

Part E: The Anchor-Loss Horizon in a Self-Modifying Machine

Does the urn's anchor-loss law $t^ \approx B/h_{\blacksquare}$ emerge in a concrete self-modification loop? An exact study of the "Sidis Machine": a 16-bit self-rewriting register whose state is its program. State space $2^{16} = 65,536$ — fully enumerated; nothing is sampled.*

The machine

State $x = (\text{opcode} \mid \text{operand} \mid \text{data} \mid \text{data})$, four nibbles. Sixteen opcodes; ops 5–6 let data rewrite code, op 12 lets code rewrite data, and a clock rule (opcode always increments mod 16) prevents fixed points. After each self-modification the post-write string selects its own vulnerable bit $j = \text{operand XOR data}$, which chance flips with $p = 0.3$. The machine chooses where it is vulnerable; chance chooses whether. Entropy injection: $h_{\blacksquare_micro} = h_{\blacksquare}(p) = 0.881$ bits/step. The noisy reachable set from the anchor $x_{\blacksquare} = 0x0000$ is 29,052 states — 44.3% of the space, a genuinely mixing world.

Measured constants

Quantity	Value	Reading
$h_{\blacksquare_micro}$ (channel)	0.881 bits/step	raw entropy injection
$h_{\blacksquare_coarse}$ (equilibrium audit)	2.086 bits/step	cost of recording macrostate transitions
h_s ramp	$0.88 \rightarrow 2.43 \rightarrow 2.62 \rightarrow 2.26$ bits ($s=1, 20, 60, 400$)	record cost rises as the distribution spreads
$KL(q_s \parallel p)$ decay	$8.31 \rightarrow 0.56 \rightarrow 0.051 \rightarrow 0.003$ nats ($s=1, 20, 300, 500$)	anchor's statistical signature fades
t_{mix}	218 steps	time to equilibrium of the coarse marginal

The audit cost roughly doubles between the raw channel and the audit interface: coarse-graining into 272 macrostates (opcode $\times$ popcount) makes the deterministic skeleton partially unpredictable too. $h_{\blacksquare}$ is a property of the (system, audit interface) pair, not of the system alone.

E2 — The reversed-claim check

Overlap of the memory with the true future: 0.943. Overlap of the reversed reconstruction (pulling the future state back through the successor map) with the true past: 0.981. Retrodictive claims are more accurate than predictive ones — Sidis's time-mirror survives in record space.

E3 — The Bayesian anchor detector

s (distance from anchor)	Co-oriented: $P(\text{boundary in my past})$	Counter-oriented (pseudo-living)
0	1.000	0.000
20	0.627	0.373
80	0.557	0.443
200	0.511	0.489
400	0.511	0.489

Confidence that the special boundary lies in the agent's own past falls below 0.75 at $s = 20$, below 0.55 at $s = 100$, and asymptotes to chance. The counter-oriented agent's confidence rises in mirror image. Provenance is a decaying signal, not a permanent fact.

E4 — Does t^* emerge the way the urn predicts?

Record model	B (bits, $L=8$ chain)	t^* measured ($\text{conf} < 0.75$)	Urn prediction $B/h_{\blacksquare}$
log-everything	64.7	20 steps	8.0 steps

log-change-only	7.9	20 steps	3.8 steps
optimal (Kraft)	7.9	20 steps	3.8 steps

The horizon is real — but it does not equal B/h . The measured t^* (20 steps; soft horizon 100 at confidence 0.55) is longer than the urn’s storage-side prediction and shorter than the mixing time (218). The resolution:

$t^* = \min(\text{storage horizon } B/h, \text{detector decay } t_{\text{det}})$. The urn law bounds how long the trajectory fits the budget; the detector bounds how long the anchor stays distinctive at the audit interface. Coarse-graining discards exactly the information that distinguishes near-anchor states from equilibrium ones, so the detector side binds first.

The RSI translation

An auditor who sees only a coarse summary of a self-modifying system loses verified provenance at the detector’s horizon, not the storage horizon — and the detector horizon is shorter. The binding constraint on verified descent is the resolution of the audit interface, not the size of the log. A coarse audit trail loses the anchor at t_{det} even with an infinite budget; a fine one loses it at B/h even with perfect logging. Both horizons are real; the $\min()$ rules. Practical corollary: alignment auditing of self-modifying systems must budget for interface resolution (what the auditor can distinguish) at least as carefully as for log volume.

Generated 18 August 2026. Artifacts: part_E_self_modifying_machine.py; E1_hbar_profile.csv, E2_overlap.csv, E3_anchor_detector.csv, E4_horizon.csv. Companions: multibit_memories_report.pdf (Part D), phase_diagram_report.pdf (Part F).