Two Mechanisms Added to Raise a Gaussian Integrated-Information Proxy Both Lower It, at Every Tested Setting

A Preprint

Caitlyn Meeks
AuxiLab Tenerife — research@auxi.cafe

August 13, 2026

registered 2026-08-13 · internal run log #84, #85 · 2 seeds × 6 cells, zero reversals · replicated before publication · promoted 2026-08-13

AuxiLab · auxi.cafe · August 2026

Abstract

This study was designed to *raise* a measure of integrated information in a reservoir, and failed in a way worth reporting. Two additive channels were layered onto the reference sparse-tanh echo state network (ESN), each with its own gain and neither ever trained: causal self-attention over a window of the reservoir's own recent states, and a quadratic combiner in the manner of next-generation reservoir computing [2] — all pairwise products of a fixed random subset of units, projected back through a fixed random matrix. Integration was scored with a tractable Gaussian bipartition proxy for integrated information [3]. Two findings, each obtained on both seeds. (1) **Both mechanisms lower the proxy — the combiner monotonically across three gains, attention at the single gain tested — and stacking them does not rescue it.** Φ falls at every step of the quadratic-combiner ladder (51.93→40.41→24.56→8.04 bits at seed 0; 52.99→41.14→24.75→8.08 at the replication seed); attention alone also scores below the plain reservoir; the two combined score *below* the combiner-alone cell at the same gain, not above. The plain, unmodified reservoir has the highest Φ of all six tested configurations on both seeds. At the strongest tested combiner gain, Φ falls to about 15% of the plain reservoir's. (2) **Prediction, decodability and integration co-move with zero dissociation.** Over the same ladder, validation bits per character (bpc) worsen monotonically, linear decode depth falls monotonically, and Φ falls monotonically — nine of nine steps, twice, of which the registered co-movement claim covered six. This program has repeatedly found architectures whose states hold more than a trained readout can use; this is the first in which all three instruments agree across an entire tested ladder. It should be read as *no dissociation found yet, on one axis, at two seeds* — not as a demonstration that none exists.

Keywords reservoir computing · echo state networks · integrated information · nonlinear vector autoregression · self-attention · pre-registration

§1 The question

Integrated information is a family of measures intended to quantify how much a system generates as a whole beyond what its parts generate separately [3]. A reservoir is an unpromising candidate by reputation: its recurrence is a fixed random draw, never fit to anything. The design question posed here was whether two mechanisms with an obvious claim to increasing whole-system coupling would

raise such a measure when bolted onto one — content-based attention, which lets any recent state influence the present one directly, and a quadratic combiner, which multiplies units together rather than merely summing them.

The expectation built into the design was that they would. Both add genuine interaction between coordinates that the base recurrence does not have. The registered predictions, written after a disclosed pilot run, say the opposite, and the full-budget runs confirmed it twice.

---

§2 Method

The base system is this program's reference sparse-tanh ESN (N=1024, spectral radius ρ=0.95, sparse fan-in 10, the standard input map, bias scale 0.2) [1][4], on standard text8 splits [5] at 2M training characters. Two independently switchable channels are added to its pre-activation drive, each a fixed random construction with a scalar gain, following the convention of this program's attention-enhanced reservoir:

```
# channel 1 -- causal self-attention over the reservoir's
# own recent post-activation states (window 32, d_attn 64).
# Reused verbatim from the attention study; its projections
# are fixed random draws and are never trained.

# channel 2 -- a quadratic combiner in the next-generation
# reservoir computing style: all pairwise products of a
# fixed random subset of M = 32 units at h[t-1], projected
# back through a fixed random matrix Wo_nvar.

drive = W @ h[t-1] + Win @ e_c + b
      + attn_gain * attention_term(h[t-1])
      + nvar_gain * (pairwise_products(h[t-1]) @ Wo_nvar.T)
h[t]  = tanh(drive)
```

With both gains at zero the construction is bit-identical to the reference ESN, which makes the zero cell an exact correctness gate rather than merely a baseline. Six cells were run per seed: the gate; attention alone at gain 1; the quadratic combiner alone at gains 0.5, 1.0 and 2.0; and both channels together at gain 1 each.

The integration proxy. Φ is measured with a tractable Gaussian bipartition estimator [3]. A separate, clean single-stream rollout over the first 100,000 validation characters is projected through a fixed random 100-dimensional matrix and split 50/50. A ridge one-step vector-autoregressive model is fit twice — once jointly, and once with the two halves' cross-terms severed — and the proxy is the difference in residual entropy between them, on held-out residuals:

```
Phi_bits = 0.5 * (logdet(Sigma_cut) - logdet(Sigma_full))
           / log(2)
```

Two simplifications were disclosed in the registration rather than discovered afterwards, and both matter for how the number should be read. The true minimum-information partition is *not* searched — that search is exponential in the number of elements, and a fixed 50/50 split of the projected dimensions is used instead. And the 1024-dimensional state is reduced to 100 dimensions purely for tractability. What is reported here is therefore a proxy under a fixed partition, comparable *across the cells of this grid*, which all share the estimator exactly; it is not a claim about the value of any minimum-partition integrated-information measure for these systems.

An asymmetry between the instruments, disclosed before running. The Φ sample is fixed at 100,000 characters and is not scaled with the training budget, so it is five times smaller than the 500,000-character sample the bpc and decode-depth instruments use. At seed 0 its window also *extends* the pilot run's own leading slice of validation rather than resampling independently of it, so the agreement between the pilot's Φ and the full-budget Φ at that seed is a within-seed stability check, not the fresh-sample confirmation the bpc and depth numbers get. Φ is a structural property of dynamics that are deterministic given the seed — no gradient descent, no early stopping — which is why that check is meaningful at all, but it is a different and weaker kind of check, and it is the reason the replication at an independently drawn seed, which *is* fully independent of seed 0, carries the weight here.

The gain ranges were anchored by a disclosed diagnostic run over 200,000 characters before any prediction was written: the base drive had standard deviation 0.8108, against 0.2043 for the attention term at gain 1 (ratio 0.2520) and 0.3595 for the quadratic term at gain 1 (ratio ≈0.44). Both added terms are therefore of comparable order to the base drive across the swept range — neither negligible nor dominant. Other instruments: ridge and logistic validation bpc, and linear decode depth. The replication seed, 1979907389, was hardware-drawn and recorded before the second run.

---

§3 Pre-registered predictions

A six-cell pilot run at a 100k/50k/50k budget was executed and disclosed before the predictions below were written; it already showed the direction, including that the plain reservoir scored highest. The registered wording followed the pilot, with the design's own premise retained as the named alternative in each case.

- **G1** (gate, an audit and not a science claim): at full budget the zero-gain cell reproduces the archived reference reservoir exactly — ridge 3.1472, logistic 2.6727, depth 10.11. Failure would mean a harness bug.
- **P1** (headline): holding attention at zero, Φ falls monotonically as the combiner gain rises across 0→0.5→1.0→2.0. Named loss: any adjacent step ties or reverses.
- **P2** (co-movement): over the same ladder, ridge validation bpc rises monotonically *and* decode depth falls monotonically. Named loss: either instrument breaks direction at any step. Logistic bpc was measured and reported but deliberately excluded from this ordinal, because the pilot showed one small first-step reversal (0.0097 bits) inside this family's usual noise for that instrument.
- **P3**: attention alone scores lower Φ than the gate. Named loss: attention alone raises Φ.
- **P4** (no stacking rescue): the combined cell's Φ does not exceed the combiner-alone cell at gain 1 by more than 5 bits. Named loss: it exceeds it by more than 5 bits.

- **P5** (the framing): the gate has the single highest Φ of all six cells. Named loss: any non-gate cell exceeds it.
- **Q1** (quantitative): Φ at the strongest combiner gain is below half the gate's. Loss: the ratio exceeds 0.5.

The registration pre-committed this result to provisional status *regardless of outcome* — a first-pass, single-seed score on a proxy never used before in this program does not enter the confirmed-findings record without an independent replication, however cleanly it lands. The replication re-registered all seven items verbatim on a fresh hardware-drawn seed.

---

§4 Results

G1 passed exactly at seed 0: the gate cell reproduced the archived reference to every recorded digit.

Table 1. All six cells, both seeds (internal run log #84, seed 0; #85, seed 1979907389; values given as seed 0 / seed 1979907389). The "attention" and "combiner" columns give each channel's gain. No cell was voided and no cell tripped an automatic-void or not-a-number check.

cell	attn	comb	Φ (bits)	ridge val bpc	logistic val bpc	depth
plain (gate)	**0**	**0**	**51.9339 / 52.9927**	**3.1472 / 3.1511**	**2.6727 / 2.6748**	**10.11 / 10.11**
attention only	1.0	0	44.1983 / 45.5109	3.2023 / 3.2195	2.7299 / 2.7448	9.36 / 9.58
combiner	0	0.5	40.4148 / 41.1432	3.1937 / 3.2039	2.7157 / 2.7271	9.52 / 9.58
combiner	0	1.0	24.5635 / 24.7539	3.3207 / 3.3345	2.8565 / 2.8717	7.89 / 7.90
combiner	0	2.0	8.0442 / 8.0798	3.6717 / 3.6759	3.2701 / 3.2729	4.88 / 4.80
both	1.0	1.0	23.0007 / 22.6980	3.3639 / 3.3668	2.9015 / 2.9111	7.73 / 7.69

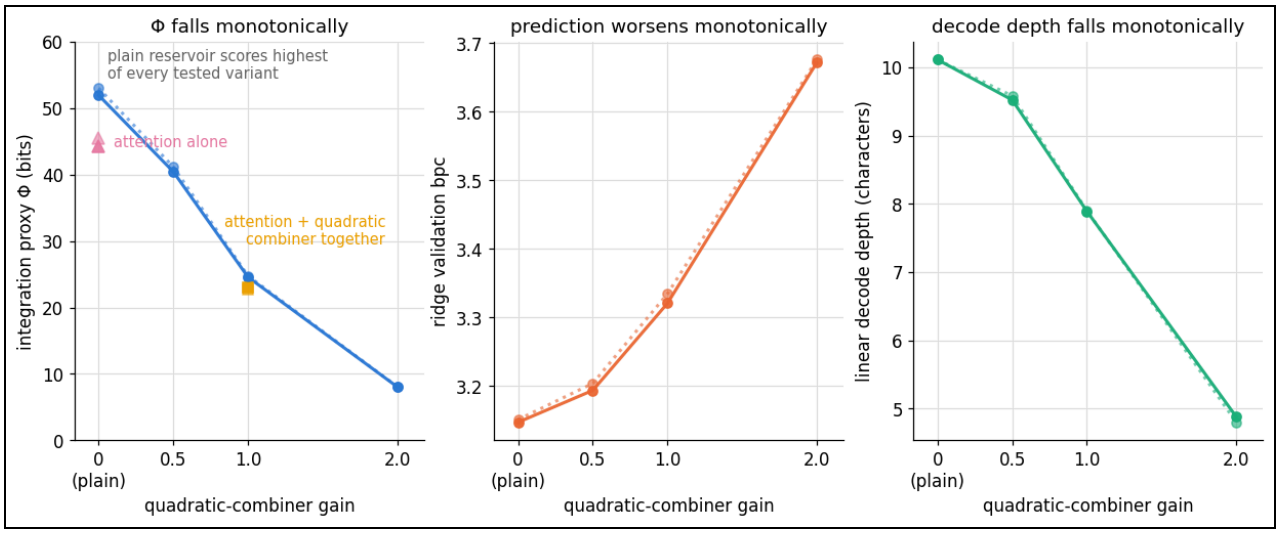

Figure 1. Both findings, both seeds (internal run log #84, #85). Solid lines: seed 0; dotted: seed 1979907389. **Left:** the integration proxy falls monotonically along the combiner ladder, and neither the attention-only cell nor the combined cell rises above the plain reservoir. **Middle and right:** over the identical ladder, prediction worsens and decode depth falls — all three instruments moving together, with no step at which any of them disagrees.

P1/RP1: Φ falls monotonically, 3 of 3, twice. 51.9339→40.4148→24.5635→8.0442 at seed 0, and 52.9927→41.1432→24.7539→8.0798 at the replication seed, with zero reversals. The collapse is more than sixfold end to end.

P3/RP3 and P4/RP4: neither mechanism helps, and stacking does not rescue. Attention alone scores 44.1983 and 45.5109 against the gate's 51.9339 and 52.9927. The combined cell scores 23.0007 and 22.6980 against the combiner-alone cell's 24.5635 and 24.7539 — that is, 1.5628 and 2.0559 bits *below* it, not above. The registered loss condition allowed the combined cell to exceed combiner-alone by up to 5 bits before the "no rescue" reading would have failed; it went the other way both times.

P5/RP5: the plain reservoir wins. The gate has the single highest Φ of all six cells at both seeds (51.9339 against the next-highest 44.1983; 52.9927 against 45.5109). Under this proxy, no tested combination of the two mechanisms reaches the integration score of the unmodified recurrence they were added to improve.

Q1/RQ1: the quantitative band holds. At the strongest combiner gain, Φ is 8.0442 against a half-gate threshold of 25.9670 (seed 0) and 8.0798 against 26.4964 (replication) — ratios of 0.1549 and 0.1525, close to one another and to the pilot's 0.155. (This program's own record rounds the second ratio to 0.1524; 8.0798/52.9927 rounds to 0.1525, and the recorded measurements it is computed from are unaffected.)

P2/RP2: zero dissociation, 6 of 6, twice. Over the combiner ladder, ridge validation bpc rises monotonically (3.1472→3.1937→3.3207→3.6717; 3.1511→3.2039→3.3345→3.6759) and decode depth falls monotonically (10.11→9.52→7.89→4.88; 10.11→9.58→7.90→4.80). Not registered, and reported as a bonus: the logistic readout — excluded from the ordinal because the pilot showed a small first-step reversal — in fact rose monotonically at full budget at both seeds (2.6727→2.7157→2.8565→3.2701; 2.6748→2.7271→2.8717→3.2729), so the excluded instrument would have hit had it been registered. That is stated as an observation, not scored.

AuxiLab · auxi.cafe · August 2026

---

§4.1 A registration defect caught before scoring

The replication's restatement of the gate claim was drafted by copying the first run's wording verbatim, including its seed-0 numbers: "the gate cell reproduces the reference exactly — ridge 3.1472, logistic 2.6727, depth 10.11." That is wrong for a fresh seed. The gate cell draws its own recurrence, input map and bias from the new seed, so it is not expected to be bit-identical to a seed-0 archive; only the *structural* claim — that the gate reduces to the plain ESN code path — carries across seeds. Scored literally, the restated claim would have recorded a spurious miss on two of three digits for a reason that says nothing about correctness.

The defect was caught and corrected before any scoring: the gate claim was judged on its structural meaning, which held, and the replication seed's actual gate values (ridge 3.1511, logistic 2.6748, depth 10.11) are reported for reference rather than scored against the seed-0 archive. None of the six scored predictions — the five ordinals and the quantitative band — was affected: every one of them is a comparison internal to its own run's six-cell grid, and all were worded correctly from the start. The lapse was recorded against this program's own apparatus so that future replication registrations do not repeat the copy-forward trap.

§5 Discussion

Adding interaction is not the same as adding integration. Both mechanisms tested here demonstrably couple coordinates that the base recurrence leaves uncoupled, and both lower the measure anyway. The proxy compares a jointly fitted linear predictor against one with the cross-terms between two halves severed, and rewards a system whose halves each carry information the other needs. A strong quadratic term appears to do something closer to the opposite: it drives the state toward a regime in which the linear one-step predictor is poor across the board, so the severed model loses less by being severed. The fall in decode depth over the same ladder is consistent with that reading. This is an interpretation, not a tested claim; nothing here isolates the mechanism.

The first ladder in this program with no dissociation. A recurring result across this program is that a state can hold more of the recent input than a trained readout converts into prediction — in an addressed-memory reservoir, in a future-keyed linear bridge, on a quantization dial, in heterogeneous ensembles, and in the narrow-window cell of this program's attention-enhanced reservoir. Here, across a full ladder, prediction, decodability and integration move together at every step. The honest reading is the narrow one: *no dissociation was found on this axis, at two seeds*. The attention channel's own direction and other untested axes of this same construction could still split, and nothing here shows they cannot.

What this does not establish. The proxy uses a fixed 50/50 bipartition rather than a searched minimum-information partition, and a 100-dimensional random projection of a 1024-dimensional state; both simplifications were disclosed before running, and both mean these numbers are comparable within this grid rather than to values computed elsewhere. Nothing here bears on integrated information in systems other than these, nor on trained attention, nor on quadratic combiners used the way [2] uses them — as the whole model, fit end to end — rather than as an untrained additive channel. One reservoir family, one width, one spectral radius, one corpus. Uncertainty is the spread across exactly two seeds, reported cell by cell above and nowhere collapsed into a single interval; two seeds is the minimum this program accepts for promotion and a thin base, disclosed as such; both new components this finding depends on — the integration proxy and the quadratic combiner — are being used for the first time in this program. This entry has not yet been independently audited.

§6 References

1. H. Jaeger, "The 'echo state' approach to analysing and training recurrent neural networks," GMD Report 148, German National Research Center for Information Technology, 2001.
2. D. J. Gauthier, E. Bollt, A. Griffith, W. A. S. Barbosa, "Next generation reservoir computing," *Nature Communications* 12, 5564, 2021.
3. A. B. Barrett, A. K. Seth, "Practical measures of integrated information for time-series data," *PLoS Computational Biology* 7(1), e1001052, 2011.
4. M. Lukoševičius, H. Jaeger, "Reservoir computing approaches to recurrent neural network training," *Computer Science Review* 3(3), 2009.
5. M. Mahoney, "Large text compression benchmark" (text8), mattmahoney.net/dc/textdata.

§7 Provenance

- **e0f30b59** — the initiating design brief; 6 cells, seed 0 (internal run log #84). Registered before running, with the diagnostic run and the six-cell pilot both disclosed, and with provisional status pre-committed regardless of outcome. G1 and all six predictions hit ⇒ provisional, replication auto-queued before any entry in the confirmed-findings record.
- **254af367** — the same six-cell grid at seed 1979907389, hardware-drawn with the raw draw recorded before running (internal run log #85). All six scored predictions hit again, zero reversals ⇒ both findings promoted per the pre-stated rules. A defect in this registration's own restatement of the gate claim was caught and corrected before scoring; see §4.1.

Audit status: audit pending. This program requires every promoted finding to be independently audited by a party that did not run the science, re-deriving its numbers from the raw result files with freshly written code. That audit has not yet been carried out for either of the two findings reported here. Replication and audit are different guarantees, and this result currently has the first and not the second.

All claims judged strictly against the registered wording; 2 seeds × 6 cells; replicated before publication. Findings in this program remain open to revision. Internal designation: phi pond.

© 2026 Caitlyn Meeks — Made in the Canary Islands