

The Internalization Program: What We Have Tried

J-space / carry-whiteboard experiments, items 1–29

Nils & Claude — 2026-07-16/17 — gemma-4-E2B-it (frozen), DGX Spark

Scored record: `results-loop/PROTOCOL_UNIFIED.md` — every item pre-registered before running

1. Substrate and architecture

All experiments run on a **frozen** gemma-4-E2B-it. The J-lens (averaged Jacobian readout, reproducing the 2026 workspace paper) locates a *workspace band* at layers 14–30: the depth range whose states are verbalizable, persistent, and low sensor/motor. The only trainable component anywhere is a $\sim 2.4\text{M}$ -parameter **merge adapter** $m(e, s)$ that folds a layer-30 band-exit state s back into the layer-14 band entrance alongside the anchor embedding e . Two loops reuse it:

- **Vertical (settle):** each prompt position re-descends the band $k=2$ times on its own state.
- **Diagonal (carry):** each post-prompt position’s band input is $m(e_t, s_{t-1})$ — an RNN through the band across positions. Optional pause tokens (`<unused0>`) are inert compute slots.

GSM8K items are bucketed by base-model capability: *easy* (solved bare, 93.1%), *hard* (needs CoT), *drop* (unsolved even with CoT at harvest). Base: 10.9% overall.

2. The prior arc (items 1–20): looping a frozen band, and its limits

Before the internalization program, ~ 20 pre-registered items established what band-looping can and cannot do (MBPP unless noted; “hard” = base-model-unsolvable bucket, $n=28\text{--}56$):

The retrofit works, and the recurrence is load-bearing (items 1–6). The unified merge adapter lifted MBPP hard from 3.6% ($k=0$) to 28.6% at $k=2$ (family level $\approx 37\text{--}46$ across seeds). A same-size no-recurrence FF adapter reaches only 17.9% — the loop, not the 1.6M new weights, does the work. Mixed-task training interfered in both directions; GSM gained nothing from day one.

Where to loop is not negotiable (items 7–10). Band-location ablation: L14–30 hard 43.6% vs. early L2–12 at 23.6% (overall destroyed) and shifted L6–22 at 21.8%; several late bands are *structurally null* on E2B (KV sharing makes $k>0 \equiv k=0$ bit-identically). Anchor sweep: entrances below L14 collapse monotonically (13/12/11 $\rightarrow$ 34/31/25% overall); the L9 full-attention discriminator confirmed it is the *lens boundary*, not layer type, that gates the retrofit. The lens’s “where” claim survived its falsification attempts.

Adapter-class factorial (items 11–15): the simple merge wins. Alternatives, all controlled: unconstrained RecurrentAdapter ($\rho \rightarrow 4.5$, substrate damage, $4\times$ params bought nothing); Parcae-constrained ($\rho < 1$ certified — trains robustly, gains nothing; *stability and fidelity are independent dials*); tied-alpha (isolates the cause: a *free* B matrix costs ~ 17 easy points, anchoring protects; learned α never moves — hand-tuned 0.3 was optimal); per-depth adapters (content gets *depth-stranded*; weight-sharing is load-bearing); h2048 / noise- s_0 single-cell records (50–54 hard) resolved as lucky seeds by pre-registered rule — *only seed means are levels*. Dynamics deflation: the “fixed point” is an output-stable orbit, not a state fixed point (state cosine keeps drifting at 10^{-3}); no free ACT signal exists; per-item convergence depth shows no difficulty gradient.

Transfer is task-local (item 16): HumanEval tie $\rightarrow$ LiveCodeBench trained-hurts $\rightarrow$ GSM8K trained-content actively toxic (easy 93 $\rightarrow$ 28–45%, hard $\sim 1\%$). A three-point monotone ladder.

The GSM boundary (item 17) and the gate program (items 18–20). GSM-only training in the correct regime: hard $\leq 8.7\%$, no usable gain — the failure is structural (supervision density + state-evolution bottleneck), which set up item 21. Gates: every learned halting head (E1a/b/c) loses to E0’s frozen class-balanced logistic probe on the pre-loop state; E1c solves easy *routing* (95.9% at 0.11 mean iters — 82% compute saved) but regresses hard recall. Item 20’s oracle bound: 59.6% overall at 0.24 mean iters — hard items are *depth-diverse* (18/28 solvable at some k , ≤ 13 at any single k); gate quality is worth ~ 9.6 points and remains unclaimed.

3. The hybrid baseline (item 21) — the number to beat

Dense self-distilled supervision: the model’s own verified terse scratchpads ($\sim 30\text{--}60$ tokens, answer-checked, 427 items) trained through the carry architecture. **Result: 57.4%** overall ($5\times$ the prior GSM best of 12.1). Matched feed-forward control (same supervision, no recurrence): 54.7. The recurrence’s specific, significant

edge is *reach*: drop-bucket discordants 20–6, McNemar $p=0.0094$. Conclusion: scratchpad supervision carries the bulk; the carried state extends reach into problems unreachable at labeling time.

4. The internalization ladder and its controls (items 22–24)

Question: can visible scratchpad steps be deleted and their computation absorbed into the latent chain? Front-first deletion, each deleted step replaced by 10 pauses, warm-started curriculum.

item	manipulation	best cell	verdict
22	ladder $d=1/2/3$	31.6 / 18.4 / 19.1	breaks at $d=1$ (−25.8); plateau ≈ 19
23a	3× training steps	(val only)	no recovery; overfit — time not binding
23b	3× pauses (30/step)	29.3	bandwidth not binding
24	band-LoRA r16 (loop-only)	stopped	fit unchanged — expressivity not binding

Positive half of item 22: the pause-only rung ($\approx$ rung C) at 19.1 is *double* the cold-trained answer-only floor (9.4) and paired-significantly above base ($p=0.0025$): curriculum buys real latent capability — just a fraction of a visible step’s worth. The break is structural: capacity, time, positions, and band expressivity all fail to move it.

5. Microscopy interlude — photographing the whiteboard

Single-prompt lens tables (every band layer × every position; carry vs. feed-forward control; easy/hard/drop specimens; published as an interactive artifact). Findings that shaped everything after:

- **Prompt processing** ends with roles (“price/discount”), a chosen route (“Multiply”), and a coarse answer magnitude (“6/sixty/Sixty”) already on the board — magnitude is a base-model capability, one hop deep (on the hard specimen the base latches the *intermediate*, e.g. “thirty”).
- **Exact digits appear only just-in-time**, 1–2 positions before emission ($P=0.997$ at the “=” before the result exists in text). The board natively carries *plans*, *magnitudes*, *completion-state* — not digit registers.
- **The drop-item divergence (Charlotte, 430):** at identical prefixes, the carry board keeps “plus” alive and *defers* (re-expands the sum, correct 132); the feed-forward board commits a wrong digit at 99.3% confidence (155 → 124). The recurrence’s reach = knowing it isn’t done.
- Carry and FF boards agree ≥ 85 –92% at L26–30 on solved items: everything token-legible is computed locally; the carry’s contribution is not lens-visible where both succeed.

6. Lens-supervised internalization (items 25–28)

The ladder failed blind (output CE only); these supply the latent chain with process-level information. The lens readout is differentiable, so “what the board should hold” becomes a trainable loss.

item	supervision	matched	ablated	outcome
25 lens-teach	pause $j \leftrightarrow$ deleted-step token j (lens-CE)	31.2	—	lens-CE 10.3 → 1.9: board <i>writes</i> the step; accuracy flat. Writing $\neq$ computing.
26 staging	pre-“=” positions must hold the line result	24.2 / 15.6	—	Actively harmful: forcing thought before operands are readable corrupts the JIT schedule.
27 burst	zero pauses; 10 in-place band iterations at prompt end; iteration $i \leftrightarrow$ token i	34.0	32.4	best nominal; paired n.s. ($p=.45/.58$). Pause tape is dead weight (zero-pause $\geq$ tape).
28 teacher state	burst endpoint pulled onto teacher’s “step finished” L30 state (co-sine)	30.1	29.7	distillation succeeds geometrically (cos-dist .113 → .044); function does not follow; easy damaged (48.3).

Uniform conclusion: **encoding is free; consumption is the wall**. Every representational target was hit; none transferred. The write side is exhausted across eight axes.

7. Trajectory teacher-forcing (item 29) — the first significant positive, with a twist

Nils’s design: measure the full-CoT teacher run’s L30 at 10 evenly spaced waypoints $T_1..T_{10}$; since both the input to the merge (T_{i-1}) and the desired band output (T_i) are known, the burst trains as **ten independent supervised transitions** — dense gradient on the loop’s *dynamics*, no BPTT chain, $T_0 \equiv$ the student’s own settled anchor state.

Arm 1 (tf-only, $d=1$): 39.1% — clears the pre-registered reopen threshold (36.6); every bucket a $d=1$ record (easy .69, hard .433, drop .25); paired vs. the blind reference: 50–31 discordants, **McNemar** $p=0.045$. Vals best-in-series.

The twist: the ablation (same adapter, burst off) also scores 39.1 (8–8, $p=1.0$). The test-time loop is causally inert — as it has been in every experiment of the program. The gain is a **training-signal effect**: the transition regressions teach the merge operator better state-folding, which pays off at every ordinary carry step under visible tokens. The transition loss also acts as an internalization gauge: it plateaus at ~ 0.085 (from 0.116) — the autonomous operator absorbs $\approx 25\%$ of the token-driven dynamics.

Still running at time of writing: the tf+fr arm (exposure-gap term) and job 2 (answer-only output, waypoints across the *entire* CoT — compressed-rehearsal internalization vs. the 19.1 plateau).

8. Side results banked along the way

- **Cross-model lens survey complete (5 scales):** E4B, 26B-A4B (30 layers, not 48), and 31B share an early-persistence / terminal-motor profile that breaks with E2B/12B’s sensor→workspace→motor layout; no E2B-style band found on the larger models yet.
- **LoopLoRA validated:** loop-only band deltas leave $k=0$ bit-exact with trained weights loaded.
- Ops: H100 pipeline harvested and node destroyed; exp4 save bug fixed mid-pipeline; ~ 14 pre-registered items scored in ~ 36 h.

9. Standing conclusions and open moves

1. The recurrent loop *at inference* has never been causal — across pauses, bursts, supervision, distillation, capacity, and time. Its proven inference-time value is confined to the hybrid regime: reach on drop items alongside a visible scratchpad (item 21).
2. The whiteboard’s native cargo is plans, coarse magnitudes, and completion-state; it is not a digit register, and it cannot be made one by write-side training.
3. Teacher-trajectory transition supervision is the best adapter training signal found (+7.5, $p=0.045$) — a discovery about *gradients*, not architecture.
4. **Open, in order:** (a) item 30 candidate: trajectory TF applied to rung A itself (nothing deleted) — move the 57.4 headline; (b) seed replication of the 39.1 before any strong claim; (c) the read-side clamp (write teacher states in at inference, measure consumption causally); (d) job 2 verdict on compressed-rehearsal internalization.