Geometric Entropy and Retrieval Phase Transitions in Continuous Dense Associative Memory

Tatiana Petrova   Evgeny Polyachenko   Radu State

Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg • tatiana.petrova@uni.lu

Interdisciplinary Centre for Security, Reliability and Trust

ICML International Conference On Machine Learning

ICML 2026 • July 7–11, 2026 • COEX, Seoul, South Korea

Why finite-temperature associative memory?

- **Dense Associative Memory** (modern Hopfield nets) stores an exponential number of patterns, $M = e^{\alpha N}$.
- It is mathematically equivalent to self-attention in Transformers — a tractable theory for attention-style memory.
- Existing theory is essentially zero-temperature.

Our question

How does retrieval survive **noise** and **thermal fluctuations**, and what is set by the **kernel** vs. by the **geometry**?

Setup: two kernels on the N-sphere

States and patterns live on $\sum_i x_i^2 = N$; **alignment** $\phi = \frac{1}{N}\mathbf{x}\cdot\boldsymbol{\xi}$ ($\phi \approx 1$ on a pattern).

Gaussian — Log-Sum-Exp (LSE)

$$H_{\mathrm{LSE}} \propto -\frac{1}{\beta_{\mathrm{net}}} \ln \sum_{\mu} e^{-\frac{\beta_{\mathrm{net}}}{2} d^2(\mathbf{x}, \boldsymbol{\xi}^{\mu})}$$

global support

Epanechnikov — Log-Sum-ReLU (LSR)

$$H_{\mathrm{LSR}} \propto -\frac{1}{\beta_{\mathrm{net}}} \ln \sum_{\mu} \max\left[\epsilon, 1 - \frac{d^2}{2\sigma^2}\right]$$

finite support

$\beta_{\mathrm{net}} = 1/\sigma^2$ sets *kernel sharpness* — not the thermal bath.

Free energy = energy − temperature × entropy

Replica computation of the disorder-averaged free energy gives

$$\langle f \rangle \approx \underbrace{u(\phi)}_{\text{kernel}} - T \underbrace{s(\phi)}_{\text{geometry}}$$

Kernel-independent geometric entropy

$$s(\phi) = \frac{1}{2} \ln\left(1 - \phi^2\right)$$

Depends *only* on the spherical geometry — never on the kernel. High alignment is entropically *expensive* in high dimensions.

$$u_{\mathrm{LSE}} = 1 - \phi, \qquad u_{\mathrm{LSR}} = -b^{-1} \ln[1 - b(1 - \phi)], \quad b = N\beta_{\mathrm{net}}$$

Crosstalk, the noise floor, and stability

- At exponential load the Gaussian-crosstalk picture fails: random states form a noise floor (Random Energy Model), max spurious alignment $\phi_{\text{max}} = \sqrt{2\alpha}$.

Retrieval is thermodynamically stable iff

$$\min_{\phi, q} \left[u(\phi) - T\, s(\phi, q) \right] \le u_{\text{noise}}$$

As $T\uparrow$ entropy lifts the retrieval free energy; as $\alpha\uparrow$ the noise floor drops $\Rightarrow$ a sharp first-order retrieval transition.

The two kernels behave qualitatively differently

LSE (Gaussian)

$$\alpha_c^{\mathrm{LSE}}(T) = \frac{1}{2}[1 - f_{\mathrm{ret}}(T)]^2$$

A finite critical temperature at every load: interference is always present.

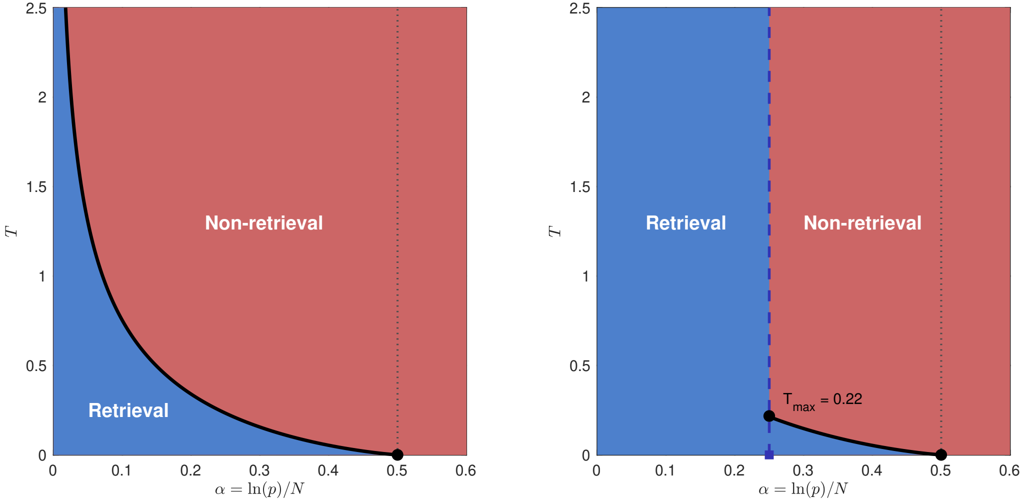
LSR (Epanechnikov)

$$\alpha_{\mathrm{th}}(b) = \frac{1}{2}\left(1 - b^{-1}\right)^2$$

Below α_{th} the basin is isolated — **perfect retrieval at any** T. No analog in LSE.

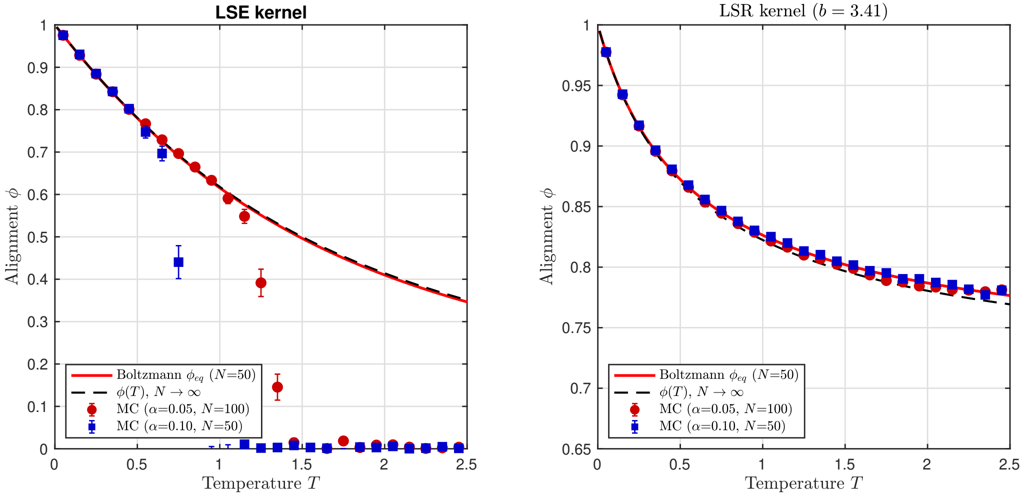
Both share the zero-temperature capacity $\alpha_c(0) = \frac{1}{2}$ — exponential storage is a property of the **geometry**.

Phase diagrams in the (α, T) plane



LSE (left): critical line $\alpha_c(T)$ at all loads. **LSR** (right): perfect retrieval below α_{th}=0.25 (b=3.41);

Monte Carlo validation



Metropolis sampling on the N-sphere (N=50, 100; M=148 patterns) matches the thermodynamic-limit

Takeaways

- **Geometry, not the kernel**, fixes the entropy and the zero-T capacity $\alpha_c(0) = \frac{1}{2}$.
- **LSE:** thermal robustness at all loads, but ever-present interference $\Rightarrow$ finite T_c.
- **LSR:** a support threshold eliminates interference — perfect retrieval at *any* temperature.
- Kernel choice is a **robustness vs. capacity** trade-off for attention-like memory.

Thanks! • tatiana.petrova@uni.lu • SnT, University of Luxembourg