STABLE LOCI ADDRESSING: A SHARED-STATE ARCHITECTURE FOR REFERENTIAL STABILITY IN EVOLVING LONG-TERM LLM MEMORY

Anonymous authors
Paper under double-blind review

ABSTRACT

Long-term memory for large language models (LLMs) is usually treated as a retrieval problem: given a query, find the relevant stored information. But a memory that changes over time faces another problem. Once a computation refers to a record, later updates should not make that reference silently point elsewhere. We call this requirement *referential stability* and show that exact revisitation requires retained information about the earlier choice. We introduce STABLE LOCI ADDRESSING (SLA), which unifies persistent indirection, rebinding, and structured locus operations within one shared-state memory interface without a separate per-record reverse-locator layer. Shared writes can therefore affect other references. We derive exact RLS and Delta one-write laws, an exact cumulative native-Delta law, and a necessary-and-sufficient Top-1 certificate. Experiments span semantic conflict, learned cue embeddings, long mutation sequences, LongMemEval content and structural evolution, cross-bank maintenance, and zero-trainable-parameter integration with a frozen DeltaNet. Raw semantic keys can initialize correctly yet fail after relocation. Verified dereference prevents wrong-record returns, and the certificate tracks failure. On 70 LongMemEval-S knowledge-update questions, stale predecessor evidence lowers answer accuracy by 68.6 percentage points, while SLA exactly matches updated-pointer semantics. Together, these results show static retrieval accuracy is insufficient for evolving shared memory and establish referential stability as a distinct, testable correctness criterion.

1 INTRODUCTION

Long-term LLM memory is usually framed around semantic retrieval: retrieving useful information beyond the active context (Liu et al., 2024; Khandelwal et al., 2020; Lewis et al., 2020; Borgeaud et al., 2022; Wu et al., 2022; Wang et al., 2023). Persistent memories, however, are also revised, reorganized, and revisited (Wu et al., 2025; Xu et al., 2025; Fang et al., 2026; Shi et al., 2026; Xu et al., 2026). Fast-weight, delta-rule, and test-time memories make this distinction especially important because they retain history in shared state that changes online (Schlag et al., 2021; Yang et al., 2024; Behrouz et al., 2025). Once semantic retrieval has selected a record, what makes a later access still refer to *that same record* after its content or location changes?

We call this requirement *referential stability*. Semantic relevance and storage location can change while logical identity remains fixed, so an existing reference must still identify its original record even when a later query favors another. Two histories may establish different referents yet lead to the same query q_t and candidate state $\mathcal{C}_t$. Current retrieval alone cannot recover the earlier choice. Section 3.1 proves a retained-information lower bound that applies to identifiers, pointers, dictionary entries, retained handles, and equivalent historical state. Separating identity from location also prevents silent aliasing when a slot is reused, as in DNC de-allocation/reuse (Csordás & Schmidhuber, 2019).

With mutable placement, indirection is unavoidable. STABLE LOCI ADDRESSING (SLA) unifies persistent addressing, rebinding, and structured locus operations within one shared-state memory interface, avoiding a separate per-record reverse-locator layer and its synchronization with evolving memory. In SLA, a protocol-stable key recovers a candidate locus, and a forward registry verifies the occupant before returning the payload (Fig. 1). Shared writes can perturb other references, so correct

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.

initialization does not imply stability after later updates. We derive exact target and collateral laws, their cumulative native-Delta form, and conditions for correct decoding.

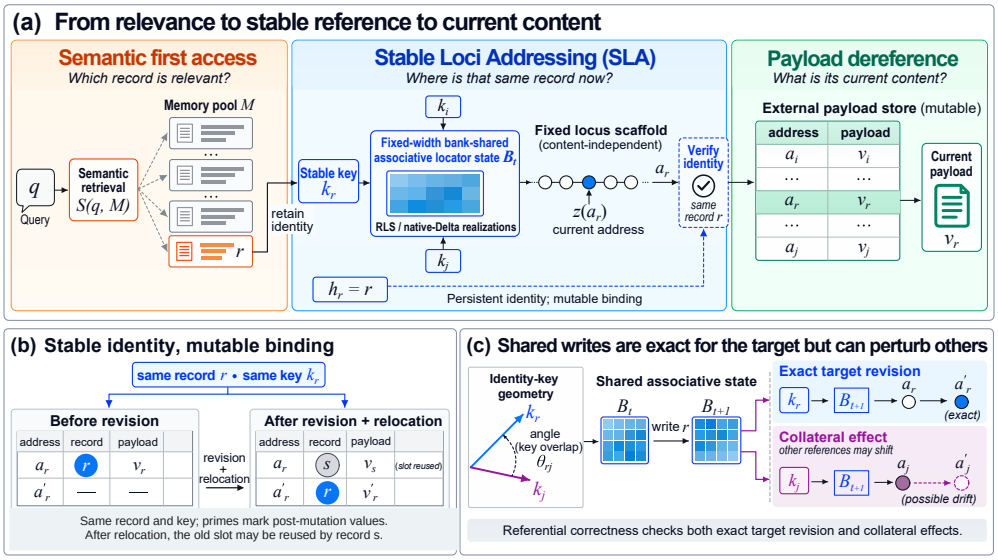


Figure 1: **SLA separates relevance, persistent identity, and current content.** (a) After semantic retrieval selects record r, SLA uses its stable key and bank-shared binder to recover the address, verify identity, and return the current payload. (b) Subscripts identify records, and primes mark updated values. Record r retains k_r as its address changes from a_r to a'_r and its payload from v_r to v'_r. Another record s with payload v_s can reuse the vacated slot. (c) Under native Delta, a shared write updates the target exactly while inducing overlap-dependent collateral change in other references.

Our contributions are threefold.

- **Referential stability as a memory contract.** We separate current relevance from established denotation, formalize identity through payload and locator changes, and prove a realization-independent retained-state lower bound.
- **Shared-state indirection and certification.** SLA unifies persistent addressing, rebinding, and structured locus operations within one shared-state memory interface without a separate reverse-locator layer. Exact RLS/Delta laws, cumulative native-Delta dynamics, a necessary-and-sufficient Top-1 certificate, verified dereference, and conservative maintenance characterize this regime.
- **Evidence across evolving memories.** Initial lookup accuracy can mask mutation-induced failure: raw semantic keys lose references after relocation, the certificate tracks the transition, and cross-bank stress validates maintenance. LongMemEval isolates stale-state effects with matched SLA/pointer controls, while frozen models and systems measurements connect the contract to outputs and cost.

SLA operates after semantic retrieval establishes a logical record identity. Retrieval, consolidation, forgetting, conflict resolution, and memory learning determine what exists and is selected, while SLA preserves established references as memory changes.

2 Related Work

Retrieval and evolving long-term memory. Retrieval-augmented and memory-augmented large language models (LLMs) extend usable context by selecting, organizing, or compressing external evidence (Khandelwal et al., 2020; Lewis et al., 2020; Borgeaud et al., 2022; Wu et al., 2022; Wang et al., 2023; Mohtashami & Jaggi, 2023; Sarthi et al., 2024; Xiao et al., 2024; Fountas et al., 2025). LongMemEval separates indexing, retrieval, reading, update, and abstention (Wu et al., 2025). Newer

agent-memory systems explicitly consolidate, restructure, revise, or revisit stored experience (Zhong et al., 2024; Gutiérrez et al., 2025; Xu et al., 2025; Fang et al., 2026; Xu et al., 2026; Shi et al., 2026; Tang et al., 2026; Ko et al., 2026). These methods primarily focus on what should be stored, organized, or retrieved at the current time. Referential stability instead asks a complementary question: once a computation has established which logical record it denotes, what retained state preserves that denotation through subsequent content and placement changes? Our history-distinguishability result in Section 3.1 formalizes why current semantic content alone cannot, in general, recover an established referent.

Indirection and shared associative state. Classical indirection separates stable program identity from mutable placement (Denning, 1970). A pointer or dictionary stores the current locator explicitly for each record, while a forward registry scan reconstructs it from address-indexed state at linear bank-local cost. Distributed working memory supports pointer-like binding (Kriete et al., 2013), and Memory Networks, NTMs, and DNCs provide addressable memory (Sukhbaatar et al., 2015; Miller et al., 2016; Graves et al., 2014; 2016). DNC slot reuse also shows how physical reuse can alias logical identity if the two are not separated (Csordás & Schmidhuber, 2019). MemGPT manages storage tiers (Packer et al., 2023). Fast weights, DeltaNet, and Titans instead keep writable history in shared online state (Schlag et al., 2021; Yang et al., 2024; Behrouz et al., 2025). Key–value theory further distinguishes storage fidelity from retrieval discriminability (Gershman et al., 2025). These systems suggest three ways to represent locator state: an explicit reverse locator, a forward scan, or a shared writable representation. SLA studies the third case, where one update can change the state used by other references and therefore introduces interaction, capacity, and decoding questions that do not arise from independent pointer updates.

Structured distributed state. Sinusoidal/RoPE phase structure, holographic/vector-symbolic representations, and fixed or cognitive spatial scaffolds motivate shared algebraic operations on fixed-width states (Vaswani et al., 2017; Shaw et al., 2018; Su et al., 2024; Plate, 1995; Kleyko et al., 2022; Chandra et al., 2025; Whittington et al., 2020; Sharma et al., 2022). Episodic-memory and method-of-loci studies also distinguish durable scaffold state from newly bound content (Maguire et al., 2003; Dresler et al., 2017; Wagner et al., 2021; Huet et al., 2025; Li et al., 2026). We use these ideas for relation-state realizations after establishing a persistent reference. Symbolic coordinates and adjacency provide exact controls that separate relation semantics from representation choice.

3 Stable Loci Addressing

3.1 Referential state and shared mutable indirection

Contract and necessity. At semantic-access time τ, an upstream selector may establish $\mathrm{Resolve}(q_\tau, \mathcal{C}_\tau) = h_r$. A fresh resolution and a later revisit of that established reference are different operations.

Definition 1 (Referential stability)**.** *An established reference to active record r is referentially stable if every later revisit returns the current state of the same logical record despite admissible changes to payloads, locators, other records, and semantic rankings. After retirement, revisit returns $\perp$ rather than another record.*

This contract is history-conditioned. Let the current semantic observation be $O_t = (q_t, \mathcal{C}_t)$ and let $S_t \in \mathcal{S}$ denote any state retained from the earlier resolution event.

Theorem 1 (Retained-state necessity under history indistinguishability)**.** *If two admissible histories have the same current observation O_t but their established references denote different active logical records, no zero-error revisit rule whose input is only O_t can be referentially stable on both. More generally, if m such histories require m distinct referents, any deterministic exactly stable mechanism must realize at least m distinguishable retained states, so $|\mathcal{S}| \geq m$. A finite binary encoding therefore needs at least $\lceil \log_2 m \rceil$ bits in the worst case.*

Proof. A current-state-only rule receives the same input on indistinguishable histories and therefore cannot return two different records with probability one. If fewer than m retained states are available, two of the m referent classes share both O_t and S_t by the pigeonhole principle and must receive the same output, contradicting exact stability. □

An explicit identifier carried into a later query, a pointer, or any equivalent token of the earlier choice is itself history-bearing state. The theorem therefore establishes the need for retained identity without privileging pointer, scan, or SLA realizations.

State and operations. SLA retains an opaque handle h_r as logical identity and canonical cue metadata c_r that deterministically regenerates a protocol-stable key $k_r = \kappa_{\mathcal{P}}(c_r)$. Its logical state is

$$h_r \equiv r, \qquad x_r = (h_r, \ell_r, v_r, u_r), \qquad \ell_r^{\mathrm{SLA}} = (s_r, a_r), \tag{1}$$

where v_r is the external payload and u_r is optional relation state. This tuple describes logical state, not an additional handle-indexed table. Resolve selects a record, Reinstate and Dereference revisit it, and Insert, Revise, and Retire mutate memory. With ϱ denoting the memory realization, the interface is

$$\begin{aligned}
&\mathrm{Resolve}(q, \mathcal{C}_t) \to h_r \text{ or } \bot, && \mathrm{Reinstate}_\varrho(h_r) \to \ell_r \text{ or } \bot,\\
&\mathrm{Dereference}_\varrho(h_r, \ell_r) \to v_r \text{ or } \bot, && \mathrm{Insert}_\varrho(\eta, v, u) \to h_{\mathrm{new}} \text{ or } \bot,\\
&\mathrm{RevisePayload}_\varrho(h_r, v_r') \to \mathsf{ok} \text{ or } \bot, && \mathrm{ReviseLocator}_\varrho(h_r, \ell_r') \to \mathsf{ok} \text{ or } \bot,\\
&\mathrm{Retire}_\varrho(h_r) \to \mathsf{ok} \text{ or } \bot.
\end{aligned} \tag{2}$$

Here η supplies admissible canonical-cue and routing metadata for a new record. Control state and payload storage are represented separately as

$$\mathfrak{M}_\varrho = (\Theta_\varrho, \{\mathcal{V}_s\}_s), \tag{3}$$

where Θ_ϱ contains scope, active-handle and occupancy state, immutable metadata for reconstructing keys from canonical cues, the forward registry, and the shared locator state for the chosen realization. Retained handle, cue, and scope metadata never store the current bank-local address. They therefore cannot act as a hidden reverse locator. A shared binder $B_s \in \mathbb{R}^{d_z \times d_k}$ is fitted only to stable key–locus associations,

$$B_s^\star = \arg\min_B \sum_{r \in \mathcal{R}_s} \|Bk_r - z(a_r)\|_2^2 + \lambda \|B\|_F^2, \tag{4}$$

while $\mathcal{D}_s$ stores the occupant identity and $\mathcal{V}_s$ the external payload. For row-stacked $K \in \mathbb{R}^{n_s \times d_k}$ and $Z \in \mathbb{R}^{n_s \times d_z}$, the static dual-ridge solution is

$$B_s^\star = Z^\top (KK^\top + \lambda I_{n_s})^{-1} K. \tag{5}$$

Insertion-only RLS yields the same ridge solution (Appendix B.5). The primary setting uses at most 1,024 records/bank, $d_k = 4096$, and $d_z = 16$. Each unit-norm locus $z(a)$ concatenates eight pairs $[\cos(2\pi a/p_m), \sin(2\pi a/p_m)]/\sqrt{8}$, where $(p_1, \ldots, p_8) = (5, 7, 11, 13, 17, 19, 23, 29)$. These codes uniquely identify addresses within a bank.

Reinstatement decodes

$$\tilde{z}_r = B_{s_r} k_r, \qquad \hat{a}_r = \arg\max_{a \in \mathcal{A}_{s_r}^{\mathrm{all}}} \tilde{z}_r^\top z(a), \tag{6}$$

and verifies the candidate through $\mathcal{D}_{s_r}(\hat{a}_r) = h_r$ before returning $\mathcal{V}_{s_r}(\hat{a}_r)$. The binder locates the candidate, while the forward registry checks identity. Active-handle state determines liveness, routing selects the authoritative bank, and mutations become visible as complete committed states. Together, these checks prevent wrong-record returns when control metadata is trusted. The binder analysis describes localization and drift across references.

Payload revision changes only the verified external payload. Same-bank locator revision moves registry and payload occupancy together with the binder association. A failed revision leaves the public state unchanged. If semantic retrieval is payload-derived, its index is refreshed upstream rather than conflated with post-retrieval mutation. Writing X for the memory's induced logical state, successful revisions make the following changes:

$$\begin{aligned}
&\text{Payload revision:} && X'_{\mathrm{payload}} = X[r : v_r \leftarrow v'], && \mathcal{V}'_{s_r}(a_r) = v',\\
&\text{Locator revision:} && X'_{\mathrm{locator}} = X[r : \ell_r \leftarrow (s_r, a_r')], && \big(\mathcal{D}'_{s_r}(a_r'), \mathcal{V}'_{s_r}(a_r')\big) = (h_r, v_r),\\
& && \big(\mathcal{D}'_{s_r}(a_r), \mathcal{V}'_{s_r}(a_r)\big) = (\bot, \bot), && \mathcal{A}_{s_r}^{\mathrm{occ}\,\prime} = \big(\mathcal{A}_{s_r}^{\mathrm{occ}} \setminus \{a_r\}\big) \cup \{a_r'\}.
\end{aligned} \tag{7}$$

Retirement invalidates the handle before slot reuse. Rebinding repeats the complete locator move, including registry and payload updates. The contract assumes atomic, linearizable commits and leaves unsynchronized writes and crash recovery to the surrounding system. Appendix B gives the complete interface, state-factorization table, trust boundary, and merge/failure semantics.

3.2 ONE-WRITE AND SEQUENCE-LEVEL INTERFERENCE

Locator revision moves a record to a free same-bank address a'_r while preserving h_r, c_r, k_r and logical identity. Let $\epsilon = z(a'_r) - Bk_r$. For covariance-preconditioned RLS with $P = P^\top \succ 0$ and native Delta,

$$B'_{\mathrm{RLS}} = B + \epsilon \left(\frac{Pk_r}{k_r^\top Pk_r} \right)^\top, \tag{8}$$

$$B'_\Delta = B + \epsilon k_r^\top, \qquad \|k_r\|_2 = 1. \tag{9}$$

Both write the target exactly, while a non-target key k_j changes by

$$(B'_{\mathrm{RLS}} - B)k_j = \epsilon \frac{k_r^\top Pk_j}{k_r^\top Pk_r}, \qquad (B'_\Delta - B)k_j = \epsilon (k_r^\top k_j). \tag{10}$$

Proposition 1 (Exact target relocation and realization-specific non-target change)**.** *Under Eqs.* (8)–(9)*, both realizations recover $z(a'_r)$ exactly for k_r, and Eq.* (10) *gives the induced non-target change.*

Corollary 1 (Overlap conditions)**.** *RLS induces zero non-target change on Bk_j when $k_r^\top Pk_j = 0$, whereas Delta does so when $k_r^\top k_j = 0$. In addition,*

$$\|(B'_{\mathrm{RLS}} - B)k_j\|_2 \le \|\epsilon\|_2 \sqrt{\frac{k_j^\top Pk_j}{k_r^\top Pk_r}}, \tag{11}$$

while Delta non-target change is exactly proportional to $|k_r^\top k_j|$.

For independent normalized Gaussian keys, $k_r^\top k_j$ is approximately zero-mean with variance $1/d_k$, so $\mathbb{E}|k_r^\top k_j| \approx \sqrt{2/(\pi d_k)} = 0.01247$ at $d_k = 4096$ and typical Delta collateral displacement scales as $O(d_k^{-1/2})\|\epsilon\|_2$. Operational failure also depends on decoding margin: for unit-norm loci with pre-write margin $\mu_j > 0$, any perturbation δ_j with $\|\delta_j\|_2 < \mu_j/2$ preserves the intended Top-1. RQ2 tests both the vector law and margin-mediated failure regime.

In exact arithmetic, the single-write law extends without stochastic assumptions. A *protected segment* tracks record j after insertion or a direct write, ending before its next direct write, retirement, migration, or bank rebuild. Migration and rebuilding start new segments for active records. Under native-Delta external writes $B_t = B_{t-1} + \epsilon_t k_{r_t}^\top$, $r_t \neq j$:

Theorem 2 (Exact cumulative interference)**.** *For every prefix u of the segment,*

$$\Delta_j(u) := (B_u - B_\tau)k_j = \sum_{t=\tau+1}^{u} \epsilon_t (k_{r_t}^\top k_j). \tag{12}$$

Telescoping $B_u - B_\tau = \sum_{t=\tau+1}^{u} \epsilon_t k_{r_t}^\top$ and right-multiplying by k_j proves the identity for the realized sequence, without iid, zero-mean, or random-walk assumptions. If $g_{j,a}^{(\tau)} = (B_\tau k_j)^\top [z(a_j) - z(a)]$, the exact current competitor gap is

$$g_{j,a}(u) = g_{j,a}^{(\tau)} + \Delta_j(u)^\top [z(a_j) - z(a)]. \tag{13}$$

Corollary 2 (Exact sequence-level Top-1 certificate)**.** *The intended locus remains unique Top-1 exactly when $\min_{a \neq a_j} g_{j,a}(u) > 0$. If $\mu_j^{(\tau)} = \min_{a \neq a_j} g_{j,a}^{(\tau)} > 0$ and loci are unit norm, $\|\Delta_j(u)\|_2 < \mu_j^{(\tau)}/2$ is sufficient.*

The norm condition and the more conservative triangle envelope discard score geometry. Appendix B.5 gives proofs, complete segment semantics, and worst-case diagnostics.

3.3 PRETRAINED AND STRUCTURED EXTENSIONS

A dedicated SLA Delta matrix persists alongside a frozen DeltaNet cache. It changes only on explicit memory mutations, while DeltaNet's recurrent activations evolve normally. A retained handle supplies verified evidence to the frozen model,

$$h_r \to (s_r, k_r) \to B_{s_r}^{\Delta} k_r \to \hat{a}_r, \quad \mathcal{D}_{s_r}(\hat{a}_r) = h_r \Longrightarrow v_r = \mathcal{V}_{s_r}(\hat{a}_r) \to \mathrm{LM}_\theta(\mathrm{Prompt}(v_r)). \tag{14}$$

Thus SLA integrates with a frozen pretrained DeltaNet through its persistent cache lifecycle with zero new trainable parameters (Appendix D.4).

Verified references also support downstream relation actions. We compare additive, commuting multi-axis, non-commutative branching, phase-based, symbolic, adjacency, and learned random-code realizations. Address-coupled relations use fixed layouts, whereas dynamic trajectory and session relations use coordinates independent of storage (Appendix C.4).

4 EXPERIMENTS

The four questions address identity, interference, natural evolution, and system trade-offs. Independent units are revision trials, controlled conditions or seeds, routes, and benchmark questions. Checkpoint and prefix counts give exhaustive within-unit coverage, not additional independent samples. Paired 95% intervals use 10,000 unit-level resamples (Appendix A).

4.1 RQ1: REFERENTIAL IDENTITY UNDER SEMANTIC CONFLICT AND MUTATION

After target resolution, we make a distractor payload semantically preferable. Payload RAG ranks the established target first in only 0.000–0.005 of cases, whereas Stable-random and Full SLA reinstate it at 1.000. Target hit@8 is 1.000 for all, ruling out simple evidence absence. A matched nested multiplicity control keeps the target semantically above every distractor and all methods at 1.000. Free-form re-address remains an upstream semantic stage: across four paraphrase families, resolver/address Top-1 is 0.9545 overall and strict all-four reinstatement 0.866.

Table 1 summarizes target revision, continuous mutation, independent preservation audits, and matched RLS/Delta transfer. Obsolete addresses sometimes remain lower in Top-8 (at most 0.014), but never change verified Top-1 or return the obsolete payload.

Table 1: **RQ1 controlled evidence.** Complete strata, checkpoint schedules, and preservation audits are in Appendix C.1.

Evidence	Result
Semantic conflict	Payload-RAG Top-1 is 0–0.005, versus 1.0 for stable reinstatement. Target hit@8 is 1.000 for all.
Target revision	After relocation, overwrite, and eight rebindings, the current target and payload are exact. Obsolete Top-1 and payload return are 0.
Checkpoint preservation	Each matched audit preserves 4,103,500/4,103,500 non-target checks. Independent replication with different key geometry agrees.
Continuous lifecycle	Each RLS/Delta backend preserves 7,500/7,500 target and 3,920,000/3,920,000 non-target checks, including 87,500 inserted-record observations. Wrong/fail/mismatch $= 0$.
Matched backend transfer	With identical keys/checkpoints, RLS $\leftrightarrow$ native Delta recovers the target exactly and preserves 4,103,500/4,103,500 collateral checks per realization.

4.2 RQ2: NATIVE-DELTA INTERFERENCE FROM ONE WRITE TO LONG SEQUENCES

A 190-condition signed-overlap intervention (two bank sizes and five seeds) matches Eq. (10) to FP32 precision, including the predicted sign reversal for negative overlap. Target recovery remains exact. The correlated non-target remains correct through $|\rho| = 0.32$. In the 1,024-slot stress, the first Top-1 and verified-dereference degradation appears at $+0.50$. All 145,350 zero-overlap non-target checks remain exact. The content-independent keys used in the main experiments are much less correlated: across 500 frozen banks (65.4M pairs), their mean/max absolute overlap is 0.01247/0.0895.

We also test raw normalized 1,024-D BGE-M3 and E5-large-v2 cue embeddings against the 4,096-D SHA control across five seeds and bank sizes 128–1,024. Both encoders are frozen, with no whitening or fitted transform. Their maximum absolute cosines are 0.9942 and 0.9967, compared with the earlier 500-bank SHA reference maximum of 0.0895. Both initialize perfectly through load 768. At 1,024, BGE-M3 achieves 5103/5120 Top-1 and 5119/5120 Top-8, while E5 achieves 5093/5120 and 5120/5120. After relocation, non-target Top-1 and Top-8 degrade for both learned sources at every tested load. Registry verification yields 273,927 fail-closed outcomes among 658,245 non-overlapping checkpoint observations, with zero wrong-record returns. The exact certificate agrees

with observed Top-1 on all 1,024,725 prefixes across the three sources, with maximum projection residual 2.58×10^{-5}. The law therefore applies across normalized key sources, but reliable updates depend on geometry and margin (Appendix B.4). This identifies identity-key geometry as the design variable. Learned keys should control coherence and margin while preserving or explicitly rebinding established identities.

Frozen continuous and LongMemEval-V2 replays preserve Top-1 and match the exact certificate at every prefix (Table 2), with maximum FP32 residual below 9.1×10^{-7}. The norm bound certifies 95.9% of continuous and 95.4% of natural prefixes, compared with 26.9% and 33.6% for the triangle bound. Failure of either sufficient bound alone does not imply a decoding failure. Appendix C.3 reports segment counts, coverage, headroom, and residual distributions.

Table 2: **RQ2 sequence certificate.** Prefixes are checked exhaustively within protected segments. Repeated checks within a segment are not independent samples.

Trace	Prefixes	Exact cert.	Top-1	Min. gap
Continuous stress	6,519,500	100%	100%	0.1311
LongMemEval-V2	1,092,979	100%	100%	0.1364

Cross-bank high-churn stress. Across eight banks with capacity 1,024 and initial load 512/bank, a serialized 20,000-operation trace includes 3,000 migrations and 11,000 Delta writes. An atomic move publishes a verified destination, changes routing, and clears the source registry and payload. The source binder stays unchanged. Every migrated target is correct before and after commit. Source-bank non-targets are preserved in all migrations, with zero wrong-record returns or retired-handle aliases. Accumulating destination writes cause 444 collateral misses in 384 migration events. Without maintenance, checkpoint Top-1 and Top-8 first fail at operations 5,700 and 16,300, despite loads of only 468–548/bank. The exact certificate tracks this loss of margin.

Migration ends the source segment, and the committed destination write starts a new baseline for later destination-bank interference. On the identical trace, a candidate write that violates the sufficient norm condition $\|\Delta_j\|_2 < \mu_j/2$ triggers a rebuild of the affected bank before publication. This single untuned rule replaces 1,744 candidate writes, while 9,256 commit normally. Rebuilds fit current active bindings, exclude retired or migrated-away records, and verify the whole bank before commit. Every rebuild restores all active bindings. All 201 checkpoints retain Top-1/Top-8 and verified payloads, with zero active misses, wrong returns, or rebuild-verification failures, at $2.13\times$ baseline runtime. The exact score certificate diagnoses decoding, while the norm bound conservatively triggers maintenance (Appendix C.2).

4.3 RQ3: Natural content evolution and pretrained inference

Natural knowledge updates and downstream effect. Of LongMemEval-S's 78 released knowledge-update questions, 70 meet frozen eligibility after excluding six abstentions and two cases lacking old turn-level provenance. Relocation, insertion, natural payload update, and rebinding preserve the same identity and current payload in all 420 state checks, with no obsolete returns, wrong returns, fail-closed misses, or registry mismatches and all 22,300 unaffected-record checks correct. Appendix D.2 details eligibility and the 210 post-update states.

Qwen3-8B receives updated SLA, an updated pointer, or the stale predecessor with question, non-target evidence, prompt, model, and checker fixed. SLA exactly matches updated-pointer evidence, inputs, outputs, and labels in all 70 cases. Both achieve 56/70 (80.0%) correctness versus 8/70 (11.4%) for stale state, a paired difference of $+68.57$ percentage points (95% CI 52.86–82.86). Updated and stale outputs differ in all 70 cases. The 55 stale-wrong/updated-correct transitions and seven reversals isolate the effect of state freshness, while the two updated locators remain equivalent (Fig. 2).

With the key already available, composing measured locator, evidence/prompt/tokenization, and common reader stages gives 5,743.498 ms/question for pointer and 5,743.529 ms/question for SLA. The increment is 0.030785 ms (0.000536%). These are composed costs for 70 frozen inputs, not directly timed pipeline totals, and semantic retrieval is not invoked. The 69-case timestamp sensitivity remains 79.7% versus 11.6%, with a local answer-containment check agreeing in direction (Appendix D.2).

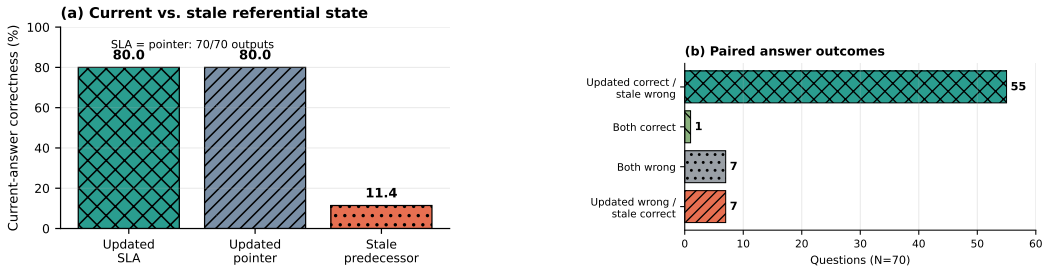


Figure 2: **RQ3 natural content evolution affects downstream answers.** (a) Updated SLA and pointer behave identically, while the stale predecessor reduces current-answer correctness. The contrast isolates state freshness. (b) Paired transitions show how individual answers change.

Structural evolution and pretrained composition. LongMemEval-V2 (Wu et al., 2026) tests storage changes with factual content fixed. Semantic/evidence policies change downstream scores, whereas Hybrid-SLA and cached-pointer locators preserve the selected evidence and score. Across 439 eligible questions, 24,369 relocations, insertions, and rebindings act on 2,475 natural seed records. Native Delta preserves all 22,275 retained checks, 369,485 non-target checks, and 4,776 valid relation transitions, with zero wrong returns. Independent Qwen3.5-9B verification reproduces baseline text, tokens, hashes, and scores for all 439 questions. The separately measured 439-question query-stage timings are reported in Appendix D.3, together with the all-451 statistics and Fig. D.4.

SLA matches updated-pointer evidence before generation. Qwen3-8B tests stale versus current content, Qwen3.5-9B tests structural preservation, Mistral-7B tests first-access sensitivity, and DeltaNet tests recurrent-model composition. The frozen 1.366B-parameter DeltaNet uses 65,536 FP32 SLA state elements and no new trainable parameters. SLA matches updated-pointer semantics in 11/11 base states and 700 comparisons across 100 held-out cases, with zero final-token logit difference under the 10^{-6} criterion. After rebinding, stale, shuffled-address, and empty/reset controls each change payload, complete input, and logits in 100/100 cases (Appendices D.1–D.4).

4.4 RQ4: Structured reuse and design trade-offs

Structured analytic, symbolic, and adjacency operators remain exact through $H = 32$ for additive, commuting, and non-commutative branching relations, while learned random-map algebraic null controls stay at or below 0.0014 Top-1. Periodic and integer outputs agree under the same hard relation decisions, whereas soft superposition and finite beams can differ. Noise exposes distinct finite-dimensional margins (Appendix C.4). The content-independent ridge sweep is exact in tested cells with $d_k \geq n_s$ and degrades when load exceeds key dimension. RQ2 shows that coherence also matters. Routing scales to 10,114 records with at most 1,024 candidates/bank.

At 1,024 slots on the matched CPU, pointer and dictionary address recovery average 0.113 and 0.143 μs. Balanced registry scan and exact SLA average 20.15 and 30.90 μs (Fig. 3). These timings measure resident-key address recovery before payload access. Separately, with scope already retained, the bank-local vector kernel is $9.21\times$ faster than repeated full-corpus exact-vector search at $N = 10{,}114$. Native Delta removes the RLS covariance matrix, reducing writable locator tensors from 64.25 to 0.25 MiB/bank ($257\times$). This ratio compares $B_s{+}P$ with B_s alone (Appendix E).

Verified-dereference boundary and fair state accounting. With a correct registry, stale/wrong candidates, empty slots, retirement/reuse, and stale routing each fail closed in 500/500 trials, with zero wrong returns. Forging both registry identity and payload yields 500/500 wrong returns, locating the trusted-metadata boundary (Appendix B.2).

Charging the same 277,304-B registry/control substrate gives 285,496 B for pointer and 539,448 B for Delta-SLA at 1,024 slots. Payload content and the optional 16-MiB SLA key cache are separate. Baseline totals exclude certificate tracking and rebuild workspace. These measurements distinguish direct-host lookup from shared-state addressing. SLA analyzes the update dynamics of interacting bindings. Appendix E gives all five methods and distinguishes logical storage from CPython overhead.

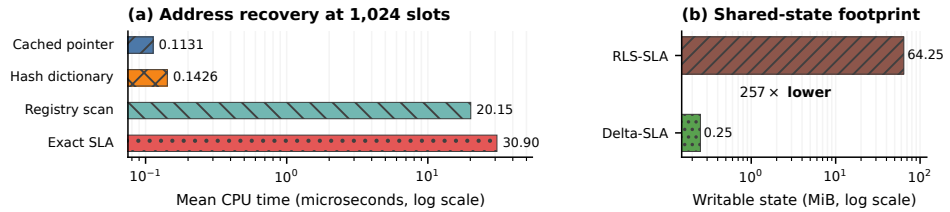


Figure 3: **RQ4 matched state–latency boundary.** (a) Mean matched-host CPU address-recovery time. (b) Writable locator-specific tensors at $d_k = 4096, d_z = 16$: RLS-SLA/Delta-SLA = 64.25/0.25 MiB (257× tensor-only reduction). Common control, registry/store, and model/cache state are excluded and accounted separately in Appendix E.

5 Discussion

Referential stability is different from current retrieval quality. Retrieval asks which record is relevant now, whereas referential stability asks whether a reference established earlier still denotes the same logical record. Theorem 1 shows that exact revisitation requires enough retained information to distinguish histories that otherwise look identical at the current step. SLA unifies associative recovery, rebinding, and locus algebra within one shared-state memory interface instead of coordinating a separate reverse-locator layer with evolving state. Shared writes then create cross-reference interactions. Proposition 1 gives the effect of one write, Theorem 2 gives the cumulative effect, and the score-space certificate states exactly when the intended address remains Top-1. Learned-key stress shows why initialization accuracy is insufficient: coherent semantic keys can decode correctly before mutation yet lose references after one relocation. Capacity therefore depends jointly on load, key geometry, decoding margin, and write history. Identity geometry need not match semantic similarity. Encoder updates must preserve key versions or rebind affected references.

Sequence analysis turns this distinction into an operational rule. The exact score-space certificate identifies loss of Top-1, while the norm-based condition supplies a conservative rebuild trigger. On the matched cross-bank trace, maintenance preserves active references at a measured runtime cost. Fixed-period controls separate correctness certification from maintenance scheduling. These guarantees assume atomic commits and trusted registry, routing, and liveness state so a mislocalized candidate fails closed rather than returning another record.

Natural studies separate content updates from structural changes across reader and recurrent-model interfaces. SLA matches updated-pointer evidence, while stale state changes answers. Systems measurements separate direct lookup, repeated vector access, reader-path cost, and maintenance across their respective timing boundaries. Together, these results make binding dynamics measurable alongside retrieval quality. Other memory components decide what exists and is selected. SLA preserves references to those records as memory changes.

6 Conclusion

Referential stability requires an established reference to keep denoting the same logical record as its content and location change. SLA unifies persistent addressing, rebinding, and structured locus operations within one shared-state memory interface without a separate reverse-locator layer. We derive exact RLS/Delta laws for individual writes, an exact cumulative native-Delta law, and a necessary-and-sufficient condition for retaining the intended Top-1 address. The experiments show that accurate initialization alone does not guarantee stable access after later writes. Highly coherent semantic keys can lose references after relocation, while the exact certificate predicts the transition and verified dereference prevents wrong-record returns. A separate sufficient norm condition guides staged maintenance that preserves active references on the matched 20,000-operation trace. Natural updates, structural changes, cross-bank migration, stale-state controls, frozen models, and measured system costs connect these mechanisms to downstream behavior. Together, the results establish referential stability as a distinct correctness criterion: retrieval selects what is relevant now, whereas referential stability preserves what an earlier reference denotes.

REPRODUCIBILITY STATEMENT

Appendices A–F specify evaluation units, key and relation protocols, benchmark/model revisions, statistical units, sampling identifiers, state-accounting boundaries, and timing boundaries. The controlled RLS/Delta and continuous verified-locator studies preserve their frozen identities, schedules, and numerical contracts. Sequence-level audits replay those frozen protocols and first reproduce the original aggregate outcomes before evaluating cumulative certificates.

The learned-key stress uses 75 frozen source/seed/load cells: protocol-stable SHA keys, raw normalized BGE-M3 cue embeddings, and raw normalized E5-large-v2 cue embeddings. No encoder is trained and no whitening, learned transform, or post-hoc threshold is introduced. The exact certificate is evaluated on all 1,024,725 prefixes across the three matched key sources. The verified-dereference boundary audit reuses 500 frozen identities per fault condition. The cross-bank baseline and certificate-maintained run share one fixed 20,000-operation serialized trace. Every operation type, target, source, destination bank, and destination address matches row-for-row, and the maintained run uses only the prespecified $\|\Delta_j\|_2 < \mu_j/2$ condition to choose between committing the candidate Delta write and rebuilding the affected bank. These stress results are deterministic coverage checks rather than inferential samples.

The LongMemEval-S natural-update set is frozen at 70 questions before mechanism execution or reader inference. Its causal reader arm uses three matched conditions per question, a fixed Qwen3-8B revision and prompt, and a question-level 10,000-draw paired bootstrap. The timing audit reuses the frozen reader measurements, verifies identical updated SLA/pointer evidence and token IDs in 70/70 cases, and measures the previously untimed evidence/prompt/tokenization stage with five warm-up rounds and 50 repetitions per question. Reported pointer/SLA totals are sums of compatible measured stages rather than a single full-pipeline timer, and semantic retrieval is not invoked by this causal-reader protocol. LongMemEval-V2 preserves its 451-question semantic selections, relation decisions, and evidence-construction protocol. Final-state Qwen3.5-9B behavior is independently re-executed on all 439 eligible questions, and the persisted per-question memory-query stages provide the 439-question timing subtotal reported in the paper.

Table F.1 records model/data revisions and central artifact identifiers. The anonymous compact review artifact mirrors the frozen natural-update, native-Delta sequence-audit, and causal-reader bundles with sanitized rows/aggregates and offline claim verifiers. The direct-access appendix separates historical and matched-host measurements, and the supplementary checksum file verifies the packaged figure/data files. Public benchmarks and pretrained models remain referenced at their original sources. Appendix G reports the address-table-free runtime audit, numerical publication check, fixed-period maintenance controls, full maintenance-state accounting, and same-dimensional Gaussian control.

ETHICS STATEMENT

The experiments use procedurally generated controlled data and public long-context or long-term-memory benchmarks. No new human-subject study or personal-data collection was conducted. A deployed system would also need access control, consent, deletion, provenance, and malicious-write policies. Because a shared associative write can affect other references through key overlap, deployment should validate key geometry and write provenance as part of that control boundary. These governance and security mechanisms are outside the evaluated record-state operations. Cognitive and method-of-loci work is used only as computational inspiration for separating durable scaffold state from newly bound content.

REFERENCES

Yushi Bai, Shangqing Tu, Jiajie Zhang, Hao Peng, Xiaozhi Wang, Xin Lv, Shulin Cao, Jiazheng Xu, Lei Hou, Yuxiao Dong, Jie Tang, and Juanzi Li. LongBench v2: Towards deeper understanding and reasoning on realistic long-context multitasks. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 3639–3664. Association for Computational Linguistics, 2025. doi: 10.18653/v1/2025.acl-long.183. URL https://aclanthology.org/2025.acl-long.183/.

Ali Behrouz, Peilin Zhong, and Vahab Mirrokni. Titans: Learning to memorize at test time. In *Advances in Neural Information Processing Systems*, volume 38, pp. 125925–125962, 2025. doi: 10.52202/085713-3786. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/a4ca07aa108036f80cbb5b82285fd4b1-Abstract-Conference.html.

Sebastian Borgeaud, Arthur Mensch, Jordan Hoffmann, Trevor Cai, Eliza Rutherford, Katie Millican, George Bm Van Den Driessche, Jean-Baptiste Lespiau, Bogdan Damoc, Aidan Clark, Diego De Las Casas, Aurelia Guy, Jacob Menick, Roman Ring, Tom Hennigan, Saffron Huang, Loren Maggiore, Chris Jones, Albin Cassirer, Andy Brock, Michela Paganini, Geoffrey Irving, Oriol Vinyals, Simon Osindero, Karen Simonyan, Jack Rae, Erich Elsen, and Laurent Sifre. Improving language models by retrieving from trillions of tokens. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 2206–2240. PMLR, 2022. URL https://proceedings.mlr.press/v162/borgeaud22a.html.

Sarthak Chandra, Sugandha Sharma, Rishidev Chaudhuri, and Ila R. Fiete. Episodic and associative memory from spatial scaffolds in the hippocampus. *Nature*, 638:739–751, 2025. doi: 10.1038/s41586-024-08392-y.

Jianlyu Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. M3-Embedding: Multi-linguality, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 2318–2335. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.findings-acl.137. URL https://aclanthology.org/2024.findings-acl.137/.

Ziyang Chen, Xing Wu, Junlong Jia, Chaochen Gao, Qi Fu, Debing Zhang, and Songlin Hu. LongBench Pro: A more realistic and comprehensive bilingual long-context evaluation benchmark, 2026. URL https://arxiv.org/abs/2601.02872.

Róbert Csordás and Jürgen Schmidhuber. Improving differentiable neural computers through memory masking, de-allocation, and link distribution sharpness control. In *International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=HyGEM3C9KQ.

Peter J. Denning. Virtual memory. *ACM Computing Surveys*, 2(3):153–189, 1970. doi: 10.1145/356571.356573.

Martin Dresler, William R. Shirer, Boris N. Konrad, Nils C. J. Müller, Isabella C. Wagner, Guillén Fernández, Michael Czisch, and Michael D. Greicius. Mnemonic training reshapes brain networks to support superior memory. *Neuron*, 93(5):1227–1235.e6, 2017. doi: 10.1016/j.neuron.2017.02.003.

Jizhan Fang, Xinle Deng, Haoming Xu, Ziyan Jiang, Yuqi Tang, Ziwen Xu, Shumin Deng, Yunzhi Yao, Mengru Wang, Shuofei Qiao, Huajun Chen, and Ningyu Zhang. LightMem: Lightweight and efficient memory-augmented generation. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/a05b72653ec5b473732129829ae04195-Abstract-Conference.html.

Zafeirios Fountas, Martin A. Benfeghoul, Adnan Oomerjee, Fenia Christopoulou, Gerasimos Lampouras, Haitham Bou Ammar, and Jun Wang. Human-inspired episodic memory for infinite context LLMs. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/c05144b635df16ac9bbf8246bbbd55ca-Abstract-Conference.html.

Samuel J. Gershman, Ila Fiete, and Kazuki Irie. Key-value memory in the brain. *Neuron*, 113(11):1694–1707.e1, 2025. doi: 10.1016/j.neuron.2025.02.029.

Alex Graves, Greg Wayne, and Ivo Danihelka. Neural Turing machines, 2014. URL https://arxiv.org/abs/1410.5401.

Alex Graves, Greg Wayne, Malcolm Reynolds, Tim Harley, Ivo Danihelka, Agnieszka Grabska-Barwińska, Sergio Gómez Colmenarejo, Edward Grefenstette, Tiago Ramalho, John Agapiou, Adrià Puigdomènech Badia, Karl Moritz Hermann, Yori Zwols, Georg Ostrovski, Adam Cain, Helen King, Christopher Summerfield, Phil Blunsom, Koray Kavukcuoglu, and Demis Hassabis. Hybrid computing using a neural network with dynamic external memory. *Nature*, 538:471–476, 2016. doi: 10.1038/nature20101.

Bernal Jiménez Gutiérrez, Yiheng Shu, Weijian Qi, Sizhe Zhou, and Yu Su. From RAG to memory: Non-parametric continual learning for large language models. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 21497–21515. PMLR, 2025. URL https://proceedings.mlr.press/v267/gutierrez25a.html.

Alexis Huet, Zied Houidi, and Dario Rossi. Episodic memories generation and evaluation benchmark for large language models. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/7ff013b7e372ba5b790352ccd6908f03-Abstract-Conference.html.

Jeff Johnson, Matthijs Douze, and Hervé Jégou. Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data*, 7(3):535–547, 2021. doi: 10.1109/TBDATA.2019.2921572.

Urvashi Khandelwal, Omer Levy, Dan Jurafsky, Luke Zettlemoyer, and Mike Lewis. Generalization through memorization: Nearest neighbor language models. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=HklBjCEKvH.

Denis Kleyko, Mike Davies, E. Paxon Frady, Pentti Kanerva, Spencer J. Kent, Bruno A. Olshausen, Evgeny Osipov, Jan M. Rabaey, Dmitri A. Rachkovskij, Abbas Rahimi, and Friedrich T. Sommer. Vector symbolic architectures as a computing framework for emerging hardware. *Proceedings of the IEEE*, 110(10):1538–1571, 2022. doi: 10.1109/JPROC.2022.3209104.

Ching-Yun Ko, Payel Das, Sihui Dai, Georgios Kollias, Subhajit Chaudhury, Aurelie C. Lozano, and Pin-Yu Chen. ImReasoner: Improving memory-based language models for reasoning-in-a-haystack tasks. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 606–622. Association for Computational Linguistics, 2026. doi: 10.18653/v1/2026.acl-long.26. URL https://aclanthology.org/2026.acl-long.26/.

Trenton Kriete, David C. Noelle, Jonathan D. Cohen, and Randall C. O'Reilly. Indirection and symbol-like processing in the prefrontal cortex and basal ganglia. *Proceedings of the National Academy of Sciences*, 110(41):16390–16395, 2013. doi: 10.1073/pnas.1303547110. URL https://doi.org/10.1073/pnas.1303547110.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems*, volume 33, pp. 9459–9474, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html.

Moufan Li, Kristopher T. Jensen, Qiong Zhang, Qihong Lu, and Marcelo G. Mattar. A neural network model of free recall learns multiple memory strategies. *Nature Machine Intelligence*, 8(8):1238–1250, 2026. doi: 10.1038/s42256-026-01274-0.

Nelson F. Liu, Kevin Lin, John Hewitt, Ashwin Paranjape, Michele Bevilacqua, Fabio Petroni, and Percy Liang. Lost in the middle: How language models use long contexts. *Transactions of the Association for Computational Linguistics*, 12:157–173, 2024. doi: 10.1162/tacl_a_00638. URL https://aclanthology.org/2024.tacl-1.9/.

Eleanor A. Maguire, Elizabeth R. Valentine, John M. Wilding, and Narinder Kapur. Routes to remembering: The brains behind superior memory. *Nature Neuroscience*, 6(1):90–95, 2003. doi: 10.1038/nn988.

Alexander Miller, Adam Fisch, Jesse Dodge, Amir-Hossein Karimi, Antoine Bordes, and Jason Weston. Key-value memory networks for directly reading documents. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1400–1409. Association for Computational Linguistics, 2016. doi: 10.18653/v1/D16-1147. URL https://aclanthology.org/D16-1147/.

Amirkeivan Mohtashami and Martin Jaggi. Random-access infinite context length for transformers. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-2378. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/ab05dc8bf36a9f66edbff6992ec86f56-Abstract-Conference.html.

Charles Packer, Sarah Wooders, Kevin Lin, Vivian Fang, Shishir G. Patil, Ion Stoica, and Joseph E. Gonzalez. MemGPT: Towards LLMs as operating systems. *arXiv preprint arXiv:2310.08560*, 2023. doi: 10.48550/arXiv.2310.08560. URL https://arxiv.org/abs/2310.08560.

Tony A. Plate. Holographic reduced representations. *IEEE Transactions on Neural Networks*, 6(3): 623–641, 1995. doi: 10.1109/72.377968.

Parth Sarthi, Salman Abdullah, Aditi Tuli, Shubh Khanna, Anna Goldie, and Christopher D. Manning. RAPTOR: Recursive abstractive processing for tree-organized retrieval. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/8a2acd174940dbca361a6398a4f9df91-Abstract-Conference.html.

Imanol Schlag, Kazuki Irie, and Jürgen Schmidhuber. Linear transformers are secretly fast weight programmers. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 9355–9366. PMLR, 2021. URL https://proceedings.mlr.press/v139/schlag21a.html.

Sugandha Sharma, Sarthak Chandra, and Ila Fiete. Content addressable memory without catastrophic forgetting by heteroassociation with a fixed scaffold. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 19658–19682. PMLR, 2022. URL https://proceedings.mlr.press/v162/sharma22b.html.

Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. Self-attention with relative position representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pp. 464–468. Association for Computational Linguistics, 2018. doi: 10.18653/v1/N18-2074. URL https://aclanthology.org/N18-2074/.

Yaorui Shi, Yuxin Chen, Siyuan Wang, Sihang Li, Hengxing Cai, Qi Gu, Xiang Wang, and An Zhang. Look back to reason forward: Revisitable memory for long-context LLM agents. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/7e0dc9ccba0f1333be13a3f9dc2b3138-Abstract-Conference.html.

Jianlin Su, Murtadha H. M. Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. RoFormer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, 2024. doi: 10.1016/j.neucom.2023.127063.

Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, and Rob Fergus. End-to-end memory networks. In *Advances in Neural Information Processing Systems*, volume 28, pp. 2440–2448, 2015. URL https://proceedings.neurips.cc/paper/2015/hash/8fb21ee7a2207526da55a679f0332de2-Abstract.html.

Zihao Tang, Xin Yu, Ziyu Xiao, Zengxuan Wen, Zelin Li, Jiaxi Zhou, Hualei Wang, Haohua Wang, Haizhen Huang, Weiwei Deng, Feng Sun, and Qi Zhang. Mnemis: Dual-route retrieval on hierarchical graphs for long-term LLM memory. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 23914–23928. Association for Computational Linguistics, 2026. doi: 10.18653/v1/2026.acl-long.1096. URL https://aclanthology.org/2026.acl-long.1096/.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30, pp. 5998–6008, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Isabella C. Wagner, Boris N. Konrad, Philipp Schuster, Sarah Weisig, Dimitris Repantis, Kathrin Ohla, Simone Kühn, Guillén Fernández, Axel Steiger, Claus Lamm, Michael Czisch, and Martin Dresler. Durable memories and efficient neural coding through mnemonic training using the method of loci. *Science Advances*, 7(10):eabc7606, 2021. doi: 10.1126/sciadv.abc7606.

Weizhi Wang, Li Dong, Hao Cheng, Xiaodong Liu, Xifeng Yan, Jianfeng Gao, and Furu Wei. Augmenting language models with long-term memory. In *Advances in Neural Information Processing Systems*, volume 36, pp. 74530–74543, 2023. doi: 10.52202/075280-3259. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/ebd82705f44793b6f9ade5a669d0f0bf-Abstract-Conference.html.

James C. R. Whittington, Timothy H. Muller, Shirley Mark, Guifen Chen, Caswell Barry, Neil Burgess, and Timothy E. J. Behrens. The Tolman–Eichenbaum machine: Unifying space and relational memory through generalization in the hippocampal formation. *Cell*, 183(5):1249–1263.e23, 2020. doi: 10.1016/j.cell.2020.10.024.

Di Wu, Hongwei Wang, Wenhao Yu, Yuwei Zhang, Kai-Wei Chang, and Dong Yu. LongMemEval: Benchmarking chat assistants on long-term interactive memory. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/d813d324dbf0598bbdc9c8e79740ed01-Abstract-Conference.html.

Di Wu, Zixiang Ji, Asmi Kawatkar, Bryan Kwan, Jia-Chen Gu, Nanyun Peng, and Kai-Wei Chang. LongMemEval-V2: Evaluating long-term agent memory toward experienced colleagues, 2026. URL https://arxiv.org/abs/2605.12493.

Yuhuai Wu, Markus Norman Rabe, DeLesley Hutchins, and Christian Szegedy. Memorizing transformers. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=TrjbxzRcnf-.

Chaojun Xiao, Pengle Zhang, Xu Han, Guangxuan Xiao, Yankai Lin, Zhengyan Zhang, Zhiyuan Liu, and Maosong Sun. InfLLM: Training-free long-context extrapolation for LLMs with an efficient context memory. In *Advances in Neural Information Processing Systems*, volume 37, pp. 119638–119661, 2024. doi: 10.52202/079017-3801. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/d842425e4bf79ba039352da0f658a906-Abstract-Conference.html.

Buqiang Xu, Yijun Chen, Jizhan Fang, Ruobin Zhong, Yunzhi Yao, Yuqi Zhu, Lun Du, and Shumin Deng. StructMem: Structured memory for long-horizon behavior in LLMs. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 122–146. Association for Computational Linguistics, 2026. doi: 10.18653/v1/2026.acl-short.12. URL https://aclanthology.org/2026.acl-short.12/.

Wujiang Xu, Zujie Liang, Kai Mei, Hang Gao, Juntao Tan, and Yongfeng Zhang. A-Mem: Agentic memory for LLM agents. In *Advances in Neural Information Processing Systems*, volume 38, pp. 20004–20031, 2025. doi: 10.52202/085713-0593. URL https://proceedings.neurips.cc/paper_files/paper/2025/hash/19909c36f51abc4856b4560aff3d36d6-Abstract-Conference.html.

Songlin Yang, Bailin Wang, Yu Zhang, Yikang Shen, and Yoon Kim. Parallelizing linear transformers with the delta rule over sequence length. In *Advances in Neural Information Processing Systems*, volume 37, pp. 115491–115522, 2024. doi: 10.52202/079017-3668. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/d13a3eae72366e61dfdc7eea82eeb685-Abstract-Conference.html.

Xinrong Zhang, Yingfa Chen, Shengding Hu, Zihang Xu, Junhao Chen, Moo Hao, Xu Han, Zhen Thai, Shuo Wang, Zhiyuan Liu, and Maosong Sun. ∞Bench: Extending long context evaluation beyond 100K tokens. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 15262–15277. Association for Computational Linguistics, 2024. doi: 10.18653/v1/2024.acl-long.814. URL https://aclanthology.org/2024.acl-long.814/.

Wanjun Zhong, Lianghong Guo, Qiqi Gao, He Ye, and Yanlin Wang. MemoryBank: Enhancing large language models with long-term memory. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 19724–19731, 2024. doi: 10.1609/aaai.v38i17.29946. URL https://ojs.aaai.org/index.php/AAAI/article/view/29946.

AI USE STATEMENT

Generative AI tools were used for limited writing and implementation assistance. Google Gemini was used to improve the organization, clarity, grammar, and English expression of author-written manuscript text. OpenAI Codex assisted with the implementation, debugging, and execution of experimental code. The research questions, conceptual contributions, algorithms, experimental design and protocols, mathematical claims and proofs, literature review, and scientific interpretation were developed by the authors. All AI-assisted code was manually reviewed and subjected to correctness testing, and all AI-assisted manuscript edits were manually reviewed and proofread. The authors take full responsibility for the final manuscript and all reported results.

A APPENDIX ROADMAP AND EVALUATION MAP

Table A.1: Evaluation map. Each row identifies the corresponding main-paper RQ, the role of the supplementary study, and its evaluation unit or coverage, paired-resampling units are specified in Appendix F.1.

Main	Study	Purpose	Unit / coverage
RQ1	Controlled payload variation	matched multiplicity and cue–payload conflict	200 families/condition
RQ1	Referential identity and revision	controlled checkpoint retargeting plus matched preservation evaluations	two 500-trial evaluations, 4,103,500 non-target checks/evaluation
RQ1	Continuous verified locator	preregistered complete free-slot locator semantics under continuous mixed mutation	500 trials, 7,500 target and 3,920,000 non-target checks/backend, 87,500 inserted-record checks
RQ2	Native Delta backend transfer	same-key, matched-checkpoint RLS/Delta revision comparison	500 trials, 4,103,500 collateral checks/backend
RQ2	Controlled Delta overlap stress	causal key-overlap intervention, vector-law fit, decoding-margin stress, content-independent key analysis	190 conditions, 500 banks, 65,408,000 natural pairs
RQ2	Learned/content-derived key stress	raw BGE-M3 and E5 cue embeddings under matched load, mutation, verified dereference, and exact certification	75 frozen cells, 1,024,725 certificate prefixes
RQ2	Sequence-level interference audit	exact cumulative identity and score-space certificate on frozen-protocol replays	331,072 protected segments, 7,612,479 prefixes
RQ2	Cross-bank high-churn stress	serialized migration/retirement mechanics plus long-horizon certificate boundary	20,000 operations, 3,000 migrations, 11,000 Delta writes

Continued on next page

Table A.1 (continued)

Main	Study	Purpose	Unit / coverage
RQ3	Realistic first access / Hybrid	selector change versus stable dereference	120 Qwen, 30 Mistral
RQ3	LongMemEval-S evidence retrieval	evidence-policy intervention	500 questions
RQ3	LongMemEval-S natural knowledge update	benchmark-defined old/new content evolution with fixed identity and collateral audit	78 released update questions, 70 frozen eligible
RQ3	LongMemEval-S causal reader	updated SLA/pointer versus stale predecessor under matched reader inputs	70 paired questions, 210 Qwen generations
RQ3	LongMemEval-V2 dynamic-memory lifecycle	natural RLS/Delta structural lifecycle, relation fidelity, collateral preservation, independent reader re-execution	451 total, 439 eligible, 2,475 seed records, 24,369 mutations, 369,485 collateral checks
RQ3	Pretrained DeltaNet integration	cache-resident SLA state, causal model dependence, held-out scale-up	1.366B frozen parameters, 11 revision states, 100 held-out cases
RQ4	Structured relation operators	additive, commuting multi-axis, and branching non-commutative execution/composition	5,000 routes/length, 256 test states, 2,048 noise trials/condition
RQ4	Structured phase/route controls	RoPE-style/FHRR execution, local reuse, relation uncertainty	500 routes/length, five FHRR seeds, 500 uncertainty routes/regime
RQ4	Banking, load, and access	load envelope, locator-state accounting, matched direct lookup, balanced registry scaling, repeated vector access	128–10,114 records, five load seeds/cell
RQ4	Verified-dereference trust boundary	binder/registry/routing fault audit and fair logical/implementation accounting	500 requests/fault, 1,024-slot accounting

Appendix B gives the formal record state, necessity result, protocol-stable key construction, the learned-key stress, verified dereference, update algebra, sequence certification, banking, and locator-state accounting. Appendix C.1 expands controlled identity/revision, the continuous verified-locator stress, cross-bank routing and maintenance, matched native Delta transfer, signed-overlap interference, and sequence-level certification. Appendix C.4 covers structured relation execution. Appendices D.1–D.3 document realistic first access, LongMemEval-S content evolution, LongMemEval-V2 structural evolution, independent reader re-execution, and pretrained DeltaNet integration. Appendix E reports systems and direct-access measurements with fair control-state accounting. Appendix F collects statistical and reproducibility details.

Table A.2: Shared-state transfer and resource summary. RLS and native Delta use matched protocol-stable identity keys, the natural lifecycle evaluation and state accounting extend the same referential-state comparison.

Study	Result
RLS $\leftrightarrow$ native Delta	Target state 1.000 at every checkpoint, 4,103,500/4,103,500 collateral checks preserved per backend
Continuous verified locator	RLS and native Delta each preserve 7,500/7,500 target and 3,920,000/3,920,000 non-target checks, 87,500/87,500 inserted-record observations, 0 wrong records
Controlled Delta overlap	Collateral law matches the identity line to FP32 precision, content-independent key overlaps remain inside the controlled low-interference region
Learned-key stress	Raw BGE-M3/E5 cue keys reach max $\lvert\cos\rvert = 0.9942/0.9967$ and leave the safe mutation regime, exact certificate agrees with observed Top-1 on 1,024,725/1,024,725 prefixes across the three matched key sources, 0 wrong-record returns

Continued on next page

Table A.2 (continued)

Study	Result
Sequence certificate	Exact score-space certificate and observed Top-1 hold on 7,612,479/7,612,479 prefixes across controlled and natural traces
Cross-bank stress	3,000/3,000 migrated targets and routed revisits exact, 0 wrong records/aliases, cumulative certificate predicts eventual unmaintained margin exhaustion
Certificate-triggered maintenance	Same 20,000 operations, 1,744/11,000 candidate writes trigger bank rebuilds, no Top-1/Top-8 failure, active miss, wrong return, or rebuild failure, $2.13\times$ baseline runtime
Natural Delta lifecycle	22,275/22,275 retained checks, 369,485/369,485 collateral, 4,776/4,776 valid relations, 0 wrong records, 439/439 independent reader re-executions
Writable locator tensors	RLS 64.25 MiB versus Delta 0.25 MiB per bank at $d_k = 4096$, $d_z = 16$, $257\times$ is tensor-only
Fair control-state accounting	At 1,024 slots, measured CPython totals (scan/pointer/hash/Delta/RLS): 277,304 / 285,496 / 332,008 / 539,448 / 67,648,312 B, payload and optional SLA key cache excluded

Table A.2 summarizes controlled RLS/Delta transfer, learned-key stress, continuous verified-locator behavior, cross-bank stress and maintenance, natural lifecycle evaluation, and matched state accounting.

B Representation, Binding, and Formal Properties

B.1 Formal state and interface semantics

At semantic-access epoch t, let $\mathcal{R}_t$ denote the current record set and $\mathcal{C}_t = \{e_r(t) : r \in \mathcal{R}_t\}$ denote the selector-specific retrieval views. A semantic selector returns ranked views

$$\widehat{\mathbf{E}}_{q,t} = \big((e_i, \sigma_i)\big)_{i=1}^{K_q} = \mathcal{S}(q, \mathcal{C}_t), \tag{15}$$

where σ_i is the retrieval score. Ingestion metadata maps each hit to its retained handle through $r_i = \iota(e_i)$. We write $\mathrm{Resolve}(q.\mathcal{C}_t) \to h_r$ (or $\perp$) for this read-only semantic selection. First access and explicit re-address use the same operation without modifying a locator. Insertion publishes a retrieval view before the new record participates in a later semantic epoch. Payload revision refreshes that view when the selector is payload-derived. This upstream index maintenance is outside the post-retrieval locator timings. The controlled revision experiments revisit retained handles rather than re-querying the semantic stage. *Persistent* refers to state retained for the memory-instance lifetime.

Theorem 1 is a realization-independent state lower bound. For any fixed current observation O_t, partition admissible histories by the active logical record that an earlier resolution established. Exact revisit must distinguish every nonempty referent class. If m such classes share O_t, any retained state alphabet with fewer than m distinguishable values merges two classes and forces the same revisit output on histories requiring different records. A later query that explicitly carries the retained handle or another history-derived identifier no longer belongs to the current-state-only setting because that identifier itself supplies the distinguishing state. The theorem therefore separates the need for identity-bearing state from the choice of pointer, scan, or shared-associative realization.

Complete state factorization. SLA specializes the current locator as

$$\ell_r^{\mathrm{SLA}} = (s_r, a_r), \qquad a_r \in \mathcal{A}_{s_r}^{\mathrm{occ}}, \tag{16}$$

and separates implementation/control state from the external payload store described in Eq. (3).

The logical record state in Eq. (1) is an induced semantic tuple, not a handle-indexed storage layout. Here, h_r is the invariant identity, ℓ_r is the realization-specific current locator, v_r is the mutable external payload, and $u_r = \perp$ when no independent relation state is materialized. In scope s, $\mathcal{D}_s(a)$ stores the occupied-slot handle and $\mathcal{V}_s(a)$ stores the payload. SLA retains canonical cue metadata c_r and uses Eq. (16), together with a fixed slot universe $\mathcal{A}_s^{\mathrm{all}}$, an occupied/free partition, and a content-independent locus code $z(a)$. The routing directory and retained cue metadata do not store the current bank-local address. Direct dictionary or pointer backends may instead materialize the full

Table B.1: SLA factorizes persistent identity from mutable state. The interface fixes which component each operation may change.

Component	Role	Contract under memory evolution
h_r	opaque logical identity	invariant under payload and locator revision
k_r	protocol-stable identity key	fixed with h_r within a protocol namespace
a_r	current bank-local address	revisable through shared binder state
v_r	external payload	independently overwritable after verified access
B_s	bank-shared mutable indirection	updated by RLS or native Delta rule
u_r	optional relation state	used by relation execution without redefining h_r

current locator in handle-indexed state. Payload overwrite changes $\mathcal{V}_{s_r}(a_r)$ while preserving identity and locator. In the primary benchmark protocol, locator revision is bank-local: it moves the handle-registry and payload-store entries to a free a'_r, updates occupancy and the binder association, and preserves r, v_r, c_r, k_r, and s_r. Cross-bank migration is treated separately as a routing transaction in Appendix C.2. Structured-route studies keep storage layout fixed, so locator revision and fixed-layout relation execution remain separate operations.

The representation-independent operations are given in Eq. (2). Retirement and reuse are specified below. The contract is defined over complete committed states. Let $X(\mathfrak{M}_\varrho)$ denote the abstract logical-state relation induced by Eq. (3). It is not stored as an additional handle-indexed table. The implementation state Θ_ϱ contains the common scope, active-handle, occupancy, and registry substrate together with realization-specific locator or relation state, while $\mathcal{V}_s$ remains the external payload store. Structural consistency requires a unique occupied a_r with $\mathcal{D}_{s_r}(a_r) = h_r$ and $\mathcal{V}_{s_r}(a_r) = v_r$. SLA stores scope, canonical cue material, and binder state, but it stores neither a handle-indexed current-address entry nor a separate persistent key-to-current-locus table. Insertion requires admissible metadata, a free slot, and no realized-key collision. Payload revision changes only the verified external payload entry. Locator revision moves the registry and payload entries, updates occupancy and binder state, and becomes visible only after verified reinstatement. Failure leaves the prior committed state unchanged. Rebinding repeats this full locator move rather than only changing the binder. Successful state transitions are summarized in Eq. (7). Cross-scope migration is not a hidden extension of REVISELOCATOR: it changes the routing state and is evaluated as an explicit serialized transaction.

Retirement/reuse is fail-closed. Define $\mathrm{Retire}_\varrho(h_r) \to \mathsf{ok}$ (or $\perp$) to remove the retrieval view, mark h_r inactive, clear its registry/payload occupancy, and free the slot without ever retargeting h_r. Later reinstatement returns $\perp$. Even if an obsolete binder projection is lazily retained and the slot is reused, active-handle and $\mathcal{D}_s(a) = h_r$ checks prevent aliasing to the new occupant. Identity safety follows from invalidation/verification. Binder cleanup or rebuild separately reclaims capacity. A merge explicitly chooses/creates a survivor and retires absorbed handles. It never silently rewrites identity.

B.2 VERIFIED-DEREFERENCE TRUST BOUNDARY

Verified dereference combines several checks. The binder reconstructs a candidate address. $\mathcal{D}_s$ verifies the candidate occupant. Active-handle state rejects retired identities. $\Sigma(h)$ selects the authoritative bank, and implementations publish mutations atomically. Under these assumptions a wrong binder candidate can turn a correct access into $\perp$, but cannot return another record's payload.

A deterministic fault audit exercised 500 frozen identities per condition using the verified-dereference rule. Table B.2 separates ordinary localization/control faults from a coordinated trust-boundary violation. Stale or shuffled binder candidates, empty slots, registry mismatch, retirement/reuse, and stale routing all fail closed with zero wrong-record returns under a correct registry. In contrast, a forged registry-plus-payload control defeats the identity gate by making a wrong slot falsely claim the requested handle. This negative control lies outside the trusted-metadata guarantee.

The aborted-mutation condition stages a private copy and discards it before publication. It validates the contract's commit boundary rather than claiming exception-atomicity for every sequential reference implementation. The guarantee is therefore explicit: no-wrong-record safety is conditional on correct

Table B.2: Verified-dereference fault boundary, 500 deterministic requests per condition. "Wrong" denotes a non-null payload from a different logical record.

Condition	Correct	Fail closed	Wrong
Control	500	0	0
Stale binder	0	500	0
Wrong occupied candidate	0	500	0
Empty candidate	0	500	0
Registry mismatch	0	500	0
Retire then slot reuse	0	500	0
Stale routing	0	500	0
Forged registry + wrong payload	0	0	500
Aborted staged mutation	500	0	0

registry, routing, liveness/occupancy, and commit-control metadata. Arbitrary coordinated corruption of those trusted controls requires a stronger integrity mechanism and is outside the evaluated contract.

For relation label λ, let T_λ be the partial semantic action on execution coordinate $\Psi_\xi(x_r)$. A representation-specific executor realizes the same supplied action when its decoder satisfies $\delta_\xi(A_{\lambda,\xi} z_\xi(u)) = T_\lambda(u)$ on the evaluated domain. Navigation follows Eq. (26). If the action is undefined, the target is not unique, or reinstatement/dereference fails, navigation returns $\perp$ without mutation. Independently materialized relation state is unchanged by locator revision. The one-dimensional periodic realization instead derives the relation coordinate from the current address and therefore uses fixed-layout route protocols. LongMemEval-V2 uses trajectory identity and trajectory-local ordinal. LongMemEval-S uses session-local chronological order, both independently of physical bank contiguity.

For a static scope, the dual ridge solution used in static evaluation is Eq. (5). Insertion-only construction uses the algebraically equivalent recursive update below. Relocation/rebinding then apply the exact key-specific correction and do not in general preserve the global batch-ridge optimum. Table B.3 places SLA beside the explicit locator endpoints under the common scoped registry.

Table B.3: Persistent-locator design space under the common scoped registry. The rows represent distinct locator-state mechanisms, measured timings and fixed-size storage are reported separately.

Realization	Locator-specific state	Revisit / address revision	Distinguishing property
Registry scan	no reverse locator	scan current bank, move common registry/store entries	exact endpoint without reverse-index state
Dictionary / pointer	explicit current locator per record	direct revisit, update one explicit locator	lowest measured direct-operation latency
SLA	retained cue plus bank-shared B_s (P when writable)	cue-key projection/ranking, rank-one binder correction plus registry/store move	shared mutable indirection with analyzable capacity and cross-key interaction

B.3 Protocol-stable deterministic key construction

The main SLA experiments use deterministic identity keys reconstructed from retained canonical cue metadata and held fixed under locator revision. Key generation does not consume the mutable payload v_r, the current address a_r, or the locus code $z(a_r)$. Each protocol in Table B.4 maps a canonical cue byte string to a 4,096-dimensional FP32 key. We use c_r for the final namespace-qualified UTF-8 byte sequence passed to the hash routine. Source-visible text before canonicalization is not c_r. For text-derived cues, canonicalization applies NFKC normalization, trims leading and trailing whitespace, collapses internal whitespace, and appends the immutable workload namespace or disambiguator. Thus $k_r = \kappa_{\mathcal{P}}(c_r)$ for protocol namespace $\mathcal{P}$. Relocation and rebinding change a_r while holding c_r and k_r fixed.

Binder construction requires distinct realized keys within a scope. Different cue bytes can in principle collide because the implemented routines reduce SHA256 to a 64-bit, or modulo-$2^{63}-1$, PRNG seed. Under a uniform-seed model, the birthday-collision probability is below 6×10^{-14} for at most 1,024 records even in the 63-bit case. Ingestion therefore rejects any realized seed collision within a binder scope.

Table B.4: Protocol-stable content-independent key instantiations used in the main experiments. All use retained cue bytes, SHA256, 4,096 FP32 Gaussian values, and L_2 normalization.

Studies	Digest-to-seed conversion	PRNG
Controlled payload variation	first 8 digest bytes, unsigned little-endian	PyTorch CPU Generator
Reinstatement, target-centric revision, primary same-run preservation evaluation, matched native Delta evaluation, routes, and LongMemEval-V2 dynamic-memory lifecycle Delta evaluation	first 8 digest bytes, unsigned big-endian	NumPy PCG64
Independent preservation replication	first 8 digest bytes, unsigned big-endian, reduced modulo $2^{63}-1$	NumPy PCG64
Realistic first access, LongMemEval-S, and systems	first 8 digest bytes, unsigned big-endian, reduced modulo $2^{63}-1$	PyTorch CPU Generator

For a NumPy-PCG64 row, let u be the unsigned big-endian integer represented by the first eight SHA256 digest bytes. Rows marked with modulo reduction replace u by $u \bmod (2^{63}-1)$ before seeding. The key is then

$$\begin{aligned} x &= \mathrm{Generator}(\mathrm{PCG64}(u)).\,\mathrm{standard_normal}(4096, \mathrm{float32}), \\ k_r &= x/\|x\|_2. \end{aligned} \tag{17}$$

The PyTorch rows use the corresponding row-specific integer seed with a CPU generator and the same FP32 Gaussian dimension and normalization. Each row defines a fixed workload-specific namespace $\mathcal{P}$, and every study uses that mapping consistently. Different rows are distinct deterministic key realizations. We do not assume that two protocols produce the same binder behavior. The target-centric revision study and the preservation replication use different fixed namespaces but share the same 512-record trial identities, update addresses, and revision schedule. Changing the key routine defines a new protocol namespace, so an established record either retains its original mapping or is explicitly rebound under the new one. The same rule applies when the key is produced by a learned encoder: model updates must not silently change the identity key of an existing handle.

B.4 Raw learned/content-derived key stress

To test whether the interference law depends on the key source, we replace the protocol-stable Gaussian keys with raw normalized cue embeddings from two frozen encoders: BGE-M3 and E5-large-v2. No encoder is trained, and no whitening, projection, regularization, or post-hoc threshold is applied. The same five seeds and bank sizes $n_s \in \{128, 256, 512, 768, 1024\}$ are used for all three key sources, giving 75 frozen condition/seed/load cells. BGE-M3 and E5 produce 1,024-dimensional keys. The SHA control remains 4,096-dimensional.

The semantic encoders produce much more coherent identity candidates than the SHA control. Table B.5 summarizes the geometry and the 1,024-record initialization result. Initial Top-1/Top-8 remains perfect through load 768. After the first target relocation, both learned sources lose non-target Top-1 and Top-8 in every tested load/seed cell. Mandatory registry verification prevents a wrong-record return: the canonical non-overlapping audit records 273,927 fail-closed outcomes over 658,245 checkpoint observations and zero wrong-record returns.

The failure does not break the Delta law. Across 1,024,725 protected prefixes, the exact score-space certificate agrees with observed Top-1 on every prefix. The maximum cumulative projection residual is 2.58×10^{-5}, and the maximum one-step Delta-equation residual is 1.42×10^{-6}. The norm and triangle sufficient bounds cover 32.69% and 9.00% of these prefixes. This experiment therefore separates algebra from operating regime: the exact law is agnostic to the normalized key source,

Table B.5: Raw learned/content-derived key geometry. Geometry entries summarize the five seeds and five bank sizes, initial Top-1/Top-8 is aggregated over the five 1,024-record banks.

Key source	Mean $\lvert\cos\rvert$	Max P99	Max $\lvert\cos\rvert$	Top-1@1024	Top-8@1024
SHA-Gaussian	0.012476	0.041066	0.075937	5120/5120	5120/5120
BGE-M3 raw	0.797987	0.925149	0.994241	5103/5120	5119/5120
E5-large-v2 raw	0.898300	0.962707	0.996713	5093/5120	5120/5120

while safe addressing depends on realized geometry, margin, load, and mutation history. These two raw semantic encoders are not safe drop-in identity-key replacements under the tested protocol.

The controlled long-context studies use 512-token blocks, at most 4,096 retrieved tokens, and deterministic reader decoding.

B.5 Writable binder updates and revision

For insertion-only RLS construction, writable banks use the recursive least-squares update algebraically equivalent to batch ridge. With $B_0 = 0$, $P_0 = \lambda^{-1} I$, $z_t = z(a_{r_t})$, and

$$w_t = P_{t-1} k_t, \qquad \gamma_t = 1 + k_t^\top w_t, \tag{18}$$

we update

$$B_t = B_{t-1} + \frac{(z_t - B_{t-1} k_t) w_t^\top}{\gamma_t}, \qquad P_t = P_{t-1} - \frac{w_t w_t^\top}{\gamma_t}. \tag{19}$$

The primary RLS experiments use $d_k = 4096$, FP32 binder state, and $\lambda = 10^{-4}$.

For RLS address relocation, $g = P k_r$ and $\epsilon = z(a'_r) - B k_r$ give

$$B'_{\mathrm{RLS}} = B + \epsilon \left(\frac{g}{k_r^\top g} \right)^\top. \tag{20}$$

The binder component of repeated rebinding uses the same correction at each locator move. Registry, payload-store, and occupancy updates follow the atomic locator semantics in Appendix B.1. The key-specific replacement enforces the revised association. Writable RLS banks retain P so subsequent insertion updates continue from the current online state rather than refitting the global batch-ridge solution.

The Delta realization uses the same L_2-normalized protocol-stable keys and locus targets but updates the shared binder directly with FLA's native Delta-rule recurrence at $\beta = 1$. In the matrix orientation used by SLA,

$$B_t = B_{t-1} + (z_t - B_{t-1} k_t) k_t^\top, \qquad \|k_t\|_2 = 1, \tag{21}$$

and relocation is the same corrective write with $z_t = z(a'_r)$. The native operator was also checked numerically, independently of retrieval: over 72 checks, the maximum FP32 target-write residual was $9.5367431640625 \times 10^{-7}$ and the maximum collateral-identity verification residual was $4.172325134277344 \times 10^{-7}$, both below the prespecified 2×10^{-6} FP32 tolerance and therefore numerically zero at the working precision.

Proof of Proposition 1. For RLS, the covariance update is the Sherman–Morrison form of $(P_{t-1}^{-1} + k_t k_t^\top)^{-1}$, so symmetry and positive definiteness are preserved from P_0. With $g = P k_r$, $k_r^\top g = k_r^\top P k_r > 0$. Multiplying Eq. (20) by k_j gives $(B'_{\mathrm{RLS}} - B) k_j = \epsilon (g^\top k_j)/(k_r^\top g)$. Symmetry of P gives the RLS term in Eq. (10). Setting $j = r$ yields $(B'_{\mathrm{RLS}} - B) k_r = \epsilon$ and therefore $B'_{\mathrm{RLS}} k_r = z(a'_r)$. For Delta, Eq. (9) gives $(B'_\Delta - B) k_j = \epsilon (k_r^\top k_j)$ directly. Since $\|k_r\|_2 = 1$, setting $j = r$ gives $(B'_\Delta - B) k_r = \epsilon$ and hence $B'_\Delta k_r = z(a'_r)$.

Proof of Corollary 1. The RLS zero-interference condition follows from $k_r^\top P k_j = 0$, and the bound follows from Cauchy–Schwarz in the P-inner product, $|k_r^\top P k_j| \le \sqrt{k_r^\top P k_r} \sqrt{k_j^\top P k_j}$. Equation (11) is a deterministic worst-case envelope. Average-case RLS scaling remains history-dependent through P. When $P \approx \alpha I$, however, the exact RLS interference coefficient reduces to

$k_r^\top k_j$ for unit-norm keys and therefore inherits the same high-dimensional low-overlap suppression. No such uniform isotropic claim is made for general history-dependent P. The Delta condition follows directly from $k_r^\top k_j = 0$ and its exact collateral expression in Eq. (10).

Proof of Theorem 2. For each external write in a protected segment, Eq. (21) gives $B_t - B_{t-1} = \epsilon_t k_{r_t}^\top$. Summing from $t = \tau + 1$ to u telescopes to $B_u - B_\tau = \sum_t \epsilon_t k_{r_t}^\top$. Right-multiplication by k_j yields Eq. (12). The identity is deterministic and sequence-specific. No distributional assumption on keys, write targets, or residuals is used. A direct write to j changes its intended state and therefore starts a new protected segment rather than being counted as collateral accumulation.

Proof and envelopes for Corollary 2. Using $B_u k_j = B_\tau k_j + \Delta_j(u)$, subtraction of the score of competitor a from the intended score gives Eq. (13). The intended locus remains the unique Top-1 exactly when every competitor gap remains positive. For unit-norm loci, $\|z(a_j) - z(a)\|_2 \le 2$, so Cauchy–Schwarz gives $g_{j,a}(u) \ge g_{j,a}^{(\tau)} - 2\|\Delta_j(u)\|_2$, proving the norm sufficient condition. A still more conservative envelope follows from

$$W_j(u) = \sum_{t=\tau+1}^{u} \|\epsilon_t\|_2 \, |k_{r_t}^\top k_j| \ge \|\Delta_j(u)\|_2, \tag{22}$$

so $W_j(u) < \mu_j^{(\tau)}/2$ is also sufficient. These three levels, exact score gap, cumulative-vector norm, and triangle envelope, must be distinguished: failure of a sufficient envelope does not imply decoding failure.

B.6 Formal properties of the periodic locus

Once stably reinstated and verified, a reference can support supplied relation actions without changing the identity contract. For $M = 8$ with coprime $(p_1, \dots, p_8) = (5, 7, 11, 13, 17, 19, 23, 29)$ and block-diagonal displacement,

$$z(a) = \frac{1}{\sqrt{M}} \bigoplus_{m=1}^{M} [\cos(2\pi a/p_m), \sin(2\pi a/p_m)], \tag{23}$$

$$R(\Delta) z(a) = z(a + \Delta), \qquad R(\Delta_2) R(\Delta_1) = R(\Delta_1 + \Delta_2), \tag{24}$$

$$R(\Delta_H) \cdots R(\Delta_1) z(a) = z\left(a + \sum_{h=1}^{H} \Delta_h \right). \tag{25}$$

More generally, a supplied relation label acts on a representation-specific execution coordinate, then reuses the same persistent-access path,

$$\begin{aligned} q' &= T_\lambda(\Psi_\xi(x_r)), \qquad & h_{r'} &= \Gamma_\xi(q'), \\ \ell_{r'} &= \mathrm{Reinstate}_\varrho(h_{r'}), \qquad & v_{r'} &= \mathrm{Dereference}_\varrho(h_{r'}, \ell_{r'}), \qquad \mathsf{Navigate}_{\lambda,\xi}(\mathfrak{M}_\varrho, r) = x_{r'}. \end{aligned} \tag{26}$$

Proposition 2 (Equivariance, composition, and local uniqueness)**.** *Equations* (24)–(25) *hold for the periodic locus map. If two addresses lie in an interval shorter than* $P_{\mathrm{CRT}} = 1{,}078{,}282{,}205$*, identical locus codes imply identical addresses.*

We compare this shared action with commuting two-axis, branching non-commutative, RoPE-style, FHRR, integer/tuple, adjacency, and learned random-code realizations. The address-coupled 1-D construction is evaluated only under fixed layout. Dynamic trajectory/session studies use relation coordinates independent of physical storage (Appendix C.4).

Proof of Proposition 2. For the evaluated periodic realization, $(p_1, \dots, p_8) = (5, 7, 11, 13, 17, 19, 23, 29)$. For period p_m, define the two-dimensional component of $z(a)$ as $r_m(a) = [\cos(2\pi a/p_m), \sin(2\pi a/p_m)]^\top$. The corresponding block of $R(\Delta)$ is the planar rotation by angle $2\pi\Delta/p_m$. The angle-addition identities give $R_m(\Delta) r_m(a) = r_m(a + \Delta)$ and $R_m(\Delta_2) R_m(\Delta_1) = R_m(\Delta_1 + \Delta_2)$. Because $R(\Delta)$ is the block-diagonal direct sum of these rotations, both identities hold for the full locus, and induction gives the H-step composition.

If $z(a) = z(a')$, equality of every phase pair implies $a \equiv a' \pmod{p_m}$ for every m. The periods are pairwise coprime, so the Chinese remainder theorem gives $a \equiv a' \pmod{\prod_m p_m}$. For two valid addresses whose difference in magnitude is strictly smaller than this product, the congruence implies $a = a'$. □

B.7 BANKING STATE

Within each document, source-order block j is initially assigned to bank $\lfloor j/1024 \rfloor$ and address $j \bmod 1024$. $\Sigma(h_r)$ stores current scope. Locator accounting separates the common substrate (Σ, active-handle/occupied-free bookkeeping, $\mathcal{D}_s$, and $\mathcal{V}_s$) from realization-specific state. Registry scanning adds no reverse index. A dictionary adds one handle-to-address mapping per record with implementation-dependent container/key overhead. The cached-pointer control adds 8 bytes per record. Under the fixed-size locator-tensor accounting used below, locator-static SLA contributes B_s (0.25 MiB/bank). Retained variable-length cue metadata is accounted separately, and an optional materialized 4,096-D FP32 key cache adds 16 KiB/cached record.

Writable RLS-SLA retains $P \in \mathbb{R}^{4096\times4096}$ (64.00 MiB) together with $B_s \in \mathbb{R}^{16\times4096}$ (0.25 MiB), for 64.25 MiB of persistent locator-specific mutable tensor state per bank. The Delta realization maintains the same B_s but no P, for 0.25 MiB under the same accounting boundary. The resulting RLS/Delta ratio is exactly 257.0 for this configuration. These totals exclude the common registry, external payload store, semantic index, language-model parameters, native token cache, temporary FLA/Triton buffers, and transient update/query tensors. The 257.0 ratio pertains specifically to persistent locator-state tensor footprint. Timing and total-system memory are reported separately.

Locator-static and locator-writable modes remain distinct. Static banks keep B_s after their update phase. RLS-writable banks retain $P + B_s$. Delta-writable banks retain B_s and continue applying the delta rule. The external payload store remains separate and can be overwritten after verified reinstatement without changing the locator.

C CONTROLLED MEMORY MECHANICS

C.1 IDENTITY AND REVISION PROTOCOLS AND ADDITIONAL RESULTS

Protocol and metrics. Each controlled payload-variation memory contains 1,024 records. The independent-strata protocol contains 200 fixed target/cue/answer families at each $C \in \{1, 10, 100, 500, 1000\}$, with C semantically near-duplicate payload distractors and the exact stable cue retained in the query. Each C uses its own cue namespace, target permutation, answer set, and semantic-family realization, so the five conditions are separate payload-multiplicity corpora rather than nested perturbations of shared examples. The fixed family instance is the evaluation unit.

A separate matched nested BGE-M3 control uses 200 families. For each family it fixes one target, one canonical cue, and one master pool of 1,000 near-duplicate payload records. $C = 1, 10, 100, 500, 1000$ uses nested prefixes of that same pool. The target remains above the strongest Payload-RAG distractor for every family and C (minimum target-minus-distractor cosine margin 0.0603). Under this target-order-preserving construction, Payload RAG, Stable-random, and Full SLA obtain Top-1, Top-8, and evidence hit@8 1.000 at every C. The paired Top-1 change from $C = 1$ is 0.000 with 95% CI 0–0 for Payload RAG, Stable-random, and Full SLA. The result therefore characterizes increasing nested multiplicity while target semantic ordering is held fixed.

The cue–payload conflict protocol is a separate fixed intervention in which the stable cue identifies one target while a held-out `intfloat/e5-large-v2` verifier makes a distractor payload semantically preferable. Five equal-size strata of 200 examples are ordered by increasing achieved distractor-over-target E5 verifier margin. Their (min/median/mean/max) margins are Q1: 0.000001/0.001699/0.001648/0.003035. Q2: 0.003053/0.004159/0.004157/0.005176. Q3: 0.005184/0.006254/0.006242/0.007267. Q4: 0.007278/0.008411/0.008440/0.009751. Q5: 0.009760/0.011647/0.012183/0.019591. The verifier is used only to construct the fixed strata. Its scores are not supplied to evaluated methods. Thus Q1–Q5 index E5 construction margin, while BGE-M3 retrieval and reader outcomes are evaluated independently of that ordering. Payload RAG supplies the BGE-M3 payload-semantic ranking endpoint. Stable-random keeps the same cue/key

binder and cosine decoder while replacing periodic address codes with fixed random unit-norm codes. Full SLA uses the periodic scaffold.

The reinstatement study uses 512-record memories and 500 held-out targets. A fixed BGE-M3 encoder (Chen et al., 2024) searches canonical cue text by cosine similarity. Each target has lexical, syntactic, indirect, and discourse paraphrases. The resolver returns a canonical cue ID, whose deterministic cue key then queries the binder. Strict reinstatement requires all four paraphrases to recover the correct memory. Pairwise agreement averages the six paraphrase pairs for each target. SLA-address relocation corrects one key-to-locus association within the binder and tests it after 1, 2, 4, 8, and 16 unrelated insertions. Overwrite changes $\mathcal{V}_{s_r}(a_r)$ at the same locus. Repeated rebinding alternates the same record between two designated loci for eight cycles. These checkpoint studies isolate target retargeting and collateral behavior in the shared binder. The preregistered continuous verified-locator experiment below instantiates the complete registry/store/occupancy move under the formal locator contract. Route protocols keep the address layout fixed, so relocation is evaluated separately from local relation semantics. Each update condition uses 500 trials. Target revision metrics report current-target recovery and obsolete-state leakage.

Non-target preservation is evaluated in two matched 500-trial evaluations. The primary same-run evaluation reproduces the frozen target-revision checkpoint definitions and target outcomes exactly. It then uses the corresponding binder states with the same trial identities, targets, relocation loci, update definitions, and deterministic key namespace as the primary revision study. The second evaluation uses an independent binder realization under its separately fixed preservation-replication key namespace listed in Table B.4. The controlled study evaluates scenario-specific checkpoint states rather than asserting one continuous insertion-to-overwrite-to-rebinding trajectory. In both evaluations, each of the 511 pre-existing non-target records is checked at all 16 reported checkpoints: pre-update, post-relocation, after cumulative insertions of 1, 2, 4, 8, and 16 records, post-overwrite, and after each of eight rebinding cycles. Inserted records are evaluated at the five cumulative-insertion checkpoints, contributing $500(1 + 2 + 4 + 8 + 16) = 15{,}500$ observations. Overwrite/rebinding scenarios do not include those inserted records. These five checkpoints therefore define the inserted-record preservation coverage. Top-1 preservation means that the evaluated record's designated current locus remains rank 1. Top-8 preservation means that it remains within the first eight decoded candidates, and payload preservation requires the evaluated lookup to retain the unchanged payload. In each evaluation, the full-sequence pre-existing-record component contributes 4,088,000 repeated observations and the insertion-checkpoint component contributes 15,500. This gives 4,103,500 evaluated non-target record–checkpoint observations in total. Every reported preservation outcome is 1.000 with zero failures. Table C.7 summarizes both evaluations.

Tables C.5, C.6, and C.7 therefore carry the controlled-detail summary directly: Table C.5 lists the paraphrase families, Table C.6 reports the exact revised-target checks, and Table C.7 records both exhaustive collateral-preservation audits.

For the payload-variation and paraphrase retrieval studies, evidence hit@8 means inclusion of the target payload among the admitted Top-8 evidence. In the conflict split, semantic controls can include the target somewhere among eight blocks even when they do not rank it first, so target-record Top-1 and evidence hit@8 measure different outcomes. Strict reinstatement groups the four paraphrases of the same target. A leakage value of zero means that no Top-1 leakage event was observed among the stated trials. Table C.1 reports the independent-strata payload-variation traces, and Table C.2 reports the same cue–payload conflict summarized in Table 1, row 1. Table C.4 reports reader and latency measurements from the same controlled study. Tables C.5 and C.6 report paraphrase reinstatement and target revision. Table C.7 reports the same-run and independent non-target preservation evaluations.

Perfect evidence hit@8 means that the correct evidence is present in the admitted Top-8 set. Downstream answer generation is measured separately by the reader metrics. Because each independent-strata C value uses a separate corpus realization, cross-C values summarize distinct distractor-count conditions. The nested control below isolates within-family growth in C. The matched nested BGE-M3 control above increases C within shared families while preserving Payload-RAG target ordering, and Payload RAG, Stable-random, and Full SLA remain at 1.000. Reader exact match (EM) and F1 are downstream generation outcomes rather than access-correctness metrics, so a correct evidence hit need not imply an exact generated answer.

Table C.1: Independent-strata payload-variation traces. Each C is a separately generated payload-multiplicity condition with 200 test families, conditions are not nested within shared families.

C	Method	Top-1	Top-8	Evidence hit@8
1	Payload RAG	0.000	0.000	0.000
1	Stable-random	1.000	1.000	1.000
1	Full SLA	1.000	1.000	1.000
10	Payload RAG	0.000	0.000	0.000
10	Stable-random	1.000	1.000	1.000
10	Full SLA	1.000	1.000	1.000
100	Payload RAG	0.000	0.000	0.000
100	Stable-random	1.000	1.000	1.000
100	Full SLA	1.000	1.000	1.000
500	Payload RAG	0.005	0.005	0.005
500	Stable-random	1.000	1.000	1.000
500	Full SLA	1.000	1.000	1.000
1000	Payload RAG	0.000	0.005	0.005
1000	Stable-random	1.000	1.000	1.000
1000	Full SLA	1.000	1.000	1.000

Table C.2: Cue–payload conflict. Q1–Q5 are equal-size strata ($n = 200$ each) ordered by increasing held-out E5-verifier distractor-over-target margin, the exact min/median/mean/max ranges are stated in the protocol text above. Top-1 uses each method's own ranking score, while target evidence hit@8 is 1.000 for all three rows in every stratum.

Method	Q1	Q2	Q3	Q4	Q5
Payload RAG Top-1	0.000	0.000	0.005	0.000	0.005
Stable-random Top-1	1.000	1.000	1.000	1.000	1.000
Full SLA Top-1	1.000	1.000	1.000	1.000	1.000

Table C.3: Reader EM on the same cue–payload conflict strata. Target evidence hit@8 is 1.000 for Payload RAG, Stable-random, and Full SLA in every stratum, so these reader values report downstream generation behavior separately from locator correctness.

Method	Q1	Q2	Q3	Q4	Q5
Payload RAG	0.190	0.080	0.235	0.170	0.190
Stable-random	0.690	0.805	0.835	0.925	0.905
Full SLA	0.515	0.505	0.665	0.760	0.760

Table C.3 shows that Stable-random and Full SLA share target Top-1 $= 1.000$, while Stable-random reader EM exceeds Full SLA by 0.145–0.300 across Q1–Q5. Because each codebook induces its own Top-8 ranking, the two reader contexts can differ in the other seven admitted records and their order even when target rank and target inclusion agree. Operational SLA dereference uses verified Top-1. The unreranked Top-8 diagnostics characterize address-code rankings rather than semantic expansion, since locus proximity encodes address geometry rather than semantic relevance. Reader EM therefore characterizes the resulting contexts separately from locator correctness.

In Table C.6, obsolete-payload leakage applies to payload overwrite, where a previous payload version exists. Relocation and rebinding change locator association without creating an obsolete payload version. Repeated rebinding also dereferences the target after every cycle. Current-payload correctness is 1.000 across all 4,000 target cycle observations.

Preregistered continuous verified-locator composition. A complementary 500-trial experiment tests the locator contract in one persistent state sequence. The initial 512 records occupy addresses 0–511. 16 online insertions use 512–527, and each target relocation is fixed before execution to a deterministic free address in 528–1023. The original target slot then becomes the alternate free endpoint for eight rebindings. RLS and native Delta use the same trial identities, keys, locus codes, registry/store semantics, and operation schedule. Only the shared-state update rule differs. The state

Table C.4: Reader EM and per-example wall-clock latency for Payload RAG, Stable-random, and Full SLA across the five independent payload-variation conditions. The corresponding access outcomes are in Table C.1.

C	Method	Reader EM	Mean latency (s)	Median latency (s)
1	Payload RAG	0.000	1.392	1.292
1	Stable-random	0.100	1.619	1.688
1	Full SLA	0.190	1.465	1.340
10	Payload RAG	0.000	1.277	1.204
10	Stable-random	0.210	1.426	1.395
10	Full SLA	0.200	1.424	1.250
100	Payload RAG	0.000	1.282	1.189
100	Stable-random	0.200	1.221	1.182
100	Full SLA	0.200	1.378	1.251
500	Payload RAG	0.000	1.277	1.186
500	Stable-random	0.140	1.192	1.155
500	Full SLA	0.175	1.282	1.192
1000	Payload RAG	0.000	1.324	1.250
1000	Stable-random	0.215	1.179	1.189
1000	Full SLA	0.250	1.246	1.196

Table C.5: Reinstatement by paraphrase family. Each family contains 500 target–paraphrase rows.

Family	Resolver Top-1	Address Top-1	Address Top-8	Evidence hit@8	Reader EM
Lexical	0.964	0.964	0.966	0.966	0.268
Syntactic	0.968	0.968	0.970	0.970	0.390
Indirect	0.894	0.894	0.898	0.898	0.318
Discourse	0.992	0.992	0.994	0.994	0.182
Overall	0.9545	0.9545	0.9570	0.9570	0.2895

Table C.6: Revision outcomes for the revised target. Panel (a) reports current-state recovery, panel (b) reports obsolete-state diagnostics. Current/new recovery is exact at every reported relocation checkpoint and rebinding cycle, while lower-ranked obsolete addresses can remain in the fixed slot-codebook Top-8.

(a) Current-state recovery

Intervention	Trials	Top-1	Top-8	Current payload
SLA-address relocation	500/checkpoint	1.000	1.000	1.000
Overwrite	500	1.000	n/a	1.000
Repeated rebinding	500/cycle	1.000	1.000	1.000

(b) Obsolete-state diagnostics

Intervention	Address Top-1	Address Top-8	Obsolete payload hit
SLA-address relocation	0.000	0.014	n/a
Overwrite	n/a	n/a	0.000
Repeated rebinding	0.000	0.010–0.014	n/a

persists through relocation, cumulative insertions of 1/2/4/8/16 records, payload overwrite, and eight rebindings. Active-handle, occupancy, occupied-slot identity, payload, Top-1, and Top-8 checks are evaluated throughout. All 500 schedules pass the free-slot preflight.

The inserted-record total combines 15,500 observations at cumulative insertion checkpoints with 72,000 observations after overwrite and the eight subsequent rebindings. For native Delta, the maximum target-write residual is 5.07×10^{-7} and the maximum collateral-equation residual is 4.28×10^{-8}. The minimum observed pre-existing non-target margin is 0.131, while target–inserted $|\cos|$ has mean/median/P95/max 0.01240/0.01033/0.03085/0.06351. The RLS maximum target-write residual is 3.65×10^{-7} and its minimum non-target margin is 0.190. These continuous results test the target/collateral mechanism under the full verified-dereference state contract.

Table C.7: Matched non-target preservation evaluations under the controlled revision checkpoint protocol. The independent unit is the revision trial ($n = 500$), observation counts are repeated record–checkpoint checks within trials. The primary same-run evaluation uses the exact target-revision checkpoint definitions and key namespace of the primary revision study, while the independent replication uses the separately fixed preservation-replication key namespace in Table B.4. Pre-existing non-targets are evaluated through all 16 checkpoints and inserted records at the five cumulative-insertion checkpoints in both evaluations.

Evaluation	Coverage	Observations	Top-1	Top-8	Payload preserved
Primary same-run	Pre-existing non-target, all 16 checkpoints	4,088,000	1.000	1.000	1.000
Primary same-run	Inserted records, insertion checkpoints	15,500	1.000	1.000	1.000
Primary same-run	Total evaluated non-target observations	4,103,500	1.000	1.000	1.000
Independent replication	Pre-existing non-target, all 16 checkpoints	4,088,000	1.000	1.000	1.000
Independent replication	Inserted records, insertion checkpoints	15,500	1.000	1.000	1.000
Independent replication	Total evaluated non-target observations	4,103,500	1.000	1.000	1.000

Table C.8: Preregistered continuous verified-locator stress under complete free-slot, single-occupant registry/store semantics. Counts are exhaustive record–checkpoint checks within 500 trials, not independent samples.

Backend	Target	Inserted full horizon	Full non-target	Wrong / fail / registry
RLS-SLA	7,500/7,500	87,500/87,500	3,920,000/3,920,000	0 / 0 / 0
FLA Delta-rule SLA	7,500/7,500	87,500/87,500	3,920,000/3,920,000	0 / 0 / 0

C.2 Cross-bank routing and high-churn stress

A separate deterministic stress test evaluates cross-bank routing and retirement semantics without changing the key/locus construction. Eight banks of capacity 1,024 begin with 512 records each. The workload contains 20,000 globally serialized operations: 5,000 revisits, 3,000 same-bank relocations, 3,000 cross-bank migrations, 3,000 insertions, 2,000 payload overwrites, 2,000 retire operations, and 2,000 retire-slot reuse operations. All locator writes use native Delta. The run contains 11,000 such writes. Mutations are evaluated under one logical commit order, so the experiment tests atomic/linearizable orchestration rather than lock-free or unsynchronized concurrency.

For a cross-bank migration, the destination bank stages a free slot, writes the key to the new locus, and verifies Top-1, registry identity, and payload before publication. The commit then changes $\Sigma(h_r)$ to the destination and clears the source registry/store entry. This commit ends the record's protected segment in the source bank. The committed destination write starts a new segment in the destination bank, so subsequent certificate accumulation uses only later writes to that authoritative bank. The old source-bank binder projection is not deleted. After routing changes it is unreachable through correct Σ and is treated as maintenance/capacity debt rather than an active reference. Retirement likewise removes routing and occupancy before slot reuse, so a stale retired handle must return $\perp$.

The stress result separates transaction safety from long-horizon availability. Cross-bank targets and routed revisits remain exact under the stated commit model, retired handles do not alias reused slots, and binder drift produces fail-closed misses rather than wrong-record returns. Without maintenance, however, many shared writes eventually consume the decoding margin even at moderate load. The sequence identity remains accurate through this transition. The source-bank segment ends at migration, and the committed destination state starts a new segment whose later interference comes only from writes to that authoritative bank.

We therefore replay the exact same 20,000 operations with the existing sufficient condition $\|\Delta_j\|_2 < \mu_j/2$ as the only maintenance trigger. Every requested native-Delta write is first staged in the same

Table C.9: Cross-bank high-churn stress. Transaction correctness and long-horizon shared-state preservation are reported separately.

Measurement	Result
Cross-bank migrations	3,000, migrated target exact 3,000/3,000, post-commit revisit 3,000/3,000
Source/destination collateral	source preserved 3,000/3,000 events, destination preserved 2,616/3,000, with 444 new fail-closed collateral misses
Identity safety	0 wrong-record returns, 0 retired-handle aliases after 2,000 reuse operations
Load envelope	468–548 active records/bank, 0 capacity-exhaustion events
Long-horizon availability	first checkpoint Top-1 failure at operation 5,700, first Top-8 failure at 16,300
Exact cumulative certificate	predicted/actual minimum gap $-0.16548/-0.16548$, maximum projection residual 6.43×10^{-7}

way as in the baseline. If all destination-bank non-targets remain inside the bound, the write is committed normally. Otherwise the candidate is not published and the affected bank is rebuilt from its current active bindings. Retired and migrated-away handles are excluded, and every rebuilt association must pass Top-1, Top-8, registry-identity, and payload verification before publication. Direct writes and successful rebuilds start new protected segments. No safety factor, threshold sweep, or periodic rebuild schedule is used.

Table C.10 compares the matched traces. The certificate intercepts 1,744 of 11,000 candidate writes. 9,256 commit as ordinary native Delta. Every rebuild restores all active records exactly. Across all 201 checkpoints there is no Top-1 or Top-8 failure, no fail-closed active miss, no wrong-record return, and no post-rebuild verification failure. Rebuild construction and verification consume 367.40 s of the 654.11 s run. Total runtime is $2.128\times$ the 307.44 s unmaintained baseline. The minimum exact score gap remains positive at 0.14646 and the maximum certificate-projection residual is 5.08×10^{-7}. The remaining control assumption is trusted registry/routing metadata at atomic/linearizable commit boundaries.

Table C.10: Certificate-triggered maintenance on the exact same 20,000-operation cross-bank trace. The unmaintained run is retained as a negative control.

Metric	Unmaintained	Certificate-maintained
First checkpoint Top-1 failure	5,700	none
First checkpoint Top-8 failure	16,300	none
Wrong-record returns	0	0
Fail-closed active misses	6,756	0
Certificate-triggered rebuilds	0	1,744
Ordinary Delta writes	11,000	9,256
Total wall time (s)	307.44	654.11

C.3 Native Delta-rule backend transfer

The realization-transfer experiment holds the controlled checkpoint definitions fixed and changes only the shared-state update rule. The stored canonical SHA256 digest is interpreted directly under the protocol's big-endian NumPy-PCG64 rule to reconstruct the original 4,096-dimensional Gaussian identity key. The RLS reference run first reproduces the original target trajectory before the matched Delta comparison. The matched Delta run then uses the same 500 trial identities, 512-record banks, target handles, checkpoint loci, cumulative insertion schedule, overwrite state, eight-cycle rebinding definitions, locus codes, and identity keys, while replacing the RLS update by FLA's native `delta_rule_recurrence` with $\beta = 1$. Table C.11 reports the matched outcome.

Both backends remain exact at every recorded checkpoint. For the native Delta operator, the original matched transfer has maximum FP32 target-write and collateral-identity contract residuals 9.54×10^{-7} and 4.17×10^{-7}, respectively. Its sampled identity-key mean/max absolute off-diagonal cosine are 0.01247/0.07934.

Table C.11: Controlled shared-state backend transfer under identical identity keys and matched revision checkpoint states. Collateral coverage includes all 511 pre-existing non-targets at all 16 checkpoints plus inserted records at their five cumulative-insertion checkpoints.

Backend	Target Top-1	Target Top-8	Payload	Collateral	Checks/backend
RLS-SLA	1.000	1.000	1.000	1.000	4,103,500
FLA Delta-rule SLA	1.000	1.000	1.000	1.000	4,103,500

Controlled signed-overlap stress. To test the collateral law causally rather than infer it from a naturally low-overlap trajectory, a preregistered stress test fixes $d_k = 4096$, $d_z = 16$, FP32, $\beta = 1$, bank sizes $n_s \in \{512, 1024\}$, five deterministic seeds, and 19 signed overlaps

$$\rho \in \{0, \pm 0.01, \pm 0.02, \pm 0.04, \pm 0.08, \pm 0.16, \pm 0.32, \pm 0.50, \pm 0.70, \pm 0.85\}.$$

For target key k_r, an orthogonal unit vector u constructs $k_j(\rho) = \rho k_r + \sqrt{1-\rho^2}u$, while other non-target keys are projected away from the controlled span. Their 145,350 preservation checks are therefore zero-overlap controls that validate the zero-interference corollary. The correlated k_j isolates pairwise margin erosion rather than simultaneous nonzero-overlap interference. All 190 conditions are valid before the target write. The actual FLA native Delta update is then applied. The analytical prediction is $\Delta_j^{\text{pred}} = \epsilon(k_r^\top k_j)$, not a surrogate update.

Table C.12: Controlled native-Delta overlap stress and natural-key operating geometry. The controlled failure counts concern the correlated non-target, all other non-targets remain exact. At extreme overlap, a wrong-address outcome denotes Top-1 landing on a different occupied slot, the formal SLA interface still requires occupant-handle verification before payload acceptance.

Measurement	Result
Collateral vector law	identity-line fit at FP32 precision, max vector residual 4.88×10^{-7}
Target write / direction	max target-write residual 6.33×10^{-7}, signed direction is 1.000 at displayed precision
Operational stress	target 190/190, controlled non-target preserved through $\lvert\rho\rvert = 0.32$, first Top-1/verified-dereference degradation at $+0.50$, first Top-8 miss at $+0.70$ ($n_s = 1024$)
Extreme-overlap failures	35 controlled-non-target Top-1 locator failures (25 wrong-address, 10 fail-closed) and 15 Top-8 misses, 145,350/145,350 other non-target checks preserved
Natural key geometry	500 banks, 65,408,000 pairs: mean/median/P90/P95/P99/max $\lvert\cos\rvert$ = 0.0124666/0.0105402/0.0256978/0.0306209/0.0402408/0.0895067

The zero-intercept fit is effectively the identity line at FP32 precision, and sign reversal follows the predicted direction. A separate preregistered five-bank source-level replay reproduces the law with maximum collateral residual 2.81344×10^{-8}. For independent normalized Gaussian keys, the predicted mean absolute overlap is 0.01247 at $d_k = 4096$, matching the measured 0.01247 across the full 500-bank analysis. The observed natural maximum is 0.0895 and remains inside the controlled low-interference region. These results support both parts of the mechanism. Native Delta follows Eq. (10), and operational degradation appears only after the imposed overlap becomes large enough to consume the decoding margin. The threshold is empirical for this codebook/state configuration, not a universal constant.

Sequence-level cumulative-interference audit. The sequence-level audit uses logging-only replays of the continuous and LongMemEval-V2 frozen protocols, requiring exact reproduction of the previously reported aggregate gates before any cumulative certificate is computed. The continuous replay reproduces 7,500/7,500 target, 3,920,000/3,920,000 non-target, and 87,500/87,500 inserted-record checks with zero wrong/fail-closed events. The LongMemEval-V2 replay reproduces 439 eligible questions, 2,475 seeds, 24,369 mutations, 22,275 retained checks, 369,485 collateral checks, 4,776 relation transitions, and zero wrong-record returns.

The sequence identity remains numerically tight in FP32: residual median/P95/P99 are $2.296157/3.170394/3.818735 \times 10^{-7}$ for the continuous trace and $1.676627/4.049620/4.859880 \times$

Table C.13: Sequence-level native-Delta certificate audit. Exact score-space coverage is the operational certificate in Corollary 2, norm and triangle rows are progressively more conservative sufficient conditions. Prefix counts are exhaustive within protected segments.

Quantity	Continuous	LongMemEval-V2
Protected segments	268,500	62,572
Evaluated prefixes	6,519,500	1,092,979
Observed Top-1	6,519,500/6,519,500	1,092,979/1,092,979
Exact-score certificate	6,519,500/6,519,500	1,092,979/1,092,979
Norm certificate	6,253,020 (95.913%)	1,042,881 (95.416%)
Triangle certificate	1,752,968 (26.888%)	367,067 (33.584%)
Minimum exact headroom	0.131139805	0.136360774
Max identity residual	9.043125×10^{-7}	7.285685×10^{-7}

10^{-7} for LongMemEval-V2. The normalized norm headroom $1 - 2\|\Delta_j(u)\|/\mu_j^{(\tau)}$ reaches -1.0081 and -0.9481, respectively, so the norm sufficient condition can fail even while every exact score gap remains positive. The triangle envelope is more conservative still. These are bound violations, not locator errors. The exact score-space condition, observed Top-1, and verified-dereference outcomes remain aligned at every evaluated prefix.

C.4 Structured relation execution, composition, and storage

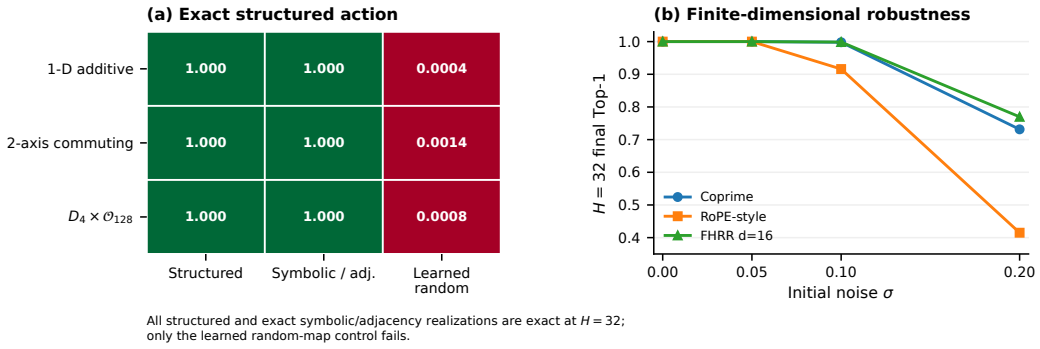


Figure C.1: **Structured relation realizations separate exact action from robustness.** (a) A heatmap-style summary shows that, at $H = 32$, structured analytic and exact symbolic/adjacency realizations achieve Top-1 1.000 across additive, commuting, and non-commutative families, while learned random-map controls remain near zero. (b) Under matched input noise, exact structured phase realizations develop distinct finite-dimensional robustness.

The operator study uses the fixed 16-dimensional locus representation with $N = 1{,}024$ and local displacements $\Delta \in \{-4, -1, +1, +4\}$. Associative binder state is FP32 as specified above. Analytic relation-state construction, transformation, and algebraic residual checks use float64 arithmetic, while decoded Top-1/Top-8 are the cross-realization correctness metrics. Periodic loci use analytic $R(\Delta)$ without fitting. The measured quantities are finite-precision decoding on disjoint test addresses and composition. The learned-map random-code relation control uses 1,024 independently sampled, L_2-normalized 16-dimensional Gaussian codes. It is an algebraic null control: one linear map is fitted per relation to test whether an equally sized unstructured representation supports the same shared global linear action without relational structure. Training, development, and test addresses are disjoint. For each relation and $K \in \{25, 50, 100, 256, 512\}$ nested training pairs, the relation-specific map is fitted by ridge least squares,

$$W_\Delta = Y^\top X (X^\top X + 10^{-4} I)^{-1}, \tag{27}$$

where rows of X and Y are source and target codes. The protocol fixes $K = 100$ before any test composition or perturbation result is evaluated. The one-step table reports every candidate fitting budget $K \in \{25, 50, 100, 256, 512\}$, while composition and perturbation use the fixed $K = 100$

maps. The one-dimensional comparisons use this $K = 100$ composition control. The two-axis and $D_4 \times \mathcal{O}_{128}$ studies use their own pre-test random-map configurations, so structured-versus-random results are evaluated within relation family. For each $H \in \{1, 2, 4, 8, 16, 32\}$, 5,000 test routes use addresses disjoint from the random-code fitting and development sets. Periodic states apply $R(\Delta_H) \cdots R(\Delta_1)$ on the same routes without fitting. Random states apply the corresponding learned operators and are L_2 normalized after every step. Complete-route success requires correct Top-1 decoding at every intermediate state. Explicit integer coordinates and four-neighbor adjacency tables follow the same routes. Noise trials add isotropic Gaussian perturbations to the initial state and test direct decoding, one step, and 8, 16, or 32 transformations over 2,048 paired trials per condition. The perturbation levels are $\sigma \in \{0, 0.01, 0.02, 0.05, 0.10, 0.20\}$. Table C.14 summarizes one-step and composed accuracy for the four realizations.

Table C.14: Analytic local-displacement execution and the unstructured-code control. One-step learned-map random-code values average the four displacements. Periodic $R(\Delta)$ is predefined, random-code composition uses the protocol-fixed $K = 100$ maps.

Method / training pairs	One-step Top-1	One-step Top-8	$H = 8$ Top-1	$H = 32$ Top-1	$H = 32$ complete
Periodic phase / 0	1.0000	1.0000	1.0000	1.0000	1.0000
Learned-map / 25	0.0000	0.0079	n/e	n/e	n/e
Learned-map / 50	0.0010	0.0079	n/e	n/e	n/e
Learned-map / 100	0.0010	0.0069	0.0014	0.0004	0.0000
Learned-map / 256	0.0000	0.0069	n/e	n/e	n/e
Learned-map / 512	0.0010	0.0089	n/e	n/e	n/e
Integer coordinate	1.0000	1.0000	1.0000	1.0000	1.0000
Adjacency table	1.0000	1.0000	1.0000	1.0000	1.0000

n/e denotes a condition outside the reported multistep protocol, not a missing experiment. Only the pre-fixed $K = 100$ learned-map realization is carried into $H = 8$ and $H = 32$ composition.

At $H = 32$, periodic minus random Top-1 is +0.9996 (95% CI 0.9990–1.0000). With $\sigma = 0.20$, direct decoding is 0.7212/0.9717 for periodic loci and 0.7725/0.9780 for random codes, so both codebooks remain directly decodable. After 32 transformations, periodic is 0.7344/0.9746 and the learned random control 0.0010/0.0078, diagnosing failure of the unstructured shared action rather than matched noise robustness. Table C.15 provides the matched structured-robustness comparison.

C.4.1 Structured phase baselines

Shared additive action is evaluated across three structured phase realizations: the coprime schedule, a RoPE-style phase code, and Fourier holographic reduced representation (FHRR) controls (Plate, 1995; Kleyko et al., 2022). For the 16-dimensional one-dimensional RoPE-style control, pair $j \in \{0, \ldots, 7\}$ uses $\omega_j = 10000^{-2j/16}$ and code block $[\cos(a\omega_j), \sin(a\omega_j)]/\sqrt{8}$. Displacement applies the matching block rotation by $\Delta\omega_j$. The two-axis 16-dimensional control uses two 8-real-dimensional subspaces, each with four pairs $\omega_j^{\text{axis}} = 10000^{-2j/8}$ for $j \in \{0, \ldots, 3\}$, and rotates only the displaced axis. For an FHRR control of real-equivalent dimension d, NumPy PCG64 via `default_rng(seed)` draws $d/2$ independent phases $\phi_j \sim \text{Unif}[-\pi, \pi)$ and uses the unit-norm real representation $z(a) = \sqrt{2/d} \bigoplus_j [\cos(a\phi_j), \sin(a\phi_j)]$. Displacement applies the matching block rotations, equivalently multiplying each complex phasor by $e^{\mathrm{i}\Delta\phi_j}$. The two-axis FHRR construction uses independent bases ϕ_j^x, ϕ_j^y and binds the coordinates in the same fixed width,

$$z(x, y) = \sqrt{2/d} \bigoplus_j [\cos(x\phi_j^x + y\phi_j^y), \sin(x\phi_j^x + y\phi_j^y)],$$

so an x or y displacement rotates each block by the corresponding $\Delta\phi_j^x$ or $\Delta\phi_j^y$. No structured baseline is fitted on test states. One-dimensional and two-axis evaluations use 500 routes per horizon. The one-dimensional noise evaluation uses 1,024 routes per condition. FHRR reports seeds 2026090501–2026090505 at real-equivalent dimensions 16, 64, and 256.

All structured phase controls are exact through $H = 32$ without perturbation in both the one-dimensional additive and two-axis commuting families. Table C.15 reports the common exact

result and a separate $H = 32, \sigma = 0.20$ one-dimensional perturbation evaluation using 1,024 matched routes per realization. This matched-baseline evaluation is distinct from the 2,048-paired-trial one-dimensional operator perturbation protocol in Fig. C.1b. The coprime point estimates are 0.731/0.977 here and 0.734/0.975 there. At 16 dimensions, the coprime and FHRR controls retain higher Top-1 than the RoPE-style schedule under this perturbation, while all three are exact in the unperturbed compositions. Across the five FHRR seeds, sample standard deviations of Top-1/Top-8 at $\sigma = 0.20$ are 0.021/0.006 at $d = 16$, 0.007/0.002 at $d = 64$, and 0.009/0.002 at $d = 256$. The higher-dimensional FHRR rows also increase mean noisy Top-1. These results separate the common exact-action behavior observed in the tested structured phase constructions from their representation-specific finite-dimensional margins. Table C.15 gives the exact/noisy point estimates, and Figure C.2 visualizes their divergence as noise increases.

Table C.15: Structured phase baselines. Unperturbed $H = 32$ columns report final Top-1 / complete-route accuracy, the noise column reports one-dimensional Top-1 / Top-8 after 32 transforms at $\sigma = 0.20$. FHRR values are means over five fixed seeds.

Representation	d	1D $H = 32$	2-axis $H = 32$	1D noise 0.20
Coprime periodic	16	1.000 / 1.000	1.000 / 1.000	0.731 / 0.977
RoPE-style phase	16	1.000 / 1.000	1.000 / 1.000	0.415 / 0.858
FHRR	16	1.000 / 1.000	1.000 / 1.000	0.770 / 0.978
FHRR	64	1.000 / 1.000	1.000 / 1.000	0.926 / 0.993
FHRR	256	1.000 / 1.000	1.000 / 1.000	0.947 / 0.992

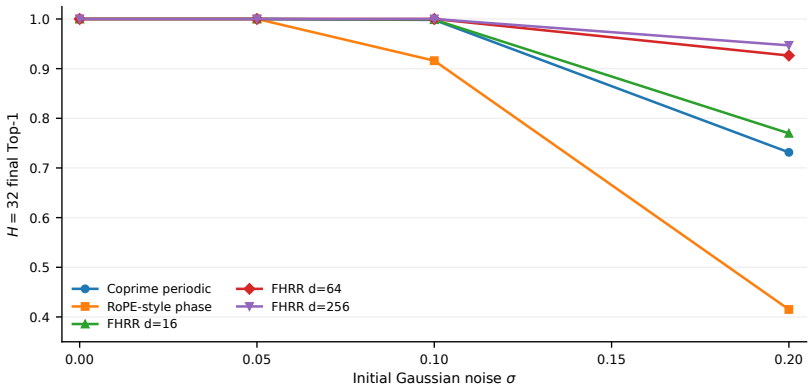


Figure C.2: **Noise reveals robustness differences that are invisible in the exact unperturbed regime.** All realizations have unperturbed $H = 32$ Top-1 of 1.000, but their curves separate as input noise increases. FHRR curves average five fixed seeds at each dimension, the coprime and RoPE-style rows are deterministic realizations.

C.4.2 Fixed-dimensional structured relation extensions

To move beyond a single additive coordinate without changing the state budget, we evaluate a commuting 1,024-state 32×32 product at $d_z = 16$. Four two-dimensional phase modules encode x with periods $(5, 7, 11, 13)$ and four encode y with periods $(17, 19, 23, 29)$, jointly normalized by $1/\sqrt{8}$. Axis operators $R_x(\pm 1)$ and $R_y(\pm 1)$ rotate only the corresponding phase blocks. The deterministic split holds out slabs only along x: $x = 0{:}15$ for training, 16:23 for development, and 24:31 for test, while all splits use $y = 0{:}31$. Transitions are retained only when both endpoints stay within one x slab. Independent 16-dimensional random codes receive one ridge-fitted 16×16 map per axis action. The protocol fixes the random control at 25 fitting pairs per axis action before test evaluation, and all two-axis test statistics condition on that map configuration. Integer (x, y) and four-neighbor adjacency execute the identical test routes. Because the random-map fit/development/test slabs are disjoint while the analytic operators have no fitting stage, the evaluation measures execution on the common test slab together with commuting composition. The protocol and paired route/outcome records are identified by SHA256 in Table F.1.

The non-commutative family uses the Cartesian state set $D_4 \times \mathcal{O}_{128}$, where $\mathcal{O}_{128} = \{0, \ldots, 127\}$ is a bounded ordinal coordinate rather than a cyclic group. Write a group element as $g = q^k f^\epsilon \in D_4$ with $k \in \{0, 1, 2, 3\}$ and $\epsilon \in \{0, 1\}$, where q is a quarter-turn and f a reflection. Let $v_0 = [\cos 0.37, \sin 0.37]^\top$, $\pi(q) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, $\pi(f) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$, and $p_1, \ldots, p_7 = (5, 7, 11, 13, 17, 19, 23)$. The fixed angle 0.37 avoids reflection-fixed axes, giving eight distinct D_4 orbit states. The complete unit-norm code is

$$z(g, t) = \frac{1}{\sqrt{8}} \left(\pi(g) v_0 \oplus \bigoplus_{m=1}^{7} [\cos(2\pi t/p_m), \sin(2\pi t/p_m)]^\top \right). \tag{28}$$

The shared actions $A_Q, A_{Q^{-1}}, A_F$ act on the two-dimensional group block and A_{T_+}, A_{T_-} on the seven ordinal phase blocks. Matrix products act right-to-left on column states. Thus the ordered sequence F then Q is represented by $A_Q A_F$, whereas Q then F is $A_F A_Q$. These generally differ. The group actions satisfy $A_Q^4 = I$, $A_F^2 = I$, and $A_F A_Q A_F = A_{Q^{-1}}$. Mathematically, $T_+(g, t) = (g, t + 1)$ is defined for $t < 127$ and $T_-(g, t) = (g, t - 1)$ for $t > 0$, with no wraparound. Train/development/test boundaries constrain only which valid relation pairs enter each data split. They do not redefine the mathematical domains of T_+ or T_-. The split is $t = 0{:}63/64{:}95/96{:}127$ with all eight group states in each, and evaluation relation pairs are retained only when both endpoints lie in the same split.

The random-code control uses seed 2026090301, pair counts $\{25, 50, 100, 256\}$, and ridge grid $\{10^{-6}, 10^{-4}, 10^{-2}, 1, 100\}$. For each relation, training transitions are sorted by source and target index and the first K are retained. Macro development Top-1 selects one global K with smaller K breaking ties, yielding $K = 50$ before any test route is scored. Ridge strength is selected per relation on development data. At $K = 50$, the selected values for (Q, Q^{-1}, F, T_+, T_-) are $(1, 10^{-6}, 10^{-6}, 100, 10^{-6})$. Mixed and order-sensitive challenge routes use seed 2026090302 with 5,000 routes per horizon. Mixed routes contain at least two relation labels for $H \geq 2$ and both a D_4 and an ordinal relation for $H \geq 4$. Challenge routes contain an adjacent ordered pair Q, F or F, Q and, for $H \geq 4$, an ordinal move, with the designated order approximately balanced. Paired route bootstraps use 10,000 draws with seed 2026090303. The symbolic-tuple control stores (k, ϵ, t) explicitly and uses the exact left-action updates

$$\begin{aligned} Q &: (k, \epsilon, t) \mapsto (k + 1 \bmod 4, \epsilon, t), & Q^{-1} &: (k, \epsilon, t) \mapsto (k - 1 \bmod 4, \epsilon, t), \\ F &: (k, \epsilon, t) \mapsto (-k \bmod 4, 1 - \epsilon, t), & T_\pm &: (k, \epsilon, t) \mapsto (k, \epsilon, t \pm 1) \end{aligned}$$

with the same bounded domain for $T_\pm$. The relation-labeled adjacency control stores the explicit target of each valid relation edge. Both receive the same source states, relation labels, and route schedules as the analytic code.

Across the four two-axis actions, structured Top-1/Top-8 are 1.0000/1.0000, whereas the fixed 25-pair random maps obtain 0.0011/0.0032. Five thousand mixed-axis test routes per $H \in \{1, 2, 4, 8, 16, 32\}$ remain inside the test region. Structured, integer, and adjacency complete-route accuracy stays 1.0000 at every H, while at $H = 32$ the random control has final Top-1 0.0014 and complete-route accuracy 0.0000. At $H = 32$, structured minus random final Top-1 is +0.9986 (95% paired-bootstrap CI 0.9974–0.9996), and the complete-route difference is +1.0000 (95% CI 1.0000–1.0000). The commutation error $\|R_x R_y z - R_y R_x z\|_2$ has mean, median, P95, and maximum all equal to 0.0.

The $D_4 \times \mathcal{O}_{128}$ family likewise achieves perfect decoded accuracy on the disjoint test states: mean one-step Top-1/Top-8 are 1.0000/1.0000 versus 0.0008/0.0056 for the development-selected random maps. All 1,200 labeled transitions from 240 interior test sources decode Top-1 correctly and each source has five distinct relation targets. Both ordered two-step sequences $Q \to F$ and $F \to Q$ decode correctly and to distinct targets on 256/256 test states. For column-state composition, the tested identities $A_Q^4 z = z$, $A_F^2 z = z$, and $A_F A_Q A_F z = A_{Q^{-1}} z$ hold on all 256 test states. All five one-step analytic actions decode their intended relation targets correctly throughout the reported test transition set. Decoded correctness is the cross-realization comparison metric, while the algebraic identity checks provide supplementary numerical checks. Five thousand ordinary mixed routes are evaluated at each $H \in \{1, 2, 4, 8, 16, 32\}$, and a second set of 5,000 challenge routes at each $H \in \{2, 4, 8, 16, 32\}$ requires an adjacent ordered Q, F or F, Q pair (and, for $H \geq 4$, an ordinal move). At $H = 32$, structured analytic, symbolic-tuple, and adjacency realizations retain final Top-1 and complete-route accuracy 1.0000 on both route sets. The random control reaches 0.0010/0.0000

on mixed routes and 0.0008/0.0000 on challenge routes. On the challenge set, structured minus random final Top-1 is +0.9992 (95% CI 0.9984–0.9998), with complete-route difference +1.0000 (95% CI 1.0000–1.0000). Table C.16 summarizes representative test results.

Table C.16: Fixed-$d_z = 16$ structured relation extensions. Two-axis $H = 32$ values use mixed-axis routes, $D_4 \times \mathcal{O}_{128}$ $H = 32$ values use the non-commutative challenge routes. Random rows use their fixed pre-test map configurations.

Family / realization	One-step Top-1	One-step Top-8	$H = 32$ Top-1	$H = 32$ complete
Two-axis structured	1.0000	1.0000	1.0000	1.0000
Two-axis learned-map	0.0011	0.0032	0.0014	0.0000
$D_4 \times \mathcal{O}_{128}$ structured	1.0000	1.0000	1.0000	1.0000
$D_4 \times \mathcal{O}_{128}$ learned-map	0.0008	0.0056	0.0008	0.0000

Table C.17 compares relation-representation object sizes and operator metadata for the one-dimensional executor realizations. Locator storage is accounted separately in the main design-trade-off analysis and Appendix E.

Table C.17: Relation-representation object sizes and operator metadata for one-dimensional displacement. "State bytes" are the explicit per-record representation size if materialized, deterministic phase states can instead be generated from their coordinate, so this is not persistent-system storage. Bytes assume FP32 distributed vectors and int32 symbolic coordinates or edges.

Representation	State bytes	Dense-op bytes	Min-op bytes	Bytes / edge	Capability
Coprime phase ($d = 16$)	64	1,024	4	0	shared analytic 1D displacement
RoPE-style phase ($d = 16$)	64	1,024	4	0	shared analytic 1D displacement
FHRR ($d = 16$)	64	1,024	4	0	shared phasor displacement
Learned-map random code ($d = 16$)	64	1,024	1,024	0	fitted transform
Integer coordinate	4	4	4	0	exact symbolic arithmetic
Adjacency table	0	0	0	4	exact stored neighbor

The four local displacement labels fit in one int32. Under the representation-object accounting above, the minimal per-relation metadata is therefore 4 bytes for coprime, RoPE-style, FHRR, and integer-coordinate realizations. The coprime realization also stores the eight int64 periods once (64 shared bytes), while the RoPE-style frequency schedule is generated from the fixed base 10000. For FHRR, an explicitly materialized real-equivalent d-dimensional state uses $4d$ bytes (64, 256, and 1,024 bytes for $d = 16, 64, 256$). Its shared phasor basis uses the same $4d$ bytes once per realization. A dense real $d \times d$ operator would cost $4d^2$ bytes. The adjacency table uses four int32 edges per record, or 16 bytes per record. In the one-dimensional periodic SLA routes, $z(a)$ is generated from the current address rather than stored as an additional per-record vector. Appendix B and Table E.5 separately account for locator-only costs, including the 8-byte cached-pointer control and SLA binder components. The two-axis and $D_4 \times \mathcal{O}_{128}$ experiments use their own fixed shared operators and are reported separately in Table C.16.

C.5 Deterministic route protocol

The deterministic traversal study uses 512-record memories and 500 routes at each $H \in \{2, 4, 8, 16\}$. Non-terminal payloads specify one of five operations: next, previous, plus four, minus four, or re-address by canonical cue. Sixty percent of transitions are ± 1, 25% are ± 4, and 15% require re-addressing. Every method receives the same supplied operation and fallback cue. Full SLA uses $R(\Delta)$ for local moves and cue resolution only for re-addressing. Repeated cue search, stable-random addressing, and the periodic no-reuse control re-resolve by cue at every step, whereas the integer-coordinate control uses local integer addition. Complete-route success is 1.000 for every method at every tested hop count. The access paths nevertheless differ exactly by protocol: repeated cue search, stable-random addressing, and the periodic no-reuse control re-resolve by cue at every step, whereas Full SLA uses local relation execution except on the 15% re-address operations and the integer control uses local arithmetic. Measured wall-clock latency is reported separately in Appendix E.

C.6 CONTROLLED RELATION UNCERTAINTY

Periodic and integer representations agree whenever they receive the same hard relation decision. Soft superposition and finite beams can instead follow different trajectories. The following matched protocols separate uncertainty in the supplied relation from execution fidelity.

The deterministic route study supplies the relation label. A separate fixed-posterior intervention compares hard commitment, single-state periodic superposition, Top-1-projected superposition, and symbolic beams. At each local step, one action in $\{-4, -1, +1, +4\}$ has probability 0.7 and the other three 0.1. Five fixed tables make posterior argmax accuracy on 27,273 development steps equal to $\alpha \in \{0.70, 0.80, 0.85, 0.90, 0.95\}$. Incorrect argmax identities are fixed row-by-row and shared by every executor. Test tables cover 500 routes at each $H \in \{2, 4, 8, 16, 32\}$ with a separate route generator using cue-readdress probability 0.12, where re-address is a common hard reset. Hard periodic and integer follow the same argmax. Soft periodic propagates $\sum_\lambda p(\lambda) A_\lambda z$. The projected variant decodes Top-1 after each step. Symbolic beams retain two, four, or eight states, merging identical coordinates by summed path mass before pruning. The 136,150 evaluated step posteriors and all aggregates are included in the source package and covered by the supplementary checksum file.

Hard periodic and integer have identical Top-1/Top-8, complete-route, and payload outcomes throughout. At $H = 32$, single-state soft superposition is below hard execution in every regime. No beam width dominates hard execution across all regimes. Top-1 projection matches hard final Top-1 in four regimes and differs by 0.002 at $\alpha = 0.95$ (Table C.18). Thus common hard decisions agree across representations, while posterior aggregation can change trajectories.

Table C.18: Controlled relation uncertainty at $H = 32$ (500 routes/regime), final Top-1. α is the development-set argmax-correctness construction target, "Action" is realized test-step argmax accuracy at $H = 32$. Hard periodic and hard integer have identical decoded outcomes, "Proj." is soft periodic followed by Top-1 projection after every step. Complete $H = 2, 4, 8, 16, 32$ aggregates are included in the supplementary data.

α	Action	Hard	Soft	Proj.	Beam-2	Beam-4	Beam-8
0.70	0.699	0.356	0.330	0.356	0.354	0.362	0.358
0.80	0.799	0.468	0.416	0.468	0.470	0.478	0.482
0.85	0.848	0.530	0.464	0.530	0.534	0.512	0.520
0.90	0.899	0.622	0.514	0.622	0.614	0.590	0.602
0.95	0.948	0.762	0.582	0.760	0.744	0.692	0.696

D NATURAL-WORKLOAD EVALUATION DETAILS

D.1 REALISTIC FIRST ACCESS AND MISTRAL CONFIGURATION EVALUATION

For realistic first access, the fixed primary sample contains 60 LongBench v2, 30 LongBench Pro, and 30 ∞Bench short-answer, QA, reasoning, multiple-choice, or retrieval-style examples selected before inference with seed 20260817 and stratified by official task/category and context-length quartile (Bai et al., 2025; Chen et al., 2026; Zhang et al., 2024). Long-form summarization and tasks requiring long open generation are excluded. The evaluated 120-example sample is identified by its SHA256 in Table F.1. Contexts are divided into contiguous 512-token blocks with the BGE-M3 tokenizer. Payload RAG indexes complete blocks. The multi-view descriptor index represents each block by the normalized mean of four fixed BGE-M3 channels: a heading channel (the detected heading, or the first meaningful sentence when no heading is found), an opening-sentence channel, the final sentence, and the complete block. The channels are averaged as defined rather than deduplicated. When the heading fallback and opening sentence coincide, that text contributes to both fixed channels in both Descriptor RAG and Descriptor-index SLA. This descriptor representation is used only for first-access localization. SLA dereference instead uses each selected block's internal cue. In this protocol, the visible cue is the detected heading when one is available. Otherwise it is the ingestion routine's first meaningful sentence. The routine combines that text with the document-local block ordinal and a fixed workload identifier. It then applies the canonicalization in Appendix B and stores the resulting bytes as c_r. Those bytes are the input to deterministic identity-key generation. Because

the document-local block ordinal also determines the initial source-order address in these unrelocated natural corpora, the natural first-access protocol evaluates the end-to-end selection-to-dereference path. Invariance to locator revision is evaluated by the controlled relocation/rebinding protocols, where c_r and k_r remain fixed while a_r changes. After selection, SLA operational access uses the retained handle only to obtain coarse document/bank scope and canonical cue/key material. The binder then decodes the current bank-local address. Global block IDs are used only to verify decoded record identity and evidence equivalence. They are not handle-resolved locators and are not inputs to binder ranking. Retrieval uses normalized inner product with Top-8 ranking (Johnson et al., 2021). Evidence preserves retrieval order and is capped at 4,096 reader tokens.

Descriptor-index SLA performs multi-view descriptor retrieval before stable dereference. Hybrid SLA does not search descriptor representations. Instead, it inherits the Payload-RAG selected handles and ingestion metadata. Each handle supplies only scope and cue/key material, after which the binder reconstructs the current SLA address and dereferences the payload. The selected block ID is compared with the decoded registry handle only as an integrity/evidence-equivalence check. Qwen3-8B is the primary reader for the 120-example evaluation: revision `b968826d9c46dd6066d109eabc6255188de91218`, bitsandbytes INT8, batch size one, no sampling, Qwen thinking disabled, and task-specific generation caps of 3–128 tokens held identical across methods within each example. A separate 30-example set uses Mistral-7B-Instruct-v0.3 with deterministic decoding under the same task-specific prompt/evidence construction for its selected examples. For each example, the Hybrid stable-address stage reconstructs keys, selects the known-bank binders, projects and ranks loci, decodes addresses, and verifies block IDs for all retrieved blocks. This timing excludes semantic retrieval, bank construction, reader generation, and scoring. Across 120 examples, the mean, median, and P95 latency are 2.7336, 2.2851, and 5.5020 ms, with s.d. 1.6316 ms and range 2.0235–17.1093 ms. Descriptor RAG and descriptor-index SLA share the same descriptor-selected block IDs. With successful stable dereference, they therefore present identical evidence to the reader and have identical scores. Table D.1 reports both rows explicitly.

Table D.1: Realistic long-context scores.

Benchmark	n	Payload RAG	Descriptor RAG	Descriptor-index SLA	Hybrid SLA
LongBench v2	60	0.2833	0.2333	0.2333	0.2833
LongBench Pro	30	0.1833	0.1167	0.1167	0.1833
∞Bench	30	0.2778	0.0801	0.0801	0.2778

A paired bootstrap stratified within the prespecified 60:30:30 benchmark mixture gives Payload-RAG-minus-descriptor-index SLA difference +0.0911 (95% CI 0.0211–0.1648). Per-benchmark paired differences are +0.0500 for LongBench v2 (95% CI -0.0667–0.1667), +0.0667 for LongBench Pro (0–0.1667), and +0.1977 for ∞Bench (0.0619–0.3563).

The Mistral-7B-Instruct-v0.3 evaluation uses a separate 30-example 10:10:10 benchmark mixture and is reported independently from the primary 120-example Qwen configuration. Its macro score is 0.1667 for Payload RAG and 0.2000 for descriptor-index SLA. Using the same Payload-RAG-minus-descriptor orientation as the main-text Qwen comparison, the fixed-mixture paired point estimate is -0.0333 (95% CI -0.1000–0. Fig. D.1).

D.2 LongMemEval-S evidence retrieval and natural knowledge evolution

The LongMemEval-S evaluation uses all 500 questions in the cleaned release (Wu et al., 2025). Session order, timestamps, turns, and released evidence annotations are preserved. User turns are packed within session boundaries into chronological payload records of at most 512 BGE-M3 tokens, retaining session/turn provenance. The deterministic key namespace is tied to the memory instance and source chronology: memory-instance identifier, chronological session ordinal, official session ID, record/turn provenance, and block ordinal. It does not depend on the evaluation query text. Cached-pointer access and Hybrid SLA dereference the same semantic selection. Chronological SLA retains the four highest-ranked semantic seed records. It then round-robins over seeds with offsets $-1, +1, -2, +2, \ldots$ in session-local chronological record order until the common 4,096-token budget is full. The relation unit is the packed chronological record itself, and the reached record is dereferenced through its own locator. The four-seed operating point and offset schedule were fixed

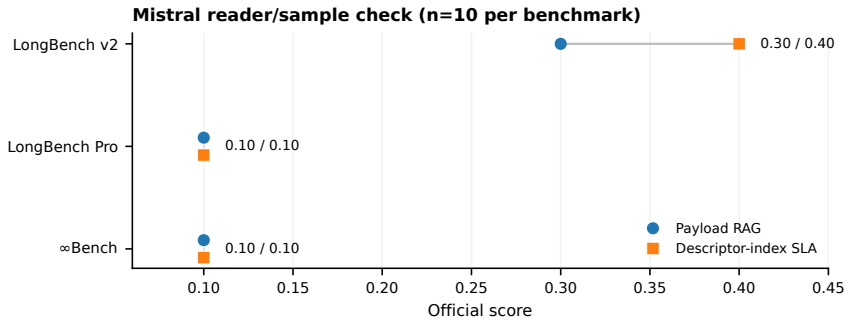


Figure D.1: **Mistral reader/sample evaluation.** Paired benchmark scores on 30 examples (10 per benchmark) show one positive difference and two ties. Macro scores are 0.1667 for Payload RAG and 0.2000 for descriptor-index SLA, Payload RAG minus descriptor-index SLA is -0.0333 (95% CI -0.1000–0).

before benchmark scoring and held constant for all 500 questions. Neither direction nor stopping consults answer annotations. Session recall, turn recall, and all-required recall follow the released annotations. The 30 abstention items have no positive evidence location and are excluded from evidence-recall denominators.

The complete cleaned release contains 30 abstention items without a positive evidence location. Turn recall uses 419 questions with positive user-turn annotations, while session recall and all-required recall use all 470 evidence-bearing questions. Table D.2 reports the corresponding evidence-retrieval and latency results.

Table D.2: LongMemEval-S evidence retrieval. Recall columns are micro averages over their stated evidence-bearing subsets, Local moves and latency are per-question means. Latency includes batch-one query embedding, exact search, and memory lookup.

Method	Session recall	Turn recall	All required	Local moves	Latency (ms)
Payload RAG	0.9624	0.9427	0.9128	0.00	4.582
Cached pointer	0.9624	0.9427	0.9128	0.00	4.581
Hybrid SLA	0.9624	0.9427	0.9128	0.00	4.767
Chronological SLA	0.8429	0.8298	0.7234	12.98	4.708

All methods perform one corpus-level semantic first-access search. Cached-pointer access and Hybrid SLA use the Payload-RAG evidence, and every binder dereference is correct. Chronological SLA retains four semantic seeds and adds alternating chronological neighbors. Relative to Payload RAG, session recall changes by -0.1195 (95% CI -0.1450 to -0.0955), turn recall by -0.1130 (-0.1409 to -0.0848), and all-required recall by -0.1894 (-0.2298 to -0.1511). Chronological expansion keeps the single initial semantic first-access search and adds 0.1268 ms (95% CI 0.1225–0.1316) within this evaluation boundary. Figure D.2 shows the overall and category-level recall changes after semantic blocks are replaced by deterministic chronological expansion.

Benchmark-defined natural knowledge-update construction. The cleaned LongMemEval-S release contains 78 questions labeled as knowledge updates. Its construction associates one `answer_session` with an earlier `session_old` (suffix `_1`) and later `session_new` (suffix `_2`). Released session order defines old before new, and 77/78 cases also have strictly increasing timestamps. The released `answer_session_ids` and turn-level `has_answer` fields provide question-linked evidence provenance, while the released evaluator requires the updated answer. Eligibility is determined before any SLA mutation or reader inference: six abstention questions and two cases without the required old turn-level annotation are excluded, producing a frozen 70-question set. The sole eligible case without a strict timestamp increase remains ordered by the benchmark-defined construction. A prespecified 69-question sensitivity excludes only that case.

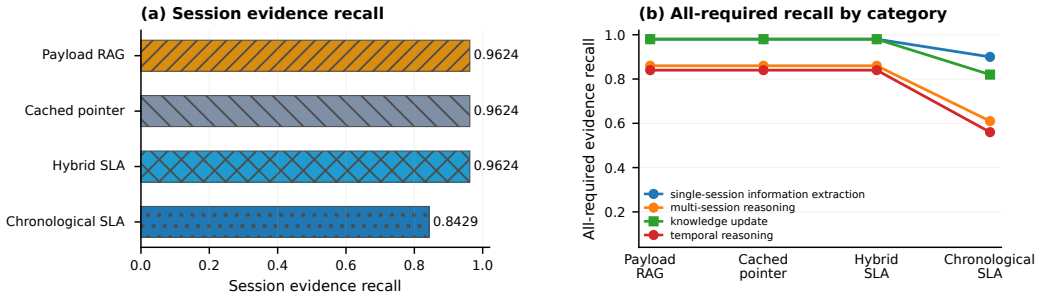


Figure D.2: **Semantic relevance is not implied by chronology.** Left, overall session-level evidence recall across 470 evidence-bearing LongMemEval-S questions. Right, all-required-evidence recall by question category. Cached pointer and Hybrid SLA inherit Payload RAG's evidence set, deterministic chronological expansion replaces semantic blocks and lowers recall.

For each eligible question, one logical handle/key is established on the old natural evidence and retained while the external payload is updated to the released later evidence. The fixed mixed lifecycle then composes payload update with same-bank relocation, unrelated insertion, and rebinding using the existing SLA mutation semantics. Across six scheduled states per question, same-handle/key and current-payload checks are 420/420, explicit-pointer equivalence is 420/420, and wrong-record, fail-closed, and registry-mismatch counts are zero. The three post-update states contribute 210 current-payload checks, all correct, with 0/210 obsolete-payload returns. Unaffected-record preservation is 22,300/22,300. These record/state counts are repeated coverage over 70 independent benchmark questions.

Table D.3: LongMemEval-S natural knowledge-update lifecycle. Eligibility is frozen before execution, payloads are exact released evidence text.

Quantity	Result
Released knowledge-update questions	78
Frozen eligible set	70 (6 abstention + 2 missing old turn-level annotation excluded)
Same handle/key, current payload	420/420, 420/420
Post-update current / obsolete payload	210/210, 0/210 obsolete returns
Pointer equivalence	420/420
Wrong / fail-closed / registry mismatch	0 / 0 / 0
Unaffected-record preservation	22,300/22,300

Paired causal reader intervention. The LongMemEval-S evidence-retrieval study above is retrieval-only. The causal reader arm therefore freezes a separate evidence-only protocol before generation on the same 70 benchmark-defined update cases, with one fixed prompt shared by all three conditions. The primary reader is Qwen3-8B revision `b968826d9c46dd6066d109eabc6255188de91218`, loaded with bitsandbytes INT8, greedy decoding, thinking disabled, batch size one, and a fixed 128-token cap. Each of the 70 frozen questions produces three conditions: UPDATED_SLA, UPDATED_POINTER, and STALE_OLD. The first two have byte-identical evidence, complete serialized prompts, and tokenized inputs in 70/70 cases. Immediate and final post-update inputs are also identical in 70/70, showing that later locator evolution does not change delivered evidence. Updated and stale inputs differ in all 70 cases. Generation completes for all 210 conditions with no inference failures. One output reaches the prespecified token cap and is retained without retry.

Timing audit for the matched reader path. The frozen UPDATED_SLA and UPDATED_POINTER conditions have identical evidence bytes, complete serialized reader inputs, and token IDs in all 70 cases. Their downstream Qwen3-8B computation is therefore common to the two locator conditions. The existing CUDA-synchronized reader intervals average 5,742.889 ms/question

when the paired identical-input runs are combined into a single common reader-cost series. A new CPU-only replay measures the disjoint pre-reader stage—evidence serialization, fixed prompt construction, chat-template rendering, and tokenization—at 0.609223 ms/question over 5 warm-up rounds and 50 timed repetitions per question. The matched locator measurements are 0.000113 ms/question for cached pointer and 0.030898 ms/question for exact SLA Top-1. The resulting composed totals are 5,743.498 and 5,743.529 ms/question, respectively. SLA adds 0.030785 ms/question, or 0.000536% of the pointer total. These totals are composed from compatible measured stages, not directly timed end-to-end wall-clock intervals. The frozen causal-reader protocol starts from preregistered target evidence, so semantic retrieval is not invoked. Zero retrieval time in this accounting means "not part of the protocol," not zero-cost search. No output, score, or accuracy result is changed by this timing audit.

Table D.4: LongMemEval-S matched timing for the frozen 70-question causal-reader path. The totals sum compatible measured stages, they are not a single directly timed full-pipeline interval. Semantic retrieval is not invoked by this evidence-frozen protocol.

Stage	Updated pointer (ms/q)	Updated SLA (ms/q)
Evidence/prompt/tokenization	0.609223	0.609223
Locator	0.000113	0.030898
Common Qwen3-8B reader	5,742.888957	5,742.888957
Composed total	5,743.498293	5,743.529078

Reader outputs are scored by the released LongMemEval answer-checking procedure using `gpt-4o-2024-08-06` at temperature zero. All 210 judgments complete successfully. UPDATED_SLA and UPDATED_POINTER produce identical output text, output token sequences, and released-answer-checker labels in 70/70 cases. Their current-answer accuracy is 56/70 (0.8000), whereas STALE_OLD is 8/70 (0.1143). The paired difference is $+0.6857$ (95% percentile-bootstrap CI 0.5286–0.8286. 10,000 question-level draws, seed 2026091601). The paired table is 55 updated-correct/stale-wrong, 1 both correct, 7 both wrong, and 7 updated-wrong/stale-correct. The exact paired-binomial two-sided p-value is 2.43×10^{-10}. Updated and stale output text differs in 70/70 cases. A transparent non-official normalized answer-containment diagnostic gives 48/70 updated versus 8/70 stale ($+0.5714$, 95% CI 0.4143–0.7286), reproducing the same direction under a deterministic local diagnostic.

The prespecified strict-timestamp sensitivity excludes only the one equal-timestamp case: UPDATED_SLA and UPDATED_POINTER are 55/69 (0.7971), STALE_OLD is 8/69 (0.1159), and the paired difference is $+0.6812$ (95% CI 0.5217–0.8261). Table D.5 summarizes the primary 70-question intervention.

Table D.5: LongMemEval-S causal reader intervention. The same 70 frozen questions are scored against the released current answer, UPDATED_SLA and UPDATED_POINTER are input-identical controls, while STALE_OLD replaces only the benchmark-defined updated state by its predecessor.

Condition / paired outcome	Correct	Rate
UPDATED_SLA	56/70	0.8000
UPDATED_POINTER	56/70	0.8000
STALE_OLD	8/70	0.1143
Updated correct / stale wrong	55	0.7857
Updated wrong / stale correct	7	0.1000

D.3 LongMemEval-V2 multimodal protocol and additional results

Benchmark and common evaluation. LongMemEval-V2 uses the fixed repository and dataset identifiers in Table F.1 over all 451 Small-tier questions, 240 web and 211 enterprise. The five abilities are static-state recall ($n = 134$), dynamic-state tracking ($n = 86$), workflow knowledge ($n = 74$), environment gotchas ($n = 29$), and premise awareness ($n = 128$). All rows share the same Qwen3.5-9B multimodal reader configuration, Qwen3-Embedding-8B retriever service,

dated GPT-5.2 judging configuration, task prompt templates, and 200,000-token context limit. Their complete reader inputs differ whenever their memory policies produce different evidence. The reference methods are the released `rag_query_to_slice`, `rag_query_to_slice_notes` (state-slice retrieval augmented with released trajectory notes), and `agentrunbook_r` modules at that repository revision. Hybrid is a matched dereference substitution for Semantic RAG. The periodic/integer/adjacency comparison is an executor-equivalence test conditioned on one fixed set of realized relation-policy decisions: semantic seeds, policy outputs, candidate order, and budgets are fixed while only execution representation changes.

Semantic RAG supplies ranked semantic seed slices from the question, optional question image, and trajectory memories. Each indexed trajectory slice is a stored record with a fixed indexing-time canonical internal cue and derived key. Neither is generated from the evaluation query. Stored metadata carries the slice identity, storage scope, trajectory identifier, and trajectory-local ordinal used by the relation policy. Hybrid SLA uses the selected seed's retained handle to obtain storage scope and canonical cue material, derives the protocol key, and queries the bank-local binder to reconstruct the current address before dereferencing the multimodal slice. No handle-resolved current address is used. Its complete reader input is therefore identical to Semantic RAG for all 451 questions.

Relation controller and executor variants. The Gated trajectory policy uses the specified Qwen3.5-9B controller on the visible question, optional question screenshot, and a ranked compact preview of the initial semantic seeds. No ability label or answer metadata enters the controller. The system instruction requests one JSON object with boolean field `use_local_relation` plus `direction` $\in \{$`both`, `predecessor`, `successor`$\}$ for predecessor/successor context around retrieved trajectory states. Gate decoding uses temperature 0.6, top-k 20, top-p 0.95, thinking enabled, and an 8,192-token gate limit. The controller is run once per question, yielding 149 active and 302 inactive decisions in the evaluated run. These realized outputs define the relation-policy decisions on which the executor comparison conditions. The controller is not rerun for different executor representations. The complete 451-row trace (question ID, gate state, direction, and post-controller fidelity counts) is retained as evaluation data, allowing periodic, integer-coordinate, and adjacency execution to be replayed against exactly the same observed decisions.

The 75/25 semantic/relation memory split and expansion schedule were fixed before scoring and held constant for all 451 questions. When the gate is active, 75% of the 200,000-token memory-context budget is reserved for highest-ranked semantic seeds and 25% for same-trajectory relation expansion. For both directions, expansion follows $-1, +1, -2, +2, -4, +4, \ldots$. One-direction decisions use the sign-restricted subsequence. Endpoint-crossing candidates are removed before execution, duplicate reached evidence is then removed, and the remaining candidates fill the relation budget in candidate order. Relations act on stored trajectory-local ordinals, while metadata maps each reached ordinal to a target record whose locator is resolved independently. The protocol generates 7,159 seed–offset candidates: 4,776 remain in trajectory bounds and all 4,776 execute correctly, while 2,383 would cross an endpoint. After deduplication and budget admission, 425 relation-derived evidence groups enter reader context, corresponding to 719 accepted primitive-displacement units.

Controller activity. The relation gate activates for 149 questions and remains off for 302. Directions are both for 77, predecessor for 18, and successor for 54. Figure D.3 summarizes released-policy task outcomes and the exact matched-substitution checks. Table D.7 reports the natural lifecycle evaluation, Table D.9 reports released-policy overall accuracy, Table D.6 reports the matched locator/executor substitutions, Table D.10 reports ability-level accuracy, and Table D.11 reports access, relation, token, and latency quantities.

Natural dynamic-memory lifecycle and matched Delta evaluation. The natural dynamic-memory lifecycle starts from the released LongMemEval-V2 records and keeps the semantic selections, realized relation-policy decisions, evidence construction, and reader workload fixed for all 451 questions. It then applies the specified writable SLA mutations. Before mutation, binder dereference matches the stored reference state for the full fixed population of 5,095/5,095 records/slices. Of the 451 questions, 439 contain parsed Hybrid-SLA seed records and are eligible for the lifecycle. Their 2,475 admitted seed-record instances are drawn from that initial population and form the mutation subset. The remaining 12 questions have no parsed Hybrid seed record and record zero

stable-seed accesses, relation attempts, and local relation items. They therefore instantiate no mutable post-retrieval seed state.

Across the 439 eligible questions, those 2,475 natural seed records undergo 24,369 same-bank relocation, insertion, and rebinding operations. The RLS lifecycle and the matched native Delta-rule evaluation use the same identity keys, mutation order, relation-policy decisions, registry/store semantics, and evidence budgets. Across the nine scheduled lifecycle states, the Delta evaluation yields 22,275/22,275 correct Top-1, 22,275/22,275 correct Top-8, and 22,275/22,275 correct payload dereferences with zero incorrect-record returns. The eligible lifecycle subset retains all 4,776 valid relation transitions. Periodic, integer-coordinate, and adjacency executors each remain exact. The same-bank non-target evaluation covers 369,485 natural record–checkpoint states and preserves all 369,485. The native Delta numerical check has maximum target-write residual 9.54×10^{-7} and collateral-identity verification residual 2.38×10^{-7} in FP32.

Evidence identity and serialized reader-input identity remain exact for all 439 eligible questions after the final lifecycle state. Qwen3.5-9B is independently re-executed from the reconstructed final-state evidence for all 439 questions under the same checkpoint, prompt/evidence pipeline, serving configuration, and decoding protocol. All 439 reruns match the corresponding baseline exactly in output text, token sequence, and output hash. Benchmark scores are identical for 439/439 questions. Low-level KV-cache placement and FlashAttention settings are outside the reported comparison boundary. Factual payloads are held fixed in this lifecycle so that it tests relocation, insertion, rebinding, retained access, and relation operations. Natural payload evolution is evaluated separately on LongMemEval-S in Appendix D.2.

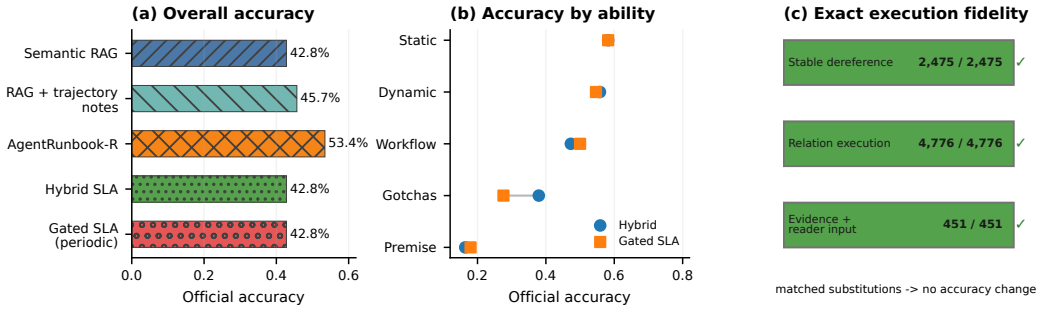


Figure D.3: **LongMemEval-V2 task outcomes and exact execution fidelity.** (a) Full-set accuracy under the released memory policies and matched SLA variants. (b) Ability-level accuracy for Hybrid and the gated policy with periodic execution. (c) A compact fidelity scorecard shows that matched locator/executor substitutions preserve stable dereference, relation transitions, and evidence/reader inputs exactly at the displayed denominators. The separate dynamic-memory lifecycle evaluation tests writable-state evolution and the independent reader re-execution.

Table D.6: Matched LongMemEval-V2 Small substitutions under one fixed set of realized relation-policy decisions. Semantic seeds, realized policy outputs, budgets, and candidate order are held fixed as specified above. The last column reports overall-accuracy difference in points over all 451 questions.

Comparison	Fixed components	Exact fidelity checks	Δ overall
Semantic RAG → Hybrid SLA	semantic seeds, candidate order, evidence budget	stable dereference 2,475/2,475, evidence and reader-input agreement 451/451	0.00
Periodic → integer execution	Hybrid-SLA seeds plus identical realized predecessor/successor decisions	in-trajectory relation execution 4,776/4,776, evidence and reader-input agreement 451/451	0.00
Periodic → adjacency execution	same fixed components	in-trajectory relation execution 4,776/4,776, evidence and reader-input agreement 451/451	0.00

Table D.7: Natural dynamic-memory lifecycle evaluation and matched native Delta-rule evaluation on LongMemEval-V2 Small. The lifecycle unit is the benchmark question, record-at-lifecycle-state counts are exhaustive within-question coverage, not independent samples.

Quantity	Result
Lifecycle eligible / total questions	439 / 451
Mutated natural seed records	2,475
Mutation operations	24,369
Delta retained Top-1 / Top-8 / payload	22,275/22,275 each
Delta wrong-record returns	0
Delta periodic / integer / adjacency relation execution	4,776/4,776 each
Delta non-target preservation	369,485/369,485
Evidence / complete reader-input identity after final lifecycle state	439/439 each
Independent reader output text / token / hash identity after final lifecycle state	439/439 each
Recomputed score identity after final lifecycle state	439/439
Delta target-write / non-target update-equation residuals	9.54×10^{-7} / 2.38×10^{-7}

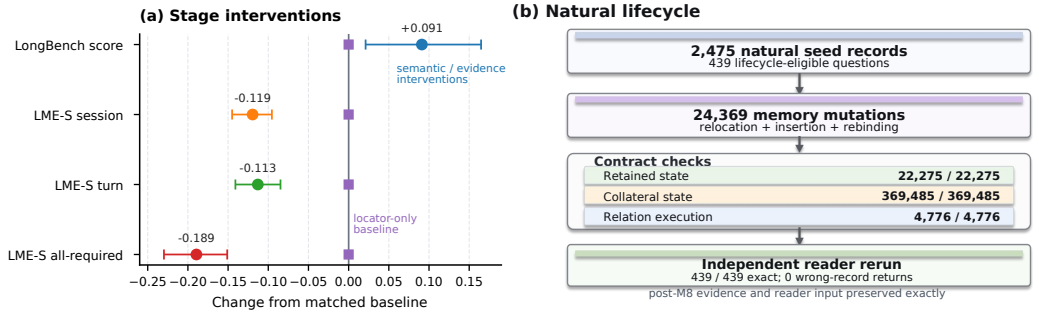


Figure D.4: **RQ3 stage isolation and structural lifecycle.** (a) Semantic/evidence interventions move LongBench or LongMemEval-S outcomes, whereas matched locator-only substitutions remain at zero change. (b) With factual payloads held fixed, 24,369 protocol-controlled operations mutate 2,475 natural LongMemEval-V2 seed records while preserving retained state, non-target state, relation execution, and the independent reader path. Natural content evolution is tested separately by the benchmark-defined LongMemEval-S update intervention in Fig. 2.

Supplementary stage-isolation visualization.

D.4 Pretrained DeltaNet model integration

The model-level experiment integrates a dedicated SLA Delta state with `fla-hub/delta_net-1.3B-100B`, revision `b4dcbbafd4fde802717bdec3008d4aba9cb3a1f8`. The loaded `DeltaNetForCausalLM` has 24 layers, hidden size 2,048, 16 heads, and 1,365,677,056 parameters. Every pretrained parameter is frozen. The SLA adapter introduces zero trainable parameters. The 65,536-element FP32 SLA matrix (262,144 bytes) persists across calls in the same cache lifecycle as the native FLA cache but is semantically distinct from token-sequence recurrent state: protocol-stable identity is maintained by the SLA channel, while learned recurrent representations remain on the model's semantic/inference path. Reads from SLA state determine the verified payload/evidence serialized into the frozen model's input, so this integration tests native-Delta-compatible persistent referential state while pretrained recurrent dynamics remain unchanged.

The base integration trajectory spans 11 revision states covering persistence, relocation, insertion, overwrite, rebinding, and reset. Correct SLA state and a proper mutable pointer resolve identical current evidence and therefore produce identical serialized model inputs and logits in 11/11 states. The three designated base perturbation checks–reset, deterministic handle-to-locus shuffling, and stale state–all change the retrieved model input and logits (3/3 total). Sequential state hashes verify

Table D.8: **RQ3 pretrained DeltaNet integration.** Correct SLA state is compared with a correctly updated explicit pointer, stale, shuffled, and reset SLA states serve as causal controls.

Test	Result
Frozen integration	1,365,677,056 pretrained parameters frozen, SLA adds zero trainable parameters and 65,536 FP32 persistent state elements.
Correct-state equivalence	pointer-identical model inputs and logits in 11/11 base states and 700/700 held-out seven-state comparisons, maximum logit difference 0 under the 10^{-6} criterion.
Deliberately incorrect states	stale pre-relocation, shuffled handle-to-address, and empty/reset states each change payload, input, and logits in 100/100 held-out cases.

that the same persistent adapter state is carried across calls and changes only at the corresponding state transitions.

The scaled validation draws from the frozen LongMemEval-V2 Small (web + enterprise) retrieval cache containing 5,095 natural-language records/slices. Eight records used by the earlier integration are excluded and non-empty payload is required, leaving 5,087 eligible records. Selection is deterministic: records are ordered by SHA256 of a fixed selection namespace concatenated with the record ID, and the first 100 are retained. The 100 identities are therefore fixed before model inference, with no outcome-based exclusion. The frozen ID list has SHA256 `9daebb3f25a6f93bb6d58c5fa38242b86d158d8714cd61b40aa3d664132ae8ec`.

Each case follows seven predefined states: the initial state, after relocation, after one unrelated insertion, after four cumulative unrelated insertions, after payload overwrite, after the first rebind, and after the final rebind. At every state, SLA is compared with a correctly updated explicit pointer. Record identity, payload/evidence identity, complete model input, and logits agree in 100/100 cases at every state, giving 700/700 comparisons for each quantity. Logit equivalence is defined by the maximum elementwise absolute difference over the final-token logit vector, with threshold 10^{-6} and no relative tolerance. The maximum observed difference is 0. After the final rebind, three incorrect states are tested once per case: a stale pre-relocation state, a shuffled handle-to-address mapping, and an empty/reset state. Each changes the retrieved payload, complete model input, and logits in 100/100 cases. The pretrained checkpoint, wrapper, identity-key construction, and mutation procedure are held fixed throughout. The integration therefore tests the complete path from persistent referential state through evidence construction to pretrained-model computation. All pretrained weights remain frozen, while protocol-stable identity keys define the persistent identity channel and learned language representations remain on the semantic/inference path.

Table D.9: Released-policy LongMemEval-V2 Small reference accuracies at the evaluated repository and dataset revisions. Accuracy is in %, Correct count is over all 451 questions. Each row retains its released memory policy, matched locator/executor substitutions are reported separately in Table D.6.

Method	Correct	Overall	Non-abst.	Web	Enterprise
No retrieval	17/451	3.77	4.64	2.50	5.21
Semantic RAG	193/451	42.79	53.25	41.25	44.55
RAG + trajectory notes	206/451	45.68	59.13	46.25	45.02
AgentRunbook-R	241/451	53.44	67.49	54.17	52.61
Hybrid SLA	193/451	42.79	53.25	41.25	44.55
Gated SLA + periodic exec.	193/451	42.79	52.63	41.25	44.55

A stable access consists of the deterministic cue key, bank-local binder query, and bank-slot codebook decode under locus-code cosine for one record. The 0.664-ms value is the mean per-question total for the complete seed stable-address stage, which may contain multiple seed accesses. It is not a per-access latency. Likewise, 8.378 ms is the mean per-question total for relation candidate construction through production of admitted relation-derived evidence. A candidate relation transition is one in-bound predecessor or successor offset evaluated before deduplication and budget admission. The 425 relation-derived items are the distinct evidence groups admitted after those filters, whereas the 719 primitive-displacement units sum the absolute displacement of admitted relation items. These quantities and timing rows therefore refer to disjoint stages of the executed access pipeline.

Table D.10: LongMemEval-V2 accuracy by memory ability. Ability denominators are shown in the header. The three Gated SLA rows use the same SLA locator and realized relation-policy decisions, only the relation executor changes.

Method	Static (134)	Dynamic (86)	Workflow (74)	Gotchas (29)	Premise (128)
No retrieval	0.0522	0.0233	0.0270	0.1379	0.0156
Semantic RAG	0.5821	0.5581	0.4730	0.3793	0.1641
RAG + trajectory notes	0.6194	0.5930	0.6216	0.3793	0.1172
AgentRunbook-R	0.7761	0.7209	0.6081	0.2414	0.1797
Hybrid SLA	0.5821	0.5581	0.4730	0.3793	0.1641
Gated SLA + periodic exec.	0.5821	0.5465	0.5000	0.2759	0.1797
Gated SLA + integer exec.	0.5821	0.5465	0.5000	0.2759	0.1797
Gated SLA + adjacency exec.	0.5821	0.5465	0.5000	0.2759	0.1797

Table D.11: Post-controller access and relation metrics. Timing rows are mean per-question stage totals over the 451-question evaluation, latency excludes relation-controller inference, reader generation, and judging.

Quantity	Value
Stable accesses, Top-1 / Top-8	2,475/2,475, 1.000/1.000
In-bound candidate relation transitions correct	4,776/4,776 (1.000)
Final relation-derived evidence	425 (0.9424/question)
Accepted primitive-displacement units	719 (1.5942/question)
Mean corpus-level semantic first-access searches	1.000/question
Mean memory tokens, Hybrid / Gated SLA	92,011.1 / 101,541.9
Semantic localization / question	98.899 ms
Seed stable-address stage / question	0.664 ms
Relation-stage construction / question	8.378 ms
Post-controller subtotal / question, Hybrid / Gated SLA	99.564 / 107.941 ms[†]

[†]Subtotals are computed from unrounded stage timings. Displayed rounded components can differ by 0.001 ms.

For the exact 439-question subset used by the independent reader rerun, recomputing statistics over the persisted per-question timing rows gives 98.341858 ms/question for semantic localization, 0.682399 ms/question for stable addressing, and 8.606919 ms/question for relation operations/evidence-group construction. The measured controlled memory-query subtotal is therefore 107.631176 ms/question. These stage timings use the same frozen question set and memory-query pipeline as the reader re-execution.

Under the realized policy decisions, Gated SLA + periodic execution and Hybrid have the same overall accuracy, with difference 0.0000 (95% CI -0.0200 to 0.0200). Ability-level changes are 0.00 points for static-state recall, -1.16 for dynamic-state tracking, +2.70 for workflow knowledge, -10.34 for environment gotchas, and +1.56 for premise awareness. Thus category gains and losses redistribute while the overall score is unchanged. The workflow change is 47.30% to 50.00% (95% CI for the paired difference 0.00–6.76). This pair changes relation expansion while keeping the semantic seeds and reader configuration fixed. Executor fidelity is isolated by holding the same realized decisions, candidate order, admission rules, and reader configuration fixed across periodic, integer-coordinate, and adjacency execution. Those three variants produce identical admitted evidence and reader inputs. The released reference-policy rows in Table D.9 instantiate broader memory policies.

E Systems and Direct-Access Measurements

The deterministic route study uses identical wall-clock timing boundaries for all compared methods. At 16 hops, mean route-evaluation latency is 1.169 s for repeated cue search and 1.043 s for Full SLA, a paired difference of -0.126 s (95% CI -0.128 to -0.124). Retrieved payload tokens and the reader configuration are held fixed across the comparison, so the timing difference reflects alternative access paths under a common protocol. Figure E.1 summarizes the storage, isolated lookup, and post-selection reuse measurements.

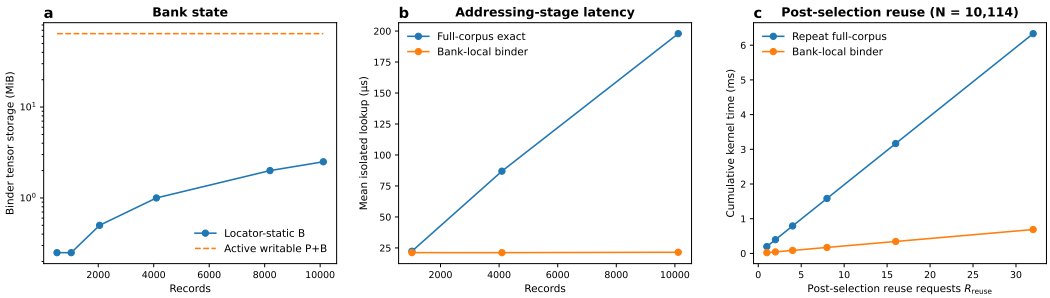


Figure E.1: **Isolated repeated-vector kernels under retained bank scope.** (a) Locator-static binder storage, writable line: one active $P + B$ bank. (b) Synthetic full-corpus exact-vector versus bank-local binder latency. The flat bank-local curve follows the fixed 1,024-candidate routed-bank width, it characterizes retained-scope access rather than an unpartitioned binder-algebra speedup. (c) Conditional on first access having retained identity, key, and scope, R_{reuse} future requests cost $R_{\text{reuse}}T_g$ with repeated full-corpus lookup or $R_{\text{reuse}}T_b$ with the bank-local binder.

The scale benchmark uses $N \in \{512, 1024, 2048, 4096, 8192, 10114\}$. The largest N comes from the untruncated 5,178,181-token LongBench-v2 context encountered in the fixed realistic-evaluation workload and is represented as 10,114 blocks of at most 512 tokens (10,113 full blocks plus one 325-token remainder). Because scope routing caps each bank at 1,024 records, increasing N increases bank count while the retained binder ranks the fixed codebook of one routed bank. GPU timings use an RTX 4060 Ti, explicit synchronization, 30 warm-up iterations, and 1,000 measured repetitions. The full-corpus comparator is an isolated resident synthetic 1,024-dimensional exact-inner-product matrix/query kernel, whereas the retained kernel computes $B_s k$ inside one routed bank. These kernels quantify repeated vector access. Table E.2 separately gives the matched Hygon address-recovery endpoints. The 4,096-dimensional key for the reused identity is resident before the timed binder region. Key construction is excluded, and one materialized FP32 key occupies 16 KiB. Conditioning on the common first access, let R_{reuse} be the number of later requests to that identity. Repeating the full-corpus kernel costs $T_{\text{repeat}} = R_{\text{reuse}}T_g$, whereas retained-scope access costs $T_{\text{SLA}} = R_{\text{reuse}}T_b$. At $N = 10{,}114$, the measured means $T_g = 197.903\,\mu$s and $T_b = 21.499\,\mu$s give 3.166 versus 0.344 ms for $R_{\text{reuse}} = 16$ and 6.333 versus 0.688 ms for $R_{\text{reuse}} = 32$. The per-reuse kernel ratio is $9.21\times$ at this tensor shape. These totals are arithmetic accumulations of the measured per-kernel means, not separately timed multi-request sequences. Table E.1 reports the GPU distributions. Reader prefill, generation, semantic encoding, key reconstruction, and the common first-access retrieval are outside these kernel timings.

Table E.1: Isolated GPU lookup distributions. Times are microseconds over 1,000 repetitions.

N	Method	Mean	Median	P95	s.d.
1,024	Full-corpus exact	22.324	22.239	22.595	0.501
1,024	SLA binder	21.206	21.078	21.719	0.670
4,096	Full-corpus exact	86.937	86.770	87.213	1.059
4,096	SLA binder	21.177	21.062	21.845	0.454
10,114	Full-corpus exact	197.903	197.383	199.087	2.510
10,114	SLA binder	21.499	21.103	22.969	2.135

At 1,024 slots the balanced scan averages 512.476 comparisons. Its mean latency across $n_s = 128, 256, 512, 768, 1024$ is 2.7017, 5.0019, 9.9820, 14.7896, and 20.1514 μs. The measured means are near-linear over the tested sizes (about 0.0194 μs per slot). The exact scan is faster than SLA at the evaluated bank sizes while using no reverse locator. Together, these measurements characterize three different state organizations: pointer/dictionary stores the current locator explicitly, scan reconstructs it from forward state, and SLA stores the mutable binding in shared associative state.

The exact SLA implementation in Table E.2 makes the same Top-1 decision as the unchanged Hygon path on 7,000/7,000 controlled states and 100,000/100,000 additional deterministic queries. It uses the same payload/failure checks and the same 262,144-byte binder, with no additional persistent

Table E.2: Matched Hygon C86 CPU address-recovery timings at a 1,024-slot bank. Query identity/key and resident state are available before timing, payload dereference is excluded. Registry scan uses the production forward representation $\mathcal{D}_s(a)$ with no reverse locator and position-balanced targets. Exact SLA uses the semantics-preserving resident implementation after exact Top-1 equivalence validation. Times are microseconds.

Operation	Repetitions	Mean	Median	P95	s.d.
Cached pointer	100,000	0.1131	0.1100	0.1300	0.0762
Hash dictionary	100,000	0.1426	0.1300	0.1800	0.1153
Registry scan, balanced	100,000	20.1514	20.2190	37.6500	11.6444
Exact SLA Top-1	10,000	30.8980	28.6900	44.0510	6.3764

per-record locator bytes. Reusing resident normalized codes and preallocated shared scratch reduces mean Top-1 latency from 46.4335 to 30.8980 μs (1.50280×). On the same Hygon host, the original production path takes 542.8389 μs because it rebuilds the codebook for each query. The resident-codebook path takes 46.4335 μs, and the validated optimized path takes 30.8980 μs. These correspond to 11.69× and 17.57× reductions relative to production. These measurements concern operational Top-1. Historical ordered Top-8 behavior is not part of this optimization comparison.

For provenance, the original Intel Core i7-14700KF microbenchmark reported 0.0531 μs for the hash dictionary, 0.0427 μs for the cached pointer, 0.0480 μs for the legacy integer transition, and 9.0186 μs for SLA binder+ranking. These cross-host measurements are kept separate from the matched Hygon comparison. On Hygon, the legacy integer transition takes 0.1755 μs, but it includes payload access and therefore does not share the handle-to-address timing boundary used in Table E.2.

Table E.3: Persistent locator-specific mutable state under the same accounting boundary at $d_k = 4096, d_z = 16$. Common registry/store state, immutable cue/key material, semantic index, model parameters, native token cache, and transient kernels are excluded from both rows and reported separately.

Realization	B (MiB)	P (MiB)	Total (MiB)	Footprint reduction vs. RLS
Writable RLS-SLA	0.25	64.00	64.25	1.0×
Writable Delta-SLA	0.25	0	0.25	257.0×

Table E.4: Fair mutable locator/control-state accounting at 1,024 occupied slots. The logical column reports storage for the tested control representation separately from the concrete CPython 3.11.16 representation. Payload content and optional key caches are excluded from all rows.

Method	Logical bytes	CPython fixture bytes	Realization-specific state
Registry scan	33,152	277,304	none
Cached pointer	41,344	285,496	8,192-B reverse locator
Hash dictionary	41,344[†]	332,008	54,704-B measured incremental reverse map
Delta-SLA	295,296	539,448	262,144-B B_s
RLS-SLA	67,404,160	67,648,312	67,371,008-B B_s+P

For the tested two-bank control representation, the logical common-state accounting is 33,152 B. It includes one 256-bit occupant identity per slot, occupancy and liveness bits, and one routing bit per handle. Payload content is excluded. The concrete CPython representation is 277,304 B, of which 181,336 B is the forward registry $\mathcal{D}_s$. The remaining 244,152 B above the 33,152-B logical accounting reflects Python objects and containers that are shared by the scan, pointer, and SLA rows.

The hash-dictionary row adds an audited 54,704-B incremental `dict[str,int]` representation while reusing the handle strings already charged to common state. Its 41,344-B logical entry is the same common accounting plus an 8,192-B address payload. It is not a Python-object lower bound. A materialized 4,096-D FP32 key cache would add 16 MiB per 1,024 records to either SLA row and is reported separately. For B routed banks, the minimal routing field grows from one bit per

handle to $\lceil \log_2 B \rceil$ bits per handle. The $257\times$ value in Table E.3 compares writable locator tensors only, not total memory. These baseline totals also exclude the per-record certificate tracking and rebuild workspace used by the maintenance experiment. †The hash-dictionary logical column charges only the address payload. The concrete CPython total uses the measured incremental reverse-map footprint.

Table E.3 gives the matched persistent locator-state accounting. An optional materialized 4,096-D FP32 key cache costs 16 KiB per record (16 MiB for 1,024 records) and is common to both realizations. On-the-fly key reconstruction is likewise outside the resident-key timing boundary in Table E.2. On active RLS-writable banks, mean insertion, SLA-address relocation, overwrite, and repeated-rebinding times are 0.694 ms, 0.295 ms, 0.103 μs, and 0.294 ms, respectively. These address-update timing rows correspond specifically to continuously RLS-writable banks. Delta fidelity and persistent-state footprint are reported separately in the realization-transfer and state-accounting results. Locator-static RLS banks release P after their address-update phase. Table E.5 reports build and storage scaling, and Figure 3 visualizes the matched Hygon locator design points.

Table E.5: Record-count and locator-static binder-component scaling under fixed-size banking. Each bank contains at most 1,024 records, bank-local Top-1, Top-8, and dereference are 1.000 at every corpus size shown. The state column reports the fixed B_s component only, Appendix B separately states variable-length cue requirements, optional key caches, writable P, and the common scoped-registry substrate.

Records	Banks	Build (s)	Ingestion (ms/binding)	Binder B (MiB)	Peak construction VRAM (MiB)
512	1	0.572	1.117	0.25	136.28
1,024	1	0.785	0.766	0.25	161.51
2,048	2	1.562	0.763	0.50	172.88
4,096	4	3.147	0.768	1.00	205.38
8,192	8	6.429	0.785	2.00	270.38
10,114	10	7.773	0.769	2.50	300.91

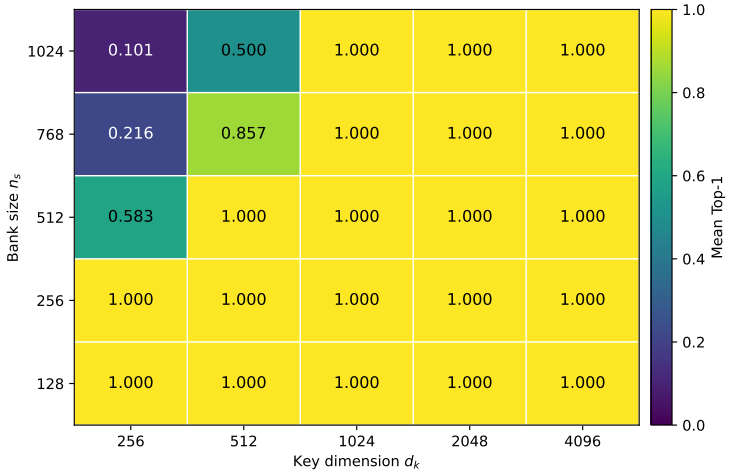


Figure E.2: **Content-independent binder recovery depends on key dimension relative to bank load.** Mean Top-1 over five deterministic Gaussian-key seeds for the float64 ridge-objective sweep. The primary FP32 operating point is $d_k = 4096$ with at most 1,024 records per bank.

Binder load–dimension sweep. A separate five-seed numerical sweep evaluates the ridge objective with content-independent Gaussian keys over bank sizes $n_s \in \{128, 256, 512, 768, 1024\}$ and key dimensions $d_k \in \{256, 512, 1024, 2048, 4096\}$. Float64 linear algebra isolates the load/rank effect from the FP32 operating-point studies. Mean Top-1 is 1.000 in every tested Gaussian-key cell with $d_k \geq n_s$. When $n_s > d_k$, recovery degrades progressively: at $d_k = 256$, mean Top-1 is 0.583, 0.216,

and 0.101 for $n_s = 512, 768, 1024$. At $d_k = 512$, it is 0.857 at $n_s = 768$ and 0.500 at $n_s = 1024$. This is a load–dimension result for the tested low-coherence key family, not a universal rule for arbitrary 1024- or 4096-dimensional keys. The learned-key stress in Appendix B.4 shows that high coherence can reduce the safe regime even when $d_k \geq n_s$.

The matched direct-operation comparison is shown in main Fig. 3. Table E.2 retains the full Hygon distributions, registry scaling, exact Top-1 optimization analysis, and historical Intel provenance.

F Statistical and Reproducibility Details

F.1 Statistical units and paired intervals

Paired bootstraps resample the natural independent test unit: target family for controlled payload variation, route for deterministic traversal, one-dimensional operator composition, two-axis composition, and $D_4 \times \mathcal{O}_{128}$ mixed/challenge routes, and benchmark question for natural evaluations. Each draw retains all paired method outcomes for that unit. Within each controlled non-target preservation evaluation, the independent unit is the revision trial ($n = 500$). The primary and independent preservation evaluations are reported separately rather than pooled. Record/checkpoint and mutation-prefix measurements are repeated within their parent trial, protected segment, or benchmark question and are not treated as independent samples.

The preregistered continuous verified-locator study uses the trial as its independent unit: 7,500 target and 3,920,000 non-target record–checkpoint checks per backend are exhaustive repeated observations over 500 continuously evolved states, including 87,500 inserted-record observations. The sequence-level audit evaluates every protected prefix available in the two frozen-protocol replays: 6,519,500 continuous prefixes and 1,092,979 LongMemEval-V2 prefixes. These counts provide deterministic certificate coverage and are not used as independent inferential samples. The LongMemEval-V2 lifecycle uses the benchmark question as its independent unit: 439 questions contain parsed Hybrid seed state and 12 contain none. Its 22,275 retained-state checks and 369,485 collateral record–checkpoint checks are exhaustive within-question coverage. Evidence and serialized reader-input identity are exact counts, and final-state Qwen3.5-9B outputs/scores are independently re-executed on all 439 eligible questions. The verified-dereference fault audit reports exact deterministic counts over 500 frozen identities per fault class. The learned-key stress contains 75 frozen key-source/seed/load cells and 1,024,725 protected prefixes. Those prefix checks are deterministic coverage, not independent samples. The cross-bank high-churn study is a single fixed serialized stress trajectory (seed 2026092201). Its operation/checkpoint counts and certificate residuals are likewise deterministic coverage rather than independent statistical samples.

The LongMemEval-S natural-update experiment freezes 70 eligible knowledge-update questions before execution. Its 420 mechanism states, 210 post-update states, and 22,300 unaffected-record checks are repeated coverage within those questions. The causal reader experiment uses the same 70 questions as paired units. UPDATED_SLA and STALE_OLD correctness differences are bootstrapped over questions with 10,000 percentile draws using seed 2026091601. An exact paired-binomial test is reported on discordant correctness pairs. The 69-question strict-timestamp analysis is prespecified and excludes only the sole equal-timestamp case. The realistic first-access interval remains a stratified paired bootstrap within the prespecified 60:30:30 LongBench v2 / LongBench Pro / ∞Bench mixture. An unstratified 120-row bootstrap gives the same point estimate and a similar interval. Unless stated as exact deterministic coverage, reported intervals are percentile 95% intervals with 10,000 draws. Deterministic evidence equivalence, pointer equivalence, integer-route equivalence, complete reader-input identity, and exact sequence certificates are reported as exact counts.

F.2 Model, data, sampling, and runtime details

Controlled reader experiments use Qwen3-8B with a 4,096-token evidence budget, 512-token blocks, and deterministic decoding. The LongMemEval-S causal reader uses the same frozen Qwen3-8B revision under its separately frozen evidence-only prompt, INT8 greedy decoding, and the released LongMemEval answer-checking procedure with `gpt-4o-2024-08-06`. Realistic tasks otherwise use their released scorers. Local controlled experiments, the continuous verified-locator stress, learned-key stress, systems measurements, LongMemEval-V2 reader/retriever inference, dynamic-memory lifecycle evaluation, sequence replays, natural-update lifecycle, causal reader, and DeltaNet

transfer/integration experiments ran across their recorded frozen environments. The learned-key stress uses the frozen BGE-M3 and E5-large-v2 revisions in Table F.1, raw L_2-normalized cue embeddings, and no learned or fitted post-processing transform. LongMemEval-V2 answer judging uses the dated GPT-5.2 service configuration in Table F.1. Its lifecycle holds fixed the 451-question semantic selections, relation-policy decisions, and evidence-construction protocol while applying the post-selection writable-state intervention described in Appendix D.3. Final-state reader behavior is verified by an independent Qwen3.5-9B re-execution on all eligible questions with exact output-text/token/hash and recomputed-score identity. Table F.1 records benchmark/model revisions, selected external-data hashes, and central execution identifiers. Artifact-manifest hashes identify the natural-update, sequence-audit, and causal-reader result bundles. The supplementary checksum file separately verifies the packaged figure/data files.

Table F.1: Model, dataset, sample, and evaluation identifiers.

Object	Version / identifier
Qwen/Qwen3-8B	`b968826d9c46dd6066d109eabc6255188de91218`
Mistral-7B-Instruct-v0.3	`c170c708c41dac9275d15a8fff4eca08d52bab71`
BAAI/bge-m3	`5617a9f61b028005a4858fdac845db406aefb181`
intfloat/e5-large-v2 verifier	`f169b11e22de13617baa190a028a32f3493550b6`
LongBench v2	`2b48e494f2c7a2f0af81aae178e05c7e1dde0fe9`
LongBench Pro	`4996884deae51f5e5d23c88da9d857fc54e5fa15`
∞Bench	`90f0394333616266d9fe85824ceaf505093cbaa5`
LongMemEval-S	`98d7416c24c778c2fee6e6f3006e7a073259d48f`
LongMemEval-S causal reader	Qwen/Qwen3-8B revision `b968826d9c46dd6066d109eabc6255188de91218`, bitsandbytes INT8, greedy decoding, thinking disabled, 128-token cap
LongMemEval-S causal evaluator	released LongMemEval answer checker with `gpt-4o-2024-08-06`, temperature 0
LongMemEval-V2 repository	`2cc8c540bdb87fe6761629b585e727e1c4704520`
LongMemEval-V2 dataset	`f152293e235517d504809563c833d7190b8c713b`
LongMemEval-V2 reader	Qwen/Qwen3.5-9B, quantized Q4_K_M GGUF serving files plus the F16 multimodal projector `mmproj-F16.gguf`
LongMemEval-V2 relation policy	Qwen3.5-9B boolean gate plus predecessor/successor/both direction, fixed 75/25 semantic/relation budget and offset schedule in Appendix D.3
LongMemEval-V2 embeddings	Qwen/Qwen3-Embedding-8B, quantized Q4_K_M GGUF serving files, last-token pooling
LongMemEval-V2 judge	`gpt-5.2-2025-12-11`, medium reasoning
Pretrained DeltaNet integration	`fla-hub/delta_net-1.3B-100B`, revision `b4dcbbafd4fde802717bdec3008d4aba9cb3a1f8`
DeltaNet integration state	1,365,677,056 frozen pretrained parameters, zero new trainable parameters, 65,536 FP32 SLA-state elements
DeltaNet held-out scale-up	LongMemEval-V2 Small (web + enterprise), 5,095 frozen source records/slices, 5,087 eligible after excluding the eight records used by the earlier integration, deterministic first-100 selection by SHA256(namespace|record ID), no outcome-based exclusion
DeltaNet scale-up identifiers	source-cache SHA256 `0830291d93c1ec6c9f7c12bf5fe2544060e18a75b9b31bb5a23d0ecfd6ffc7fe`, selected-ID-list SHA256 `9daebb3f25a6f93bb6d58c5fa38242b86d158d8714cd61b40aa3d664132ae8ec`, mutation seed 20260910
DeltaNet scale-up logit criterion	maximum elementwise absolute difference on the final-token logit vector $\leq 10^{-6}$, 700/700 comparisons satisfy the criterion, maximum observed difference 0
Delta backend runtime	Python 3.11.16, PyTorch 2.7.1+cu128, NVIDIA RTX 4500 Ada, FLA native `delta_rule_recurrence`
Controlled Delta-overlap stress	190 preregistered conditions, Experiment-B manifest SHA256 `fa1d040b7df832700ce333aad67de9ded6ade56ccb54b7d062c9b21bbdf086d1`, native-FLA source SHA256 `20ae032f4e08019b4298cc9fcf2c397d871faac0fd1f180243d9f699a9192968`
Learned/content-derived key stress	75 frozen key-source/seed/load cells, BGE-M3/E5 raw normalized cue embeddings, results SHA256 `b40d534391d19bf32414b4c205b584146d39c6d87e5eb9a95f176d538da8e082`, summary SHA256 `59588533ba6b3bc910b3b9ca5c50b8976d9cea4b45f894c36b0b8b87d7a6de34`, runner SHA256 `7559e221a079e2145f75d2732244988bba44fd3e857b2019994023040e543571`
Verified-dereference boundary audit	500 deterministic identities/fault, results SHA256 `8744fb2546ecf15131a423add34664c256046a0a69d5185a59fb18e0a3e0d6a9`, summary SHA256 `a169a273a52c9b1c89f1202f9420d537680c51bc22e6a260b3493b21a7a71371`

Object	Version / identifier
Cross-bank high-churn stress	seed 2026092201, 20,000 serialized operations, results SHA256 e46843dba855d529b3ce7bc0038877b335a7c7caf06a590489b0722274982d73, summary SHA256 17eed542904bebb393f50689830d6a5efc6c466900f85156215b6c4c194ebb43
Certificate-triggered cross-bank maintenance	same seed and 20,000-operation trace, results SHA256 1fbdcf92116d47cc5b6088f053820e11f0449722f5d568b0b0c8870e56146fd2, summary SHA256 f71de52959bf1be8d6eadbe8936799b4481150dbd10c8ae4d22003dcfa1f996b
Matched Hygon direct-access benchmark	Hygon C86 3350, Python 3.11.16, PyTorch 2.7.1+cu128, NumPy 2.4.6, Experiment-A manifest SHA256 b485b5b079791943b9bd6c3cef18fe6948442fb63792fae9211553759dcf6a88
Exact Top-1 implementation source	SHA256 3502a352296a44f80e571a0b6b2b103e08699f8fe8bb5b954f4ee6a161838242, historical Table-E.2 result SHA256 f58b2a38739eaa81077144094e9757fb9a177306f8782dc37ddd6471b33d98ab
Realistic primary sample SHA256	22e4cb884e2d0654b7cbb2fac9e1ad2daecb0a1bd3fa4dcc6276d966ade39180
Mistral configuration sample SHA256	01db8608e66ebf85bd56e5677020fa7faa97cc3b9ca934ca59b55b3efd87d64f
LongMemEval-S data SHA256	d6f21ea9d60a0d56f34a05b609c79c88a451d2ae03597821ea3d5a9678c3a442
LongMemEval-S natural-update artifact manifest	SHA256 fe4f9415f34df099b5c8a59db913356f24c31bb03ec2a55f3b0cec19d6e689e9
Cumulative-interference artifact manifest	SHA256 f7161bc9f183cbe8fea54e2d76607f77a08d18e3597fbc42e5522dc8f7c71b4a
LongMemEval-S causal-reader artifact manifest	SHA256 e61602068fb269c81d58139284acb7f1a655dc6cb7c5416a8c2c26cedbea3265
LongMemEval-S timing audit	70 frozen questions, composed compatible-stage accounting, results SHA256 d78d4847fa592e565651049f9d27e85418d66094b76f792ea733e0c2f28ce61b, summary SHA256 357afe4fcc22aa80aad7ecad7c2eca049177856957e2595077c128d1412b8a90
Anonymous review-artifact checksum manifest	SHA256 bbe5c57407fd928d6dffdf4aa003a400861e4d287f2f0cd8b30e5effcab7bf6f
V2 question-ID SHA256	28de62f7ce1ce26c59fcec26958a52e00ba084a4aaa83bdb07007c740e50a96d
V2 policy-trace SHA256	9d63085dd310d343d108718951c5ae6c06ace2d62a075faa3bd800ca9dd50c67
Dynamic-lifecycle aggregate summary SHA256	9f888b714cc4b385eeffe007169bb0f27bfb862083664c111a73fd7bd93465f9
Dynamic-lifecycle ineligible-question table SHA256	98aa8744a9d1f05afd0d03e05a0226ebea634acdedbca0f527030bd5b6e86898
Dynamic-lifecycle relation-fidelity table SHA256	2cd5e71c7bc8e8d5905a5a312723420985a458ed317ab06a96258ff36061d836
Controlled identity source revision	b140f6db6d384240f54a55c32cb27ae1250e0b5d
Two-axis protocol identifier	51bfa7f027be5c493e74462217049fc00d45cb557745d697aec93aab7b691490
Two-axis route/outcome identifier	788cf4c5f014fdcf5a79c0dd1a865e19db5e0d1657a0a16e4c63af4f2c7df860
Supplementary checksum-file SHA256	2f9f016a03e749e5bea402ba9dfad227b64d2ee4bf76a2e896c2937aadd53ad0

Execution, timing, and sampling details. Most controlled and systems runs used an i7-14700KF host with 70.50 GiB RAM and an RTX 4060 Ti (16 GiB), running Ubuntu 22.04.5, Python 3.11.7, PyTorch 2.6.0+cu124, CUDA 12.4, and NumPy 1.26.3. The original LongMemEval-V2 evaluation used Python 3.11.13, PyTorch 2.7.1+cu126, and CUDA 12.6. The matched direct-access/Top-1 optimization study used a Hygon C86 3350 host (Linux 7.0.0-30), Python 3.11.16, PyTorch 2.7.1+cu128, NumPy 2.4.6, and one PyTorch computation thread. The Delta backend, overlap stress, sequence replays, and pretrained-model integration used Python 3.11.16, PyTorch 2.7.1+cu128, and an NVIDIA RTX 4500 Ada GPU. Delta experiments use FLA's native `delta_rule_recurrence`. The LongMemEval-S causal reader uses the Qwen3-8B revision in Table F.1, bitsandbytes LLM.int8, greedy decoding, and no thinking. The released answer checker uses `gpt-4o-2024-08-06`.

LongMemEval-V2 reader/controller inference uses the quantized `Qwen3.5-9B-Q4_K_M` GGUF serving files plus the F16 multimodal projector `mmproj-F16.gguf`. Embeddings use the corresponding quantized `Qwen3-Embedding-8B-Q4_K_M` serving files with last-token pooling. The independent final-state reader re-execution uses the same Qwen3.5-9B checkpoint, eight CPU service threads, and three concurrent context slots of 262,144 tokens each. Low-level KV-cache placement and FlashAttention settings are outside the reported comparison boundary. The original V2 serving stack used llama.cpp build 10472 (commit `60eeeb608`. Binary SHA256 `455dd4eda910d9aa786262a1a7fd991d3fde0d5215a6fee5feb290c48c06af0d`). *Timing:* GPU lookup uses FP32, 30 warm-ups, and 1,000 synchronized repetitions. The matched Hygon address-recovery study uses 1,000 warm-ups, 100,000 pointer/dictionary/position-balanced-registry queries, and 10,000 exact-SLA Top-1 queries from resident query/state inputs. The historical Intel microbenchmark is retained only as provenance. *Sampling:* first-access seed is 20260817. Bootstrap seeds are 20260812 (controlled payload), 20260815 (traversal), 20260818 (1D operator), 20260817 (LongMemEval-S/Hybrid), 20260826 (LongMemEval-V2), 2026090303 (non-commutative challenge routes), and 2026091601 (LongMemEval-S causal reader). Protocol seed 20260818 uses 1D fit/dev/test addresses 0–639/640–767/768–1,023. FHRR seeds are 2026090501–05, binder-load 2026090511–15, and nested multiplicity 2026090521.

G Runtime State, Numerical Publication, and Maintenance Controls

Runtime/observer separation. An independent reference replay separates operational state from evaluation truth. Runtime state contains immutable handle-to-cue/key metadata, handle-to-bank routing, bank-local forward registry and payload storage, and the shared binder. It retains no handle-indexed current bank-local address or baseline projection. Reads compute $B_s k_r$, rank the bank codebook, and verify the occupant. They never scan the registry or consult observer truth. Control-plane mutation code may scan the authoritative bank's forward registry to recover the source slot. Expected addresses and payloads are retained only by an isolated observer used for assertions. Replaying the frozen 20,000-operation trace reproduces the unmaintained checkpoint outcomes and the 1,744 norm-triggered rebuild locations without a read-side scan fallback.

Exact arithmetic versus stored floating-point state. Equations (10) and (12) are exact-arithmetic identities. For a stored floating-point candidate, the audit therefore validates the candidate itself before publication instead of extrapolating a small measured residual into a universal future bound. Writing $\widehat{s}_a$ for a computed score and s_a for its exact-real counterpart, a standard finite-normal floating-point model gives a score-error envelope of the form

$$\begin{aligned} |\widehat{s}_a - s_a| &\le c(u)\|\widehat{B}\|_F\|k\|_2\|z(a)\|_2, \\ c(u) &= \gamma_{d_k}(u) + \gamma_{d_z}(u) + \gamma_{d_k}(u)\gamma_{d_z}(u), \qquad \gamma_n(u) = \frac{nu}{1-nu}. \end{aligned} \tag{29}$$

The reference gate computes intended-versus-competitor gaps from the stored candidate in FP64 and accepts only when the lower-bounded gap remains strictly positive after accounting for FP32 readout, FP64 score evaluation, and gap subtraction. Non-finite or uncertifiable candidates are rebuilt and rechecked or are not published. This is a committed-state guarantee under the stated arithmetic and trusted-control assumptions, not a guarantee that arbitrary loads or key geometries can always be accepted. Reduced-mantissa alternatives such as TF32, overflow, unmodeled underflow, and corrupted registry/routing/liveness metadata lie outside this arithmetic model.

Matched maintenance controls. On an independent CPU reference replay of the same 20,000 operations and keys, the unmaintained policy produces 6,756 active misses. Rebuilding every 32 bank writes uses 340 rebuilds and has no observed miss, whereas rebuilding every 256 bank writes uses 40 rebuilds and has one miss. Norm-triggered maintenance plus the numerical publication check uses 1,744 rebuilds and has no miss. All four runs have zero wrong-record returns and zero retired-handle aliases. The fixed schedules provide empirical maintenance baselines, while the guarded policy supplies a per-commit publication check. Certification and maintenance scheduling therefore address different objectives. These controls invoke no language model and are evaluated separately from the native-FLA experiment.

Maintenance state and key-generation cost. For the eight-bank, 4,096-record reference initialization, resident 4,096-dimensional FP32 identity keys occupy 64 MiB, the eight binders occupy 2 MiB, and the resident codebooks occupy 0.1875 MiB. The physical-slot-indexed FP64 drift vectors and margins used by the norm controller add 1,114,112 bytes (1.0625 MiB). An isolated staged rebuild at the initial 512-record bank load has a Python-allocator-tracked peak increment of 55,152,392 bytes. This measured increment is not an upper bound on opaque BLAS workspace. On-the-fly generation of one 4,096-dimensional Gaussian identity key from its retained canonical cue takes mean/median/P95 54.87/51.91/78.25 μs over 2,000 calls after 128 warm-ups on the reference CPU. These costs are reported separately from the resident-key address-recovery microbenchmark.

Same-dimensional Gaussian control. To separate key dimension from the extreme coherence of raw BGE-M3/E5 cues, an independent control uses normalized Gaussian keys at $d_k = 1024$ and 4096 over the same five seeds, loads $128, 256, 512, 768, 1024$, locus code, and mutation schedule as the learned-key stress. Both dimensions preserve all 13,440 initial and first-relocation references. Over all later states, the 1,024-dimensional condition records 33,854 fail-closed outcomes, concentrated at loads 768 and 1,024, whereas the 4,096-dimensional condition remains correct in all 369,880 observations. Neither returns a wrong record. This control shows that the immediate raw-semantic-key failure is not explained by dimension alone, while low coherence and correct initialization still do not guarantee an unlimited mutation horizon.