Symmetric Hopfield storage, lower endpoint: round 6 status

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

2026-11-10

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,

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

The problem file defines α_-; the run adds α_+. **Bounds at the start of the campaign:** $1 \le \alpha_- \le 2$. **Now:** $1 \le \alpha_- \le \alpha_+ \le 1.999999$, proved and independently checked, with the fresh audit still to run.

How the argument for $\alpha_- \le 1.999999$ works

Neuron i owns the row $w_i = \big((J_{ij})_{j\ne i}, h_i\big) \in \mathbb{R}^n$, and its m margins are the entries of $A_i w_i$, where $A_i \in \{-1,1\}^{m\times n}$ has rows $s_i^\mu(s_{-i}^\mu, 1)$. A row is *feasible* if $A_i w_i > 0$. Each neuron has a *centre* W_i: the coupling part of its best-margin row, scaled to unit length. If (J,h) stores D, put $t_i = \|(J_{ij})_{j\ne i}\|$ and $R_i = (J_{ij})_{j\ne i}/t_i$. Symmetry $t_i R_{ij} = t_j R_{ji}$ gives

$$\sum_{i<j}(t_i W_{ij} - t_j W_{ji})^2 \ \le\ 2\sum_i t_i^2\,\|W_i - R_i\|^2. \tag{1}$$

The proof shows that the left side is large for random centres and that the right side is small, because every feasible row lies near its centre.

Round 5 ran this argument at load $2-10^{-29}$: rows within 0.12 of their centres, mismatch at most $0.0294\,\|t\|^2$, network gap $0.03\,\|t\|^2$. The user-supplied *upper_bound_3* runs it at load $2-10^{-6}$, with two new estimates:

- **Rigidity from mean margins** (Theorem U). Every feasible unit row has mean margin in a window of width 0.00084. The window forces the cone's diameter below 0.13, so (1) bounds the mismatch by $2(0.13 + 10^{-9})^2/0.999^2\,\|t\|^2 = 0.0339\,\|t\|^2$.
- **A larger gap** (Theorems V and W). The old gap 0.03 is now too small. A pointwise bound on how far the centres leak out of the kernel of the exterior design raises the population gap to 0.0374 and the network gap to $0.0373 > 0.0339$.

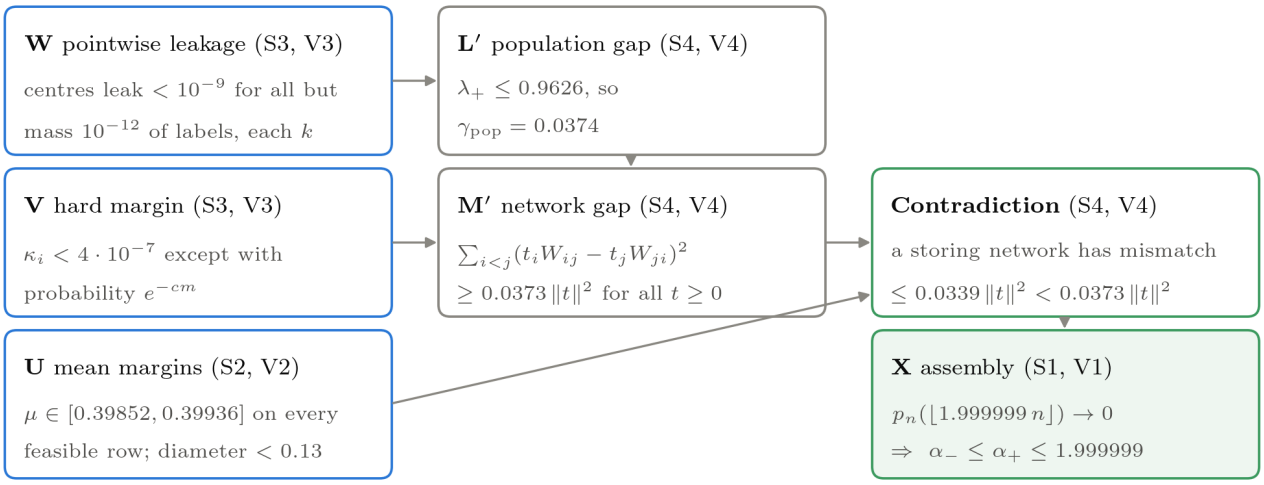


Figure 1. The checked argument for $\alpha_- \le \alpha_+ \le 1.999999$. Blue: new per-neuron estimates; grey: the round-4 transfer rerun at the new constants; green: the contradiction and the assembly. Each box names its solver and verifier.

The campaign

The user supplied *upper_bound_3* before round 6 and asked for four solvers. S1 assembled the round-5 argument into one node, then assembled the new one. S2, S3 and S4 each rebuilt one block of *upper_bound_3*, and a paired verifier checked each lane, rederiving every inference and recomputing every constant in independent exact arithmetic. All agents ran on Claude Opus 5.5. **Round 6 took 2.4 hours and cost 238.7 M tokens, about $114 at list prices** (Table 1).

Table 1. Round 6 by lane. Tokens count input, cache and output in every transcript; dollars use list prices.

Lane	Task	Round 6 result	Tokens	Cost
S1 ↔ V1	assemble both arguments	Theorems T, X	87.4 M	$41
S2 ↔ V2	*upper_bound_3* §§3.2–3.4, 4	Theorem U	30.2 M	$17
S3 ↔ V3	§5, §§6.1–6.2	Theorems V, W	34.6 M	$20
S4 ↔ V4	§§6.3–6.5	gaps 0.0374, 0.0373; contradiction	33.8 M	$19
Orchestrator	review, routing, ledger		52.8 M	$18

Round 6 results

Theorem T ($\alpha_- \le 2-10^{-29}$, assembled; S1-35)**.** $\limsup_n p_n(\lfloor(2-10^{-29})n\rfloor) \le 625\cdot 10^{12}\big(\varepsilon/(2-\varepsilon)+\bar{D}\big) < 3.2\cdot 10^{-15}$ with $\varepsilon = 10^{-29}$ and $\bar{D} < 3\cdot 10^{-162}$. Hence $\alpha_- \le 2-10^{-29}$.

Meaning. One node now carries the whole round-5 argument. Its 27 parents are proved, checked and unconditional. V1 rederived all 17 internal steps and reran 81 constants in exact and interval arithmetic. The node also settles the last convention: the normalization of degenerate rows commutes with permuting neurons.

Contribution to α_-. The first improvement of the upper bound. Theorem X supersedes it.

Theorem U (Mean margins near load 2; S2-52)**.** Let $A \in \{-1,1\}^{m\times d}$ have independent random signs with $m/d \to 1.999999$. There is $c>0$ such that for large m, with probability at least $1-e^{-cm}$, every unit w with $Aw \ge 0$ satisfies

$$0.39852 < \tfrac{1}{m}\textstyle\sum_\mu (Aw)_\mu < 0.39936.$$

Any two such w satisfy $\|w_1-w_2\|^2 \le 4(1-0.39852^2/0.39936^2) = 46543/2768896$, so the feasible unit set has diameter $\sqrt{46543/2768896} = 0.12965\ldots < 0.1297$.

Meaning. Each bound comes from an explicit dual witness, a tilted hinge loss whose expectation Gordon's comparison computes at the $1/m$ scale. The reserves are thin, $8.6\cdot 10^{-6}$ and $4.1\cdot 10^{-6}$, and $c \approx 5.7\cdot 10^{-23}$. The diameter follows because the direction of w_1+w_2 is again feasible. V2 re-decided all 18 numerical gates by its own routes.

Contribution to α_-. It replaces Theorem Q, which gave diameter 0.12 only at load $2-10^{-29}$.

Theorem V (Hard margin near load 2; S3-57)**.** For the same A, $\Pr\{\kappa(A) \ge 4\cdot 10^{-7}\} \le e^{-cm}$ with $c = 1.3\cdot 10^{-20}$, where $\kappa(A) = \max\{0, \max_{\|w\|=1}\min_\mu (Aw)_\mu\}$ is the hard margin.

Meaning. Near capacity no direction separates every pattern by more than $4\cdot 10^{-7}$.

Contribution to α_-. It bounds the leakage of Theorem W: with $\kappa < 4\cdot 10^{-7}$, the leakage is at most $625(\kappa^2+\bar{D}) \approx 10^{-10}$. Separately, Theorem P at $u' = 2-5\cdot 10^{-7}$, where $a > 1.06\cdot 10^{-10}$, keeps each regularized centre within $4.1\cdot 10^{-15}$ of the best row.

Theorem W (Pointwise leakage; S3-59)**.** Fix k and let X be the design of all but k neurons, with the bias row; P_X projects onto $\ker X$ and $u_X(y)$ is the centre for labels y. With probability tending to one over X and a uniform label y jointly, $\|(I-P_X)u_X(y)\|^2 < 1.0000001\cdot 10^{-10}$; and with probability tending to one over X, the labels with leakage above 10^{-9} have conditional mass at most 10^{-12}.

Meaning. Theorem R bounded the leakage in mean, which left a failure probability of $3.2\cdot 10^{-15}$ that never vanished. A pointwise bound removes it. S3 also proved that the mean cannot replace the pointwise bound: with two patterns the weighted leakage doubles, and Markov's inequality is sharp, so the mean route would need a mean leakage of 10^{-21}, while the available bound is about 10^{-10}.

Contribution to α_-. It raises the population gap from 0.032 to 0.0374.

Theorem X ($\alpha_- \le \alpha_+ \le 1.999999$; S1-37)**.** $p_n(\lfloor(2-10^{-6})n\rfloor) \to 0$. Hence $\alpha_- \le \alpha_+ \le 1.999999$.

Meaning. S4 rebuilt the closing sections. Good labels give the population form $\lambda_+ \le 0.9626$, so the gap is 0.0374; the round-4 transfer, rerun at the new constants, gives 0.0373 in the network; a storing network would have mismatch at most 0.0339. S1 assembled everything at the new parameters into one node with 25 checked, unconditional parents, and V1 rederived its 18 steps and 46 gates.

Contribution to α_-. The campaign's main result. It leaves 10^{23} times more room below 2 than Theorem T, the storage probability tends to zero rather than staying below $3\cdot 10^{-15}$, and it bounds α_+ as well.

Defects found in *upper_bound_3*. None is critical. Display (7) reads $< 0.13^2$, and the PDF's text layer prints it as 0.132. Section 3.2 asserts the tilted replacement, which S2 and S3 proved. Display (8) omits three hypotheses. Section 6.5 leaves the all-neuron operator-norm event implicit. Section 6.2 states the leakage bound for one exterior, where Section 6.4 needs every fixed k.

Plan for the close

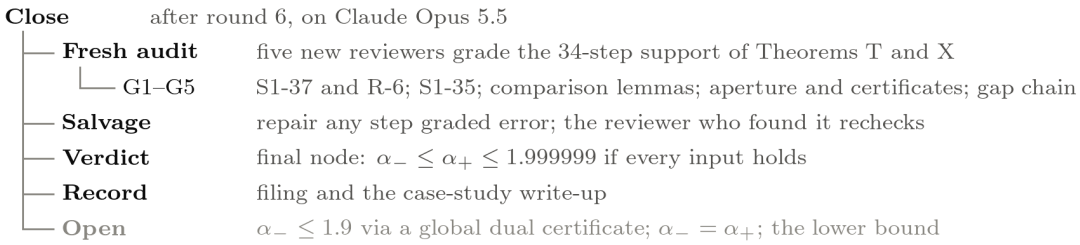


Figure 2. Closing plan. The grey line is not pursued in this campaign.

Fresh audit. Goal: an independent grade of every step behind the verdict before it is reported proved. The user scoped the audit to the dependency support of Theorems T and X, 34 steps and about 400 KB of proofs, plus the round-6 synthesis; five fresh reviewers grade it. The audit skill's default, one reviewer per 25 steps of the whole run, would have meant 17 reviewers. Each reviewer grades a step before reading the run's checks of it; a step graded *error* goes to salvage, and the reviewer who found it rechecks the repair. The rest of the run, including the own routes and the census, stays recorded as unaudited.

Verdict and record. Goal: a final node listing each input as held, salvaged or lost, then filing and the case-study write-up. The verdict is $\alpha_- \le \alpha_+ \le 1.999999$ if every input of Theorem X holds.

Open. The improvement is tiny because the argument needs every cone narrow enough that its mismatch, about $2\rho^2$, stays below the network gap, and cones are that narrow only next to load 2. *upper_bound_3* names the next target, a global dual certificate at load 1.9 (Appendix F). Endpoint equality $\alpha_- = \alpha_+$ and any lower improvement stay open.

Appendix F. A global dual certificate

Theorems T and X need each neuron's cone to be narrow, which happens only next to load 2. *upper_bound_3*, Section 7, proposes a certificate that sees the whole network at once.

Gordan's alternative. Stack every margin constraint of the network into one matrix $A(S)$ acting on the free parameters (J, h), with $N = nm$ rows. Either some (J, h) makes $A(S)(J, h) > 0$, or a nonzero $\lambda \geq 0$ satisfies $A(S)^\top \lambda = 0$ [Go1873]. The run's exact dual, Theorem B of round 1, is this alternative.

A scalar target. Put $V = \operatorname{range} A(S) \subset \mathbb{R}^N$ and

$$R(S) = \sup_{v \in V \setminus \{0\}} \frac{|\mathbf{1}^\top v|}{\|v\|_1} = \min_{A(S)^\top \lambda = 0} \|\lambda - \mathbf{1}\|_\infty .$$

The equality is linear-programming duality: the functional $v \mapsto \mathbf{1}^\top v$ on V has the same norm as its smallest ℓ_∞ extension. If $R(S) < 1$, the minimizer λ has every coordinate positive, so no network stores S.

The target theorem. Section 9 asks for $c > 0$ with

$$\liminf_n \Pr\{R(S) \leq 0.995\} \geq c \quad \text{at } m = \lfloor 1.9n \rfloor ,$$

which gives $\alpha_- \leq 1.9$. The document reports exact certificates $0.98502 < R(S) < 0.99008$ for one sample at $n = 256$, $m = 435$, load 1.699; this run has not re-certified them, and one sample is not a limit. The document also records a trap: bounding the denominator $\|v\|_1$ by the Gaussian mean absolute value fails for sparse symmetric rows, so a valid proof must couple the field's shape with the covariance alignment of J.

References

[Go1873] P. Gordan. Ueber die Auflösung linearer Gleichungen mit reellen Coefficienten. *Math. Ann.* 6 (1873), 23–28.

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

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

[UB3] User-supplied research continuation. A stronger explicit upper bound for symmetric Hopfield storage (November 9, 2026), unpublished; imported read-only with its SHA256 in the run's `sources/inherited/references-20261110/`.