ONLINE COMPRESSION OF CONTROLLED MEMORY BY AGE-GRADED REBUCKETING

Anonymous authors
Paper under double-blind review

ABSTRACT

We construct finite-memory vector approximations of controlled memories in which each past input evolves through age-dependent linear contractions. Future controls arrive online and can induce noncommuting matrix actions, so discarding a past input must preserve its response to controls that have not yet been observed. Our algorithms maintain consecutive birth intervals on an increasing age grid and merge adjacent intervals without recovering discarded writes. Under a capped log-age modulus, one vector per interval gives uniform output error ϵ with $O(\epsilon^{-1})$ vector states. With additional second-order age smoothness, positive redistribution between two endpoints improves this sufficient budget to $O(\epsilon^{-1/2})$, counting both endpoint vectors. The bounds hold for every bounded signed-vector write stream and common external control path, uniformly over elapsed time and contraction lifetime when regularity constants are fixed. The analysis couples decay-weighted local error with reciprocal-width packing. For the endpoint representation, a mean-preserving virtual birth coordinate cancels the first variation of the complete chronological matrix product. Small operator-level comparisons use dual bounds on conditional worst-write error to examine the resulting finite-budget approximations without model training.

1 INTRODUCTION

A memory may need to retain how an old input will respond to a control that arrives later. In an age-dependent recurrent memory, the same current control acts differently on contributions of different ages. Matrix-valued gates also mix channels, so the order of subsequent controls matters. Summing contributions too early loses their age information, while retaining every contribution requires storage that grows with the stream.

We ask how many vector states suffice to approximate this memory for *every* bounded write stream and common external control path. We study a class with contraction $e^{-\mu}$ in a common norm and regularity along the age coordinate. The target is a uniform error bound whose sufficient state count remains bounded as both stream length and lifetime μ^{-1} grow. This formulation includes scalar survival memory and noncommuting channel evolution. It also separates approximation of an existing memory operator from learning a finite-dimensional sequence model.

The algorithm uses age regularity to decide how much birth-time information to retain. Young intervals remain fine enough to control rapidly changing gates; older intervals can grow as their surviving mass decreases. We choose an increasing width profile for which the weighted approximation error and the reciprocal-width count are both integrable. This produces a causal calendar: records merge as they age, and no later step splits a record or reconstructs its original writes. One vector per record yields $O(\epsilon^{-1})$ sufficient states. A second representation retains two endpoint vectors and redistributes them positively when intervals merge. Under additional age smoothness, it yields $O(\epsilon^{-1/2})$ total vector states.

Interval coalescence follows a long streaming tradition (Datar et al., 2002; Cohen & Strauss, 2006), and positive mean-preserving splitting is the intrinsic-stationarity principle of dual quantization (Pagès & Wilbertz, 2012a;b). Our construction uses these principles to address a joint requirement: repeated online compression of signed-vector memory under unseen, noncommuting future controls, with an explicit state budget uniform in time and lifetime. The second-order proof must control the

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.

full ordered product after correlated redistributions; a local interpolation estimate alone does not provide that guarantee.

The contributions are:

- A merge-only age calendar and a one-vector representation with an explicit $O(\epsilon^{-1})$ sufficient budget under a first-order operator modulus.
- A positive two-endpoint representation with a complete ordered-product error analysis and an $O(\epsilon^{-1/2})$ budget under C^2 age regularity.
- A conditional induced-error evaluation for matrix memories, together with small comparisons at controlled vector budgets and interface guarantees for regular readouts and contractive feedback.

Both rates count vectors in the payload space. They are sufficient bounds for the stated classes; the second uses stronger regularity.

2 Controlled memory and access model

Let V be a nonzero finite-dimensional real normed space of dimension d, with the same norm used throughout, and let $\mathcal{C}$ be a nonempty control set. Fix $0 < \mu \le 1$, $q = e^{-\mu}$, and $a_0 > 0$. At tick $t \ge 0$, the memory first gates all old contributions, then inserts $b_t = (1-q)u_t$, where $\|u_t\| \le 1$. It starts empty. For linear gates $G(a,c) : V \to V$ defined at every real age $a \ge 0$, the exact output is

$$P_{j,t} = G(t-1-j, c_t)\cdots G(0, c_{j+1}), \quad P_{t,t} = I, \qquad y_t = \sum_{j=0}^{t} P_{j,t} b_j. \tag{1}$$

The product retains its chronological order. Exact and compressed memories receive the same u_t and external c_t, with no temporal regularity requirement.

Assumption 1 (Common contraction and first-order age modulus)**.** For all $a, a' \ge 0$ and $c \in \mathcal{C}$, with a fixed $\Lambda \ge 0$,

$$\|G(a,c)\|_{\mathrm{op}} \le q, \qquad \|G(a,c) - G(a',c)\|_{\mathrm{op}} \le q\mu\Lambda \min\left\{1, \left|\log\frac{a+a_0}{a'+a_0}\right|\right\}. \tag{2}$$

Assumption 2 (Second-order age regularity)**.** The gates are twice continuously differentiable in age and, uniformly in $a, a' \ge 0$ and $c \in \mathcal{C}$, satisfy

$$\begin{aligned} \|G(a,c)\|_{\mathrm{op}} &\le q, \qquad \|G(a,c) - G(a',c)\|_{\mathrm{op}} \le q\mu\kappa, \\ \|\partial_a G(a,c)\|_{\mathrm{op}} &\le \frac{q\mu L_1}{a+a_0}, \qquad \|\partial_a^2 G(a,c)\|_{\mathrm{op}} \le \frac{q\mu L_2}{(a+a_0)^2}, \end{aligned} \tag{3}$$

where $\kappa, L_1, L_2 \ge 0$ are finite constants.

Constants are fixed independently of ϵ, μ, t and the control path. Assumption 2 implies Assumption 1 with $\Lambda = \max\{\kappa, L_1\}$ by integrating the first derivative and using the oscillation bound. Both classes allow singular matrices and noncommuting actions. A common induced-norm contraction, rather than individual spectral radii, controls products of changing gates.

What the online algorithm can access. At tick t, the compressor may evaluate $G(a, c_t)v$ at its retained representative ages. It knows μ and regularity constants, but stores neither the discarded writes nor the past control sequence. It does not query a past birth's complete propagator $P_{j,t}$. Gate parameters and an external controller are separate from the compressed memory state. We count retained vectors, record metadata and gate applications separately.

Scale and examples. Contraction implies $\|P_{j,t}b_j\| \le (1-q)q^{t-j}$ and $\|y_t\| \le 1$. This normalization remains nontrivial: for the covered age-independent gate $G = qI$ and constant unit write $u_t = v$, $y_t = (1-q^{t+1})v$, whose limiting norm is one for every $\mu > 0$. Scalar gates $G = e^{-\lambda(a,c)}I$ embed

whenever $\mu \le \lambda \le (1+\kappa)\mu$ and λ is log-age Lipschitz. A matrix example uses the Euclidean norm and

$$G(a,c) = q[I - \alpha f(a,c)(I - P_c)], \quad f(a,c) = \tfrac{1}{2}[1 + \sin(\log(a+a_0) + 0.7c)], \quad \alpha = \min\{1, \kappa\mu\}, \tag{4}$$

where $c \in \{0,1\}$, $P_0 = \mathrm{diag}(1,0)$ and $P_1 = \frac{1}{2}\left(\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right)$. Orthogonal-projection contraction and direct differentiation give $L_1 = \kappa/2$, $L_2 = \kappa$ and $\Lambda = \kappa$. For $\kappa > 0$ and nonzero phase factors, $[P_0, P_1] \neq 0$ makes this an age-dependent, noncommuting family. Full verifications appear in Section D.

3 Age-graded coalescence and endpoint rebucketing

A record represents consecutive integer birth times $[L, R]$. After tick t, its youngest age is $a = t - R$ and its width is $w = R - L + 1$. Figure 1 shows the update and record representations. We use either of two increasing profiles, with the parameter η specified by the corresponding theorem:

$$W_1(a) = \frac{\eta}{\mu}\sqrt{\mu a + \eta^2}(1 + \mu a), \qquad W_2(a) = \frac{\eta}{\mu}(\mu a + \eta^3)^{2/3}(1 + \mu a). \tag{5}$$

An adjacent pair may merge when its union has span $w - 1 \le W_p(a)$. Intervals can have unequal widths. Increasing W_p ensures that a valid interval remains valid as it ages, so the calendar needs only merges.

One-vector representation. A record stores $v_I \in V$. Its gate is evaluated at a representative inside its current old-age interval, and merging adds record vectors. A singleton uses its exact age. Our implementation uses the log-midpoint $\beta + a_0 = \sqrt{(a + a_0)(a + w - 1 + a_0)}$; the theorem permits any representative inside the interval before each gate.

Two-endpoint representation. A record stores $v_L, v_R \in V$, each gated at its own endpoint's age. On merging into $[L', R']$, every old endpoint vector v_x is redistributed by

$$v_{L'} \mathrel{+}= \frac{R' - x}{R' - L'} v_x, \qquad v_{R'} \mathrel{+}= \frac{x - L'}{R' - L'} v_x. \tag{6}$$

The union is nondegenerate, so the denominator is positive. The weights are nonnegative, sum to one, and preserve coordinate x in their weighted mean. A singleton uses one vector and a zero second entry. This representation is updated directly; it does not claim to maintain exact moments of the uncompressed population after approximate gating.

Algorithm 1 One tick of graded memory ($p = 1$ or 2; initially empty)

Require: Tick t, shared c_t, $b_t = (1-q)u_t$, profile W_p, cutoff H
1: Gate old representative vectors at their old ages using $G(\cdot, c_t)$
2: Advance ages and insert singleton $[t, t]$ carrying b_t
3: Set the current interval to the youngest record
4: **for** each next record in order from youngest to oldest **do**
5: **if** the union has span $s - 1 \le W_p(a)$ **then**
6: Merge into the current interval by addition ($p = 1$) or Equation (6) ($p = 2$)
7: **else**
8: Close the current interval and set the next record as current
9: **end if**
10: **end for**
11: Delete every whole record with youngest age at least H
12: **return** the sum of all retained representative vectors

Calendar and storage. Zero writes still create slots. A rejected boundary stays rejected when its older side grows: the union width increases and its youngest age does not change. Thus one greedy sweep produces a maximal partition in $O(K)$ calendar work. A final retained interval may straddle H; it is not partially deleted. After deletion there are K records, containing K or at most $2K$ vectors. Insertion adds one temporary record, and merging needs $O(1)$ work vectors. Each tick uses at most one or two gate applications per old record.

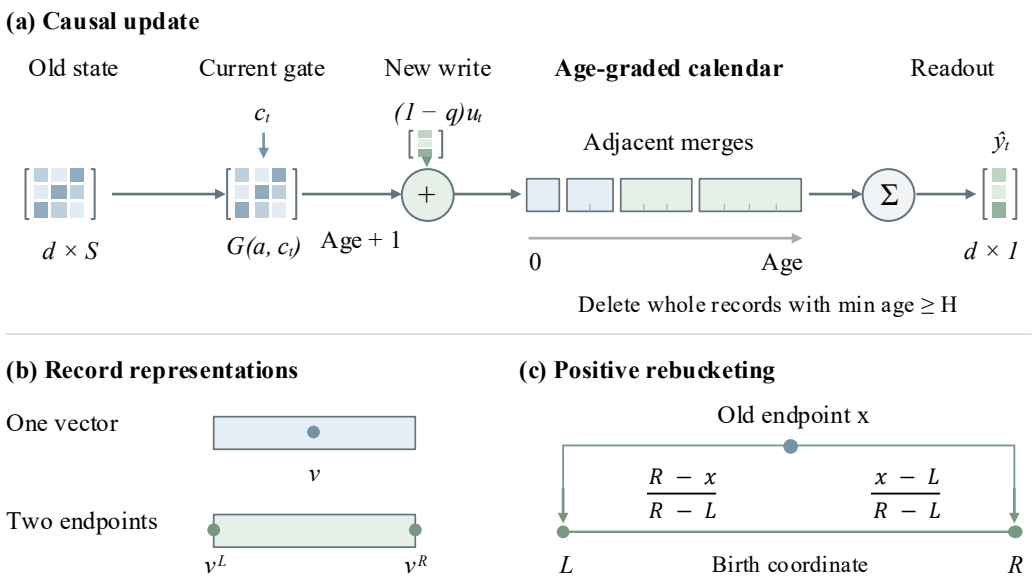

Figure 1: The online memory receives a shared current control and a signed-vector write. Retained representatives are gated before the new birth is inserted. An increasing age profile permits adjacent coalescence, followed by whole-record deletion and a sum readout. One-vector coalescence adds payloads; the two-endpoint variant redistributes them with positive weights that preserve the birth coordinate's mean. In the figure, S counts total retained vector states; panel (c) uses birth coordinates L, x, R. The interval calendar stores no discarded writes or virtual paths.

4 Uniform error and state budgets

The profiles in Equation (5) balance two costs: the gate error within an interval and the number of intervals needed to cover retained ages. The following bounds hold in exact arithmetic, for every admissible common write/control stream and all $t \geq 0$.

Theorem 1 (One-vector graded compression)**.** *Under Assumption 1, let* $0 < \epsilon < 1$ *and run the one-vector variant with*

$$\eta = \frac{\epsilon}{24(1+\Lambda)}, \qquad H = \left\lceil \frac{\log(3/\epsilon)}{\mu} \right\rceil. \tag{7}$$

Its output error and number K_t *of retained records after deletion obey*

$$\sup_{t \geq 0} \|\widehat{y}_t - y_t\| \leq 14\Lambda\eta + q^H < \epsilon, \qquad K_t \leq 2 + \frac{2\pi}{\eta \log 2}. \tag{8}$$

Hence $O_\Lambda(\epsilon^{-1})$ *vector states suffice uniformly in* $0 < \mu \leq 1$*, time and the common control path.*

Decay-weighted local error. For a record I of youngest age a, let $M_I = (1-q)\sum_{i \in I} q^i$ be its exact norm-mass bound. Set $r_I = 0$ for singletons, and $r_I = \min\{1, W_1(a)/a\}$ otherwise, with $r_I = 1$ at $a = 0$. The width invariant implies a gate defect at most $q\mu\Lambda r_I$. The central local estimate is

$$\sum_I M_I r_I \leq 14\eta. \tag{9}$$

To see why it is uniform in lifetime, write $x = \mu a$. For intermediate ages, the radius scales as $\eta x^{-1/2}$; at old ages it is bounded by a multiple of $\eta x^{1/2}$. Both are integrable against the geometric survival mass. The very young part needs a discrete argument: a nonsingleton there can exist only when $\mu \leq 2\eta^2$, and its total mass is at most $5\eta^2$.

Let R_t be the sum, over retained records, of the norm of the difference between the compressed vector and its exact atom sum. Contraction and Equation (9) give

$$R_t \leq qR_{t-1} + 14q\mu\Lambda\eta, \qquad R_t \leq 14\Lambda\eta, \tag{10}$$

because $q\mu/(1-q) \le 1$. Addition and deletion do not expand this error. Every deleted birth has age at least H, giving the remaining tail bound $(1-q)\sum_{i\ge H} q^i = q^H$.

Reciprocal-width packing. With Equation (7), $W_1'(a) < 1$ on $[0, H]$. An ineligible adjacent pair with union width s satisfies $s > W_1(a)$, and therefore

$$\int_a^{a+s} \frac{dr}{W_1(r)} \ge \log\left(1 + \frac{s}{W_1(a)}\right) > \log 2. \tag{11}$$

Pairing records while excluding the final, potentially straddling interval gives $(K_t - 2)\log 2/2 \le \int_0^H da/W_1(a) \le \pi/\eta$. The integrable profile replaces a level-count argument. The full discrete proof is in Section A.

Theorem 2 (Two-endpoint graded compression)**.** *Under Assumption 2, let* $0 < \epsilon < 1$ *and set*

$$D = 256 + 6\kappa + 324L_2 + 1080L_1^2, \qquad \eta = \sqrt{\frac{\epsilon}{3D}}, \qquad H = \left\lceil \frac{\log(3/\epsilon)}{\mu} \right\rceil. \tag{12}$$

The two-endpoint variant satisfies

$$\begin{aligned} \sup_{t\ge 0} \|\widehat{y}_t - y_t\| &\le 6\kappa\eta^3 + (324L_2 + 1080L_1^2)\eta^2 + q^H < \epsilon, \\ 2K_t &\le 4 + \frac{8\pi}{\sqrt{3}\,\eta \log 2}. \end{aligned} \tag{13}$$

Thus $O_{\kappa,L_1,L_2}(\epsilon^{-1/2})$ *total vector states suffice with the same uniformity. The constants in Equation* (12) *in fact give error less than* $2\epsilon/3$*.*

Why endpoint redistribution improves the order. Positive splitting ensures that the sum of endpoint norms is at most one. It also admits an exact proof representation: a birth j follows a virtual coordinate X_s that moves to new endpoints with the weights in Equation (6). Conditional mean preservation implies $\mathbb{E}X_s = j$. The algorithm itself is deterministic and stores no such paths. After replacing a short true-age prefix by exact gates, interpolate the remaining factors as

$$F_X(\theta) = \prod_{s=t}^{j+1} G(s - 1 - j - \theta d_s, c_s), \qquad d_s = X_s - j, \tag{14}$$

The product has $s = t$ on the left and $s = j + 1$ on the right, as in Equation (1); $d_s = 0$ on the replaced prefix. At $\theta = 0$, all undifferentiated factors are deterministic. Each term of $F_X'(0)$ is linear in one zero-mean d_s, so $\mathbb{E}F_X'(0) = 0$. Correlations between different d_s are allowed; no matrix factors are reordered.

The remaining analysis controls both the individual curvature terms and cross terms between differentiated gates. For mature true age i, the relative displacement is bounded by $r_i = 6\eta[(\mu i)^{-1/3} + 1 + \mu i]$. Weighted geometric summation gives

$$(1-q)\sum_{n\ge 0} q^n \mu \sum_{i=1}^{n-1} r_i^2 \le 648\eta^2, \quad (1-q)\sum_{n\ge 0} q^n \mu^2 \left(\sum_{i=1}^{n-1} r_i\right)^2 \le 2160\eta^2. \tag{15}$$

Taylor's integral remainder produces the coefficients in Equation (13). The young prefix costs at most $6\kappa\eta^3$: a displaced young atom must belong to a nonsingleton, forcing $\mu < 3\eta^3$. Finally, $\int_0^H da/W_2(a) \le 2\pi/(\sqrt{3}\eta)$ gives the count by the same packing argument. Section B proves each step, including the negative geometric moment estimate and all integer cases.

Readouts and feedback. Section C extends the one-vector result to a normalized concave age modulus ω, yielding the last row of Table 1, with c fixed by its regularity bound. A common L_F-Lipschitz postprocessing map transfers sum error to at most $L_F\epsilon$. Age-selective linear readouts require their own uniform norm and age-modulus bounds. Feedback is covered by a separate small-gain condition: a common-stream compression error δ becomes at most $\delta/(1-\rho)$ when the corresponding incremental loop gain $\rho < 1$. These extensions specify the additional interface assumptions explicitly. For a frozen memory module, bounded features $u_t = \phi(x_t)/\max\{1, \|\phi(x_t)\|\}$ and an input-derived control $c_t = g(x_t)$ supply a common stream to both branches. Sharing ϕ, g and the prediction head preserves the hypotheses of the replacement bound; changing the controller through each branch's own state instead requires the feedback condition.

Table 1: Regularity and sufficient state budgets for the same controlled-memory problem. Bounds are uniform in time and lifetime with fixed displayed constants. State means a vector in V; metadata and gate cost are separate.

Age dependence	Retained representation	Sufficient vectors
Age independent	One exact aggregate	1
Capped log-age modulus	One vector per interval	$O_\Lambda(\epsilon^{-1})$
C^2 bounds in Assumption 2	Two endpoint vectors	$O_{\kappa,L_1,L_2}(\epsilon^{-1/2})$
Concave modulus ω	One vector per interval	$O(1/\omega^{-1}(c\epsilon))$

5 Relation to streaming, quadrature and recurrent memory

Coalescing stream summaries and population cohorts. Exponential histograms retain a small set of chronological buckets for sliding windows (Datar et al., 2002). Time-decaying aggregates extend the question to weighted histories (Cohen & Strauss, 2006; Kopelowitz & Porat, 2008). These works establish interval aggregation and age-dependent allocation as general streaming ideas. Population methods offer a closely related numerical view: the merging EBT construction of Aye & Carlsson (2020) studies convergence with a cohort count bounded independently of the number of time steps. Adaptive history integration for growth and remodeling also coarsens stored intervals and reports plateauing history sizes (Gebauer et al., 2024). Thus neither merging old records nor bounded cohort counts alone identifies our contribution. Our bounds address a specified induced output norm for signed vector writes and arbitrary common future controls, while tracking precision and contraction lifetime explicitly. Table 2 compares the closest guarantee types.

Intrinsic stationarity and higher-order weak approximation. Pagès & Wilbertz (2012a) develop positive random splitting that preserves conditional means. Their Proposition 9 gives a second-order weak quadrature estimate, and scalar-coordinate quantization already supplies the familiar quadratic error dependence on grid resolution. Their sequential construction also propagates finite-grid weights and bounds accumulated weak error for random walks (Pagès & Wilbertz, 2012b). The weights in Equation (6) implement that established stationarity principle on the one-dimensional birth coordinate, regardless of the dimension of the vector payload. Our analysis embeds it in repeated, causal interval merges and controls a correlated virtual path inside a noncommuting propagator. The weighted derivative sums in Equation (15) and the graded packing bound together give the uniform state budget; the interpolation weights or the exponent $1/2$ by themselves are not a novelty claim.

Evolutionary compression of Volterra operators. Dölz et al. (2021) provide an evolutionary, oblivious $\mathcal{H}^2$ method for asymptotically smooth Volterra kernels, using $O(\log N)$ active history at fixed discretization orders over N steps. Their method does not require a prescribed final horizon, and its output error estimate includes a constant independent of the time interval. It requires the current kernel-block coefficients. In the present memory, those coefficients would involve ordered age-dependent gate products from past controls. A reduction must therefore show how to maintain them from the current gate oracle with the same bounded-history resources, and derive the required kernel smoothness and uniform norms. We do not assume this coefficient access in Section 2; nor do we claim that evolutionary Volterra methods require advance knowledge of future inputs.

Projection memories and selective state-space models. HiPPO maintains recurrent polynomial projections of a signal (Gu et al., 2020). HiPPO Zoo develops explicit readout, salience and multiscale memory mechanisms (Goffinet et al., 2026); SHiPPO transports polynomial projections under realized controls, including noncommuting matrix actions (Mizuguchi & Kim, 2026). These results already make controlled, noncommuting recurrent memories a meaningful model class. Our task is to approximate a given age-dependent gate family for every bounded write stream, with an explicit precision-to-state budget. It does not require the signal to be smooth or the chosen gates to close on a finite polynomial basis. The readout and feedback results in Section C serve as compatibility guarantees for this compressor; they do not introduce learned readouts or selective memory as new mechanisms.

Table 2: Closest guarantee types. Rates refer to each work's own assumptions and state units, so the rows are not an equal-assumption lower-bound comparison. N is the number of time steps.

Construction	Established guarantee	Relation to the present bound
Merging EBT (Aye & Carlsson, 2020)	Convergence and a time-step-independent cohort count	Population-specific cohort evolution
Dual quantization (Pagès & Wilbertz, 2012a;b)	Mean preservation; scalar second-order weak accuracy	Source of endpoint stationarity
Volterra $\mathcal{H}^2$ (Dölz et al., 2021)	$O(\log N)$ active history at fixed orders	Kernel-coefficient access and smoothness
Graded controlled memory	$O(\epsilon^{-1})$ or $O(\epsilon^{-1/2})$ vectors	Explicit time- and lifetime-uniform budget

6 Small operator-level validation

We test finite-budget fidelity on a two-dimensional, noncommuting projection family with $\mu = 1/64$, $\kappa = 1/2$, and $a_0 = 2$; its valid constants are $\Lambda = \kappa$, $L_1 = \kappa/2$, and $L_2 = \kappa$. The study comprises 24 comparisons: four methods, three total-vector budgets, and two predetermined control paths, each of length $T = 192$. The methods are one-vector graded coalescence (Graded-1), two-endpoint graded rebucketing (Graded-2), implemented uniform blocks, and a recent-history buffer. A count-only enumeration selects parameters before evaluating either path. All four methods attain peak counts of 12, 24, and 48 retained vectors; Graded-2 counts both endpoints, including a singleton's zero entry. These empirical parameters are distinct from the sufficient theorem budgets. For the three cohort methods, $H = 193 > T$ isolates aggregation error; recent history still discards its own old tail.

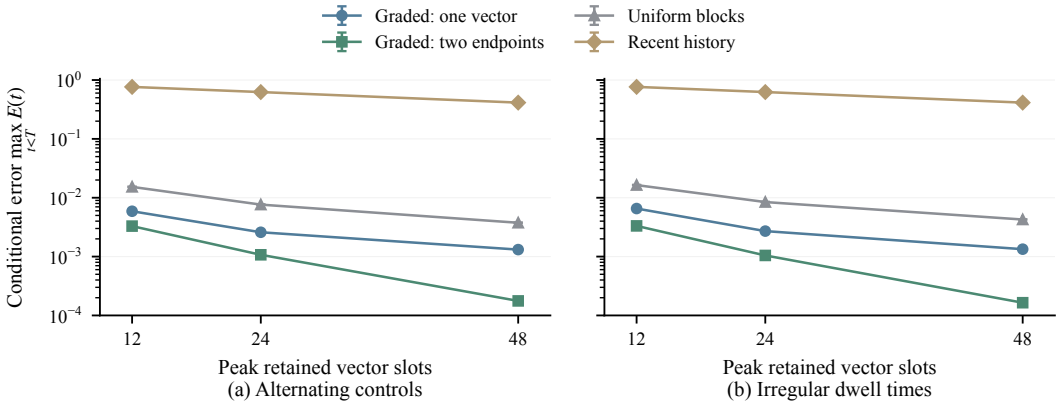

Figure 2: Finite-budget conditional error under alternating controls and irregular dwell times. Markers are lower bounds on $\max_{t<T} E(t)$; capped bars extend to the angular-grid upper bounds and are smaller than the markers at this scale. The horizontal axis counts all retained vector slots. The cohort methods retain the full 192-slot history through aggregation; the recent-history baseline truncates it. Curves connect the three measured settings and do not estimate asymptotic rates.

Conditional worst-write error. For each frozen control path and time, the offline checker maintains the exact and approximate per-birth matrices P_j and $\widehat{P}_j$. With $C_j = \widehat{P}_j - P_j$, the correct matrix metric is

$$E(t) = (1-q) \max_{\|\ell\|_2=1} \sum_{j=0}^{t} \|C_j^\top \ell\|_2. \tag{16}$$

A 512-direction circular grid supplies a lower bound and a Lipschitz discretization remainder supplies an upper bound. The resulting intervals for $\max_{t<T} E(t)$ have relative width below 0.65% in these runs. All calculations use float64; the angular remainder does not enclose arithmetic roundoff. In

each comparison, an actual replay of the writes induced by the maximizing grid direction attains its projected lower bound, with discrepancy below 7×10^{-16} against an independent full-history vector reference. The per-birth matrices are used only by this offline checker.

Finite-budget fidelity and resources. For both paths at all three budgets, the Graded-2 upper endpoint is below the Graded-1 and uniform-block lower endpoints (Figure 2). At 24 vectors, the largest upper endpoints across the two paths are 1.0792×10^{-3}, 2.7268×10^{-3}, and 8.4955×10^{-3}, respectively. This establishes a separation for the tested fixed-context operators. Table 5 in Appendix E reports the corresponding reference-implementation costs; gate evaluations and vector payloads are separated from Python allocation and elapsed time. Two additional checks use the closed-form theorem parameters at $\epsilon = 0.2$, with genuine whole-record deletion; their measured intervals lie below the stated envelopes (Appendix E).

Preserving control order. After 64 common writes $u_t = (1, 0)$, we stop writing and apply $0^{32}1^{32}$ or $1^{32}0^{32}$. The exact endpoint responses differ by 2.4070×10^{-3} in Euclidean norm. All three cohort representations preserve this nonzero order effect (Figure 3); the recent buffer has discarded the warmup writes. We reuse the 24-vector calibration, so the shorter sequence has the distinct actual peaks shown in the figure. This is an operator-fidelity mechanism check; no model is trained in this study.

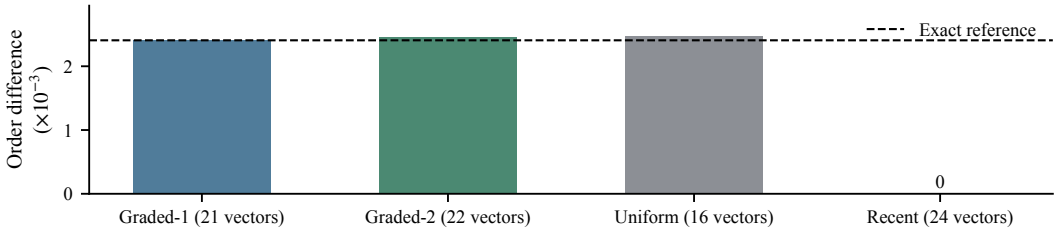

Figure 3: Equal control counts in different orders, after a shared warmup and with zero subsequent writes. Bars show $\|\widehat{y}_{0^{32}1^{32}} - \widehat{y}_{1^{32}0^{32}}\|_2$; the dashed line is the full-history response difference. Parentheses give actual vector peaks for this 128-tick check. Parameters are reused from the main comparison rather than reselected on this response.

7 Discussion and conclusion

Age regularity and contraction jointly determine a sufficient resolution for online memory. An increasing age profile supports irreversible coalescence, while the retained representation controls the approximation order. One-vector aggregation gives a first-order precision budget; mean-preserving endpoint redistribution gives a second-order budget under stronger smoothness. The resulting guarantees cover a class of controlled matrix memories with signed writes and arbitrarily ordered common controls. These are sufficient budgets; matching lower bounds remain open.

The explicit constants are conservative, and the small finite-budget studies measure a different question from asymptotic sufficiency. State counts exclude gate parameters and become real-coordinate counts after multiplication by $\dim V$. Metadata bits, floating-point error and total work across μ^{-1} ticks are separate resource questions. Extending the guarantees to gates without a common contraction norm or to feedback outside the stated small-gain regime requires additional stability analysis.

AI USE STATEMENT

Generative AI tools were used only for language polishing of the manuscript text, such as improving grammar, wording and readability. They were not used to develop the theoretical framework, formulate claims, write or check proofs, design or implement the algorithms and validation code, or interpret the results. Authors remain responsible for the final text, mathematics, code, citations and reported results.

ETHICS STATEMENT

The numerical studies use synthetic operator sequences and no personal or human-subject data. The work concerns approximation guarantees for a restricted memory class. Downstream applications require their own assessment of data, model behavior and deployment risks.

REPRODUCIBILITY STATEMENT

The main text states the operator assumptions, online access model, algorithms and explicit budgets. Appendices provide the complete proofs, operator-family calculations and numerical protocol. The accompanying source package includes portable CPU implementations, fixed configurations, saved outputs, figure and table generators, and the editable architecture figure. The numerical checks require no model training. Online memory accounting excludes the full-history reference used only for offline validation.

REFERENCES

Tin Nwe Aye and Linus Carlsson. Increasing Efficiency in the EBT Algorithm. In *Demography of Population Health, Aging and Health Expenditures*, The Springer Series on Demographic Methods and Population Analysis, pp. 289–317. Springer International Publishing, 2020. doi: 10.1007/978-3-030-44695-6_19. URL https://doi.org/10.1007/978-3-030-44695-6_19.

Edith Cohen and Martin J. Strauss. Maintaining time-decaying stream aggregates. *Journal of Algorithms*, 59(1):19–36, April 2006. ISSN 0196-6774. doi: 10.1016/j.jalgor.2005.01.006. URL http://dx.doi.org/10.1016/j.jalgor.2005.01.006.

Mayur Datar, Aristides Gionis, Piotr Indyk, and Rajeev Motwani. Maintaining Stream Statistics over Sliding Windows. *SIAM Journal on Computing*, 31(6):1794–1813, January 2002. ISSN 1095-7111. doi: 10.1137/s0097539701398363. URL http://dx.doi.org/10.1137/S0097539701398363.

J. Dölz, H. Egger, and V. Shashkov. A fast and oblivious matrix compression algorithm for Volterra integral operators. *Advances in Computational Mathematics*, 47(6):81, 2021. doi: 10.1007/s10444-021-09902-6. URL https://doi.org/10.1007/s10444-021-09902-6.

Amadeus M. Gebauer, Martin R. Pfaller, Jason M. Szafron, and Wolfgang A. Wall. Adaptive integration of history variables in constrained mixture models for organ-scale growth and remodeling. *International Journal for Numerical Methods in Biomedical Engineering*, 40(11), 2024. ISSN 2040-7947. doi: 10.1002/cnm.3869. URL http://dx.doi.org/10.1002/cnm.3869.

Jack Goffinet, Casey Hanks, and David E. Carlson. HiPPO Zoo: Explicit Memory Mechanisms for Interpretable State Space Models. In *International Conference on Machine Learning*, 2026. URL https://arxiv.org/abs/2602.21340.

Albert Gu, Tri Dao, Stefano Ermon, Atri Rudra, and Christopher Ré. HiPPO: Recurrent Memory with Optimal Polynomial Projections. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 1474–1487. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/102f0bb6efb3a6128a3c750dd16729be-Paper.pdf.

Tsvi Kopelowitz and Ely Porat. Improved Algorithms for Polynomial-Time Decay and Time-Decay with Additive Error. *Theory of Computing Systems*, 42(3):349–365, 2008. ISSN 1433-0490. doi: 10.1007/s00224-007-9031-8. URL http://dx.doi.org/10.1007/s00224-007-9031-8.

Tomoya Mizuguchi and Bum Jun Kim. SHiPPO: Recurrent Memory with Transported Polynomial Projections, 2026. URL https://arxiv.org/abs/2607.03055.

Gilles Pagès and Benedikt Wilbertz. Intrinsic stationarity for vector quantization: Foundation of dual quantization. *SIAM Journal on Numerical Analysis*, 50(2):747–780, January 2012a. ISSN 1095-7170. doi: 10.1137/110827041. URL http://dx.doi.org/10.1137/110827041.

Gilles Pagès and Benedikt Wilbertz. Dual quantization for random walks with application to credit derivatives. *The Journal of Computational Finance*, 16(2):33–60, December 2012b. ISSN 1460-1559. doi: 10.21314/jcf.2012.239. URL http://dx.doi.org/10.21314/jcf.2012.239.

A PROOF OF THE FIRST-ORDER COMPRESSION THEOREM

We give the complete proof of Theorem 1, including the discrete young-age case and the packing argument at the deletion boundary. Throughout, V is a nonzero finite-dimensional real normed space, all operator norms are induced by its fixed norm, $0 < \mu \le 1$, $q = e^{-\mu}$, $a_0 > 0$, and $0 < \epsilon < 1$. The control set $\mathcal{C}$ is nonempty. The hypotheses of Assumption 1 are, for every $a, a' \ge 0$ and $c \in \mathcal{C}$,

$$\|G(a,c)\| \le q, \qquad \|G(a,c) - G(a',c)\| \le q\mu\Lambda \min\left\{1, \left|\log \frac{a+a_0}{a'+a_0}\right|\right\}, \tag{17}$$

where $\Lambda \ge 0$. Both systems start empty and receive the same external stream (u_t, c_t), with $\|u_t\| \le 1$. The new birth is $b_t = (1-q)u_t$, inserted after all old states are gated. For a birth $j \le t$, write

$$x_{j,t} = P_{j,t} b_j, \qquad P_{j,t} = G(t-1-j, c_t) \cdots G(0, c_{j+1}), \qquad P_{t,t} = I. \tag{18}$$

Thus $y_t = \sum_{j=0}^{t} x_{j,t}$ and $\|x_{j,t}\| \le (1-q)q^{t-j}$. Products always retain this chronological order.

Set

$$\eta = \frac{\epsilon}{24(1+\Lambda)}, \qquad H = \left\lceil \frac{\log(3/\epsilon)}{\mu} \right\rceil, \qquad W_1(a) = \frac{\eta}{\mu}\sqrt{\mu a + \eta^2}(1+\mu a). \tag{19}$$

A record stores one vector and a consecutive integer birth interval. At any fixed time, equivalently write its age interval as $[a, a+w-1] \cap \mathbb{Z}$. Before a gate, select a representative α inside its current old-age interval; for a singleton it is its exact age. The selection is a fixed function of the interval and fixed parameters. One choice is $\alpha + a_0 = \sqrt{(a+a_0)(a+w-1+a_0)}$. The proof permits every such selector.

A.1 CALENDAR INVARIANTS AND MAXIMALITY

At tick t, gate each old record using $G(\alpha, c_t)$, age it by one, insert the singleton b_t, and merge adjacent records whenever their union has youngest age a and total width s satisfying $s - 1 \le W_1(a)$. A merge adds the two vectors. Continue until no adjacent pair is eligible, then delete every record whose youngest age is at least H. The output is the sum of retained vectors. A zero write still creates a singleton slot.

Lemma 3 (Width and partition invariants)**.** *For an increasing positive width profile W, this calendar retains a partition of a consecutive prefix of the nonnegative integer ages, and each record satisfies $w - 1 \le W(a)$. It can be implemented by a deterministic young-to-old greedy sweep, producing a partition in which no retained adjacent pair is eligible.*

Proof. A singleton satisfies the width bound. A merge imposes it on its union, and aging preserves it because W is increasing. The operations insert age zero and join adjacent intervals, so they preserve consecutiveness. Youngest ages increase strictly from younger to older records. Deletion therefore removes an old suffix, leaving a young prefix, whose last record may straddle H.

For maximality, grow a current interval towards older ages until adjoining the next interval fails the union test, then close it and start a new current interval. If that older interval later grows, the width across the rejected boundary only increases, whereas its youngest age remains unchanged. The rejected boundary therefore remains ineligible. Every merge reduces the number of records, so the sweep terminates, with no eligible boundary. Deleting an old suffix creates no new boundary between retained records. The sweep uses $O(K)$ calendar operations on K records. This excludes the cost of applying the linear gates. □

A.2 DECAY-WEIGHTED LOCAL GATE ERROR

The next lemma is valid for every partition satisfying the width invariant; it does not require maximality or an upper bound on the oldest retained age. For a singleton put $r_I = 0$. For a nonsingleton record of youngest age a, put

$$r_I = \begin{cases} 1, & a = 0, \\ \min\{1, W_1(a)/a\}, & a > 0, \end{cases} \qquad M_I = (1-q)\sum_{i \in I} q^i. \tag{20}$$

Here I in the sum denotes its integer age interval. Its capped logarithmic diameter, and hence every representative's capped logarithmic radius, is at most r_I, since

$$\log\left(1+\frac{w-1}{a+a_0}\right) \le \frac{w-1}{a+a_0}. \tag{21}$$

Consequently each representative gate differs from the gate at any exact age in I by at most $q\mu\Lambda r_I$.

Lemma 4 (Mass-radius estimate)**.** *If $0 < \eta \le 1/24$ and $w - 1 \le W_1(a)$ for every record, then*

$$\sum_I M_I r_I < 14\eta, \qquad \sum_I M_I \le 1. \tag{22}$$

Proof. The second inequality follows from the geometric series. Only nonsingletons contribute to the first. Write $x = \mu a$ and $d = \mu(w-1)$. Then

$$d \le \eta\sqrt{x+\eta^2}(1+x). \tag{23}$$

For young records, $x < \eta^2$ implies $d \le \sqrt{2}(1+\eta^2)\eta^2 < 2\eta^2$. A nonsingleton also has $\mu \le d$. Thus, if any young nonsingleton exists, $\mu \le 2\eta^2$, and every atom in the union of such records has $\mu i < 3\eta^2$. Their total mass, even after including all ages in that range, is at most

$$(1-q)\left(\left\lfloor\frac{3\eta^2}{\mu}\right\rfloor+1\right) \le 3\eta^2+\mu \le 5\eta^2. \tag{24}$$

If no such record exists, this contribution is zero and no restriction on μ is needed. This removes an otherwise uncontrolled integer remainder.

For intermediate records, $\eta^2 \le x \le 1$, we have $d \le 2\sqrt{2}\eta\sqrt{x} \le 2\sqrt{2}x$. Therefore any integer age i in the record satisfies $x \le \mu i \le x + d < 4x$, and

$$r_I \le \frac{2\sqrt{2}\eta}{\sqrt{x}} \le \frac{4\sqrt{2}\eta}{\sqrt{\mu i}}. \tag{25}$$

For old records, $x > 1$, we instead have

$$r_I \le \frac{\eta\sqrt{2x}(1+x)}{x} \le 2\sqrt{2}\eta\sqrt{x} \le 2\sqrt{2}\eta\sqrt{\mu i}. \tag{26}$$

The decreasing function $e^{-x}x^{-1/2}$ and $1-q \le \mu$ give the right Riemann-sum estimate

$$(1-q)\sum_{i\ge1}\frac{q^i}{\sqrt{\mu i}} \le \mu\sum_{i\ge1}\frac{e^{-\mu i}}{\sqrt{\mu i}} \le \int_0^\infty e^{-x}x^{-1/2}\,dx = \sqrt{\pi}. \tag{27}$$

Concavity of the square root, applied to the probability weights $(1-q)q^i$, gives

$$(1-q)\sum_{i\ge0}q^i\sqrt{\mu i} \le \sqrt{\frac{\mu q}{1-q}} \le 1, \tag{28}$$

where $e^\mu - 1 \ge \mu$. Summing the three disjoint record classes and enlarging the last two sums to all admissible ages proves $\sum_I M_I r_I \le 5\eta^2 + (4\sqrt{2\pi}+2\sqrt{2})\eta < 14\eta$. The argument also applies when the last record straddles the cutoff. □

A.3 Uniform output error

At the end of tick t, let

$$R_t = \sum_{I\text{ retained}}\left\|\widehat{v}_{I,t} - \sum_{j\in I}x_{j,t}\right\|, \tag{29}$$

where I now denotes the corresponding birth interval. At the next gate, insert and subtract the representative gate applied to the exact aggregate inside each record. Contraction bounds the previous-error part by q times the old record error. The remaining part is at most $q\mu\Lambda M_I r_I$, by the atom mass majorant and (17). Births are identical. Merging two records replaces their errors by their

sum and cannot increase the sum of norms. Deletion removes nonnegative summands. Lemma 4 therefore gives

$$R_t \le qR_{t-1} + 14q\mu\Lambda\eta, \qquad R_{-1} = 0, \qquad R_t \le 14\Lambda\eta\frac{q\mu}{1-q} \le 14\Lambda\eta. \tag{30}$$

Every previously deleted birth has current age at least H, even after further gates. Its norm remains bounded by $(1-q)q^i$. Hence

$$\begin{aligned}\|\widehat{y}_t - y_t\| &\le R_t + (1-q)\sum_{i\ge H} q^i \le 14\Lambda\eta + q^H \\ &\le \left(\frac{14\Lambda}{24(1+\Lambda)} + \frac{1}{3}\right)\epsilon < \epsilon.\end{aligned} \tag{31}$$

The use of sums of atom norms makes this argument valid for arbitrary signed-vector writes, including cancelling aggregates. It uses neither positivity of matrices nor reordering of their products.

A.4 Packing in reciprocal-width measure

We first record the geometric argument in a form shared by both algorithms.

Lemma 5 (Packing a maximal retained partition)**.** *Let W be increasing, positive, and differentiable, with $W'(a) \le 1$ on $[0, H]$. For a maximal retained partition produced by the union test $s-1 \le W(a)$ and whole-record deletion at youngest age H, its retained record count satisfies*

$$K \le 2 + \frac{2}{\log 2}\int_0^H \frac{da}{W(a)}. \tag{32}$$

Proof. The claim is immediate for $K \le 2$. Otherwise, order records from young to old and associate $[a, a+w-1] \cap \mathbb{Z}$ with the half-open slot cell $[a, a+w)$. Pair consecutive records among the first $K-1$, omitting any unpaired record. There are $N = \lfloor (K-1)/2 \rfloor \ge (K-2)/2$ pairs. Each pair ends at or before the youngest age of a later retained record, so its entire slot cell lies in $[0, H]$. The possibly straddling final record is deliberately excluded. These cells have disjoint interiors.

A pair starting at a with total slot width s is ineligible, so $s - 1 > W(a)$ and in particular $s > W(a)$. The derivative bound implies $W(a+v) \le W(a) + v$ for $0 \le v \le s$. Therefore

$$\int_a^{a+s} \frac{dr}{W(r)} \ge \int_0^s \frac{dv}{W(a)+v} = \log\left(1 + \frac{s}{W(a)}\right) > \log 2. \tag{33}$$

Sum over the pairs to obtain $((K-2)/2)\log 2 \le \int_0^H da/W(a)$. Closed versus half-open endpoints do not change these integrals. Deletion preserves pair ineligibility because it removes only a suffix and creates no retained boundary. □

Differentiate the profile in (19), writing $x = \mu a$:

$$W_1'(a) = \eta\left[\frac{1+x}{2\sqrt{x+\eta^2}} + \sqrt{x+\eta^2}\right] \le \frac{1}{2} + \frac{3}{2}\eta\sqrt{x} + \eta^2. \tag{34}$$

For $0 \le a \le H$, $x \le \log(3/\epsilon) + 1$, so

$$W_1'(a) \le \frac{1}{2} + \frac{\epsilon}{16}\sqrt{\log(3/\epsilon)+1} + \frac{1}{576} < 1. \tag{35}$$

Indeed $\epsilon\sqrt{\log(3/\epsilon)+1}$ increases on $(0, 1]$ and is at most $\sqrt{\log 3 + 1} < 1.45$. The derivative bound is needed only on $[0, H]$, not for arbitrarily old ages. Moreover,

$$\begin{aligned}\int_0^H \frac{da}{W_1(a)} &= \frac{1}{\eta}\int_0^{\mu H} \frac{dx}{\sqrt{x+\eta^2}(1+x)} \\ &\le \frac{1}{\eta}\int_0^\infty \frac{dx}{\sqrt{x}(1+x)} = \frac{\pi}{\eta},\end{aligned} \tag{36}$$

where $x = v^2$ reduces the last integral to $2\int_0^\infty (1+v^2)^{-1}dv = \pi$. Lemma 5 yields

$$K_t \le 2 + \frac{2\pi}{\eta \log 2} = 2 + \frac{48\pi(1+\Lambda)}{\epsilon \log 2}. \tag{37}$$

Together with (31), this proves Theorem 1.

The same derivative and packing proof also applies whenever $0 < \eta \le e/24$ and $H = \lceil \log(3/e)/\mu \rceil$, for any $0 < e < 1$, a form used in Appendix C. The count is measured after deletion. Insertion adds at most one temporary record before merges; record reuse needs only a constant number of work vectors. In dimension d, the retained vector payload is dK_t real coordinates, in addition to interval metadata. Fixed Λ gives a sufficient $O_\Lambda(\epsilon^{-1})$ vector bound, uniformly in time, $0 < \mu \le 1$, and every admissible common external stream. If $\Lambda = 0$, the gates are age independent and the untruncated exact sum itself has a one-vector recursion. When $W_1(0) < 1$, the union test simply leaves young records singleton; no continuous relaxation forces an invalid integer merge.

B Proof of the second-order compression theorem

We prove Theorem 2 under its additional smoothness assumptions. Keep the common-norm, empty-start, normalized-write and chronological-product conventions of Appendix A. In particular, $0 < \mu \le 1$, $q = e^{-\mu}$, $a_0 > 0$, and $0 < \epsilon < 1$. Assumption 2 requires that G is twice continuously differentiable in every real age $a \ge 0$ and that, uniformly over all allowed c and ages $a, a' \ge 0$,

$$\begin{aligned} \|G(a,c)\| &\le q, & \|G(a,c) - G(a',c)\| &\le q\mu\kappa, \\ \|\partial_a G(a,c)\| &\le \frac{q\mu L_1}{a+a_0}, & \|\partial_a^2 G(a,c)\| &\le \frac{q\mu L_2}{(a+a_0)^2}. \end{aligned} \tag{38}$$

The nonnegative constants κ, L_1, L_2 are finite and fixed independently of μ and ϵ. Every age comparison holds at a common context; there is no temporal regularity assumption on the external context stream. Neither commutativity nor invertibility of the gates is required.

Define

$$\begin{gathered} D = 256 + 6\kappa + 324L_2 + 1080L_1^2, \qquad \eta = \sqrt{\frac{\epsilon}{3D}}, \qquad H = \left\lceil \frac{\log(3/\epsilon)}{\mu} \right\rceil, \\ W_2(a) = \frac{\eta}{\mu}(\mu a + \eta^3)^{2/3}(1+\mu a). \end{gathered} \tag{39}$$

In particular $\eta < 1/16$ and $\eta \le \sqrt{\epsilon}/16$.

B.1 Endpoint updates and stability

A record stores a consecutive integer birth interval $[L, R]$ and two vectors v_L, v_R at its endpoints. After tick t, its age interval is $[t-R, t-L]$, with width $w = R - L + 1$. A singleton stores its vector in one entry and zero in the other. When a merge creates the nondegenerate union $[L', R']$, an old endpoint vector v at coordinate x contributes

$$\frac{R'-x}{R'-L'}v \quad \text{to } L', \qquad \frac{x-L'}{R'-L'}v \quad \text{to } R'. \tag{40}$$

Contributions arriving at the same endpoint are added. This formula is applied only to unions with $L' < R'$, so no denominator vanishes.

At tick t, gate every old endpoint at its actual endpoint age, $v_x \leftarrow G(t-1-x, c_t)v_x$; advance the ages and insert the singleton $[t,t]$ with $b_t = (1-q)u_t$. Apply a young-to-old maximal merge sweep using the union test $s - 1 \le W_2(a)$ and update (40). Delete records whose youngest age is at least H, then sum the retained endpoint vectors. All decisions depend only on the calendar and parameters, including at zero writes.

Lemma 3 applies because W_2 is positive and increasing. Thus the retained intervals form a young prefix, with

$$w - 1 \le W_2(a) \tag{41}$$

for every record. The last retained interval may straddle H. The coefficients in (40) are nonnegative, sum to one, and have coordinate-weighted sum x. Hence a merge preserves the zeroth and first birth-coordinate moments of the *current endpoint representation*. This does not identify these moments with inaccessible exact-population moments after approximate propagation.

Let N_t be the sum of norms of all retained endpoint vectors. A gate contracts it by q. Positive splitting preserves the sum of norms of each split vector, while adding vectors and deleting records cannot increase the sum. Consequently

$$N_t \le qN_{t-1} + 1 - q, \qquad N_{-1} = 0, \qquad N_t \le 1 - q^{t+1} \le 1. \tag{42}$$

This reasoning allows arbitrary signed-vector writes.

B.2 An exact virtual-path representation

Fix an original birth j that is retained at time t. Its actual record history depends only on the calendar. Initialize a virtual birth coordinate at j. At each rebucketing into $[L, R]$, replace the current coordinate x by L or R with the respective probabilities in (40). Let X_s be the resulting coordinate just before the gate at tick s, for $j + 1 \le s \le t$. It remains inside the actual record containing birth j. In particular $s - 1 - X_s \ge 0$. Each transition has conditional mean equal to its incoming coordinate, so

$$\mathbb{E}X_s = j \qquad (j + 1 \le s \le t). \tag{43}$$

The probabilities depend only on geometry; they are not reweighted by gates, vector signs, or vector norms. Expanding the finite sequence of split-and-add operations gives

$$\widehat{y}_t = \sum_{j \in J_t} \mathbb{E}\left[G(t - 1 - X_t, c_t) \cdots G(j - X_{j+1}, c_{j+1})\right] b_j, \tag{44}$$

where J_t is the set of retained original births. To see this by induction, insertion adds a deterministic coordinate at its own birth; a gate left multiplies each corresponding path product; a split replaces a path by its two extensions with the displayed probabilities; endpoint addition collects the terms; whole-record deletion removes all original births of that record. Thus every algorithmic operation preserves the expansion. The random coordinates are only a proof device; the implementation is deterministic and stores neither paths nor samples.

B.3 Discrete initial ages

A gate of true old age $i \ge 0$ is called initial when $x = \mu i < 3\eta^3$. This is a deterministic prefix of a birth's gate history, not a path-dependent stopping rule. At that gate, let $y = \mu a \le x$ be the normalized youngest age of its record and $d = \mu(w - 1)$ its normalized span. The width invariant implies

$$d \le \eta(y + \eta^3)^{2/3}(1 + y). \tag{45}$$

A nonzero virtual displacement requires a nonsingleton record, and then $\mu \le d$. If this happens at an initial gate, $y < 3\eta^3$ gives

$$\mu \le d < 4^{2/3}(1 + 3\eta^3)\eta^3 < 3\eta^3, \tag{46}$$

using $\eta < 1/16$. If $\mu > 3\eta^3$, all initial gates are therefore singletons and have zero displacement. Otherwise, there are at most $3\eta^3/\mu + 1$ nonnegative integer ages in the initial prefix. The age-oscillation bound in (38), divided by q, then bounds its pathwise cumulative discrepancy by

$$D_{\text{init}} \le \mu\kappa(3\eta^3/\mu + 1) \le 6\kappa\eta^3. \tag{47}$$

The bound remains valid when all these gates are exact. Crucially, the singleton argument removes a possible additive $\kappa\mu$ loss.

For a fixed virtual path and a birth of age $n = t - j$, denote its ordered product in (44) by $P_{j,t}(X)$. Replace only the initial factors by their exact true-age factors, and call the result $\widetilde{P}_{j,t}(X)$. Ordered-product telescoping, with all other factors bounded by q, gives

$$\|P_{j,t}(X) - \widetilde{P}_{j,t}(X)\| \le q^n D_{\text{init}}. \tag{48}$$

This is the usual telescoping difference in the original order and uses no matrix inverse.

B.4 MATURE RELATIVE DISPLACEMENTS

Suppose $x = \mu i \geq 3\eta^3$. Necessarily $y \geq \eta^3$: if $y < \eta^3$, (45) would give

$$x \leq y + d < \left[1 + 2^{2/3}(1 + \eta^3)\right] \eta^3 < 3\eta^3, \tag{49}$$

a contradiction. Every age a_θ on the segment between the true and virtual age lies in the record and is at least a; the displacement is at most $w - 1$. For $\eta^3 \leq y \leq 1$,

$$d \leq 2^{5/3}\eta y^{2/3}, \qquad x \leq y + d \leq (1 + 2^{5/3})y < 5y, \tag{50}$$

where $\eta y^{-1/3} \leq 1$. It follows that

$$\frac{w - 1}{a_\theta + a_0} \leq \frac{d}{y} \leq 2^{5/3}5^{1/3}\eta x^{-1/3} < 6\eta x^{-1/3}. \tag{51}$$

For $y \geq 1$, the alternative bound is

$$\frac{d}{y} \leq 2^{5/3}\eta y^{2/3} \leq 4\eta x^{2/3}. \tag{52}$$

Since $x^{2/3} \leq 1 + x$, both regimes are covered by

$$r_i = 6\eta \left[(\mu i)^{-1/3} + 1 + \mu i\right], \qquad i \geq 1. \tag{53}$$

Every mature gate has $i \geq 1$. These bounds hold throughout every interpolation segment, including in a record straddling H, because only the width invariant was used.

B.5 CANCELLATION IN THE COMPLETE CHRONOLOGICAL PRODUCT

For a fixed virtual path put $d_s = X_s - j$ at mature gates and $d_s = 0$ at initial gates. Interpolate only the mature factors:

$$F_X(\theta) = \prod_{s=t}^{j+1} G(s - 1 - j - \theta d_s, c_s), \qquad 0 \leq \theta \leq 1. \tag{54}$$

The decreasing-s product order is the order in (18). Its endpoints are $F_X(0) = P_{j,t}$ and $F_X(1) = \widetilde{P}_{j,t}(X)$. Each summand of $F'_X(0)$ has one differentiated factor, with all remaining factors evaluated at deterministic true ages in their original positions. It is therefore linear in one d_s. By (43),

$$\mathbb{E}F'_X(0) = 0. \tag{55}$$

No independence of displacements at different gates is needed, and no matrices have been reordered.

At a mature gate of true old age i, differentiating with respect to θ gives norms bounded by $q\mu L_1 r_i$ and $q\mu L_2 r_i^2$ for the first and second derivatives. This follows from (38) and the displacement bound at every interpolation age. Every undifferentiated factor has norm at most q. The product rule consequently gives

$$\|F''_X(\theta)\| \leq q^n \left[\mu L_2 \sum_{i=1}^{n-1} r_i^2 + \mu^2 L_1^2 \left(\sum_{i=1}^{n-1} r_i\right)^2\right]. \tag{56}$$

The first sum bounds terms with one twice-differentiated factor. The square bounds all ordered pairs of once-differentiated factors; its diagonal terms are an unnecessary but valid enlargement. Initial factors have zero interpolation displacement, so extending the sums to all positive old ages only enlarges the bound. Empty sums, including for $n = 0, 1$, are zero.

Taylor's integral formula and (55) imply

$$\|\mathbb{E}\widetilde{P}_{j,t}(X) - P_{j,t}\| \leq \frac{q^n}{2} \left[\mu L_2 \sum_{i=1}^{n-1} r_i^2 + \mu^2 L_1^2 \left(\sum_{i=1}^{n-1} r_i\right)^2\right]. \tag{57}$$

At finite t the virtual path space is finite, so expectation and Taylor's integral can be interchanged directly.

B.6 GEOMETRIC-AGE SUMMATIONS

Let N and I be nonnegative integer-valued variables with probabilities $\mathbb{P}(N = n) = \mathbb{P}(I = n) = (1 - q)q^n$, and put $X = \mu I$. Expressions involving r_I or negative powers of X below are restricted to $I \geq 1$. Define

$$J_2 = \mathbb{E}_N \left[\mu \sum_{i=1}^{N-1} r_i^2 \right], \qquad J_{11} = \mathbb{E}_N \left[\mu^2 \left(\sum_{i=1}^{N-1} r_i \right)^2 \right]. \tag{58}$$

All summands are nonnegative, so Tonelli's theorem permits the exchanges of infinite sums. Direct summation gives

$$J_2 = \mu \sum_{i \geq 1} q^{i+1} r_i^2 = \frac{\mu q}{1-q} \mathbb{E}_I[r_I^2 \mathbf{1}_{\{I \geq 1\}}] \leq \mathbb{E}_I[r_I^2 \mathbf{1}_{\{I \geq 1\}}], \tag{59}$$

because $\mu q/(1 - q) = \mu/(e^\mu - 1) \leq 1$. Cauchy–Schwarz yields $(\sum_{i=1}^{n-1} r_i)^2 \leq n \sum_{i=1}^{n-1} r_i^2$, and the exact geometric tail identity is

$$(1 - q) \sum_{n=i+1}^{\infty} n q^n = q^{i+1} \left(i + \frac{1}{1-q} \right). \tag{60}$$

Consequently

$$\begin{aligned} J_{11} &\leq \mu^2 \sum_{i \geq 1} q^{i+1} \left(i + \frac{1}{1-q} \right) r_i^2 \\ &= \frac{\mu q}{1-q} \mathbb{E}_I \left[\left(X + \frac{\mu}{1-q} \right) r_I^2 \mathbf{1}_{\{I \geq 1\}} \right] \leq \mathbb{E}_I[(X+2) r_I^2 \mathbf{1}_{\{I \geq 1\}}]. \end{aligned} \tag{61}$$

The last inequality uses $\mu/(1 - q) = \mu + \mu q/(1 - q) \leq 2$.

For $z \geq 0$, $\mathbb{P}(X > z) = q^{\lfloor z/\mu \rfloor + 1} \leq e^{-z}$. Thus X is stochastically dominated by a mean-one exponential variable, and for $p > 0$, $\mathbb{E}X^p \leq \Gamma(p + 1)$. In particular,

$$\mathbb{E}X \leq 1, \quad \mathbb{E}X^2 \leq 2, \quad \mathbb{E}X^3 \leq 6, \quad \mathbb{E}X^{1/3} \leq 1. \tag{62}$$

The last inequality also follows from concavity. Negative powers require a separate argument. Since $e^{-x} x^{-2/3}$ is decreasing and integrable,

$$\mathbb{E}[X^{-2/3} \mathbf{1}_{\{I \geq 1\}}] \leq \mu \sum_{i \geq 1} e^{-\mu i} (\mu i)^{-2/3} \leq \int_0^\infty e^{-x} x^{-2/3} \, dx = \Gamma(1/3) < 3. \tag{63}$$

For the strict last inequality, integration by parts gives $\Gamma(1/3) = 3\Gamma(4/3)$, while strict concavity gives $\Gamma(4/3) = \mathbb{E}Z^{1/3} < 1$ for a nondegenerate mean-one exponential Z.

Equation (53) and $(a + b + c)^2 \leq 3(a^2 + b^2 + c^2)$ give

$$r_i^2 \leq 108\eta^2[(\mu i)^{-2/3} + 1 + (\mu i)^2]. \tag{64}$$

Combining (59)–(64),

$$J_2 \leq 108\eta^2(3 + 1 + 2) = 648\eta^2. \tag{65}$$

Expanding $(X^{-2/3} + 1 + X^2)(X + 2)$ produces $X^{1/3} + 2X^{-2/3} + X + 2 + X^3 + 2X^2$. The same estimates give

$$J_{11} \leq 108\eta^2(1 + 6 + 1 + 2 + 6 + 4) = 2160\eta^2. \tag{66}$$

B.7 OUTPUT ERROR AND WHOLE-RECORD DELETION

Multiply (48) and (57) by $\|b_j\| \leq 1 - q$ and sum over retained original births. Enlarge the nonnegative age sums to all $n \geq 0$. Equations (47), (65), and (66) show that the retained error is at most

$$6\kappa\eta^3 + \tfrac{1}{2} L_2 J_2 + \tfrac{1}{2} L_1^2 J_{11} \leq 6\kappa\eta^3 + (324L_2 + 1080L_1^2)\eta^2. \tag{67}$$

Every birth deleted previously has current true age at least H. Its exact product norm is at most q^n, so the omitted exact tail is at most $(1 - q) \sum_{n \geq H} q^n = q^H$. Virtual endpoint coordinates do not change the original birth's record membership or the whole-record deletion rule. We conclude that

$$\|\widehat{y}_t - y_t\| \leq 6\kappa\eta^3 + (324L_2 + 1080L_1^2)\eta^2 + q^H. \tag{68}$$

Since $\eta < 1$, the retained terms are at most $(D - 256)\eta^2 < \epsilon/3$, while $q^H \leq \epsilon/3$. The displayed parameters therefore give error strictly below $2\epsilon/3$, and in particular below ϵ, uniformly in t.

B.8 COUNTING ALL ENDPOINT VECTORS

With $x = \mu a$, differentiation gives

$$W_2'(a) = \eta \left[\frac{2(1+x)}{3(x+\eta^3)^{1/3}} + (x+\eta^3)^{2/3} \right] \leq \frac{2}{3} + \frac{5}{3}\eta x^{2/3} + \eta^3. \tag{69}$$

Here $(x+\eta^3)^{1/3} \geq \eta$ bounds the constant numerator term, $x/(x+\eta^3)^{1/3} \leq x^{2/3}$ bounds the other term, and $(x+\eta^3)^{2/3} \leq x^{2/3} + \eta^2$ follows from subadditivity. For $a \in [0, H]$, $x \leq \log(3/\epsilon) + 1$. The function $\sqrt{\epsilon}[\log(3/\epsilon) + 1]^{2/3}$ increases on $(0, 1]$, since its logarithmic derivative has sign $1/2 - 2/[3(\log(3/\epsilon) + 1)] > 0$. Its maximum is $(\log 3 + 1)^{2/3} < 1.65$. Thus

$$W_2'(a) < \frac{2}{3} + \frac{5 \cdot 1.65}{48} + \frac{1}{4096} < 1 \quad (0 \leq a \leq H). \tag{70}$$

Lemma 5 now applies, including its exclusion of the possibly straddling final record and its half-open integer-slot convention. Its reciprocal-width integral satisfies

$$\int_0^H \frac{da}{W_2(a)} = \frac{1}{\eta} \int_0^{\mu H} \frac{dx}{(x+\eta^3)^{2/3}(1+x)} \leq \frac{1}{\eta} \int_0^\infty \frac{x^{-2/3}}{1+x}\, dx = \frac{2\pi}{\sqrt{3}\eta}. \tag{71}$$

For completeness, substitute $x = v^3$ in the last integral to obtain $3\int_0^\infty (1+v^3)^{-1} dv$. The partial-fraction identity

$$\frac{1}{1+v^3} = \frac{1}{3(v+1)} + \frac{-v+2}{3(v^2-v+1)} \tag{72}$$

has antiderivative $\frac{1}{3}\log(v+1) - \frac{1}{6}\log(v^2-v+1) + \frac{1}{\sqrt{3}}\arctan((2v-1)/\sqrt{3})$. Its endpoint difference is $2\pi/(3\sqrt{3})$, proving the integral value. Consequently

$$K_t \leq 2 + \frac{4\pi}{\sqrt{3}\eta \log 2}, \qquad 2K_t \leq 4 + \frac{8\pi}{\sqrt{3}\eta \log 2}. \tag{73}$$

Each retained record has at most two vectors, including the zero second entry of a singleton. This proves the total-vector assertion of Theorem 2. Insertion creates at most one transient record; rebucketing with reused storage requires a constant number of work vectors.

For fixed κ, L_1, L_2, the bound is $O_{\kappa,L_1,L_2}(\epsilon^{-1/2})$, uniformly in time, lifetime, and every admissible common external stream. It is also uniform over finite-dimensional instances that satisfy the same constants in their specified common norm; a d-dimensional instance has at most $2dK_t$ retained real coordinates. The second-order rate relies on (38), not merely the first-order modulus. Singular gates, zero writes, and discrete thresholds below one cause no exception to the proof. The results count exact-arithmetic vector state, not metadata bit widths or total work over a physical lifetime.

C REGULARITY, READOUT, FEEDBACK, AND EVALUATION INTERFACES

This appendix extends the first-order construction and gives the correct conditional worst-write functional for both constructions. Let V be a nonzero finite-dimensional real normed memory space, $a_0 > 0$, $z(a) = \log(a + a_0)$, $q = e^{-\mu}$, and $0 < \mu \leq 1$. Unless explicitly replaced below, the gates satisfy (17). All systems start empty and give every tick a calendar slot. The merge calendar and representative selection depend on ages and fixed parameters, and all operator norms use the stated input and output norms.

The common tool is Lemma 4: for every width-valid partition of the first-order profile,

$$\sum_I M_I r_I \leq 14\eta, \qquad \sum_I M_I \leq 1. \tag{74}$$

The radii may be the capped actual representative log radii, since these are bounded by (20). The propagation proof uses only this local estimate and contraction. The derivative and packing proof remains valid for $\eta \leq e/24$, $0 < e < 1$, and $H = \lceil \log(3/e)/\mu \rceil$, yielding $K_t \leq 2 + 2\pi/(\eta \log 2)$.

C.1 AGE-SELECTIVE LINEAR READOUTS

Let U be a nonzero finite-dimensional real normed output space. For every age and every allowed query ξ, let $R(a,\xi) : V \to U$ be linear, with

$$\|R(a,\xi)\| \le M, \qquad \|R(a,\xi) - R(a',\xi)\| \le \Lambda_R \min\{1, |z(a) - z(a')|\}, \tag{75}$$

where $M \ge 1$ and $\Lambda_R \ge 0$ are finite. Both systems receive the same external query ξ_t. The exact queried output is $y_t^R = \sum_{j=0}^{t} R(t-j,\xi_t)x_{j,t}$; the approximation applies R to each retained record vector at a representative of its current age interval.

Proposition 6 (Queried-output guarantee)**.** *For $0 < \epsilon < 1$, set*

$$\delta = \epsilon/M, \qquad \eta = \frac{\delta}{24(1+\Lambda+\Lambda_R/M)}, \qquad H = \left\lceil \frac{\log(3/\delta)}{\mu} \right\rceil. \tag{76}$$

The first-order algorithm has queried-output error below ϵ and

$$K_t \le 2 + \frac{48\pi(M + M\Lambda + \Lambda_R)}{\epsilon \log 2}. \tag{77}$$

The assertion is uniform in time, $0 < \mu \le 1$, and all admissible common external write, control, and query streams, for fixed M, Λ, Λ_R.

Proof. The retained propagation discrepancy is at most $14\Lambda\eta$ by (30); applying the representative readout multiplies it by at most M. Insert and subtract this representative readout applied to the exact aggregate in each record. The remaining current-query discrepancy is at most $\Lambda_R \sum_I M_I r_I \le 14\Lambda_R\eta$. The exact omitted queried tail has norm at most Mq^H. Therefore

$$\|\widehat{y}_t^R - y_t^R\| \le 14(M\Lambda + \Lambda_R)\eta + Mq^H < \epsilon. \tag{78}$$

To check the strict inequality, divide by M and use the same calculation as (31), with target δ and constant $\Lambda + \Lambda_R/M$. This consistent normalization also gives $\eta \le \delta/24$, so the derivative and packing proof with target $e = \delta$ applies. Substituting (76) into $2 + 2\pi/(\eta \log 2)$ gives (77). □

For a common L_F-Lipschitz postprocessing map $F : U \to U'$, $\|F(\widehat{y}_t^R) - F(y_t^R)\| \le L_F\epsilon$ follows immediately. For a linear map this uses its induced operator norm. Such postprocessing alone does not change future control streams.

C.2 CONCAVE AGE MODULI

Proposition 7 (Regularity-to-state budget)**.** *Let $\omega : [0,1] \to [0,1]$ be continuous, strictly increasing and concave, with $\omega(0) = 0$ and $\omega(1) = 1$. Replace the age-modulus condition by*

$$\|G(a,c) - G(a',c)\| \le q\mu\Lambda\, \omega\big(\min\{1, |z(a) - z(a')|\}\big), \tag{79}$$

and retain common-norm contraction. Set

$$\eta = \frac{1}{24}\omega^{-1}\left(\frac{\epsilon}{3(1+\Lambda)}\right), \qquad H = \left\lceil \frac{\log(3/\epsilon)}{\mu} \right\rceil. \tag{80}$$

The first-order calendar with this η achieves error below ϵ, with

$$K_t \le 2 + \frac{48\pi}{\log 2}\left[\omega^{-1}\left(\frac{\epsilon}{3(1+\Lambda)}\right)\right]^{-1}. \tag{81}$$

Proof. Use the capped representative log radius $r_I \in [0,1]$, with singleton radius zero. Add the missing mass $1 - \sum_I M_I$ at radius zero. Jensen's inequality and (74) give

$$\sum_I M_I \omega(r_I) \le \omega\left(\sum_I M_I r_I\right) \le \omega(14\eta), \tag{82}$$

provided $14\eta < 1$. Thus the retained-error recursion becomes $R_t \le qR_{t-1} + q\mu\Lambda\omega(14\eta)$, and

$$R_t \le \frac{q\mu}{1-q}\Lambda\omega(14\eta) \le \Lambda\omega(14\eta). \tag{83}$$

Concavity with normalized endpoints gives $\omega(r) \ge r$, so $\omega^{-1}(v) \le v$ and the chosen η satisfies $\eta \le \epsilon/72 < \epsilon/24$. In particular $14\eta < 1$, and $14\eta \le \omega^{-1}(\epsilon/[3(1+\Lambda)])$. Monotonicity yields

$$\Lambda\omega(14\eta) \le \frac{\Lambda\epsilon}{3(1+\Lambda)} < \epsilon/3. \tag{84}$$

Add the tail $q^H \le \epsilon/3$. The packing estimate with target $e = \epsilon$ applies because $\eta \le \epsilon/24$, and substitution gives (81). Strict monotonicity ensures that the ordinary inverse used here exists on $[0, 1]$. □

For $\omega(r) = r^\alpha$, $0 < \alpha \le 1$, the sufficient count is $O_{\alpha,\Lambda}(\epsilon^{-1/\alpha})$. For fixed modulus and constants, (81) is uniform in lifetime, time and the common external stream. It is a sufficient bound, not a necessity statement.

C.3 A CLOSED-LOOP TRANSFER UNDER A QUANTITATIVE STABILITY MARGIN

Use the first-order gate assumptions and the selective readout (75). Let the context set be a metric space with metric $d_{\mathcal{C}}$. The queries ξ_t and inputs s_t are common external streams. Both branches start with empty memory and $y^R_{-1} = \widehat{y}^R_{-1} = 0$. For $t \ge 0$, they generate writes and controls by

$$\begin{aligned} u_t &= f_t(y^R_{t-1}, s_t), \quad c_t = h_t(y^R_{t-1}, s_t), \\ \widehat{u}_t &= f_t(\widehat{y}^R_{t-1}, s_t), \quad \widehat{c}_t = h_t(\widehat{y}^R_{t-1}, s_t). \end{aligned} \tag{85}$$

All maps are defined on encountered arguments. Assume every f_t has range in the unit ball of V, and uniformly in t, s,

$$\begin{aligned} \|f_t(v,s) - f_t(v',s)\|_V &\le L_f\|v - v'\|_U, \\ d_{\mathcal{C}}(h_t(v,s), h_t(v',s)) &\le L_h\|v - v'\|_U, \\ \|G(a,c) - G(a,c')\| &\le q\mu L_T d_{\mathcal{C}}(c,c'). \end{aligned} \tag{86}$$

Put

$$s_\mu = L_f + \frac{\mu q}{1-q}L_T L_h, \qquad s_0 = L_f + L_T L_h. \tag{87}$$

Proposition 8 (Closed-loop transfer)**.** *Suppose a memory approximation has queried-output error at most δ_{out} uniformly over* every *admissible frozen stream of unit-ball writes, controls, and common external queries, including each realized stream generated by the compressed branch. If $Ms_\mu < 1$, then the branches in* (85) *satisfy*

$$\sup_{t\ge 0}\|\widehat{y}^R_t - y^R_t\|_U \le \frac{\delta_{\text{out}}}{1 - Ms_\mu}. \tag{88}$$

Proof. For the realized compressed branch, run an auxiliary exact memory on its generated writes and controls, with the same external queries. Write its states as $\bar{x}_{j,t}$ and output as $\bar{y}^R_t$. By the uniform common-stream hypothesis, $\|\widehat{y}^R_t - \bar{y}^R_t\| \le \delta_{\text{out}}$. This comparison does not apply the common-stream theorem to two different control paths.

Define $D_t = \sum_j \|\bar{x}_{j,t} - x_{j,t}\|_V$ and $e_t = \|\widehat{y}^R_t - y^R_t\|_U$. Since both exact branches use the same query at each true age, $\|\bar{y}^R_t - y^R_t\| \le MD_t$, and therefore $e_t \le \delta_{\text{out}} + MD_t$. For each old birth, subtract its two gated states by first applying $G(a, \widehat{c}_t)$ to the old state difference, then charging $[G(a,\widehat{c}_t) - G(a,c_t)]x_{j,t-1}$. Contraction, (86), and $\sum_j \|x_{j,t-1}\| \le 1$ bound these terms. The new normalized births differ by at most $(1-q)L_f e_{t-1}$. It follows that

$$D_t \le qD_{t-1} + be_{t-1}, \qquad b = q\mu L_T L_h + (1-q)L_f = (1-q)s_\mu. \tag{89}$$

Hence $D_t \le (q + Mb)D_{t-1} + b\delta_{\text{out}}$. Histories start empty, so $D_{-1} = 0$. The condition $Ms_\mu < 1$ is precisely $q + Mb < 1$; the geometric series gives $D_t \le b\delta_{\text{out}}/(1 - q - Mb)$. Substitution into $e_t \le \delta_{\text{out}} + MD_t$ proves (88). □

For a sufficient condition uniform in μ, assume $Ms_0 < 1$ and set

$$\delta_{\mathrm{out}} = (1 - Ms_0)\epsilon, \qquad \delta = \delta_{\mathrm{out}}/M,$$
$$\eta = \frac{\delta}{24(1 + \Lambda + \Lambda_R/M)}, \qquad H = \left\lceil \frac{\log(3/\delta)}{\mu} \right\rceil. \tag{90}$$

Since $\mu q/(1-q) \le 1$, $s_\mu \le s_0$ and $Ms_\mu < 1$. Proposition 6 gives common-stream error below δ_{out} for every required realized stream, so Proposition 8 yields closed-loop error at most $\delta_{\mathrm{out}}/(1-Ms_\mu) \le \epsilon$. Its retained count is

$$K_t \le 2 + \frac{48\pi(M + M\Lambda + \Lambda_R)}{(1 - Ms_0)\epsilon \log 2}. \tag{91}$$

Fixed regularity constants and a fixed positive margin $1 - Ms_0$ give a uniform $O(\epsilon^{-1})$ state budget. The queries here are external; feedback-dependent queries would need an additional cross-query bound. Unit-ball writes and open-loop contraction alone do not imply the margin.

C.4 Exact conditional worst-write evaluation

Fix an external control/query path and a write-independent calendar and representative selector. Linearity gives coefficients $C_{t,j} : V \to U$ such that

$$\widehat{y}_t - y_t = (1 - q) \sum_{j=0}^{t} C_{t,j} u_j. \tag{92}$$

The coefficients may include the queried readout. For a deleted birth, its approximate coefficient is zero, so $C_{t,j}$ is minus its exact coefficient. Hold the external path fixed and define

$$E(t) = \sup_{\|u_j\|_V \le 1 \ (0 \le j \le t)} \left\| (1 - q) \sum_{j=0}^{t} C_{t,j} u_j \right\|_U. \tag{93}$$

Proposition 9 (Dual conditional error identity)**.** *For every nonzero finite-dimensional real normed input and output space,*

$$E(t) = (1 - q) \sup_{\|\ell\|_{U^*} \le 1} \sum_{j=0}^{t} \|C_{t,j}^* \ell\|_{V^*}. \tag{94}$$

Here $C_{t,j}^*$ is the dual linear map.

Proof. Finite-dimensional norm duality gives $\|u\|_U = \sup_{\|\ell\|_{U^*} \le 1} \ell(u)$. Substitute this in (93) and interchange the two suprema over their product set. For fixed ℓ, the writes are chosen independently in their unit balls. Each summand has supremum $\sup_{\|v\|_V \le 1} \ell(C_{t,j} v) = \|C_{t,j}^* \ell\|_{V^*}$. This proves the identity. Compactness of all unit balls in finite dimension and continuity also give attainment of the extrema. □

When $U = V$ with the same norm and $C_{t,j} = c_{t,j} I$, this reduces to $E(t) = (1 - q) \sum_j |c_{t,j}|$. More directly, choose any unit vector v and writes $u_j = \mathrm{sign}(c_{t,j}) v$ to attain it, with arbitrary unit-ball writes at zero coefficients. In the general matrix case, $(1 - q) \sum_j \|C_{t,j}\|$ is only an upper bound. For example, in Euclidean dimension two, the coefficients $C_1 = \mathrm{diag}(1, 0)$ and $C_2 = \mathrm{diag}(0, 1)$ have unnormalized conditional error $\sqrt{2}$ but operator-norm sum 2. Likewise, separately maximized retained and tail errors need not add exactly: they must share the same dual direction in (94). The inequality $E_{\mathrm{total}} \le E_{\mathrm{retained}} + E_{\mathrm{tail}}$ always holds.

Corollary 10 (A certified angular bracket)**.** *Let $U = (\mathbb{R}^2, \|\cdot\|_2)$ and retain the stated norm on V. Define*

$$\ell(\theta) = (\cos\theta, \sin\theta), \qquad F(\theta) = (1 - q) \sum_j \|C_{t,j}^* \ell(\theta)\|_{V^*},$$
$$L = (1 - q) \sum_j \|C_{t,j}\|_{V \to \ell_2^2}. \tag{95}$$

For any integer $N \ge 1$, put $\theta_k = 2\pi k/N$, $k = 0, \ldots, N-1$, and $h = 2\pi/N$. Then

$$\max_{0 \le k < N} F(\theta_k) \le E(t) \le \max_{0 \le k < N} F(\theta_k) + Lh/2. \tag{96}$$

Proof. The reverse triangle inequality and equality of induced primal and dual operator norms give

$$|F(\theta) - F(\theta')| \le L\|\ell(\theta) - \ell(\theta')\|_2 \le L|\theta - \theta'|. \tag{97}$$

The support functional is nonnegative and positively homogeneous in the dual direction. Its supremum over the dual unit ball therefore equals its maximum over the circle, also when all coefficients vanish. Every circular direction is at circular distance at most $h/2$ from the full grid. Apply (97) to a maximizing direction, using 2π-periodicity across the angular endpoint, to obtain the upper bound. The lower bound holds because grid directions are allowable dual vectors. □

If V is also Euclidean, $C^*_{t,j}$ is represented by $C^{\mathsf{T}}_{t,j}$ and its input dual norm is Euclidean. Euclidean output alone does not justify this simplification. For example, with $V = (\mathbb{R}^2, \|\cdot\|_\infty)$, Euclidean output and $C = I$, the unnormalized exact value is $\sqrt{2}$, and the input dual norm in (95) is ℓ_1. These offline coefficients and direction grids are evaluation state, not online compressor state. The identity conditions on an external path; the full feedback map is generally nonlinear in its original writes.

D Operator families satisfying the assumptions

This appendix verifies scalar and noncommuting examples in the stated common norms. A bound on each matrix's spectral radius alone is not the common-norm contraction assumption used by the theorems.

The second-order assumptions also imply the first-order modulus. Put $r = |\log((a+a_0)/(a'+a_0))|$. Integrating the first-derivative bound along the age interval gives $\|G(a,c) - G(a',c)\| \le q\mu L_1 r$. Combining it with the oscillation bound and the inequality

$$\min\{\kappa, L_1 r\} \le \max\{\kappa, L_1\} \min\{1, r\} \tag{98}$$

proves that $\Lambda = \max\{\kappa, L_1\}$ is a valid first-order constant.

D.1 Bounded scalar hazards

Suppose, for all ages and controls,

$$\begin{aligned} \mu \le \lambda(a,c) &\le (1+\kappa)\mu, \\ |\lambda(a,c) - \lambda(a',c)| &\le C_{\mathrm{reg}}\kappa\mu \left|\log \frac{a+a_0}{a'+a_0}\right|, \end{aligned} \tag{99}$$

with finite fixed $\kappa, C_{\mathrm{reg}} \ge 0$. Set $G(a,c) = e^{-\lambda(a,c)}I$. The lower hazard bound gives $\|G\| \le q$. By the mean value theorem on $[\mu, \infty)$, $\|G(a,c) - G(a',c)\| \le q|\lambda(a,c) - \lambda(a',c)|$. Writing $r = |\log((a+a_0)/(a'+a_0))|$, the range and modulus in (99) give

$$|\lambda(a,c) - \lambda(a',c)| \le \mu \min\{\kappa, C_{\mathrm{reg}}\kappa r\} \le \mu\kappa \max\{1, C_{\mathrm{reg}}\} \min\{1, r\}. \tag{100}$$

Thus Assumption 1 holds with $\Lambda = \kappa \max\{1, C_{\mathrm{reg}}\}$ in every norm on V.

For a second-order family, let $z = \log(a + a_0)$ and

$$\lambda(a,c) = \mu[1 + \kappa\rho(z,c)], \qquad 0 \le \rho \le 1, \quad |\partial_z \rho| \le B_1, \quad |\partial_z^2 \rho| \le B_2, \tag{101}$$

where ρ is twice continuously differentiable in z and all displayed bounds are uniform over controls and attained z, with fixed finite B_1, B_2. Its range gives contraction and $\|G(a,c) - G(a',c)\| \le q\mu\kappa$. Differentiating z gives

$$|\lambda'| \le \frac{\mu\kappa B_1}{a+a_0}, \qquad |\lambda''| \le \frac{\mu\kappa(B_1+B_2)}{(a+a_0)^2}. \tag{102}$$

Since $G' = -\lambda' G$ and $G'' = [(\lambda')^2 - \lambda'']G$, and $\mu \le 1$, Assumption 2 holds with

$$L_1 = \kappa B_1, \qquad L_2 = \kappa(B_1 + B_2) + \kappa^2 B_1^2. \tag{103}$$

These verifications are norm independent because every gate is a scalar multiple of the identity.

D.2 A NONCOMMUTING SMOOTH MATRIX FAMILY

Take $V = \mathbb{R}^2$ with the Euclidean norm and $\mathcal{C} = \{0, 1\}$. Let

$$P_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad P_1 = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \qquad \alpha = \min\{1, \kappa\mu\}, \tag{104}$$

and define

$$\begin{aligned} f(a, c) &= \tfrac{1}{2}[1 + \sin(\log(a + a_0) + 0.7c)], \\ G(a, c) &= q[(1 - \alpha f(a, c))I + \alpha f(a, c)P_c] = q[I - \alpha f(a, c)(I - P_c)]. \end{aligned} \tag{105}$$

Both P_c are orthogonal projections, so $\|P_c\|_2 = \|I - P_c\|_2 = 1$. Since $0 \le \alpha f \le 1$, the bracket is a convex combination of two contractions and $\|G\|_2 \le q$. Also $\alpha \le \kappa\mu$ and $|f(a, c) - f(a', c)| \le 1$, which give the age-oscillation bound $q\mu\kappa$. Direct differentiation yields

$$|\partial_a f(a, c)| \le \frac{1}{2(a + a_0)}, \qquad |\partial_a^2 f(a, c)| \le \frac{1}{(a + a_0)^2}. \tag{106}$$

Thus Assumption 2 holds with $L_1 = \kappa/2$ and $L_2 = \kappa$. Moreover,

$$|f(a, c) - f(a', c)| \le \min\{1, \tfrac{1}{2}|z(a) - z(a')|\}, \tag{107}$$

so Assumption 1 holds with $\Lambda = \kappa$ as well.

For $\kappa > 0$ and ages a, b for which $f(a, 0)f(b, 1) > 0$, the commutator is

$$[G(a, 0), G(b, 1)] = q^2\alpha^2 f(a, 0)f(b, 1)[P_0, P_1], \qquad [P_0, P_1] = \frac{1}{2}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \neq 0. \tag{108}$$

Hence the covered class includes age-dependent, channel-dependent contractions with noncommuting control actions. The Euclidean norm is part of this example; changing the norm requires a new contraction check.

D.3 A ROTATION-AND-CONTRACTION FAMILY

Another first-order example in Euclidean $\mathbb{R}^2$ uses planar rotation $Q(\theta)$ and $c \in [-1, 1]$:

$$\begin{aligned} r(a, c) &= \tfrac{1}{2}[1 + \sin(\log(1 + a/a_0) + \pi c/2)], \\ G(a, c) &= qQ(0.7c)\begin{pmatrix} e^{-\mu r(a,c)} & 0 \\ 0 & e^{-\mu(1-r(a,c))} \end{pmatrix}. \end{aligned} \tag{109}$$

Each diagonal entry lies in $[e^{-\mu}, 1]$; orthogonality of the rotation therefore gives $\|G(a, c)\|_2 \le q$. At fixed c, $|r(a, c) - r(a', c)| \le \min\{1, \frac{1}{2}|z(a) - z(a')|\}$. The mean value bound for each diagonal exponential, followed by the Euclidean diagonal operator norm and orthogonality, gives

$$\|G(a, c) - G(a', c)\|_2 \le q\mu \min\{1, |z(a) - z(a')|\}. \tag{110}$$

Thus the first-order result applies with $\Lambda = 1$. The contraction acts differently on the two channels and varies with both age and control; the common-norm proof does not need a shared eigenbasis.

E OPERATOR VALIDATION PROTOCOL AND COMPLETE RESULTS

Operator family and access. The input and output spaces are both $\mathbb{R}^2$ with the Euclidean norm. Set $q = e^{-\mu}$, $\alpha = \min\{1, \kappa\mu\}$, and

$$P_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad P_1 = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

$$f(a, c) = \frac{1 + \sin(\log(a + a_0) + 0.7c)}{2}, \quad G(a, c) = q[I - \alpha f(a, c)(I - P_c)], \qquad c \in \{0, 1\}.$$

Each P_c is an orthogonal projection and $0 \le \alpha f \le 1$, so $\|G\|_2 \le q$. The bounds $|f| \le 1$, $|f'| \le [2(a + a_0)]^{-1}$, and $|f''| \le (a + a_0)^{-2}$ give the first-order constant $\Lambda = \kappa$ and second-order constants $L_1 = \kappa/2$, $L_2 = \kappa$, uniformly for $0 < \mu \le 1$. The projections do not commute. In the experiments, each method receives the same current binary control and gate oracle; there is no trained encoder or hidden historical control state.

Frozen paths and storage-only parameter selection. The main runs fix $\mu = 1/64$, $\kappa = 1/2$, $a_0 = 2$, $T = 192$, and $H = 193$. One control path alternates at every tick. The other alternates between 0 and 1 with dwell lengths $(1, 9, 2, 17, 4, 3, 23, 5, 11, 2, 7)$ repeated until T ticks. Both paths are specified in the protocol before measuring errors. For each graded method, we enumerate the 281 values $\eta = 2^{-12+14i/280}$, $i = 0, \ldots, 280$. For each vector cap, we select the smallest enumerated η whose count-only calendar has a peak below or equal to the cap. This procedure does not assume monotonicity of the discrete peak count and does not use controls, writes, or errors. It knows the declared test horizon. Uniform blocks use the smallest positive integer width satisfying the same count constraint; recent history uses the largest feasible integer length. The resulting parameters are

Method	12 vectors	24 vectors	48 vectors
Graded-1 (η)	0.25	0.1051120519	0.03983001960
Graded-2 (η)	0.5743491775	0.3186401568	0.125
Uniform (width)	16	8	4
Recent (length)	12	24	48

These values are practical storage-constrained choices on a declared finite grid, not optima over all possible schedules and not instantiations of the conservative theorem error budgets. Error intervals are computed directly for each selected configuration. No parameters are adjusted after evaluating the two control paths.

Implemented baselines. Uniform blocks group birth slots according to $\lfloor j/A \rfloor$; the current incomplete block is updated at every tick. Each block stores one vector and gates it at its current physical log-midpoint, just as Graded-1 does for its unequal-width intervals. Adjacent graded records are merged by a young-to-old maximal sweep. Graded-2 gates both endpoint entries and uses the positive barycentric remapping at each merge. A singleton allocates two entries, one of which is zero. Recent history keeps the latest L singleton writes and gates each at its true age. Its truncation is part of that baseline; $H > T$ eliminates truncation only for the three cohort compressors.

Offline matrix oracle. At each time, an independent full-history reference gates each birth matrix at its true age. A separate coefficient implementation expands the linear compressed update along its actual record history. Neither the full-history reference nor these expanded matrices are stored by the online implementation. In the common Euclidean norm, define

$$F(\theta) = (1-q)\sum_j \|C_j^\top \ell(\theta)\|_2, \qquad \ell(\theta) = (\cos\theta, \sin\theta), \qquad L = (1-q)\sum_j \|C_j\|_2.$$

The reverse triangle inequality implies $|F(\theta) - F(\theta')| \le L|\theta - \theta'|$. For $N = 512$ evenly spaced circular directions,

$$\max_k F(2\pi k/N) \le E(t) \le \max_k F(2\pi k/N) + \pi L/N.$$

Taking separate maxima over $t < T$ gives the reported interval for the conditional worst-write error over the finite horizon. At the time and direction maximizing the grid lower bound, we construct $u_j = C_j^\top \ell / \|C_j^\top \ell\|_2$, taking $u_j = 0$ if its denominator vanishes. Actual online and independent exact-vector replays attain the lower bound in projection onto ℓ. All 24 replays agree with the coefficient calculation to within 6.67×10^{-16} in Euclidean norm. These computations use float64 and include the analytic grid remainder but not an interval-arithmetic bound for roundoff.

We also record retained-only and deleted-only angular intervals. For matrix gates, independently maximizing the two components does not in general make their sum equal the total error. In the main cohort runs the deleted component is zero; recent history has exact retained states and a nonzero deleted tail. Table 3 gives all main intervals.

Two checks with theorem parameters and deletion. Separately from the storage-constrained comparison, we use $\epsilon = 0.2$, $\mu = 1/64$, the irregular path, and $T = 192$ for both theorem parameterizations. Both choose $H = 174$. Graded-1 uses $\eta = 0.005555555556$, reaches 132 vector slots, and deletes nine whole records. Its conditional-error upper endpoint is 0.0157354, below the theorem envelope 0.1048461. Graded-2 uses $\eta = 0.01168213139$, reaches 192 vector slots,

and deletes four whole records. Its upper endpoint is 0.0142017, below the envelope 0.0972824. The separate retained-only upper endpoints are 4.894×10^{-5} and 3.690×10^{-8}. These are two implementation checks of sufficient budgets, with no empirical asymptotic claim.

Control-order check. The common warmup is 64 alternating-control ticks with $u_t = (1, 0)$. The next 64 writes are zero; the two continuations are $0^{32}1^{32}$ and $1^{32}0^{32}$. Let $\Delta = y_{0^{32}1^{32}} - y_{1^{32}0^{32}}$. The exact response has $\|\Delta\|_2 = 0.002407031513$. We reuse the parameters selected for the 24-vector cap in the 192-tick experiment. Actual peaks on this shorter sequence are shown separately in Table 4. The check tests preservation of a particular controlled response, rather than an expressivity lower bound for other recurrent model classes or dominance for every write stream.

Resource accounting and reproducibility. For $d = 2$ float64 vectors and two int64 endpoints per cohort, the logical packed payload is $16V + 16K + 16$ bytes, where V counts vector entries and K counts records. A recent circular buffer uses $16L + 16$ bytes, since birth indices follow from its position and counters. These are explicit packed encodings, not the Python object size. End-of-tick peaks exclude the temporary insertion record and constant working vectors; traced allocations include online temporaries. Python's `tracemalloc` measures traced Python and NumPy allocations, not process RSS or every native allocation. It is run separately from timing. Three timed replays per path follow one warmup replay; timings include current gating, insertion, merging, deletion, and sum readout. They exclude the offline coefficient checker and output-history storage. The supplementary JSON records all timing samples, allocation measurements, gate counts, environment versions, and input hashes.

The full work consists of 24 main parameter/path configurations (4608 problem ticks), two theorem-parameter checks (384 ticks), eight control-order sequence replays, and 24 constructed-write verification replays. Count-only calibration and repeated CPU timing/allocation replays are auxiliary work, not additional independent test settings. No model is fitted and no GPU is used. All scripts, configurations, frozen control paths, CSV/JSON outputs, and figure generators are included in the supplementary source tree.

Method	Control	Vectors	Lower	Upper	Retained upper
Graded-1	fast	12	0.0058939	0.00593101	0.00593101
Graded-1	irregular	12	0.00655005	0.00659164	0.00659164
Graded-1	fast	24	0.00258951	0.00260582	0.00260582
Graded-1	irregular	24	0.00270947	0.00272676	0.00272676
Graded-1	fast	48	0.0013141	0.00132225	0.00132225
Graded-1	irregular	48	0.00133631	0.00134478	0.00134478
Graded-2	fast	12	0.00329425	0.00331559	0.00331559
Graded-2	irregular	12	0.00332164	0.0033428	0.0033428
Graded-2	fast	24	0.00107245	0.00107911	0.00107911
Graded-2	irregular	24	0.00104546	0.00105209	0.00105209
Graded-2	fast	48	0.000176501	0.000177592	0.000177592
Graded-2	irregular	48	0.000164482	0.00016553	0.00016553
Uniform	fast	12	0.0152927	0.015387	0.015387
Uniform	irregular	12	0.0164066	0.0165109	0.0165109
Uniform	fast	24	0.00764056	0.00768769	0.00768769
Uniform	irregular	24	0.00844185	0.00849548	0.00849548
Uniform	fast	48	0.00375477	0.00377794	0.00377794
Uniform	irregular	48	0.00426257	0.0042896	0.0042896
Recent	fast	12	0.764068	0.768757	0
Recent	irregular	12	0.764939	0.769654	0
Recent	fast	24	0.624269	0.6281	0
Recent	irregular	24	0.624602	0.628452	0
Recent	fast	48	0.413156	0.415692	0
Recent	irregular	48	0.413005	0.41555	0

Table 3: All 24 main comparisons, with actual total-vector peaks. Lower and upper refer to $\max_{t<T} E(t)$ for the indicated fixed path. The retained upper column isolates error in births still represented at that time.

Method	Vectors	$\|\widehat{\Delta}\|_2$	$\|\widehat{\Delta} - \Delta\|_2$	Max. endpoint error
Graded-1	21	0.00240465	3.05376e-06	1.08501e-05
Graded-2	22	0.00245953	5.29051e-05	0.000230418
Uniform	16	0.00246884	6.20683e-05	0.000263627
Recent	24	0	0.00240703	0.227095

Table 4: Control-order responses with reused parameters. The last column is the larger of the two endpoint response errors against full history.

Method	Records	Payload (B)	Traced (KiB)	Gates/tick	μs/tick
Graded-1	24	784	13.5	17.19	112.1
Graded-2	12	592	11.1	17.11	64.8
Uniform	24	784	13.5	12.38	79.8
Recent	24	400	13.5	22.44	144.3

Table 5: Resources at 24 peak retained vectors and $T = 192$. Payload assumes float64 vectors and int64 metadata; traced allocation includes online temporaries and excludes the offline checker. Gates/tick is the mean count. Time is the median of six Python CPU measurements (three per control path), with tracing disabled; these timings characterize this reference implementation.