A stronger explicit upper bound for symmetric Hopfield storage

Research continuation for Milton Lin

November 9, 2026

Abstract

For iid unbiased binary memories, arbitrary real symmetric zero-diagonal interactions, free biases, and simultaneous strict storage, the strengthened argument developed here gives

$$\mathbb{P}\{\text{storage at } m = \lfloor 1.999999\, n \rfloor\} \longrightarrow 0, \qquad \boxed{\alpha_- \leq 1.999999}.$$

The proof uses the elementary probability and dependent full-network transfer machinery of the supplied manuscript *Upper_bound.pdf*, whose relevant steps were independently audited, together with two new estimates: a tilted-hinge bound on the mean margin of every feasible unit row, and a shifted-hinge upper bound on the hard margin. These remove the main quantitative losses in the earlier $2-10^{-29}$ argument. The deficit from two increases by 10^{23}, but remains only 10^{-6} in absolute terms. This is a working research proof with independent mathematical and exact-arithmetic checks, not a published capacity theorem or a claim that the capacity is close to two. Separately, exact finite certificates were produced up to $n = 768$ near load 1.7; those support the global-dual research direction but do not prove an asymptotic bound at 1.7 or 1.9.

Contents

1 The problem and the completed result **2**

2 Critical assessment of the four reports **2**

3 Probability and transfer inputs retained from the manuscript **4**

4 New estimate I: a direct coefficient-space aperture **5**

5 New estimate II: a sharp hard-margin upper bound **7**

6 Closing the dependent full-network argument **8**

7 Executed global-dual strategy for a macroscopic bound **10**

8 What the finite executions actually established **12**

9 The next best steps, after this completed improvement **12**

10 Validation and limits **14**

1 The problem and the completed result

Let $S = (s_i^\mu) \in \{-1, 1\}^{n\times m}$ have independent uniform entries. Its columns are the memories. Strict storage means that there exist $J \in \mathbb{R}^{n\times n}$ and $h \in \mathbb{R}^n$ such that

$$J = J^\top, \quad J_{ii} = 0, \qquad s_i^\mu \left(\sum_{j\neq i} J_{ij} s_j^\mu + h_i \right) > 0 \quad \text{for every } i, \mu. \tag{1}$$

Write $p_n(m)$ for its probability and

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

The distinct-pattern model in the round reports has the same asymptotics: at $m = O(n)$, the probability of a repeated column is at most $\binom{m}{2}2^{-n} = o(1)$.

Theorem 1.1. *In the model above,*

$$p_n(\lfloor (2 - 10^{-6})n \rfloor) \longrightarrow 0.$$

Consequently $\alpha_- \leq 2 - 10^{-6} = 1.999999$. *If* $\alpha_+ := \inf\{u : p_n(\lfloor un \rfloor) \to 0\}$, *then also* $\alpha_+ \leq 1.999999$.

The proof is assembled in Section 6. Its inherited inputs are explicitly identified in Section 3; the new probability arguments appear in Sections 4–5. Thus the scope of the result is transparent: it strengthens the supplied manuscript's proof framework, including its dependent transfer theorem. It is not based on extrapolating the numerical experiments.

A smaller conclusion would suffice for the user's endpoint: any fixed u with $\limsup_n [1 - p_n(\lfloor un \rfloor)] > 0$ implies $\alpha_- \leq u$, by monotonicity in the number of patterns. A sharp-threshold theorem is unnecessary for this implication.

2 Critical assessment of the four reports

2.1 What should be retained

Report	Useful contribution	What it does not establish
Round 1	Exact Gordan dual; separation of individual-neuron feasibility from symmetry-only failure; projection-network limitation.	No fixed-load improvement of the upper bound. Repairing a particular construction is mainly a lower-bound direction.
Round 2	Exterior balancing and restricted certificates; extensive-support obstruction.	Exterior balance does not solve the coupled interior equations. Projection-correction failure is not optimal-network failure.
Round 3	Corrected threshold framework; exact-certificate checks; proposed tests of extensive restricted sets.	Exchangeability does not identify a fixed-subset probability with the probability that some subset fails. Sharpness does not locate the threshold.
Upper-bound manuscript	Positive-map reciprocity gap; regularized centers; dependence-preserving transfer; a serious working sub-two proof.	The selected-center statement at 1.9 is not an all-feasible-row result. Its previous aperture estimate is quantitatively very inefficient.

The exact dual is fundamental. Collect the $n(n-1)/2$ edge weights and n biases into θ, and let $\mathcal{A}(S)\theta$ be the nm margins. Gordan's alternative gives

$$\text{nonstorage} \iff \exists \lambda \geq 0,\ \lambda \neq 0 : \quad \sum_{\mu} \lambda_{i\mu} s_i^{\mu} = 0, \quad \sum_{\mu} (\lambda_{i\mu} + \lambda_{j\mu}) s_i^{\mu} s_j^{\mu} = 0 \quad (i < j). \tag{2}$$

The first equations include free biases; dropping them would change the problem.

Round 2's theorem that small active supports disappear at each fixed $u < 2$ is useful route selection. It does not imply storage: failure may be collective. Likewise, the corrected limitation on projection-plus-complement families rules out that family under its stated starting-set assumptions, not all symmetric weights.

2.2 The strongest audit conclusions

No decisive mathematical flaw was found in the checked positive-map, replacement, coercivity, projection-regularity, or full-network-transfer arguments of [U]. In particular:

1. The transfer does not replace dependent neuron rows by independent ones. It conditions on a common exterior, omits a fixed number of neurons, proves finite-feature response, then uses a separate uniform sampling theorem.
2. The scalar response factor cancels exactly when each internal coefficient row is normalized. A common exterior is used for the two-edge moment calculation, so the required conditional independence is legitimate.
3. The fourth-moment sampling theorem controls arbitrary nonnegative row scales. Its finite-rank approximation and moment-body argument address the possibility of a maximizing vector concentrating on few vertices.
4. The order of limits matters: regularization parameters are fixed first; a sufficiently large finite sample is fixed; network size tends to infinity; sampling tolerances are then reduced.

These checks support continuing the proof, rather than rejecting its tiny constant as evidence of a logical error. They are independent mathematical checks, not a formal proof-assistant certification.

Two interpretations need correction. First, if every fixed set of a given size has the same probability of failing, the event that *some* such set fails can still have a much larger probability: there are exponentially many choices. Second, the diameter obstruction in [U], Section 25, blocks its particular diameter-based sufficient comparison at 1.9; it does not exclude every possible comparison to an intelligently chosen center.

2.3 What the literature permits

Hillar–Sohl-Dickstein–Koepsell discuss the matching symmetric, hollow, free-threshold problem, record the lower bound one and Cover upper bound two, and report numerical behavior between roughly 1.5 and 1.7 while leaving exact capacity open [H]. Gardner–Gutfreund–Yekutieli's analytic calculation is for strong dilution [G]. Aguirre-López–Pastore–Franz explicitly identify the fully connected many-pattern problem as requiring further work because correlated loop contributions obstruct a straightforward application of the diluted analysis [A].

No published rigorous sub-two theorem for the exact model was located in this search. That is a search finding, not proof that none exists. Nardulli–Pasquariello's original 1991 paper was identified, but its complete derivation was not obtained for an assumption-by-assumption audit; the historical attribution

of capacity two is therefore not treated here as a checked theorem [N]. Results for Hebbian weights, nonzero diagonals, approximate retrieval, or independently trained rows do not settle (1).

3 Probability and transfer inputs retained from the manuscript

The new row arguments use the following elementary inputs from [U], Sections 12–20. They are stated here to make every dependency explicit. All constants are fixed before dimension tends to infinity.

3.1 Heavy coordinates

Let $A \in \{-1,1\}^{m\times d}$ be iid, $m/d \to u \in (1,2)$, and $k = \lceil m^{4/5} \rceil$. For any $q_0 > 1 - u/2$, with probability at least $1 - e^{-cm}$ every unit vector with $Aw \geq 0$ has squared mass less than q_0 on its k largest coordinates. This is Theorem 14.2 of [U]. Its proof retains a Rademacher head of dimension $o(m)$, replaces only small tail coefficients, and applies a clipped Gaussian auxiliary witness.

3.2 Replacement on deterministic domains

For a fixed head J of size k, let T_J be a deterministic compact subset of the unit sphere whose tail coordinates satisfy $\|w_{J^c}\|_\infty \leq k^{-1/2}$. For an empirical hinge loss, replacing the tail matrix entries by Gaussians changes the expected minimum over T_J by $o(1)$, uniformly in J and the domain. The same is true after a fixed translation of the hinge or addition of a fixed linear term.

For completeness, the extension is immediate from the proof: smooth the hinge at scale $m^{-1/40}$, use a deterministic m^{-3} net of logarithmic size $O(m \log m)$, and soften the minimum with inverse temperature $m^{41/40}$. Each replaced coefficient is $O(m^{-2/5})$. Summing the third-derivative replacement errors over $O(m^2)$ entries gives $O(m^{-3/20})$, and the soft-minimum error is $O(m^{-1/40} \log m)$. A translation leaves the derivative bounds unchanged; a linear tilt changes only the first derivative by a fixed constant. No data-dependent domain is used.

3.3 One-sided concentration

If f is K-Lipschitz and

$$F(A) = \min_{w\in T_J} \frac{1}{m} \sum_{\mu=1}^{m} f((Aw)_\mu),$$

the current minimizer gives, after a sign flip,

$$\left(F(A^{(\mu j)}) - F(A)\right)_+ \leq \frac{2K|w_j(A)|}{m}.$$

The sum of squared positive increments is at most $4K^2/m$. The one-sided entropy estimate of [U]therefore gives

$$\mathbb{P}\{F(A) \leq \mathbb{E}F(A) - s\} \leq \exp\left(-\frac{ms^2}{8K^2}\right). \tag{3}$$

Neither convexity of T_J nor nonnegativity of F is needed. There are only $\binom{d}{k} = \exp(o(m))$ heads, so a fixed positive expectation reserve survives their union.

3.4 Gaussian min–max comparison

Condition on the Rademacher head C. For tail coefficients b and a compact maximizing set of p, the expected primary minimum–maximum involving $p^\top Gb$ is at least its Gaussian auxiliary with

$$p^\top Gb \quad \rightsquigarrow \quad \|b\|\, g^\top p + \|p\|\, h^\top b,$$

up to the scalar-augmentation error $\mathbb{E}|\xi| \sup \|b\|\, \|p\|$. With the $1/m$ normalization this is $O(m^{-1/2})$ here. Convexity of the sets is not required for this one-sided expectation comparison. The Gaussian interpolation proof is supplied in [U], Lemma 12.1.

3.5 Positive selectors and the dependent network

The remaining imported inputs are positive-orthant coercivity at every fixed $u < 2$; the ordinary lower singular-value estimate $s_{\min}(X) \geq \sqrt{m}/25$ when $d/m \in [.49, .51]$; entrywise regularity of the projection onto $\ker X$; the positive-map inequality; and the full-center transfer theorem. Their exact uses are detailed in Section 6. They continue to apply at $u = 1.999999$: its limiting aspect ratio lies in the stated interval, its kernel projection has diagonal limit $1 - 1/u < 1/2$, and it remains strictly below two.

4 New estimate I: a direct coefficient-space aperture

Theorem 4.1. *Let A be iid Rademacher with $m/d \to u = 1.999999$. Except with probability e^{-cm} for sufficiently large m, every unit vector with $Aw \geq 0$ satisfies*

$$L \leq \mu_A(w) := \frac{1}{m} \mathbf{1}^\top Aw \leq U, \qquad L = .39852, \quad U = .39936. \tag{4}$$

Consequently the entire unit feasible cone has diameter less than .13.

4.1 Keep the head inside the comparison

Fix

$$q_0 = 10^{-6}, \quad k = \lceil m^{4/5} \rceil, \quad c_* = 1700, \quad M = 12, \quad t = 1/24000, \quad a_* = .7071069581.$$

Then $q_0 > 1 - u/2$, $tc_*M < 1$, and $a_*^2 > 1/u$. For each k-element head define

$$T_J = \{w : \|w\| = 1,\ \|w_J\|^2 \leq q_0,\ \|w_{J^c}\|_\infty \leq k^{-1/2}\}.$$

On the heavy-coordinate event every feasible unit vector lies in at least one of these deterministic compact sets. Define

$$f_+(z) = z_- + tz, \qquad f_-(z) = z_- - tz, \qquad F_{J,\pm}(A) = \min_{w \in T_J} \frac{1}{m} \sum_\mu f_\pm((Aw)_\mu).$$

For a feasible w the corresponding losses are $t\mu_A(w)$ and $-t\mu_A(w)$. It therefore suffices to prove $F_{J,+} > tL$ and $F_{J,-} > -tU$ simultaneously for every head.

Both losses are convex and $(1+t)$-Lipschitz. Their exact dual intervals are

$$f_+(z) = \max_{p \in [t-1, t]} pz, \qquad f_-(z) = \max_{p \in [-1-t, -t]} pz.$$

Replace only the tail entries by Gaussian variables. Write $w = (a, b)$ and $r = \|b\|$. The comparison and $h^\top b \geq -r\,\|h\|$ lower-bound the auxiliary by

$$\min_{\substack{\|a\|^2+r^2=1 \\ \|a\|^2\leq q_0,\ r\geq 0}} \max_{p\in I_\pm^m} \frac{1}{m}\left[p^\top Z - r\,\|h\|\,\|p\|\right], \quad Z = Ca + rg, \tag{5}$$

where $I_+ = [t-1, t]$ and $I_- = [-1-t, -t]$. Dropping the flat-tail restriction enlarges the minimizing domain, as required for a lower bound.

4.2 Uniform moments for the head plus Gaussian tail

For each fixed (a, r), a coordinate Z_μ is symmetric, variance one, and one-subgaussian. Put $\mu_0 = 1/\sqrt{2\pi}$. Then

$$\mathbb{E}Z_\mu = 0, \qquad \mathbb{E}(Z_\mu)_-^2 = \tfrac{1}{2}, \qquad \mu_0(1-q_0) \leq \mathbb{E}(Z_\mu)_- \leq \mu_0(1+q_0). \tag{6}$$

To prove the lower first-moment bound, condition on the Gaussian and use Jensen in the independent mean-zero head: $\mathbb{E}(rG + C_\mu a)_+ \geq r\mu_0 \geq (1-q_0)\mu_0$. For the upper bound, $f_r(x) = \mathbb{E}(rG + x)_+$ satisfies

$$f_r(0) = r\mu_0, \quad f_r'(0) = \tfrac{1}{2}, \quad 0 \leq f_r''(x) \leq \mu_0/r.$$

Taylor's upper bound and $\mathbb{E}(C_\mu a)^2 = 1 - r^2$ give $\mathbb{E}f_r(C_\mu a) \leq \mu_0/r \leq \mu_0(1+q_0)$.

Let $V_\mu = \min\{(Z_\mu)_-, M\}$. Integrating the subgaussian tail gives

$$\mathbb{E}V_\mu \geq \mu_0(1-q_0) - e^{-M^2/2}/M, \qquad \mathbb{E}V_\mu \leq \mu_0(1+q_0),$$
$$\mathbb{E}V_\mu^2 \leq \tfrac{1}{2}, \qquad \mathbb{E}[(Z_\mu)_- V_\mu] \geq \tfrac{1}{2} - 2e^{-M^2/2}.$$

These empirical averages, and the average of Z_μ, concentrate uniformly in (a, r) to any fixed positive tolerance. Indeed the parameter dimension is $k + 1 = o(m)$; fixed-resolution nets have size $\exp(o(m))$. The relevant functions are Lipschitz, and the only unbounded one, $z_- \min(z_-, M)$, grows at most linearly in a subgaussian variable. Bernstein or Hoeffding bounds at the net points, together with $\|[C, g]\|_{\mathrm{op}} = O(\sqrt{m})$, give uniform exponential concentration. Independently, $\|h\|/\sqrt{m} \leq a_*$ with exponentially high probability.

4.3 Two explicit dual witnesses

Choose

$$p_+ = t(\mathbf{1} - c_* V), \qquad p_- = t(-\mathbf{1} - c_* V).$$

They lie in their respective intervals because $tc_* M < 1$. Their normalized payoffs have first terms

$$\frac{p_\pm^\top Z}{mt} = \pm\frac{1}{m}\sum_\mu Z_\mu + c_*\frac{1}{m}\sum_\mu (Z_\mu)_- V_\mu,$$

and squared norms

$$\frac{\|p_\pm\|^2}{mt^2} = 1 \mp 2c_*\frac{1}{m}\sum_\mu V_\mu + c_*^2\frac{1}{m}\sum_\mu V_\mu^2.$$

Safe rational moment envelopes are

$$\mu_L = .3989418, \qquad \mu_U = .3989427, \qquad v_L = \tfrac{1}{2} - 10^{-20}.$$

Uniform concentration therefore gives limiting auxiliary lower bounds tB_+ and tB_-, where

$$B_+ = c_* v_L - a_* \sqrt{1 - 2c_*\mu_L + c_*^2/2} > .39852,$$
$$B_- = c_* v_L - a_* \sqrt{1 + 2c_*\mu_U + c_*^2/2} > -.39936.$$

Every displayed strict gate is checked with rational arithmetic. For example the first is equivalent to

$$a_*^2(1 - 2c_*\mu_L + c_*^2/2) < (c_* v_L - .39852)^2,$$

with positive square-root target. The enclosure for μ_0 uses Machin's identity and alternating arctangent series; the exponential tail is enclosed by a finite positive Taylor sum. Numerical floating-point agreement is not the certificate.

4.4 Expectations, including the negative auxiliary

The minus auxiliary can be negative, since $0 \notin I_-$. Its exceptional-event contribution cannot be discarded by nonnegativity. For either sign, the absolute auxiliary is bounded by

$$(1+t)\left(\frac{\|C\|_{\mathrm{op}}}{\sqrt{m}} + \frac{\|g\|}{\sqrt{m}} + \frac{\|h\|}{\sqrt{m}}\right),$$

which has uniformly bounded second moment. Cauchy–Schwarz therefore makes the exceptional-event contribution $o(1)$. This justifies the lower expectation bounds.

The scalar augmentation and partial replacement errors tend to zero uniformly in J. The strict inequalities for $B_\pm$ leave fixed positive reserves, so (3) and the union over $\exp(o(m))$ heads prove both desired inequalities with exponentially high probability. Intersect with the heavy-coordinate event to obtain (4).

4.5 A linear statistic gives the diameter directly

If w_1, w_2 are feasible unit vectors, their sum is nonzero because its mean margin is at least $2L > 0$. Its normalized direction is also feasible. Thus

$$2L \le \mu_A(w_1 + w_2) = \|w_1 + w_2\| \; \mu_A\left(\frac{w_1 + w_2}{\|w_1 + w_2\|}\right) \le U \, \|w_1 + w_2\| \, .$$

By the parallelogram identity,

$$\|w_1 - w_2\|^2 \le 4\left(1 - \frac{L^2}{U^2}\right) = \frac{46543}{2768896} = .01680922649\ldots < .13^2. \tag{7}$$

This is already a coefficient-space bound. The large coordinates were retained throughout; no condition-number conversion is needed. This completes the proof of Theorem 4.1.

5 New estimate II: a sharp hard-margin upper bound

Theorem 5.1. *For the same iid matrix and aspect ratio, put*

$$\kappa(A) = \max\left\{0, \max_{\|w\|=1} \min_{\mu} (Aw)_\mu\right\}.$$

For sufficiently large m, $\mathbb{P}\{\kappa(A) \ge 4 \cdot 10^{-7}\} \le e^{-cm}$.

Proof. Set $\kappa_0 = 4 \cdot 10^{-7}$ and use the same heads T_J. A vector with margin at least κ_0 is feasible, so the heavy-coordinate theorem applies. Consider

$$F_J(A) = \min_{w \in T_J} \frac{1}{m} \sum_{\mu} (\kappa_0 - (Aw)_\mu)_+ .$$

It is enough to show $F_J(A) > 0$ for all heads. A fixed translation preserves replacement and concentration. Its Gaussian auxiliary is bounded below by

$$\min_{(a,r)} \max_{p \in [0,1]^m} \frac{1}{m} \left[p^\top Z - r \|h\| \|p\| \right], \qquad Z = \kappa_0 \mathbf{1} - Ca + rg.$$

Write $Z_0 = -C_\mu a + rG$. It is symmetric with variance one, and $\mathbb{E}(Z_0)_+ \geq \mu_0 \sqrt{1 - q_0}$. Convexity in the shift implies

$$\mathbb{E}(Z_0 + \kappa_0)_+^2 \geq \tfrac{1}{2} + 2\kappa_0 \mu_0 \sqrt{1 - q_0}.$$

Clip at $M = 12$. The clipped-square loss is at most

$$2e^{-(M-\kappa_0)^2/2} \left(1 + \frac{\kappa_0}{M - \kappa_0} \right) < 10^{-20};$$

this deliberately loose bound suffices. Uniform concentration to tolerance 10^{-9} gives, for $b_m = m^{-1} \|\min(Z_+, M)\|^2$,

$$b_m \geq \tfrac{1}{2} + 2\kappa_0 \mu_0 \sqrt{1 - q_0} - 10^{-9} - 10^{-20}.$$

Also $\|h\|^2 / m \leq 1/u + 10^{-9}$ with exponentially high probability. Let $a_m = r \|h\| / \sqrt{m}$. Then uniformly

$$b_m - a_m^2 \geq \tfrac{1}{2} + 2\kappa_0 \mu_0 \sqrt{1 - q_0} - \frac{1}{u} - 2 \cdot 10^{-9} - 10^{-20} > 6 \cdot 10^{-8}.$$

The accompanying exact checker even replaces $\sqrt{1 - q_0}$ by $1 - q_0$ and retains a reserve exceeding $6.7153 \cdot 10^{-8}$.

Choose $p = \min(Z_+, M)/M$. Its payoff is at least

$$\frac{b_m - a_m \sqrt{b_m}}{M} = \frac{\sqrt{b_m}(b_m - a_m^2)}{M(\sqrt{b_m} + a_m)} \geq \frac{b_m - a_m^2}{2M} > 2.5 \cdot 10^{-9}.$$

Here the auxiliary is nonnegative even on exceptional events, since $p = 0$ is admissible. Gaussian comparison and replacement imply $\mathbb{E} F_J(A) \geq 2 \cdot 10^{-9}$ eventually, uniformly over the heads. Concentration and the $\exp(o(m))$ union establish the theorem. □

This new bound is linear in $2 - u$ before squaring. The manuscript's one-feature lift only supplied a squared-margin bound of order $2 - u$. That difference is decisive for retaining almost the entire population reciprocity gap at the new load.

6 Closing the dependent full-network argument

6.1 Regularized positive selectors

Let X contain the exterior neuron rows and the constant bias feature. For an independent label vector y, write $B = XD_y$, $D_y = \operatorname{diag}(y)$, and

$$\Delta_m = \{\pi \geq 0 : \mathbf{1}^\top \pi = 1\}, \qquad \kappa(B) = \min_{\pi \in \Delta_m} \|B\pi\| .$$

When positive, this equals the maximum unit hard margin: the closest convex-hull point $g^* = B\pi^*$ satisfies $B^\top g^* \geq \|g^*\|^2 \mathbf{1}$. Choose the unique positive optimizer

$$\pi^{\mathrm{reg}} = \arg \min_{\pi>0,\, \mathbf{1}^\top \pi = 1} \left\{ \|B\pi\|^2 + \eta m \|\pi\|^2 - \frac{\tau}{m} \sum_\mu \log(m\pi_\mu) \right\}.$$

Set $f_X(y) = D_y \pi^{\mathrm{reg}}$ and $u_X(y) = f_X(y)/\|f_X(y)\|$. Then $y_\mu u_X(y)_\mu > 0$ and $\|u_X(y)\| = 1$.

Positive coercivity may be applied at the slightly larger load $2 - 5 \cdot 10^{-7}$, allowing $a_0 = 10^{-12}$. This handles the harmless finite number of omitted features. Fix the mixing, ridge, and barrier parameters all equal to 10^{-80}. The manuscript's comparator calculation gives

$$U_{\mathrm{reg}} = 200/a_0^2 = 2 \cdot 10^{26}, \quad D < 3 \cdot 10^{-54}, \quad \kappa_{\min} > 7 \cdot 10^{-13}, \quad r_{\mathrm{reg}} := 2\sqrt{D}/\kappa_{\min} < 5 \cdot 10^{-15}.$$

Here D bounds excess squared residual, and r_{reg} bounds the distance of the regularized normalized row from the unit hard-margin center.

6.2 Pointwise leakage for almost all labels

The signed design XD_y is iid Rademacher in its joint marginal law, including the bias feature. Theorem 5.1 and $s_{\min}(X) \geq \sqrt{m}/25$ imply, with joint probability tending to one,

$$\ell_X(y) := \|(I - P_X)u_X(y)\|^2 \leq 625\big[(4 \cdot 10^{-7})^2 + D\big] < 10^{-9}, \tag{8}$$

where P_X projects onto $\ker X$. Define $p_X = \mathbb{P}_y\{\ell_X(y) > 10^{-9} \mid X\}$. Since $\mathbb{E}_X p_X \to 0$, Markov's inequality gives $\mathbb{P}_X\{p_X > 10^{-12}\} \to 0$.

This step uses a high-probability margin bound. Substituting a mean leakage bound directly into an inequality weighted by an arbitrary test function would be invalid. The conditional exceptional-label argument is essential.

6.3 From the positive-map inequality to a population gap

Projection regularity gives $\max_\mu (P_X)_{\mu\mu} \leq .500001$ and $\max_{\mu\neq\nu} |(P_X)_{\mu\nu}| \leq 10^{-16}$ with probability tending to one. For a nonnegative test function g on the label cube, define

$$M_g = \mathbb{E}[g(Y)u_X(Y)Y^\top], \qquad h_X(x,y) = (u_X(x)^\top y)(u_X(y)^\top x).$$

The checked positive-map inequality in [U], Theorem 3.1, states

$$\mathbb{E}g^2 - \mathrm{tr}(M_g^2) \geq \left(\sqrt{\tfrac{3}{80}\mathbb{E}g^2} - \|(I - P_X)M_g\|_F \right)_+^2.$$

Also $\langle g, T_{h_X} g \rangle = \mathrm{tr}(M_g^2)$ and $\mathbb{E}|h_X(X_1, X_2)|^4 \leq 105$.

If g is supported on the good labels of (8), Fourier contraction gives $\|(I - P_X)M_g\|_F^2 \leq 10^{-9}\mathbb{E}g^2$. The good-label loss therefore exceeds

$$(\sqrt{3/80} - \sqrt{10^{-9}})^2 = .03748775355\ldots > .03748.$$

With $p_X \leq 10^{-12}$, the fourth moment bounds the squared cross-block operator norm by $11 \cdot 10^{-6}$, and the bad-block norm by $4 \cdot 10^{-6}$. Bounding the whole positive-cone quadratic maximum by .9626 reduces to the positive determinant

$$.00008 \cdot .962596 - .000011 = .00006600768 > 0.$$

Thus a population reciprocity gap $\gamma_{\mathrm{pop}} = .0374$ holds on exterior cores with probability tending to one.

6.4 Transfer retains dependence and arbitrary row scales

Let W_i be each normalized internal regularized coefficient row, with $W_{ii} = 0$. The full-center transfer theorem of [U], Sections 21–24, applies because the regularizers are fixed, coercivity supplies all required moments, and every fixed omitted set leaves a good common core with probability tending to one. It gives

$$\inf_{t\geq 0,\, \|t\|=1} \sum_{i<j} (t_i W_{ij} - t_j W_{ji})^2 \geq .0373 \tag{9}$$

with probability tending to one.

Its mechanism is worth recording. A finite set of omitted neuron labels is independent conditional on its exterior. Smooth response to adding each feature gives its limiting kernel entries. The row normalization cancels the scalar response factor. A bounded fourth moment permits uniform sampling of the positive-cone quadratic maximum, which transfers the population gap back to the full matrix. None of these steps assert unconditional independence of full neuron rows.

6.5 The deterministic contradiction

The aperture and heavy-coordinate estimates hold simultaneously for all n row designs by their exponential tails. If a storing network existed, its unit full coefficient rows, including biases, would lie within .13 of the corresponding hard-margin centers. Internal normalization costs at most b_0^{-1} with $b_0 = .999$, because each bias coordinate has squared mass below q_0 and the regularization error is negligible.

Write $R_i = J_i / \|J_i\|$ and $t_i = \|J_i\|$. Both signs occur in each neuron row with probability tending to one, so strict feasibility forces $t_i > 0$. Symmetry gives $t_i R_{ij} = t_j R_{ji}$. Hence

$$\sum_{i<j} (t_i W_{ij} - t_j W_{ji})^2 \leq 2 \sum_i t_i^2 \|W_i - R_i\|^2 \leq \frac{2(.13 + r_{\mathrm{reg}})^2}{.999^2} \|t\|^2 .$$

But

$$\frac{2(.13 + 10^{-9})^2}{.999^2} = .03386770205\ldots < .0373,$$

contradicting (9). All required events have probability tending to one, proving Theorem 1.1.

7 Executed global-dual strategy for a macroscopic bound

The new near-two proof is useful progress, but a bound such as 1.9 still calls for a different global estimate. We executed the positive-dual route as well.

7.1 An exact scalar target

Let $V = \mathrm{range}(\mathcal{A}(S)) \subset \mathbb{R}^N$, $N = nm$, and define

$$R(S) = \sup_{v \in V \setminus \{0\}} \frac{|\mathbf{1}^\top v|}{\|v\|_1}.$$

Finite-dimensional convex duality gives

$$R(S) = \min_{\mathcal{A}^\top \lambda = 0} \|\lambda - \mathbf{1}\|_\infty . \tag{10}$$

Indeed, for any such λ, $\mathbf{1}^\top v = (\mathbf{1}-\lambda)^\top v$, so the right side bounds the left. Conversely, the norm of the linear functional $v \mapsto \mathbf{1}^\top v$ on V with its ℓ^1 norm equals its minimum ℓ^∞ extension norm, yielding equality. In particular, $R(S) < 1$ implies a strictly positive dual and nonstorage. The criterion is stronger than necessary: Gordan permits nonnegative duals with zeros.

7.2 A two-sided certifiable optimization

For $q \in (0,1)$ define, directly on V,

$$F_q(v) = \sum_{a=1}^{N} \left[q \log\cosh(v_a) + v_a \right].$$

Since $\log\cosh x \geq |x| - \log 2$,

$$F_q(v) \geq (q - R(S)) \left\| v \right\|_1 - qN \log 2.$$

If $q > R(S)$ the function is coercive on V and has a minimizer. Its stationarity is

$$\lambda = \mathbf{1} + q \tanh(v) \in V^\perp, \qquad 1 - q < \lambda_a < 1 + q.$$

If $q < R(S)$, a direction with negative total margin and ratio exceeding q makes F_q unbounded below along its ray. A finite rational or integer θ with

$$\frac{|\mathbf{1}^\top \mathcal{A}\theta|}{\left\| \mathcal{A}\theta \right\|_1} > q$$

certifies a lower bound on R. Thus both sides of a numerical bracket admit deterministic witnesses. Optimizer convergence flags are not used as proofs.

7.3 Exact repair of an approximate positive dual

Suppose $\widehat{\lambda} > 0$ and $r = \mathcal{A}^\top \widehat{\lambda}$. If $\mathcal{A}^\top \mathcal{A} \succeq gI$ with $g > 0$, then

$$\lambda = \widehat{\lambda} - \mathcal{A}(\mathcal{A}^\top \mathcal{A})^{-1} r \quad \text{satisfies} \quad \mathcal{A}^\top \lambda = 0.$$

The correction obeys the sharper bound

$$\left\| \lambda - \widehat{\lambda} \right\|_\infty \leq \left\| \lambda - \widehat{\lambda} \right\|_2 \leq \frac{\left\| r \right\|_2}{\sqrt{g}}. \tag{11}$$

The last inequality follows from $r^\top (\mathcal{A}^\top \mathcal{A})^{-1} r \leq \left\| r \right\|^2 / g$. Thus a floor exceeding this correction proves a positive dual exactly. The bound improves the looser row-norm estimate in the supplied manuscript.

No dense nm by $n(n+1)/2$ design matrix is needed. Matrix products use $S \odot (JS + h\mathbf{1}^\top)$ and its adjoint. To certify g, put $U = (S; \mathbf{1}^\top)$ and note

$$\left\| \mathcal{A}\theta \right\|^2 = \left\| JS + h\mathbf{1}^\top \right\|_F^2 \geq \lambda_{\min}(UU^\top)(\left\| J \right\|_F^2 + \left\| h \right\|^2) \geq \lambda_{\min}(UU^\top) \left\| \theta \right\|^2 .$$

An integer approximation to $(UU^\top)^{-1}$ and an exact residual norm provide a rational positive lower bound. All decisions are checked with arbitrary-precision integers or fractions; bounded integer matrix products are checked for overflow.

8 What the finite executions actually established

The following are individual, reproducible samples. They are not estimates of limiting storage probability.

n	m	Load m/n	Independently certified conclusion
256	384	1.50000000	Strict symmetric storage
256	409	1.59765625	Strict symmetric storage
128	217	1.69531250	Symmetry-only nonstorage
256	435	1.69921875	Symmetry-only nonstorage
256	460	1.79687500	Symmetry-only nonstorage
512	870	1.69921875	Nonstorage; individual-row control not claimed
768	1305	1.69921875	Nonstorage; individual-row control not claimed

For the symmetry-only rows, every neuron has an independently checked integer strict separator, while the full network has an exactly repairable positive dual. At $n = 256, m = 435$, independently checked witnesses give the conservative bracket

$$.9850169512647 < R(S) < .9900704938806.$$

Further projection polishing strengthens several upper bounds; the complete exact fractions are retained in the accompanying results. We use conservative independently verified bounds in this note.

The seeds for the four near-1.7 instances, in increasing n, are 20390810, 20518810, 20774810, and 21030810. The array generator, stored arrays, SHA-256 hashes, integer witnesses, inverse residual data, and two independent checkers are included. A failed dual search is recorded as unresolved until a separate primal or dual certificate decides it.

Restricted service was tested on the predetermined first neurons. In the $n = 128$ and 256 symmetry-only examples, the first $3n/4$ neurons are exactly served, whereas the first $7n/8$ admit a restricted nonstorage certificate. This does not refute a proposed asymptotic 3/4 theorem, but it shows that the chosen fraction is a substantive hypothesis and that full failure need not already appear there. The larger restricted fraction or full network is a more defensible initial proof target.

No confidence interval for an asymptotic transition is reported: there are too few independent samples, loads differ slightly after rounding, and no justified finite-size scaling law has been established. Exact certification resolves each finite instance, not the limiting quantifier.

9 The next best steps, after this completed improvement

9.1 Priority 1: prove a global ratio bound at a macroscopic load

A concrete sufficient theorem is

$$\exists c > 0 : \quad \liminf_{n\to\infty} \mathbb{P}\{R(S) \le .995\} \ge c, \qquad m = \lfloor 1.9n \rfloor. \tag{12}$$

It immediately implies $\alpha_- \le 1.9$. A fixed gap .005 is convenient for robust certification, not logically necessary: a nonnegative dual with nonvanishing probability would suffice. The new F_q formulation and exact repair turn candidate proofs into measurable inequalities instead of optimizer status claims.

The central inequality must retain the joint dependence between the covariance-aligned numerator and the absolute-field denominator:

$$\left|\operatorname{tr}(JSS^\top) + h^\top S\mathbf{1}\right| \le q \sum_{i,\mu} |s_i^\mu((JS)_{i\mu} + h_i)|\,,$$

uniformly in symmetric hollow J and arbitrary h. Separate crude numerator and denominator bounds lose the relevant structure.

A simple counterexample blocks replacing the denominator uniformly by the Gaussian mean absolute value. A symmetric cycle with two equal incident weights per neuron has

$$\frac{\mathbb{E}|\varepsilon_1 + \varepsilon_2|}{\sqrt{2}} = \frac{1}{\sqrt{2}} < \sqrt{\frac{2}{\pi}}.$$

The tempting resulting bound $\alpha_- \le \pi/2$ would therefore rest on a false uniform premise. A valid argument must couple field shape to covariance alignment or exclude such sparse rows for a justified reason.

9.2 Priority 2: a constructive positive dual with maximum-coordinate control

Prove that a specific candidate $\widehat{\lambda}(S)$ has a positive floor and a structured residual small enough for (11), with nonvanishing probability at $u = 1.9$. Direct control of the maximum-coordinate correction may improve substantially on the general Euclidean bound. The finite code already produces all quantities that such a theorem must estimate.

An exponential or log-cosh potential is a useful construction, but proving existence of its finite stationary point cannot be skipped: coercivity is precisely the missing global probability statement. Nor should one impose $YS^\top$ diagonal for $Y = S \odot \lambda$; that stronger condition decouples into individual-neuron obstructions and discards the skew-symmetric component needed below load two.

9.3 Priority 3: compare all row choices jointly

The alternative target is a positive lower bound on

$$\inf_{\substack{R_i \text{ feasible unit internal rows} \\ t \ge 0,\, \|t\|=1}} \sum_{i<j} (t_i R_{ij} - t_j R_{ji})^2,$$

with the bias constraints retained in the definition of feasible rows. This can permit much wider row cones than the current diameter argument. A result about only one selected center per neuron does not settle this infimum.

9.4 Work to deprioritize

Sharp-threshold proofs do not locate the upper endpoint and are not needed for the desired upper bound. Projection-family limitations do not constrain optimal networks. Fixed small certificates disappear at sub-two loads. Further optimization of the near-two decimals has diminishing value now that the main losses have been removed. The next substantive advance should attack (12) or its constructive equivalent.

10 Validation and limits

The new lemmas were derived and checked independently, including dual-interval signs, deterministic-domain uniformity, the negative-auxiliary expectation step, and the conditional bad-label argument. Two separately written arithmetic checks agree. The main script checks 22 gates with exact rational arithmetic. Numerical primal and dual certificates were independently checked without trusting the optimizer or importing its certificate routines.

The inherited full-network-transfer theorem remains a named dependency on the supplied research manuscript. Its relevant proof was audited, but it is not replaced by a peer-reviewed citation or formal proof-assistant output. The complete source manuscript is included as a reference in the reproducibility package so the chain can be examined.

The asymptotic dimension thresholds are not quantified and may be enormous. The new theorem does not predict feasibility at $n = 256$ or 768, does not identify a limiting threshold, and does not prove $\alpha_- \le 1.9$. It establishes a concrete stronger near-two theorem and supplies a tested global-dual method for the remaining macroscopic problem.

References

[U] Supplied manuscript, *Toward a capacity upper bound of 1.9: A self-contained improvement to* $\alpha_- \le 2 - 10^{-29}$, uploaded as `Upper_bound.pdf`. Main inherited inputs: Sections 2–4, 7, 12–24. Unpublished working proof.

[R1] Supplied *Symmetric Hopfield storage, lower endpoint: round 1 status*, November 9, 2026.

[R2] Supplied *Symmetric Hopfield storage: round 2*, November 9, 2026.

[R3] Supplied *Symmetric Hopfield storage, lower endpoint: round 3 status*, November 9, 2026.

[H] C. Hillar, J. Sohl-Dickstein, K. Koepsell, *Efficient and optimal binary Hopfield associative memory storage using minimum probability flow*, arXiv:1204.2916, 2015 revision. https://arxiv.org/abs/1204.2916.

[G] E. Gardner, H. Gutfreund, I. Yekutieli, *The phase space of interactions in neural networks with definite symmetry*, J. Phys. A 22 (1989), 1995–2008. https://doi.org/10.1088/0305-4470/22/12/005.

[A] F. Aguirre-López, M. Pastore, S. Franz, *Satisfiability transition in asymmetric neural networks*, J. Phys. A 55 (2022), 305001. https://arxiv.org/abs/2204.03577.

[N] G. Nardulli, G. Pasquariello, *Domains of attraction of neural networks at finite temperature*, J. Phys. A 24 (1991), 1103–1117. https://doi.org/10.1088/0305-4470/24/5/024. Bibliographic identity checked; full derivation not audited here.