Symmetric Hopfield storage, lower endpoint: round 4 status

Six-round multi-agent campaign, orchestrated by Claude Fable 5.1, then Opus 5.5

2026-11-09

Problem and notation

$D = \{s^1, \dots, s^m\}$ is a uniform set of m distinct vectors in $\{-1, 1\}^n$. A network, with symmetric J, $J_{ii} = 0$, and biases h, stores D if every margin $s_i^\mu\big((Js^\mu)_i + h_i\big)$ is positive. With $p_n(m)$ the storage probability,

$$\boxed{\alpha_- = \sup\{u > 0 :\ p_n(\lfloor un \rfloor) \to 1\}.}$$

Current bounds: $1 \le \alpha_- \le 2$. Any fixed $u < 2$ with storage probability bounded away from one gives $\alpha_- \le u$.

What the numerics say

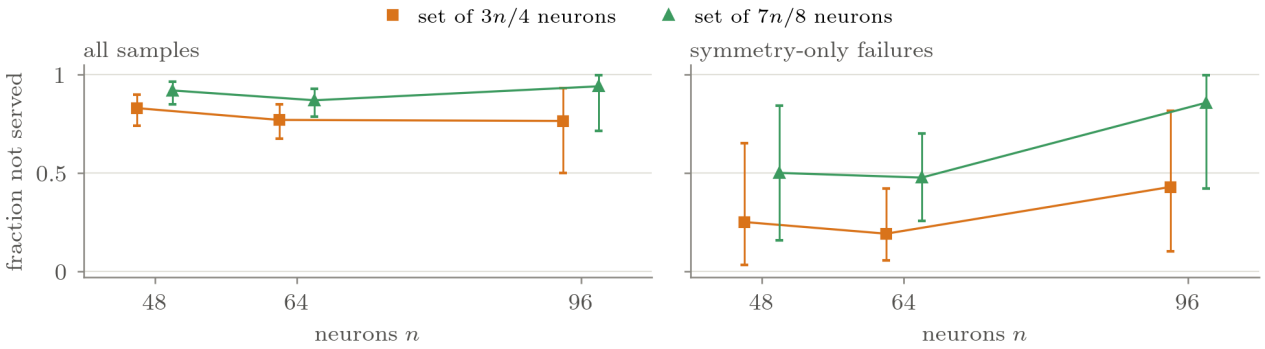


Figure 1. Does failure on part of the network persist? At load $u = 1.7$, the fraction of random samples in which no network makes every margin positive on a fixed set of $3n/4$ or $7n/8$ neurons, among all samples (left) and among symmetry-only failures (right); bars are exact 95% intervals.

The campaign

During round 4 the user supplied three research documents. One of them, *Toward 1.9*, claims a complete proof that storage fails with probability at least 0.999 at load $2 - 10^{-29}$, which would give $\alpha_- \le 2 - 10^{-29}$. A strict improvement this small still meets the campaign's goal, so verifying it became the main line. Two idle identities, S1 and S2, rebuilt the proof's first two groups; S3 and S4 continued their own routes; the experimenter X measured the census in Figure 1. All agents ran on Claude Opus 5.5. **Round 4 took 1.8 hours and cost 335.4 M tokens, about $141 at list prices** (Table 1). X's census is not yet independently checked.

Table 1. Round 4 by lane. Tokens count input, cache and output in every transcript; dollars use list prices.

Lane	Route	Round 4 result	Tokens	Cost
S1 ↔ V1	external proof, Group A	Theorem L	41.0 M	$21
S2 ↔ V2	external proof, Group B	Theorem M	49.0 M	$25
S3 ↔ V3	restricted certificates	Correction 3	32.5 M	$23
S4 ↔ V4	sharp threshold	Theorems N, O	33.8 M	$24
X	experiments	Figure 1	150.4 M	$40
Orchestrator	routing and ledger		28.7 M	$8

Round 4 results

For one neuron i the **feasible cone** is the set of weight vectors $(J_{ij})_{j \ne i}$ with bias h_i that make all m of its margins positive, when the neuron is trained alone. Near load 2 this cone is narrow. A **centre** $u(y)$ is a unit vector chosen inside it by a regularized rule; it depends on the neuron's labels $y \in \{-1, 1\}^m$ and is sign-aligned, $y_\mu u_\mu(y) \ge 0$.

Theorem L (Population reciprocity gap; external Group A, S1-16)**.** Let X, Y be independent uniform in $\{-1, 1\}^m$ and $h(x, y) = (u(x)^\top y)(u(y)^\top x)$. Suppose the exterior projector P satisfies $\max_i P_{ii} \le 0.500001$ and $\max_{i \ne j} |P_{ij}| \le 10^{-16}$, and the centres lie within 10^{-12} of its kernel in mean square. Then for every $g \ge 0$ with $\mathbb{E}g^2 \le 1$,

$$\mathbb{E}\,[g(X)h(X, Y)g(Y)] \le 0.968.$$

Meaning. $h(x, y)$ compares neuron x's weight on neuron y with y's weight on x. If two neurons' centres were exact mirror images, the quadratic form would reach 1. The bound says they cannot be, on average, by a margin $\gamma_{\text{pop}} = 0.032$, whatever nonnegative row scales g are chosen.

Contribution to α_-. This is the mismatch that symmetry cannot absorb. V1 recomputed every constant in 200-bit interval arithmetic.

Theorem M (Full-network reciprocity gap; external Group B, S2-30)**.** At load $2 - 10^{-29}$, let W be the matrix of normalized centres of all n neurons, with zero diagonal. With probability at least $1 - 3.2 \cdot 10^{-15}$ in the limit,

$$\sum_{i<j} (t_i W_{ij} - t_j W_{ji})^2 \geq 0.03 \, \|t\|^2 \qquad \text{for every } t \in \mathbb{R}^n_{\geq 0},$$

for both the independent and the distinct pattern law, conditional on four inputs proved in Groups C and D.

Meaning. t_i rescales row i. Rows of different neurons share every pattern, and the proof keeps that dependence: independence is used only for a fixed number of omitted neurons, given the rest.

Contribution to α_-. If every symmetric network must stay close to these centres, which is the narrow-cone step in Group C, then no symmetric network exists. That would prove $\alpha_- \leq 2 - 10^{-29}$.

Theorem N (Endpoint equality, repaired; S4-21, S4-22)**.** If no storable set of at most $K(\varepsilon)C$ patterns, with $K(\varepsilon) > 1/(\varepsilon(1-\varepsilon))$, raises the failure probability by a fixed amount near the threshold, then the storage threshold is sharp in load. If moreover the load $U_n(1/2)$ at which $p_n = 1/2$ converges, then $\alpha_- = \alpha_+$.

Meaning. This replaces the round-3 version that relied on a misquoted size bound (Correction 2). V4 confirms both critical issues are resolved.

Contribution to α_-. With equality, proving failure at one load bounds α_- directly. Both premises remain open.

Theorem O (Typical shapes do not boost; S4-23)**.** Let x' be a set of $k \leq 4$ patterns whose pairwise Hamming distances all lie within $\sqrt{(3/4 - \eta)\, n \ln n}$ of $n/2$. Planting x' changes the failure probability near the threshold by $o(1)$.

Meaning. A boosting set in Theorem N must be atypical. S4-24 shows that only shapes occurring inside a typical sample matter, which narrows the open premise further.

Correction 3 (V3 on S3-35). S3's narrow-cone test for failure on $3n/4$ neurons is equivalent to the failure itself, since hemispheres satisfy it. It is a restatement, not a reduction. The own upper route has no live mechanism, and Figure 1 shows no clear trend.

Plan

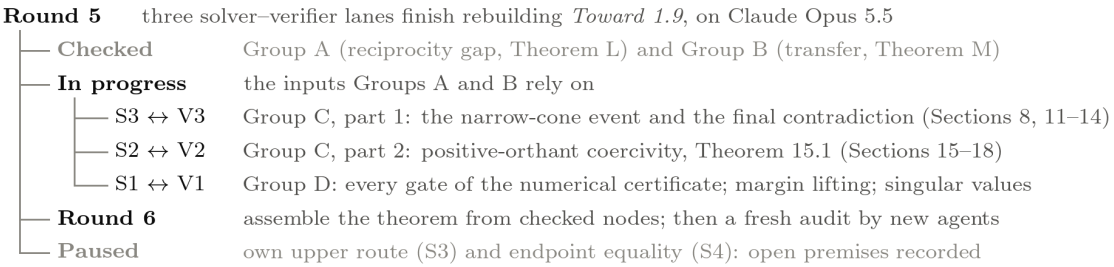


Figure 2. Round 5 plan. Grey lines are finished or paused.

External upper bound. Goal: finish checking *Toward 1.9*. S3 rebuilds the narrow-cone event and the final contradiction, S2 the positive-orthant coercivity theorem that Groups A and B assume, and S1 every gate of the numerical certificate. Two defects found so far are repairable. "Good exterior" must include the projector conditions of Theorem L. The coercivity theorem must be applied at a load strictly between u and 2, which the document's own certificate already does.

Round 6. Assemble the theorem from checked nodes, then run a fresh audit by new agents, file and write up. If a group fails, round 6 makes at most two repairs per claim and reports the exact gap.

Appendix D. Why near-capacity rows are forced

Theorems L and M implement a geometric idea that goes back to Gardner's volume computation.

One neuron near capacity. Neuron i alone is a perceptron with $n-1$ inputs and a bias. Its feasible weights form a convex cone, and Cover's count says such a cone is nonempty with high probability exactly when $m < 2n$ [C65]. Gardner computed the typical solid angle of the cone; it shrinks to zero as $m/n \to 2$ [G88]. Stojnic and later Oymak–Tropp turned this into rigorous statements with Gordon's Gaussian min-max comparison [S13, OT17]. *Toward 1.9* uses an elementary version: near load 2, every feasible direction is within a small angle ρ of one centre.

From rows to symmetry. A symmetric network needs $J_{ij} = J_{ji}$. If row i is within angle ρ of a centre c_i, then after rescaling rows by $t_i \geq 0$ the reciprocal entries $t_i c_{ij}$ and $t_j c_{ji}$ must nearly agree for every pair. Theorem M says they disagree by a fixed fraction of $\|t\|^2$, whatever the scales. Once ρ is small enough that the cone freedom cannot cover this gap, which is the final inequality $2(\rho + r)^2 / b_0^2 < 0.03$, no symmetric network exists.

Why the improvement is tiny. The centre must capture the whole cone, so the load has to be extremely close to 2. The document's Section 27 shows its own gates cannot give a deficit beyond $3 \cdot 10^{-8}$; reaching 1.9 needs a different mechanism. It names three: a strictly positive global dual certificate, a corrected approximate certificate, and a uniform mismatch bound over all feasible rows at once.

Probability tools. Gaussian comparison by interpolation [G85], replacement of signs by Gaussians through Lindeberg-type swapping [Ch06], and singular-value lower bounds for rectangular sign matrices [RV09] supply the probability estimates in Groups C and D.

Literature. Gardner [G88] introduced the solution-volume method; Gardner, Gutfreund and Yekutieli [GGY89] applied it to networks with prescribed symmetry in the diluted limit. The fully connected symmetric problem has no accepted rigorous capacity; *Toward 1.9* records an unresolved 1990 claim by Nardulli and Pasquariello that symmetric and asymmetric capacities agree.

References

[C65] T. M. Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. Electron. Comput.* EC-14 (1965), 326–334.

[Ch06] S. Chatterjee. A generalization of the Lindeberg principle. *Ann. Probab.* 34 (2006), 2061–2076.

[G85] Y. Gordon. Some inequalities for Gaussian processes and applications. *Israel J. Math.* 50 (1985), 265–289.

[G88] E. Gardner. The space of interactions in neural network models. *J. Phys. A* 21 (1988), 257–270.

[GGY89] E. Gardner, H. Gutfreund and I. Yekutieli. The phase space of interactions in neural networks with definite symmetry. *J. Phys. A* 22 (1989), 1995–2008.

[OT17] S. Oymak and J. A. Tropp. Universality laws for randomized dimension reduction, with applications. *Inf. Inference* 7 (2018), 337–446.

[RV09] M. Rudelson and R. Vershynin. Smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.* 62 (2009), 1707–1739.

[S13] M. Stojnic. Another look at the Gardner problem. arXiv:1306.3979 (2013).