Finite-Size Scaling for Dense Associative Memory Phase Boundaries on the N-Sphere: Metastability and the Limits of Linear Extrapolation

1 Introduction

We study the retrieval phase boundaries of Dense Associative Memory (DAM) models on the N-sphere $S^{N-1}(\sqrt{N})$ using GPU-accelerated Metropolis–Hastings Monte Carlo simulations. Two energy kernels are considered:

- **LSE** (Gaussian / log-sum-exp): $E(\mathbf{x}) = -\frac{1}{\beta_{\text{net}}} \log \sum_{\mu=1}^{P} e^{-\beta_{\text{net}}(N - \boldsymbol{\xi}^\mu \cdot \mathbf{x})}$,
- **LSR** (Epanechnikov / log-sum-ReLU, $b = 2 + \sqrt{2}$): $E(\mathbf{x}) = -\frac{N}{b} \log \sum_{\mu=1}^{P} \max\Big(0,\ 1 - b + \frac{b}{N} \boldsymbol{\xi}^\mu \cdot \mathbf{x}\Big)$.

Here $\boldsymbol{\xi}^1, \ldots, \boldsymbol{\xi}^P$ are random i.i.d. patterns on $S^{N-1}(\sqrt{N})$, and the number of patterns scales exponentially: $P = e^{\alpha N}$, where $\alpha = \ln P/N$ is the load parameter.

The goal is to map the order parameter $\varphi(\alpha, T)$ — the overlap with the target pattern — across the (α, T) plane, and compare the observed phase boundary with the theoretical prediction derived in the $N \to \infty$ limit.

2 Monte Carlo Setup

2.1 Adaptive-N approach

For each value of α, the number of neurons is determined by

$$N(\alpha) = \left\lfloor \frac{\ln P(\alpha)}{\alpha} \right\rceil, \tag{1}$$

where $P(\alpha)$ is interpolated linearly between $P_{\min}$ and $P_{\max}$ across the α grid ($\alpha \in [0.01, 0.55]$, step 0.01; $T \in [0.05, 2.50]$, step 0.05).

2.2 Simulation parameters

Each simulation uses:

- $N_{\text{eq}} = 5{,}000$ equilibration steps (no measurements),
- $N_{\text{samp}} = 5{,}000$ sampling steps with overlap accumulation,
- Adaptive step size $\sigma = \max(0.1,\ 2.4/\sqrt{N})$,
- Initialization near the target pattern: $\mathbf{x}_0 = \boldsymbol{\xi}^1 + 0.05\,\boldsymbol{\eta}$, projected onto $S^{N-1}(\sqrt{N})$.

The overlap is averaged over multiple independent trials (each with freshly generated random patterns).

3 Finite-Size Scaling: Approach and Motivation

3.1 The problem at finite N

The theoretical phase boundaries are derived in the thermodynamic limit $N \to \infty$. At any finite N, the phase transition is broadened:

- The overlap φ does not jump discontinuously at the critical point but changes gradually over a region of width $\sim 1/\sqrt{N}$.
- The apparent critical temperature $T_c^{\text{eff}}(N)$ differs from the theoretical $T_c(\infty)$.

Since $N = \ln P/\alpha$, at a given α one can increase N only by increasing P. Even increasing P by orders of magnitude yields only modest gains in N: for instance, at $\alpha = 0.10$, going from $P = 1{,}000$ to $P = 300{,}000$ changes N from 69 to 110 (a factor of 1.6×).

3.2 Three P-scales

We ran Monte Carlo sweeps at three progressively larger P-scales:

	P range	Trials	N at $\alpha = 0.01$	N at $\alpha = 0.55$
Scale 1	200 – 5,000	256	530	15
Scale 2	2,000 – 50,000	256	760	20
Scale 3	10,000 – 300,000	64	921	23

Table 1: Parameters for the three P-scales used in finite-size scaling.

3.3 Linear extrapolation to $N \to \infty$

For each grid point (α, T), we have three values of φ at three different N values. The finite-size scaling ansatz assumes

$$\varphi(N) = \varphi_\infty + \frac{c}{N} + O(1/N^2), \tag{2}$$

where φ_∞ is the thermodynamic-limit value. We perform a least-squares linear fit of φ vs. $1/N$ and extrapolate to $1/N = 0$.

4 Results Across Scales

4.1 Scale 1: $P = 200$–$5{,}000$ (maps1.png)

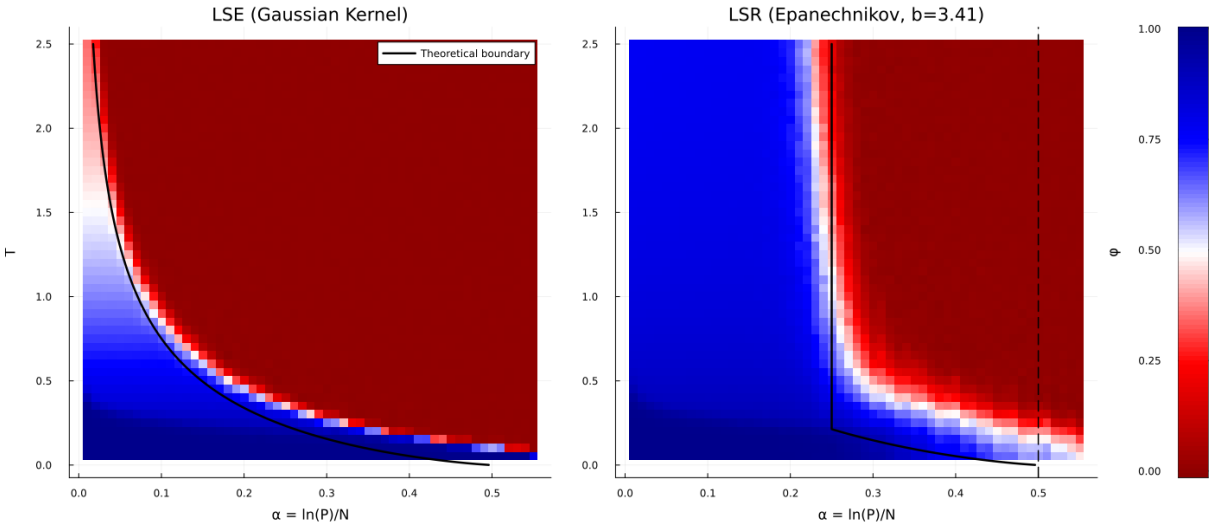

Figure 1: Scale 1 heatmaps. Left: LSE kernel. Right: LSR kernel ($b = 2 + \sqrt{2}$). Black curves show the theoretical phase boundaries. The transition is broad, especially at large α where N is small ($N \leq 31$). Despite the broadening, the apparent boundary roughly follows the theoretical curve.

At Scale 1, N ranges from 530 (at $\alpha = 0.01$) to 15 (at $\alpha = 0.55$). The transition is visibly broadened but the phase boundary location is reasonable. At $\alpha = 0.10$, $N = 69$, and the transition occurs at $T \approx 0.5$–0.8, close to the theoretical $T_c \approx 0.75$.

4.2 Scale 2: $P = 2{,}000$–$50{,}000$ (maps2.png)

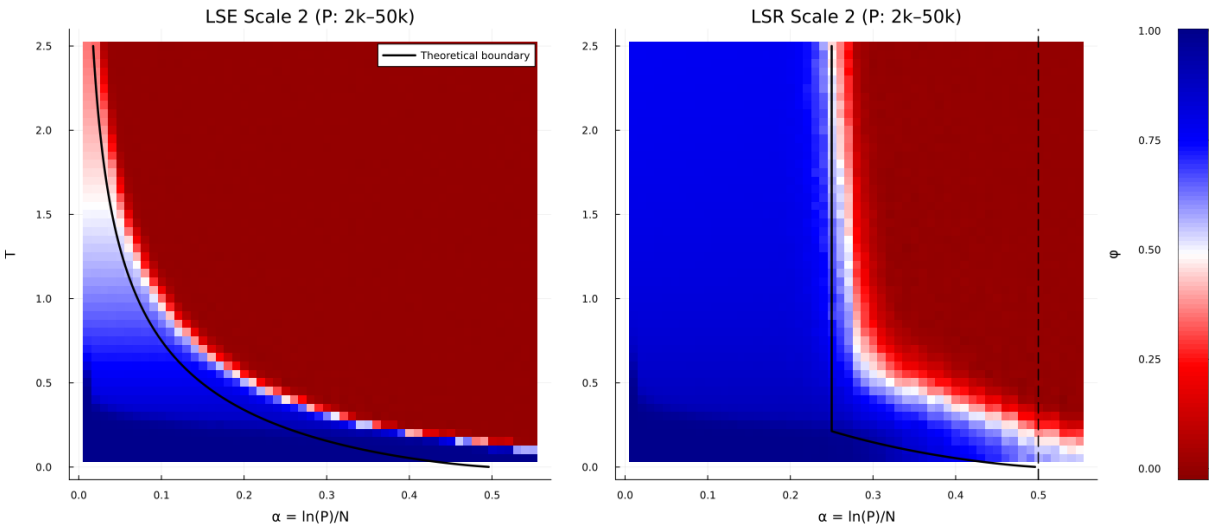

Figure 2: Scale 2 heatmaps. The transition has sharpened somewhat compared to Scale 1, but the apparent boundary has *shifted upward* in T. The retrieval region (red/warm colors) extends to higher temperatures than predicted by theory.

At Scale 2, N increases to 760 at $\alpha = 0.01$ and 40 at $\alpha = 0.25$. The transition is sharper, but the apparent critical temperature T_c^{eff} is *higher* than at Scale 1. For example, at $\alpha = 0.10$:

$\varphi(T = 1.0) = 0.33$ (Scale 2) vs. 0.065 (Scale 1), even though $T = 1.0$ is above the theoretical $T_c = 0.75$.

4.3 Scale 3: $P = 10{,}000$–$300{,}000$ (maps3.png)

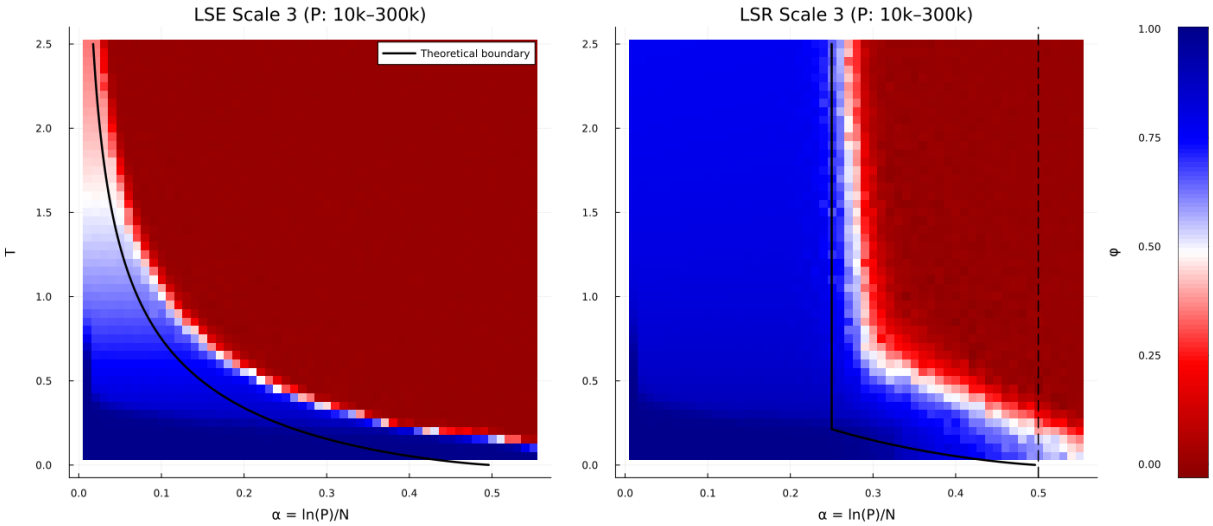

Figure 3: Scale 3 heatmaps. The boundary has shifted even further from the theoretical curve. The retrieval region extends well above T_c.

Scale 3 has the largest N values (921 at $\alpha = 0.01$, 47 at $\alpha = 0.25$), yet the match with theory is *worse*, not better. At $\alpha = 0.10$, $T = 1.0$: $\varphi = 0.49$ (Scale 3), indicating the system appears to be in the retrieval phase at a temperature well above $T_c = 0.75$.

4.4 Extrapolated map (maps.png)

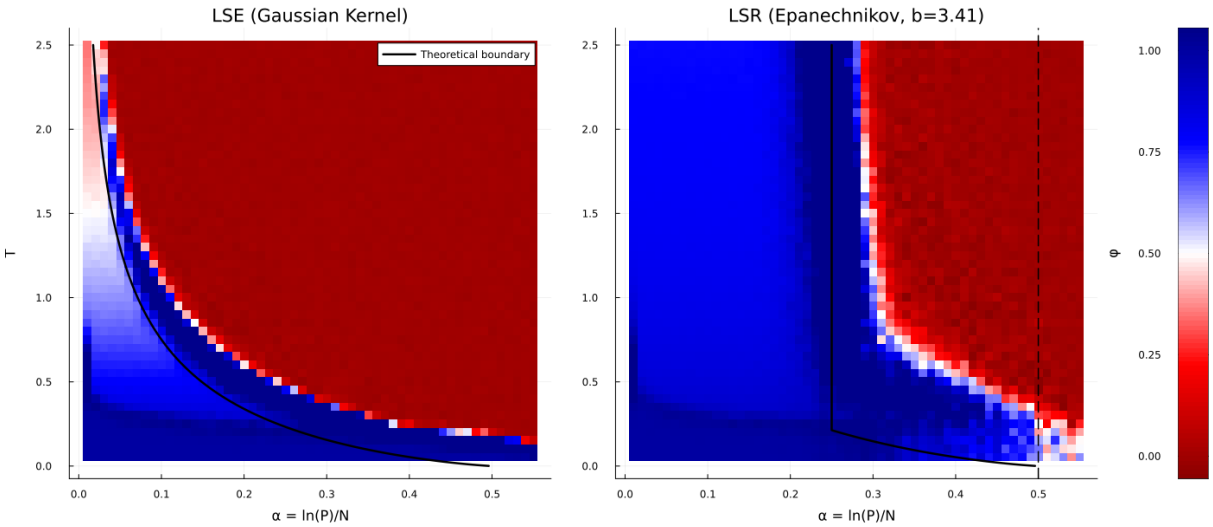

Figure 4: Extrapolated heatmaps ($N \to \infty$ via linear fit). The extrapolation amplifies the metastability artifact, producing unphysical $\varphi > 1$ values (clamped to 1.05) in the critical region. 234 LSE points and 473 LSR points exceed $\varphi = 1.0$.

The linear extrapolation produces 234 unphysical points ($\varphi > 1$) for LSE and 473 for LSR, concentrated in the critical region just above the theoretical boundary.

5 Diagnosis: Metastability and Critical Slowing Down

5.1 The first-order nature of the transition

The retrieval–disordered phase transition in DAM models on the N-sphere is *first order*: the order parameter φ jumps discontinuously at the critical point in the thermodynamic limit. This means the free energy landscape has two distinct minima (retrieval state with $\varphi > 0$ and disordered state with $\varphi \approx 0$) separated by a *free energy barrier* ΔF.

For a first-order transition, this barrier grows with system size:

$$\Delta F \sim N^{\gamma}, \quad \gamma > 0. \tag{3}$$

The Kramers escape time from the metastable state scales as

$$\tau_{\text{escape}} \sim e^{\Delta F/T} \sim e^{N^{\gamma}/T}, \tag{4}$$

which grows *exponentially* with N.

5.2 Initialization bias

Our Monte Carlo chains are initialized near the target pattern: $\mathbf{x}_0 \approx \boldsymbol{\xi}^1$. This places the chain squarely in the retrieval state. If the equilibrium state is actually the disordered phase (above T_c), the chain must cross the free energy barrier to reach it.

- At **small** N (Scale 1): The barrier is low. With $N_{\text{eq}} = 5{,}000$ steps, the chain can cross the barrier and equilibrate to the disordered state. Result: $\varphi \approx 0$ above T_c (correct).
- At **larger** N (Scales 2, 3): The barrier is higher. With the same $N_{\text{eq}} = 5{,}000$ steps, the chain *cannot* escape the metastable retrieval state. Result: φ remains elevated above T_c (incorrect — metastability artifact).

5.3 Quantitative evidence

Table 2 shows φ across scales at points that theory predicts to be in the disordered phase ($T > T_c$).

α	T	T_c (theory)	Scale 1	Scale 2	Scale 3	Extrap.
0.05	1.50	1.30	0.228	0.423	0.489	0.908
0.10	0.80	0.75	0.601	0.672	0.678	0.824
0.10	1.00	0.75	0.065	0.328	0.491	>1.05
Correctly equilibrated points (far above T_c):						
0.10	1.50	0.75	−0.003	0.002	0.001	0.011
0.10	2.50	0.75	0.000	−0.001	−0.001	−0.002

Table 2: Overlap φ at points above the theoretical T_c. In the critical region (top rows), φ *increases* with N, opposite to the expected finite-size scaling. Far above T_c (bottom rows), the chain equilibrates correctly.

The pattern is clear: φ monotonically *increases* with N in the critical region just above T_c, because the larger system is more deeply trapped in the metastable retrieval state. The linear extrapolation (2) then projects this increasing trend to $1/N = 0$, producing the unphysical $\varphi > 1$.

5.4 Why the boundary shifts upward

The visual effect on the heatmap is that the apparent phase boundary shifts to higher T with increasing N. This is the opposite of what standard finite-size scaling predicts (boundary converging toward theory). The mechanism is:

1. In the **retrieval phase** (below T_c): φ is high and nearly N-independent. Correct.
2. In the **critical region** ($T \lesssim T_c + \Delta T$): the MC chain remains trapped in the retrieval state at larger N. The apparent φ stays high $\Rightarrow$ boundary appears at higher T.
3. **Deep in the disordered phase** ($T \gg T_c$): thermal fluctuations are strong enough to kick the chain out of the metastable state even at large N. $\varphi \approx 0$ for all scales. Correct.

The width of the "metastability band" above T_c where the MC fails to equilibrate *grows* with N, explaining the worsening match with theory.

6 Why Linear Extrapolation Fails

The finite-size scaling ansatz $\varphi(N) = \varphi_\infty + c/N$ assumes that the MC correctly samples the equilibrium Gibbs distribution at each N. When metastability violates this assumption, the measured $\varphi(N)$ is *not the equilibrium value* but a metastable artifact:

$$\varphi_{\text{measured}}(N) = \varphi_{\text{eq}}(N) + \Delta\varphi_{\text{meta}}(N), \tag{5}$$

where $\Delta\varphi_{\text{meta}}(N)$ increases with N in the critical region. The linear fit then extrapolates the sum, and since $\Delta\varphi_{\text{meta}}$ grows with N (decreases with $1/N$), the extrapolated φ_∞ overshoots.

The severity depends on the region:

- **Deep retrieval:** $\Delta\varphi_{\text{meta}} \approx 0$ (system is in correct state). Extrapolation valid.
- **Critical region:** $\Delta\varphi_{\text{meta}}$ dominates. Extrapolation fails catastrophically ($\varphi > 1$).
- **Deep disordered:** $\Delta\varphi_{\text{meta}} \approx 0$ (thermal fluctuations overcome barrier). Extrapolation valid.

7 Possible Remedies

7.1 Parallel tempering (replica exchange)

Run multiple replicas at different temperatures simultaneously. Allow replicas to exchange configurations, enabling the system to escape metastable states by "borrowing" thermal energy from higher-temperature replicas. This is the most robust approach for first-order transitions.

7.2 Dual-initialization protocol

For each (α, T) point, run the MC from two starting conditions: (a) near the target pattern (testing retrieval stability) and (b) from a random point on S^{N-1} (testing whether the system spontaneously finds the retrieval state). If the two give different φ, the system is in the metastable/hysteresis region and N_{eq} is insufficient.

7.3 Binder cumulant analysis

Instead of extrapolating φ directly, compute the Binder cumulant $U_4 = 1 - \langle\varphi^4\rangle / (3\langle\varphi^2\rangle^2)$ at each scale. The crossing point of $U_4(T)$ curves for different N gives T_c without requiring equilibration in the correct phase.

7.4 Adaptive equilibration

Scale N_{eq} with N, e.g., $N_{\text{eq}} \propto N^2$ or even exponentially. This is computationally expensive but addresses the root cause.

7.5 Simulated annealing initialization

Instead of starting near the target, start from a random configuration and slowly cool the system (decrease T) in a pre-equilibration phase. This avoids the initialization bias entirely, but may fail to find the retrieval state at low T if the basin of attraction is narrow.

8 Conclusions

The finite-size scaling approach—running MC simulations at three P-scales and linearly extrapolating φ vs. $1/N$—was designed to remove finite-N artifacts and converge to the theoretical $N \to \infty$ phase boundary. However, the approach fails in the critical region due to **metastability**:

1. The retrieval–disordered transition is first order, with a free energy barrier that grows with N.
2. MC chains initialized near the target pattern become exponentially harder to equilibrate at larger N when $T > T_c$.
3. The measured φ in the critical region increases with N (opposite to equilibrium finite-size scaling), and the linear extrapolation amplifies this artifact.

The result is that larger P-scales produce *worse* agreement with theory near the phase boundary, and the extrapolated map contains unphysical values ($\varphi > 1$). Addressing this requires fundamentally different MC strategies (parallel tempering, dual initialization) rather than simply increasing system size.