps, objective, patience, work counts and event budget determine halt. Disjoint release at a completed checkpoint adds $v_{D\cup E} = v_D + \|d_E\|^2$, or Gram/modewise statistics for identical later dynamics. Permission and addition preserve the relation; subsequent rollback cannot cross this boundary. At corresponding interruption points, same-context $\mathsf{Dec}(\mathsf{Enc}(\xi_C)) = \xi_C$ inserts the identity. Finite internal checking returns to the next event. Induction over events, releases and finite restores proves Theorem 1.

C.2 Zero additional metric shift

Theorem 1 gives equal predictors, hence equal fixed output metrics and selection with fixed event tie-breaks. For any external comparator h_{cmp}, the difference decomposes as $Q_V(h_C) - Q_V(h_{\mathrm{cmp}}) = [Q_V(h_F) - Q_V(h_{\mathrm{cmp}})] + [Q_V(h_C) - Q_V(h_F)]$, whose last term is zero in exact arithmetic.

Corollary C.1 (Logit and selection stability). *If $\|z_i^C - z_i^F\|_\infty \le \delta_i$, then $|\mathcal{L}_C - \mathcal{L}_F| \le 2n^{-1}\sum_i \delta_i$. A full top-two gap above $2\delta_i$ preserves the label. If every selectable checkpoint has CE error at most ε_Q, a unique full optimum separated by more than $2\varepsilon_Q$ is selected unchanged.*

Proof. The exponential sums differ by a factor in $[e^{-\delta_i}, e^{\delta_i}]$; log-sum-exp and the true logit each change at most δ_i. Average for CE. Class and checkpoint gaps change at most twice their error bound. Computed CE additionally needs evaluation-error allowances. □

Scalar distinguishability. A common allowed trial with $u_P \neq 0$, $t \neq 0$ and active projection produces $Ru_P/\sqrt{\|u_P\|^2 + t^2 v_i}$. Different hidden energies yield different candidates, so a complete summary for that trial class must distinguish them. This necessity concerns jointly admissible states in that trial class.

C.3 Gram classes, continuous dimension, and momentum

Fix independent permission, retained state and controls. Allowed reads are $q_u(Z) = \|Zu\|^2$ for all $u \in \mathbb{R}^p$; updates $Z^+ = ZA$ depend on these reads and controls. Polarization gives cross queries $\langle Zu, Zv\rangle = (q_{u+v} - q_u - q_v)/2$.

Equal Grams give equal adaptive reads and updates $A^\top GA$, hence equal subsequent events by induction. Conversely, $G_1 - G_2 \neq 0$ is detected by e_i if a diagonal differs, or $e_i + e_j$ otherwise. Outputting that probe or thresholding its values distinguishes the states, proving the complete-class iff.

For $m \geq p$, suppose the permitted domain contains the continuous Cholesky section $Z(G) = [R(G)^\top, 0]^\top$, $R(G)^\top R(G) = G$, over an open positive-definite set $U \subset \mathbb{R}^d$, $d = p(p+1)/2$. Distinguishability makes every continuous complete summary an injection $U \to \mathbb{R}^r$. If $r < d$, appending zeros gives a continuous injection into a proper subspace of $\mathbb{R}^d$, contradicting invariance of domain. Gram entries attain d. Rank-restricted domains, discontinuous encodings and smaller query classes need not obey this bound.

For projected heavy-ball, $h^+ = \alpha_h h + \nabla f$, $d^+ = \kappa(d - \eta_{\text{step}} h^+)$, and independently inactive coordinates obey $\nabla_D f = \mu d_D$. With $Z = [d_D, h_D]$,

$$A = \begin{pmatrix} \kappa(1 - \eta_{\text{step}}\mu) & \mu \\ -\kappa\eta_{\text{step}}\alpha_h & \alpha_h \end{pmatrix}, \qquad G = \begin{pmatrix} v & c_h \\ c_h & w_h \end{pmatrix}, \qquad G^+ = A^\top GA. \tag{30}$$

For $z = (1 - \eta_{\text{step}}\mu, -\eta_{\text{step}}\alpha_h)^\top$, the hidden unprojected norm is $z^\top Gz$. For $q = (\kappa(1 - \eta_{\text{step}}\mu) - 1, -\kappa\eta_{\text{step}}\alpha_h)^\top$, the residual correction is $[\mu - (1-\nu)M]q^\top Gq/2$. Objective, pairing, and step reads are $\mu G_{11}/2$, $\mu e_1^\top Gq$, and $q^\top Gq$.

When $\alpha_h \neq 0$, querying the unprojected norm at three distinct legal steps identifies its quadratic polynomial

$$n(\eta_{\text{step}}) = v - 2\eta_{\text{step}}(\mu v + \alpha_h c_h) + \eta_{\text{step}}^2(\mu^2 v + 2\mu\alpha_h c_h + \alpha_h^2 w_h). \tag{31}$$

Its three coefficients recover v, c_h, w_h. This gives an operational necessity statement for that query family. Momentum requires its own permission: relative acceptance alone need not decrease the objective. For example, $F(d) = d^2/2$, $d = 1, h = -10, \alpha_h = 1/2, M = 4, \nu = 1/2$ yields $d^+ = 2$ and $F^+ > F$ despite residual $-1/2$. The global-ball permission below avoids this issue.

D Finite-spectrum hidden dynamics

Let the hidden regularizer be $R_D(y) = y^\top Qy/2$, with fixed $Q = \sum_{\ell=1}^K h_\ell P_\ell \succeq 0$, distinct h_ℓ, and orthogonal projectors P_ℓ. There are no unrepresented retained–hidden regularization cross terms. Independent permission removes hidden task gradients, projection is onto the same Euclidean displacement ball, and other functional hidden reads depend only on the spectral energies.

Theorem D.1 (Spectral-energy continuation)**.** *Set $v_\ell = \|P_\ell y\|^2$, $t_\ell = 1 - h_\ell/M$, and*

$$\kappa = \min\left\{1, \frac{R}{\sqrt{\|u_P\|^2 + \sum_\ell t_\ell^2 v_\ell}}\right\}, \qquad \psi_\ell = \kappa t_\ell. \tag{32}$$

The K energies close the factor candidate, objective, pairing, step, and relative-acceptance reads, with

$$v_\ell^+ = \psi_\ell^2 v_\ell, \qquad \mathcal{A}_F - \mathcal{A}_P = \tfrac{1}{2}\sum_\ell [h_\ell - (1-\nu)M](\psi_\ell - 1)^2 v_\ell. \tag{33}$$

Complete control snapshots give the same continuation and restoration statement as Theorem 1.

Proof. Orthogonality and $P_\ell u_D = t_\ell P_\ell y$ give the projection and update. Hidden objective, pairing and squared-step terms are $\frac{1}{2}\sum_\ell h_\ell(\psi_\ell^2 - 1)v_\ell$, $\sum_\ell h_\ell(\psi_\ell - 1)v_\ell$ and $\sum_\ell(\psi_\ell - 1)^2 v_\ell$. Combining them with (6) gives (33). The tolerance uses $f_P + \frac{1}{2}\sum_\ell h_\ell v_\ell$. With all reads closed, the event induction of Theorem 1 applies. Merge energies only within identical subsequent modes. □

Proposition D.2 (Identifiable modes require separate information)**.** *Suppose the permitted program class contains an inactive-ball continuation at a fixed sufficiently large M and exposes the current and following $K-1$ hidden norms. Then its observations determine every v_ℓ. On an open permitted energy section, any continuous complete summary requires at least K real coordinates.*

Proof. Choose $M > \max_\ell h_\ell/(1-\nu)$, so the hidden relative residual is nonpositive and all hidden-only trials may be accepted. The distinct numbers $a_\ell = (1 - h_\ell/M)^2$ lie in $(0, 1]$. Future norms satisfy $r_t = \sum_\ell a_\ell^t v_\ell$, $t = 0, \dots, K-1$. The Vandermonde determinant $\prod_{i<j}(a_j - a_i)$ is nonzero, hence these reads recover all energies. The continuous-dimension argument is the open-section injection proof used for Theorem 2, now in $\mathbb{R}^K$. □

A single total energy cannot usually reproduce even the next norm. If $v = \sum v_\ell > 0$ and $\bar{h} = \sum h_\ell v_\ell / v$, then

$$\sum_\ell (1 - h_\ell/M)^2 v_\ell - (1 - \bar{h}/M)^2 v = \frac{v}{M^2}\sum_\ell \frac{v_\ell}{v}(h_\ell - \bar{h})^2. \tag{34}$$

Matching norm and quadratic energy still loses decay variance. Fixed K allows arbitrary within-mode dimension; coordinate-distinct modes can remove that saving. Necessity concerns the complete class, not every trajectory.

An independent, nonempty permission class. For zero-anchor softmax atoms whose accepted factor candidates obey $\|\Theta\| \le R$, Cauchy–Schwarz gives

$$|s(b_c - b_y)a^\top x_i| \le \sqrt{2}|s|\|b\|\|a\|\|x_i\| \le |s|R^2\|x_i\|/\sqrt{2}. \tag{35}$$

Hence $w > |s|R^2 n^{-1}\sum_i \|x_i\|/\sqrt{2}$ keeps a current zero gate at zero for any dynamics staying in that ball, including spectral and momentum updates. For nonzero fixed anchors use $\|\Theta_w\| + R$. This global-ball permission defines the prescribed-zero structural tests without the radial CE assumption.

E Reliable arithmetic and finite-precision execution

Checkpoint permission. Treat stored inputs as exact dyadics and outward-enclose contrasts, loss, powers, logs and subtractions: upper a, E, lower w, upper cost, then lower margin. Ordinary rounding can underestimate $1 - (-2^{-53})$ as 1. Permission for these inputs and all-future machine execution are different objects.

Paired guard-distance diagnostic. For recorded guards, $|g_C - g_F| < |g_F|$ implies equal signs: when $g_F > 0$, $g_C \ge g_F - |g_C - g_F| > 0$; reverse the inequality for $g_F < 0$. At zero use the actual boundary rule. Fixed-mode guards are nonzero (minimum 4.396×10^{-15}); Section 5's error ratios describe observed pairs, not unknown real-arithmetic error.

Machine inactivity. For normal round-to-nearest operations, no overflow, unit roundoff u, $M \ge \mu > 0$ and $0 \le \widehat{\kappa} \le 1$, put $\gamma_4 = 4u/(1-4u)$ and $\varepsilon_d = 2\gamma_4$. Expanding $\widehat{x}^+ = \mathrm{fl}(\widehat{\kappa}\,\mathrm{fl}[x - \mathrm{fl}(\mathrm{fl}(\mu x)/M)])$ gives $\widehat{x}^+ = \psi x + r$, $\psi = \widehat{\kappa}(1 - \mu/M) \in [0, 1]$, and $|r| \le \widehat{\kappa}\gamma_4(1 + \mu/M)|x| \le \varepsilon_d|x|$. Summing $\|e_{t+1}\| \le \|e_t\| + \varepsilon_d(1+\varepsilon_d)^t\|a_0\|$ bounds ray deviations after H accepted-factor ancestors by $\Delta_H\|a_0\|$ and $\Delta_H\|b_0\|$, where $\Delta_H = (1+\varepsilon_d)^H - 1$. Include provisional accepted factors before gate queries. Let $\underline{m}_{\mathrm{ray}}$ rigorously lower-bound the ideal common-ray forcing throughout this same scope, using a CE envelope valid for the probabilities under consideration, including future rounding allowances when needed. With $K_0 = \sqrt{2}|s|\|a_0\|\|b_0\|n^{-1}\sum_i\|x_i\|$ and verified forcing-evaluation error ε_q, the condition

$$\underline{m}_{\mathrm{ray}} - K_0(2\Delta_H + \Delta_H^2) - \varepsilon_q > 0 \tag{36}$$

ensures machine inactivity when the endpoint branch writes literal zero: expand the perturbed bilinear contrast, then induct. All constants require valid enclosures.

Machine branch preservation. On a common branch, an L_t-Lipschitz reduced map and local implementation errors η_t^j give $e_{t+1}^j \le L_t e_t^j + \eta_t^j$, $j \in \{F, C\}$. A K_t-Lipschitz guard with evaluation errors ξ_t^j has matching signs whenever

$$|\widehat{g}_t^C| > K_t(e_t^F + e_t^C) + \xi_t^F + \xi_t^C. \tag{37}$$

The triangle inequality proves the claim; applying it at acceptance, projection, endpoint, reset, stop, cap and rollback guards closes branch induction. It gives equal discrete branches and bounded errors, not bitwise identity of different arithmetic graphs. Unresolved guards require retention before deletion or a separately stored recovery path.

F Frozen protocols and full experimental denominators

F.1 Data, learner selection, and statistical units

Letter uses seeds 301–303, both kernels, width eight, $s = .125$, $\mu = 2 \times 10^{-5}$, $\lambda = .04$, $\beta = .03$, zero anchors and $R = 1.25\|\Theta_0\|$. Both standardizations use fitting data only; all 45 development candidates, including the checkpoint-selected zero-release reference, remain in the artifact. CIFAR seeds 401–403 resplit one 500-image DINOv2 pool with $\mu = 2 \times 10^{-6}$, $\beta = .03$; all rows and the official test were previously exposed, so these are not independent population validation.

Additional Letter uses partition seed 2026092411 and class-stratified sampling after excluding audited used IDs and exact-feature duplicates. Its 2,730 selected rows/signatures are mutually disjoint and outside that union; "unused" is archive-relative. Each class supplies 15 shared-holdout rows and 25 fit/five validation rows per group. Numeric holdout data stay masked until validation locks. Source/warm stages use 1,500/LR-5 and 600/LR-4 steps. Two kernels are paired within three groups sharing one holdout, not independent replicates.

F.2 Release events, complete records and heterogeneous costs

The complete event ledger includes zero-release jobs. Geometric/fresh/compiled release 6/6/6 Letter atoms and 10/16/14 CIFAR atoms, counted across jobs.

Same-checkpoint permission. All 20 fresh authorizations match unique (job, sweep, atom) release keys. CE alone gives a positive recorded lower bound at 6/6 Letter and 12/14 CIFAR events; both bounds are positive for atom 7 at sweep 10 in both split-403 kernels. A nonpositive bound does not imply negative forcing. `analysis/permission_pairs.csv` lists the actual-event denominator and both bounds.

Using per-atom bytes 1,432/3,160 (Letter/CIFAR), the integral is $\sum_j b_j(1200 - \tau_j)_+$; full-direction denominators are 82,483,200/182,016,000. Compiled calls comprise 8+6 and 24+14 compilation/authorization calls; objective enclosures fall 336→14 and 597→25. Reactivation and failed-fresh retries occur only in controlled tests.

Table S1: Descriptive split variation after summing paired kernels. Retention compares compiled/fresh; storage and median-time ratios compare compiled/full. NA: both integrals zero.

Data	Split	Direction retention	Complete saving	Time ratio
Letter	301	NA	−1.60%	1.137
	302	100.00%	−0.89%	1.266
	303	94.04%	+0.53%	1.422
CIFAR	401	91.19%	+5.73%	1.562
	402	51.61%	+0.88%	1.358
	403	96.51%	+21.86%	1.374

Paired resource improvements. Each CIFAR job improves integral, complete six-pause bytes and median time over geometric (time reductions 21.8736–43.3925%). Without 403, integral rises 31.8777% and time falls 37.1659%. Letter splits instead cost 3,035/3,240/3,716 extra bytes versus geometric. All task and leave-one-split rows are in `analysis/`.

Release-integral decomposition. With common endpoint H and atom bytes b_j, $I_p = \sum_j b_j(H - \tau_j^p)_+$. Set $\tau_j = H$ for an unreleased atom. Partition job–atom keys into common releases $J_\cap$, fresh-only J_f, and compiled-only J_c. Subtraction gives

$$I_f - I_c = \sum_{j \in J_\cap} b_j(\tau_j^c - \tau_j^f) + \sum_{j \in J_f} b_j(H - \tau_j^f) - \sum_{j \in J_c} b_j(H - \tau_j^c). \tag{38}$$

The identity follows per atom, allowing negative common-event terms for earlier compiled releases. At $H = 1200$, Letter loses 128,880 byte–sweeps through delay. CIFAR loses 2,812,400 through delay and 726,800 through missed atom-5 releases at split 401 (fresh sweeps 1090/1080): 79.46%/20.54% of its 3,539,200 loss.

First positive hint versus authorization. Replay matches all 20 saved authorization clocks bitwise. All released episodes satisfy Proposition B.4 at $h = 10$. For a common event,

$$\tau_j^c - \tau_j^f = (k_{*,j} - \tau_j^f) + (\tau_j^c - k_{*,j}). \tag{39}$$

Keep the first term signed. The second is pure scanning since the first triggered check succeeds. Split 402 (Table S2) assigns $26/600 = 4.33\%$ to scanning: before-hint/scan losses are 1,813,840/82,160 byte–sweeps. Over all common events they are 94,512/34,368 (Letter) and 2,663,880/148,520 (CIFAR), separate from fresh-only losses.

Table S2: Split 402: first positive replayed hint k_*, next scan and actual outward authorization. EC/Eu mean entropy-core/Euclidean. All costs were compiled at sweep 10.

Kernel	Atom	Fresh release	k_*	Authorization	Scan wait	Lower bound
EC	7	840	974	980	6	.014789
Eu	7	840	971	980	9	.014746
EC	6	940	1098	1100	2	.016187
Eu	6	940	1091	1100	9	.016460

Initial budgets .07460/.07404 tighten to .05486–.05468 at authorization; unreleased Letter 301 ends at hints $-.27604/-.19074$. Missing episode contrasts and untriggered fresh optima prevent assigning all pre-hint waiting to stale budgets. Only actual outward checks authorize release.

F.3 Continuation, restoration, and mechanism coverage

Table S6's natural and additional-Letter records replay 163,979 and 97,927 own-compact suffix trials in twelve and six workers. Tests cover reactivation, false hints, failed authorization/factors/gates, rollback, abort, twelve identity changes and 24 codec cases. Additional Letter matches 28,759 trials per arm, with four releases in two jobs, candidate/logit errors 6.18×10^{-13}/5.51×10^{-14} and holdout CE difference $\le 6.67 \times 10^{-16}$.

Three heavy-ball seeds have outward ball margin > 1.4562, 342 matching decisions, nine restores and error $\le 4.45 \times 10^{-15}$. Every Gram-component omission changes the first candidate. Complete saving is 70.31% for this prescribed-dormant construction; full settings are in the diagnostic ledger.

Conditional coverage. The three compact policies match 145,617 branch comparisons; compiled contributes 48,539. Factor inputs use preceding completed energy; gates use $v^+ = (\kappa(1 - \mu/M))^2 v$ in original operation order. All 14,382 nonrelease end-sweep values match bitwise; release adds energy. Every sweep commits. Positive energy occurs in 32/72 records (6 Letter, 26 CIFAR).

NZ rejections split into 1,390 factor/5,727 gate trials. Unconditioned totals include 7,179/12,560 rejections and 12,334 accepted projected factors. Natural full/compact guard differences and additional-Letter energy masks are unavailable. Structural tests start with permission; six dimension conditions overlap the original spectrum design.

Table S3: Nonzero-energy (NZ) coverage. Projections count NZ factor trials; rejections count both blocks. Structural rows are factor-only; counts are not independent replicates.

Execution	Total	NZ factor	NZ gate	NZ projected	NZ rejected
Letter compiled	26,939	2,780	1,909	2,780	1,909
CIFAR compiled	21,600	5,200	10,408	4,421	5,208
Spectral modes	1,559	1,559	–	1,442	791
Fixed-mode dimension	2,316	2,316	–	2,169	1,164

Stored-trace read erasure. For each nonzero-energy factor trial, recover its input v_t from the preceding completed state and set $\widehat{f}_t = F_{\text{previous}} - .04\sum_j \gamma_j$. The archived residual, step and projection determine

$$\widehat{g}_t = \widehat{\mathcal{A}}_t - 10^{-12}\max(1, |\widehat{f}_t|), \qquad \widehat{g}_t^{\text{erase}} = \widehat{g}_t - \tfrac{1}{2}[\mu - .5M_t][\kappa_t(1-\mu/M_t)-1]^2 v_t. \tag{40}$$

All 7,980 reconstructed decisions and stored-dyadic rational classifications match. Minimum flipping-pair distance is 1.096×10^{-9}; both tolerance scales are one, so Appendix C supplies the tolerance witness. Letter has 1,390 accepted/1,390 rejected trials; CIFAR has 5,200 accepted. Every $D_v < 0$. Accepted margin/correction minima (medians) are .005419 (3.27674) and 3.06674 (19.69750); 100 Letter acceptances satisfy (29). The CSV retains zero-release split 301. This algebraic counterfactual may overlap the executed intervention and is not pooled with it or treated as a full/compact guard pair.

F.4 Strong quality checks and their negative results

Each comparator selects by own-validation CE, breaking ties by fewer steps, lower learning rate/ridge. Same-width rates are $\{1, 4, 10\}$; dense rates $\{1, 5\}$, ridges $\{0, 10^{-4}, 10^{-3}\}$, zero initialization and unpenalized bias. All 27 paths per cohort complete. CIFAR selectable steps are $\{0, 150, 1500, 6000, 12000, 24000, 36000, 48000\}$ (width eight) and $\{1500, 6000, 12000, 18000, 24000\}$ (dense); 250-step diagnostics are excluded. Selected warm pairs are $(4, 36000), (10, 12000), (1, 12000)$; dense uses rate 5/ridge 0 at 24,000/24,000/12,000 steps, two at an improving boundary. Tail–warm/dense CE gaps .031772/.056798 exceed .03; the dense accuracy gap 1.33 points exceeds one. New-class/source checks pass, but the conjunction fails. Exposed-test tail/warm/dense CE .175152/.167464/.137011 is exploratory.

Additional Letter continues its 600-step warm model to $\{600, 1500, 3000, 6000, 12000, 24000, 48000\}$; dense selects from $\{1500, 6000, 12000, 24000\}$. All paths select before their maximum steps. Table S4's same-width rule passes; the strict rule fails CE, overall and new-class accuracy, while passing the source comparison. The tail–dense validation CE gap is .116223. After locking selection, shared-holdout tail/same-width/dense CE is 1.073343/1.123557/.961941 and accuracy 73.68/71.62/74.53%; no holdout reselection occurs.

Table S4: Additional-Letter rules (conjunctions). T, S, D, W_0, W_{600}: tail, selected same-width/dense, source/warm. A: accuracy fraction.

Metric	Same-width	Separate strict joint
CE	$\mathcal{L}_T \le \mathcal{L}_S + .05$	$\mathcal{L}_T \le \min(\mathcal{L}_S, \mathcal{L}_D) + .03$
Accuracy	$A_T \ge A_S - .01$	$A_T \ge \max(A_S, A_D) - .01$
New class	$A_T^{\text{new}} \ge A_{W_{600}}^{\text{new}}$	$A_T^{\text{new}} \ge A_S^{\text{new}} - .01$
Source comparison	No additional conjunct	$A_T \ge A_{W_0}$
Observed status	Pass	Fail

G Complete resources and structural validation

G.1 Timing and complete-record accounting

Serial single-thread BLAS timing uses 48 warmups and 144 measurements; retained runs match scientific terminal states. It includes certification, slicing and synchronization, but excludes feature ex-

traction, warm training, full observers, pause copying and I/O. Permission time falls 71.0105→3.2660 s (Letter) and 50.7875→2.3371 s (CIFAR). The diagnostic ledger preserves per-job times, environment and superseded overlapped runs.

Complete records include factors/anchors, gates, summary, base, scale, budgets/steps, previous objective, stopping/permission/cache controls and framing; features are external and identity-bound. Full-process peak RSS includes data, interpreter, temporaries and I/O: compiled/full ratios are 1.1537 (1.0903–1.2644) and 1.0718 (1.0393–1.0982). Partial array-byte–seconds ratios 1.2697/1.4227 exclude metadata/temporaries.

Table S5: Signed complete-record savings in bytes. Early: 40/160; late: 900/1100; uniform: all six pauses. These are diagnostic, not measured arrival distributions.

Data	Pauses	Raw	zlib-0	zlib-6	zlib-9
Letter	Early	$-28,724$	$-28,724$	$-5,514$	$-5,515$
	Uniform	$-65,441$	$-76,893$	$-6,800$	$-6,830$
	Late	$-6,497$	$-17,949$	$+5,013$	$+4,986$
CIFAR	Early	$+54,945$	$+10,705$	$+34,929$	$+34,938$
	Uniform	$+241,993$	$+71,353$	$+141,706$	$+141,664$
	Late	$+119,473$	$+43,633$	$+65,746$	$+65,713$

Section 6's cost improves for uniform residence h iff $a_{\mathrm{mem}} h \Delta B > b_{\mathrm{time}} \Delta T$. For $\Delta B, \Delta T > 0$, divide by $a_{\mathrm{mem}} \Delta B$ to obtain the break-even residence.

Device and pipeline diagnostics. The diagnostic ledger retains three-layer, A800 and ordinary-head DINOv2 results. Persistent bytes decrease without lower GPU peaks; these microbenchmarks do not measure natural release or suffix restoration.

G.2 Fixed-mode dimension and record restoration

Spectral tests use $K = 1, 2, 4, 8$, seeds 11/29/47, 120 normalized six-dimensional features and three-class teacher labels. Two zero-gated atoms per mode have independent ball permission, verified with feasibility by dyadic checks; a retained 3×6 head supplies task gradients. `artifacts/diagnostic_ledger.md` lists all settings of the fixed 64-sweep reference.

Fixed-mode tests preserve task and mode energies at 18/180/1,800 coordinates: six tasks ($K = 1, 4$, three seeds), eighteen conditions including six spectral settings. Candidate errors are $\leq 3.33 \times 10^{-16}$/$2.22 \times 10^{-16}$ (summary/lazy). At dimension 180, full/lazy/summary bytes are 19,337/19,908/10,161 ($K = 1$), 45,500/45,670/10,339 ($K = 4$), with the same six records per row as Table 3.

Lazy retains basis, energies and multipliers. Complete NumPy ZIP/canonical-JSON plus zlib-6 records include controls, permission and identity metadata; external data/modes are bound and audit shadows excluded. Each representation passes 18 injected rollbacks. Billed bytes round-trip exactly before replay; Table S6 separates coverage. Hashes check consistency, not authenticity.

Table S6: Actual worker loaders and record/scope coverage. `REPRODUCE.md` specifies bound environments and commands.

Record	Entry point	Positive / negative coverage	Billed in
Natural compiled	`decode_checkpoint` + validation	72 + 36 restores; original scope controls	Tables 4, S5
Spectral capsule	`load_compact`	24 restores; 96 format-specific mutations	Separate format
Fixed-mode compact	`load_benchmark`	36 restores / 2,265 trials; same-entry 96 mutations	Table 3
Fixed-mode lazy	`load_lazy`	36 restores; no mutation claim	Table 3