Phase Kernel Lifts Capacity of Dense Associative Memory

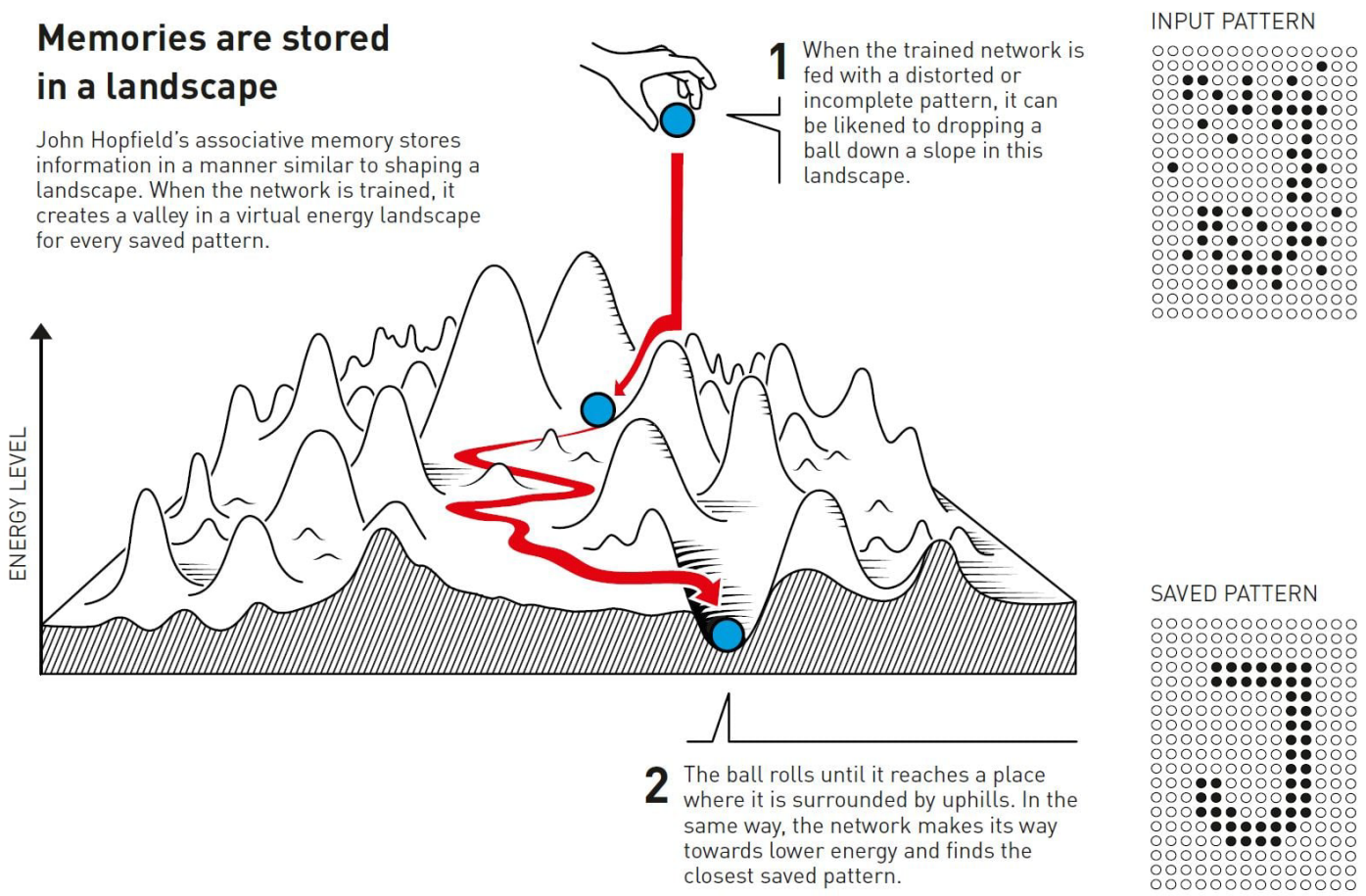
Yifei Zhao,
Ying Tang

Paper

Lab

Institute of Fundamental and Frontier Sciences
University of Electronic Science and Technology of China
Chengdu 611731, China

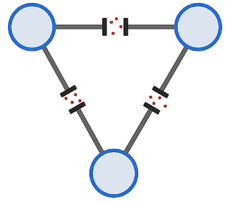
Hopfield’s 1982 idea: memory as a stable state



THE HOPFIELD IDEA

Recall is a process.

Recurrent links shape stable states.



PAIRWISE LINKS

01	**Store**	Stored patterns become stable network states.
02	**Cue**	A partial or noisy cue starts the updates.
03	**Settle**	Updates settle at a stable attractor.

Stable attractors may be spurious.

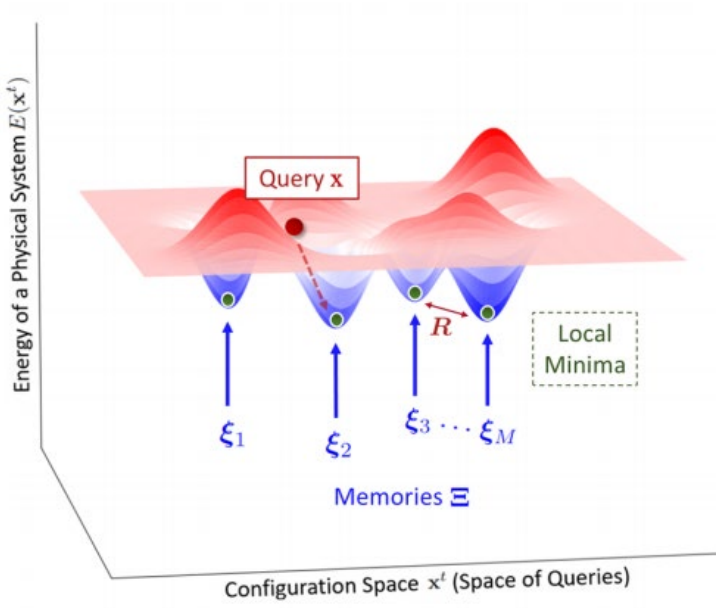
Landscape: © Johan Jarnestad/The Royal Swedish Academy of Sciences, NobelPrize.org (Physics 2024). Pairwise sketch: supplied report, slide 3.

From attractors to attention: the evolution of memory retrieval

1982 — **PAIRWISE ATTRACTORS**

2016–17 — **DENSE INTERACTIONS**

2021 — **SOFTMAX RETRIEVAL**



Landscape schematic · report slide 4

Hopfield Energy	=	New Energy	⇒	Update Rule	=	Transformer
$-\exp\left(\mathrm{lse}\left(1, \boldsymbol{\xi}^T \boldsymbol{X}\right)\right)$		$-\mathrm{lse}(\beta, \boldsymbol{\xi}^T \boldsymbol{X}) + \frac{1}{2}\boldsymbol{\xi}^T \boldsymbol{\xi} + c$		$\mathrm{softmax}\left(\beta\ \boldsymbol{\xi}^T \boldsymbol{X}\right) \boldsymbol{X}^T$		$\mathrm{softmax}\left(\frac{1}{\sqrt{d_k}}\ \boldsymbol{Q}\boldsymbol{K}^T\right) \boldsymbol{V}$

Energy → update rule → attention

Modern Hopfield retrieval

$$q_{new} = X\ soft\max(\beta X^T q)$$

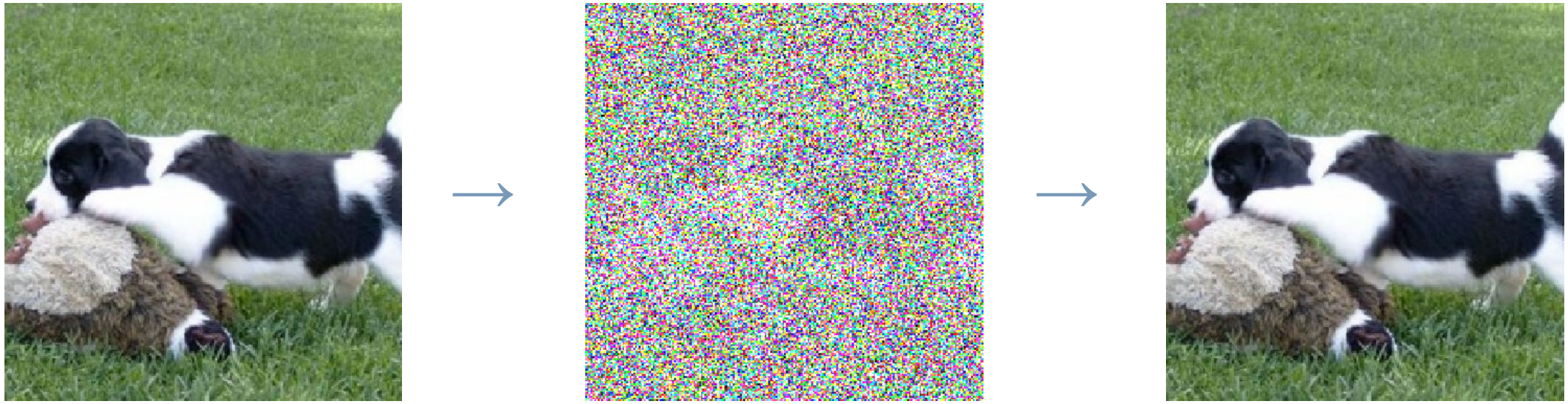
X holds memories as columns; β sets retrieval selectivity.

This attention equivalence does not guarantee convergence for every Transformer layer.

Krotov & Hopfield, NeurIPS (2016); Demircigil et al., J Stat Phys (2017); Ramsauer et al., ICLR (2021). Figures: supplied report, slides 3–4.

Associative memory retrieves stored content from noisy cues

The task is content-addressable retrieval: recover an existing memory without its index.



Stored image → **Noisy query** → **Retrieved state**

Two questions: how many patterns remain retrievable, and how robust is retrieval to noise?

Hopfield JJ. PNAS 79, 2554–2558 (1982). doi:10.1073/pnas.79.8.2554.
Illustration: current manuscript, Figure 2 (one unchanged example; separate demonstration setting).

1982–2016: attractor memories become nonlinear and dense

1982

PNAS

J. J. Hopfield

Neural Networks and Physical Systems with Emergent Collective Computational Abilities

Energy minimization turns stored patterns into attractors.

2004

Physical Review Letters

T. Nishikawa, Y.-C. Lai & F. C. Hoppensteadt

Capacity of Oscillatory Associative-Memory Networks with Error-Free Retrieval

Phase coupling provides a route to stable associative recall.

2016

NeurIPS

D. Krotov & J. J. Hopfield

Dense Associative Memory for Pattern Recognition

Higher-order interactions increase storage capability.

Hopfield, PNAS 79:2554–2558 (1982); Nishikawa et al., PRL 92:108101 (2004); Krotov & Hopfield, NeurIPS 29 (2016).

2017–2024: exponential capacity, attention and memory separation

2017

Journal of Statistical Physics

M. Demircigil et al.

On a Model of Associative Memory with Huge Storage Capacity

Exponential interactions support exponential storage capacity.

2021

ICLR

H. Ramsauer et al.

Hopfield Networks is All You Need

Continuous retrieval connects associative memory to attention.

2024

ICML

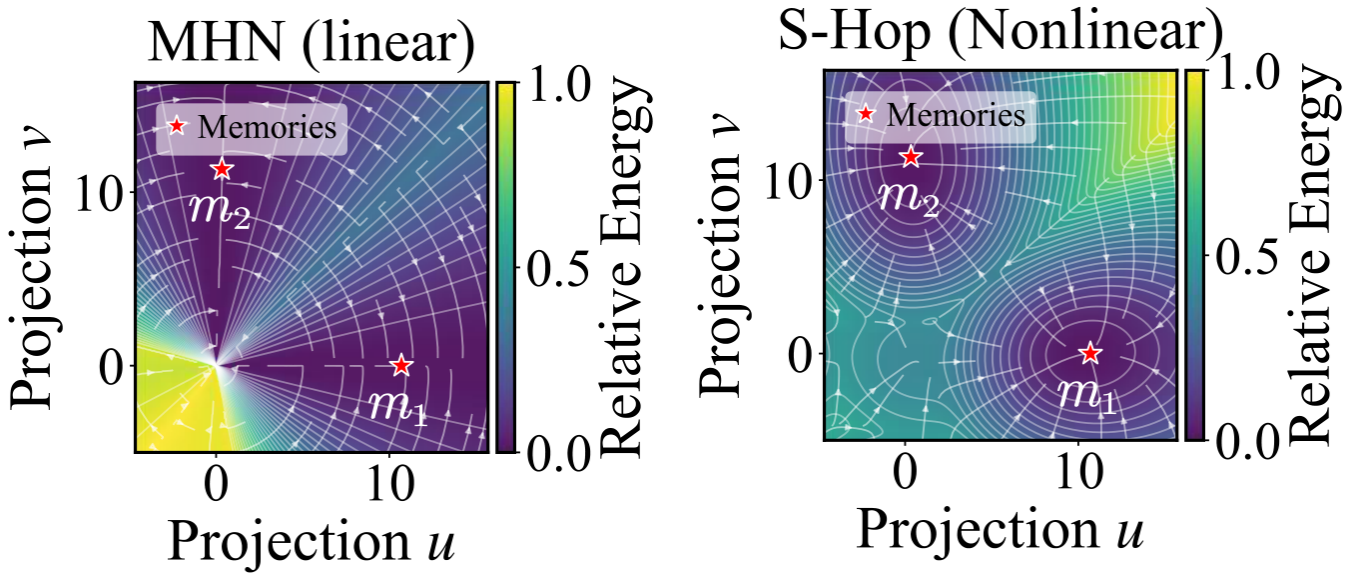
D. Wu, J. Y.-C. Hu, T.-Y. Hsiao & H. Liu

Uniform Memory Retrieval with Larger Capacity for Modern Hopfield Models

U-Hop learns a feature map to improve memory separation.

Demircigil et al., J Stat Phys 168:288–299 (2017); Ramsauer et al., ICLR (2021); Wu et al., ICML 235:53471–53514 (2024).

The open problem: high capacity with reliable noisy retrieval



Illustrative contours; color scales are independently normalized.

Existing route: U-Hop

Learn a projection that separates memories.

This work: phase kernels

Change both similarity and state dynamics.

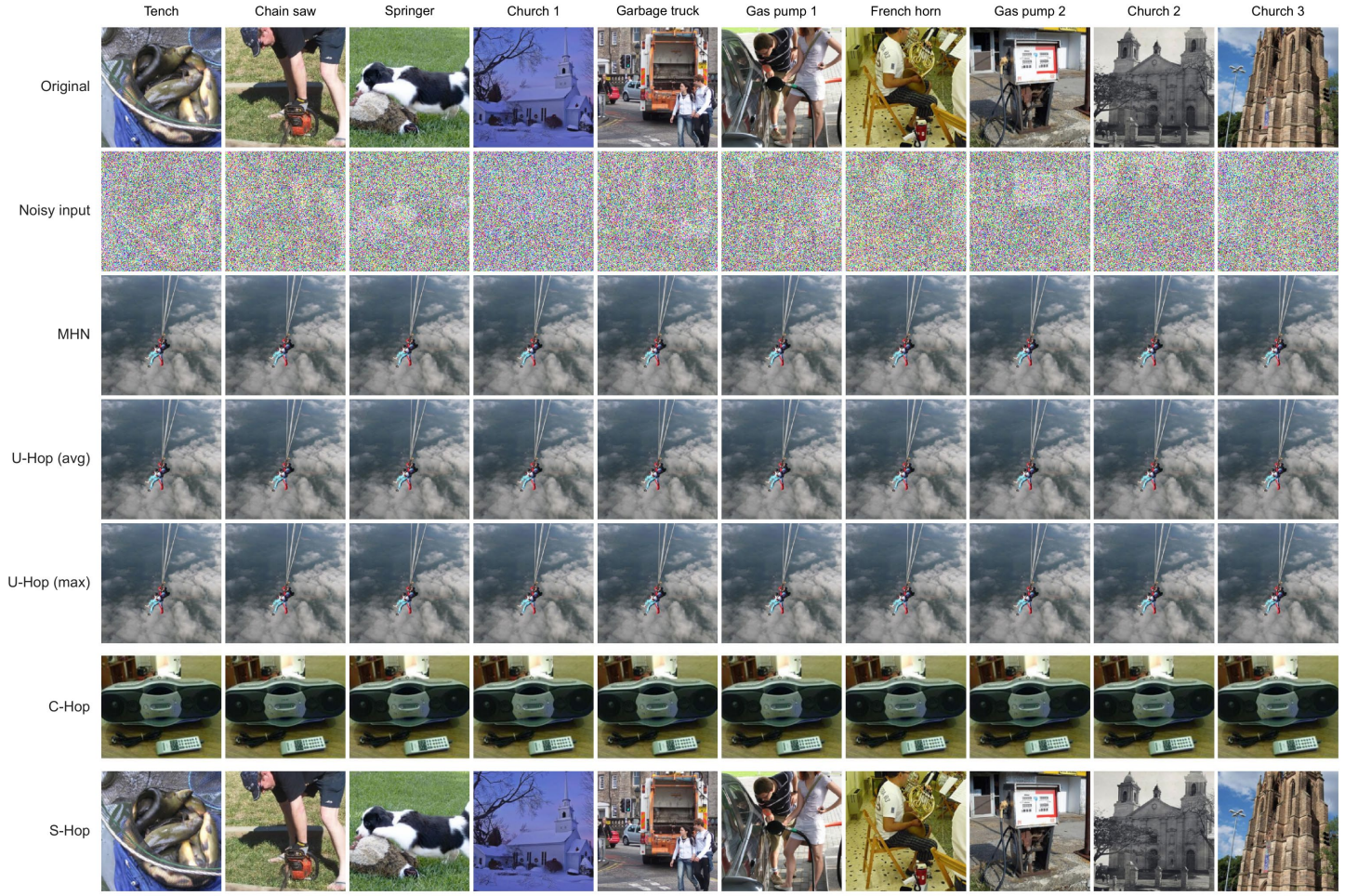
Practical question

Does this improve finite-data retrieval?

The target is useful retrieval at finite load, while preserving the exponential capacity order.

Zhao Y, Tang Y. Phase Kernel Lifts Capacity of Dense Associative Memory. Manuscript (2026). Figure 1b (illustrative energy contours). Ramsauer H et al. Hopfield Networks is All You Need. ICLR (2021). arXiv:2008.02217.

Visual retrieval examples from Gaussian-corrupted image queries



From noisy cues to stored images

Imagenette
10 illustrated cases

Gaussian noise: $\sigma = 1$
Absolute value, then clip to [0, 1]

500 memories
20 updates
No PCA or normalization

S-Hop couples a nonlinear representation to its state update

1 Embed → 2 Compare → 3 Aggregate → 4 Update

Map input and memories from d coordinates to 2d features.

Compute similarity scores and softmax weights.

Combine the embedded memories using those weights.

Apply the energy-derived phase-space transformation.

S-Hop feature pair for one coordinate

$$[\ \sqrt{(\varphi^2 + \tanh\varphi + 1)},\ \ \sqrt{(\varphi^2 + 1)}\ ]$$

Changing the kernel alone is insufficient: the matching update is part of the method.

The theory separates storage, local convergence and energy descent

Storage capacity	**$N \leq \exp(\gamma d)$, with $\gamma > 0$** Independent random phases and admissible kernel conditions.
Local convergence	**Stable fixed-point dynamics** Requires memory separation and local contraction.
S-Hop energy	**Monotone descent under the derived update** An energy guarantee does not by itself ensure target identity.

Theory explains conditional scaling and stability; image experiments measure practical retrieval.

Zhao Y, Tang Y. Phase Kernel Lifts Capacity of Dense Associative Memory. Manuscript (2026). Section 3 and theoretical appendices.

The main comparison fixes queries, update count and success criteria

Method	Representation / retrieval role
MHN	Dot-product features; modern Hopfield update
U-Hop (public avg)	Learned projection; public average-separation objective
S-Hop	Nonlinear phase map; energy-derived state update
C-Hop	Cosine-kernel structural control

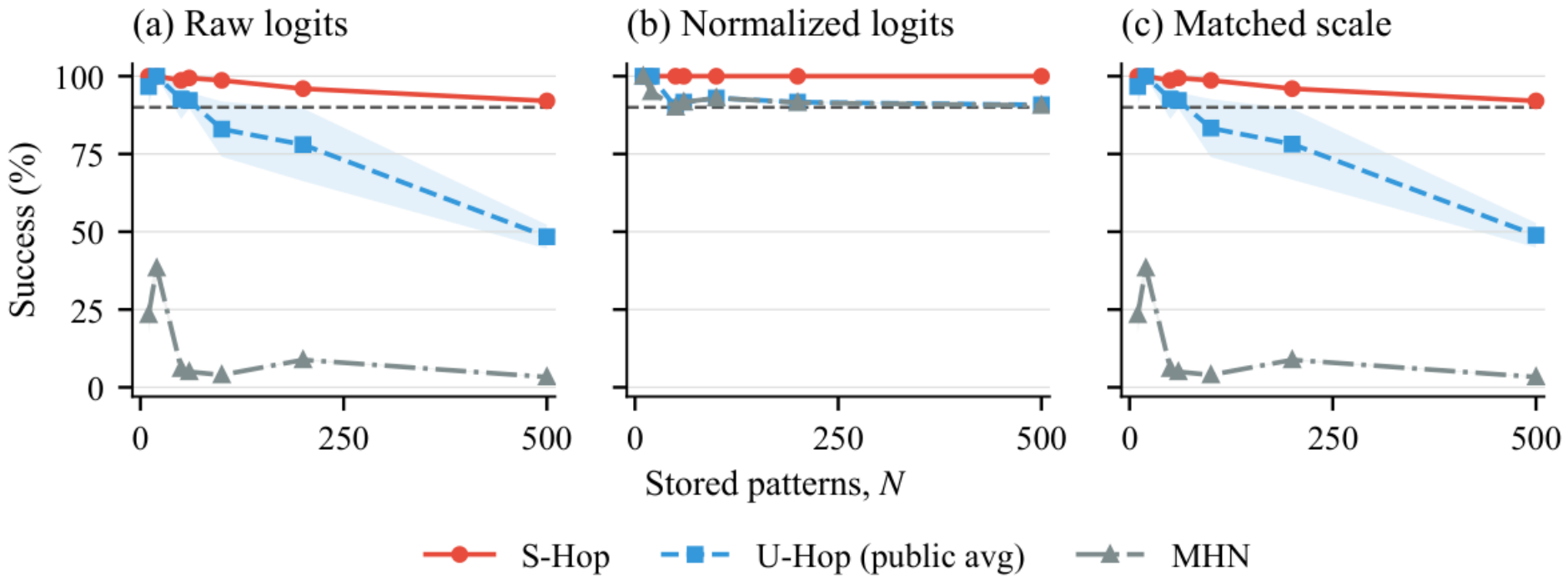
Noisy query: $q = \mathrm{clip}(|x + \varepsilon|, 0, 1), \quad \varepsilon \sim N(0, 0.3^2)$

One update · $\beta = 1$ · No PCA · Shared memories and queries · Three seeds

Success: MSE < 0.1. Passing load: at least 90% success. This is not pixel-exact recovery.

CIFAR-10: S-Hop leads across three score-scale settings

CIFAR-10 · one update · σ = 0.3 · public U-Hop avg · three-seed means and sample SD

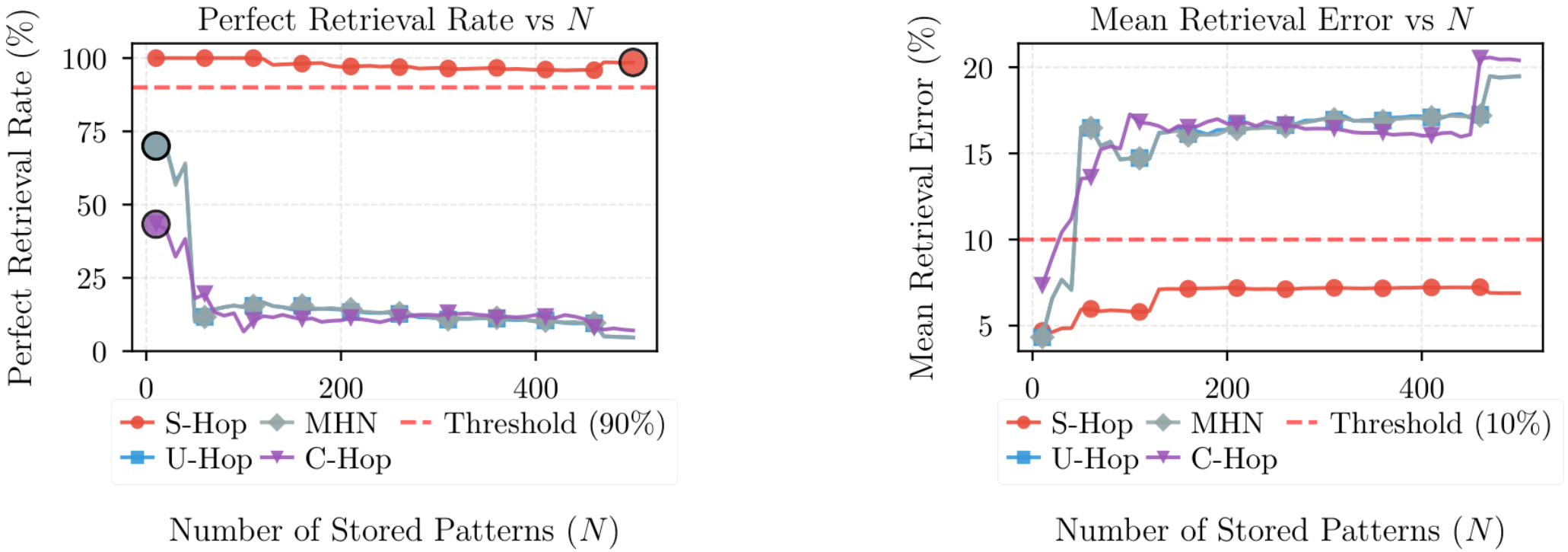


Raw scores, N = 500: S-Hop 92.07 ± 0.42% | U-Hop 48.40 ± 3.83% | MHN 3.33 ± 0.12%

ImageNet: success remains above 90% through 500 memories

Fixed subset: first 500 class-sorted ImageNet images, all tench · 224 × 224 · σ = 0.3 · one update

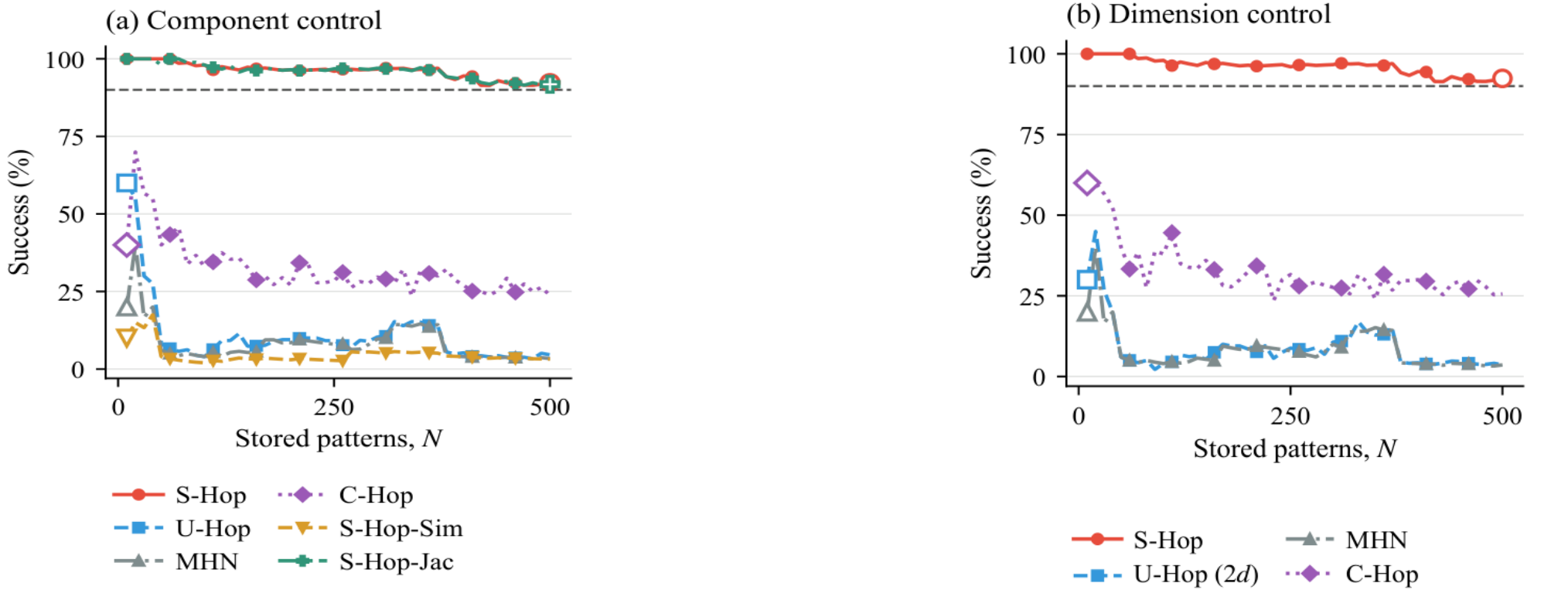
Left: success = MSE < 0.1. Right: plotted error = 100 × MSE.



At N = 500: S-Hop 98.53 ± 0.12% vs U-Hop 4.67 ± 0.12% MSE-threshold success.

The matching update matters beyond similarity or extra dimensions

CIFAR-10 · one update · σ = 0.3 · seed 42 · U-Hop max in these component controls



Similarity-only retrieval fails to match S-Hop.

A learned 2d linear embedding is also insufficient.

The Jacobian and inverse local field are restored jointly; low MSE does not by itself guarantee target identity.

Phase kernels link representation, dynamics and retrieval capacity

01	**A coupled design**	Phase features reshape the energy; retrieval needs the matching update, not the kernel alone.
02	**Conditional guarantees**	Exponential capacity order and local convergence, under stated conditions on the kernel.
03	**Measured gains**	Success stays above 90% through 500 memories; 98.53% on 224×224 ImageNet.

Next step Correlated memories, stricter identity criteria, and retrieval stability over many updates.

Zhao Y, Tang Y. Phase Kernel Lifts Capacity of Dense Associative Memory. Manuscript (2026). Sections 3–6 and control experiments.