Dense Associative Memory for Pattern Recognition

Dmitry Krotov
Simons Center for Systems Biology
Institute for Advanced Study
Princeton, USA

John J. Hopfield
Princeton Neuroscience Institute
Princeton University

Hopfield Networks is All You Need

Hubert Ramsauer* Bernhard Schäfl* Johannes Lehner* Philipp Seidl*
Michael Widrich* Thomas Adler* Lukas Gruber* Markus Holzleitner*
Milena Pavlović[‡,§] Geir Kjetil Sandve[§] Victor Greiff[‡] David Kreil[†]
Michael Kopp[†] Günter Klambauer* Johannes Brandstetter* Sepp Hochreiter[*,†]

Large Associative Memory Problem in Neurobiology and Machine Learning

Dmitry Krotov
MIT-IBM Watson AI Lab
IBM Research

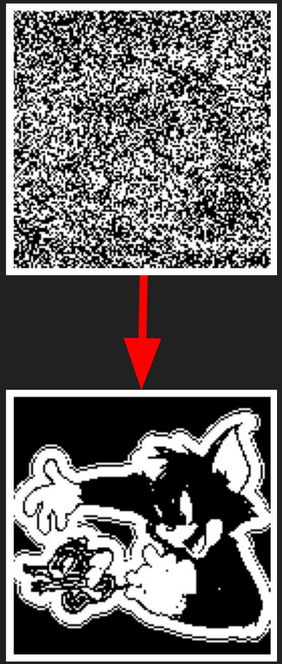
John Hopfield
Princeton Neuroscience Institute
Princeton University

Neural networks and physical systems with emergent collective computational abilities

(associative memory/parallel processing/categorization/content-addressable memory/fail-soft devices)

J. J. Hopfield

Division of Chemistry and Biology, California Institute of Technology, Pasadena, California 91125; and Bell Laboratories, Murray Hill, New Jersey 07974

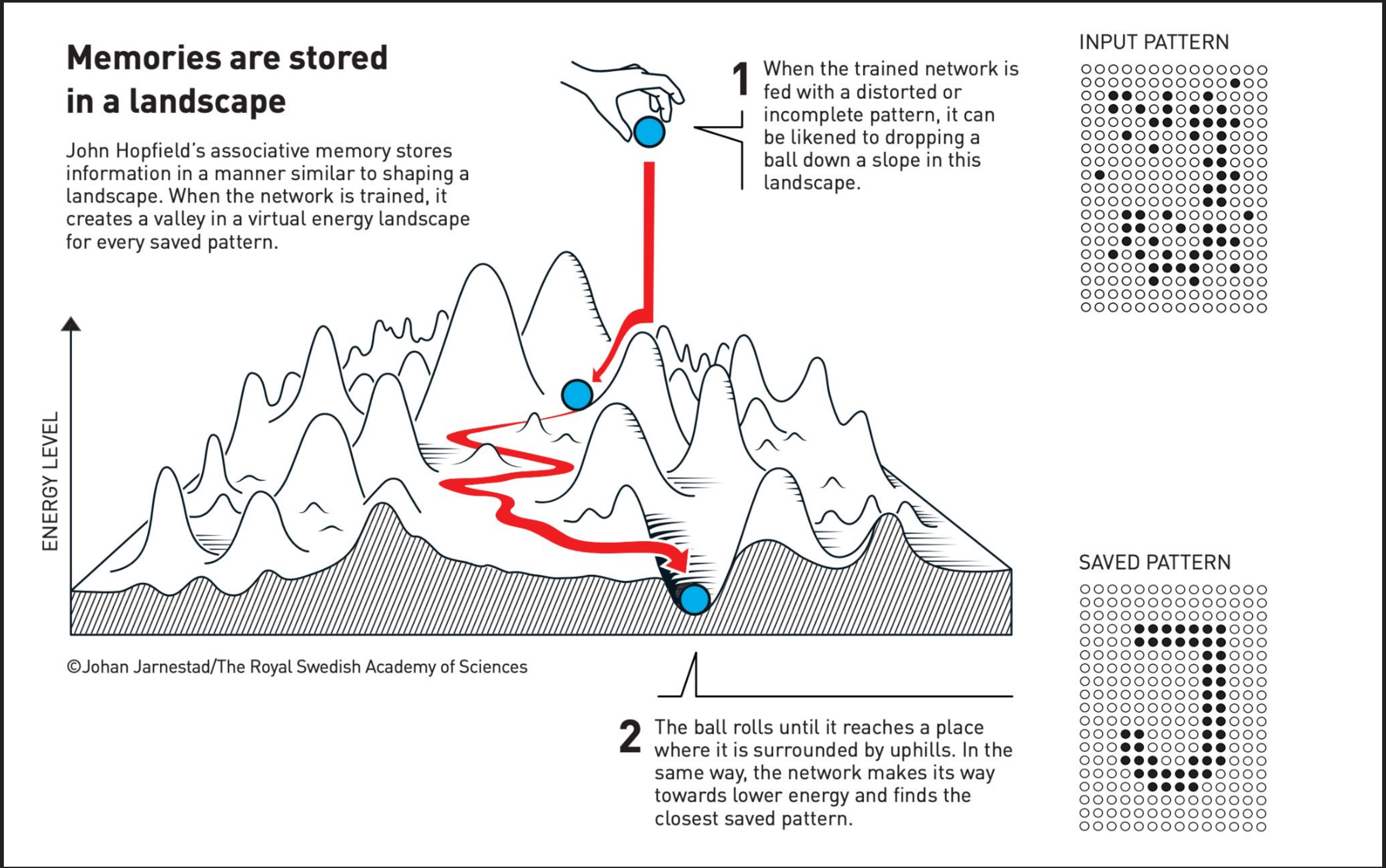


Hopfield Networks++

Neel, Forrest, Simon, Miles

Outline

1. Recap, energy formulation, and storage capacity
2. Problems with the original network
3. Generalized Energy & dense associative memories
4. Biology: leaky integrate-and-fire and real biology



https://www.nobelprize.org/uploads/2024/10/fig3_fy_en_24.pdf

1. Energy function and weight matrix

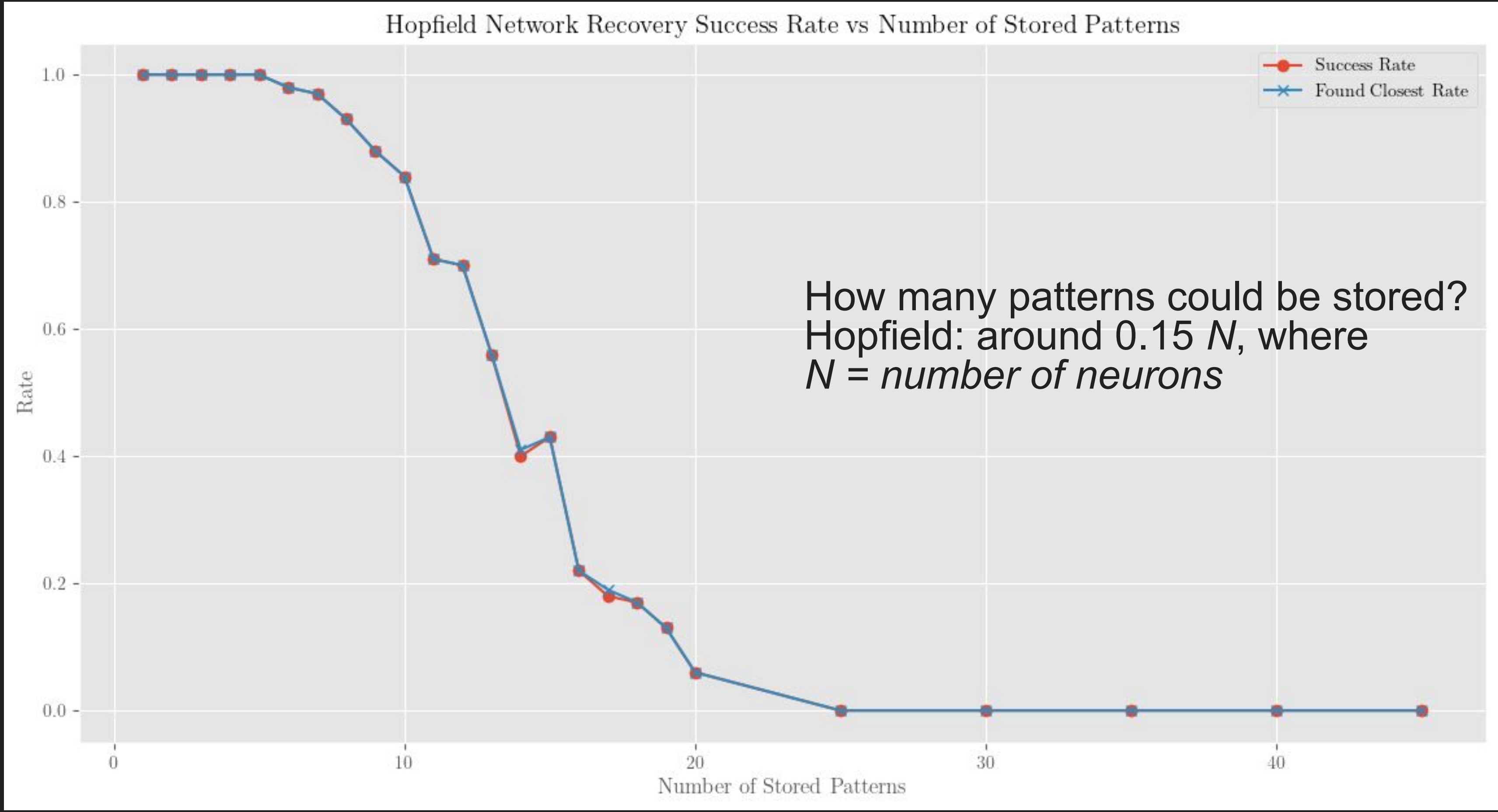
- Want: high overlap with stored patterns implies low energy
- Overlap could be computed using the dot product!
- Energy function is equal to the sum of squares of the dot products of the current pattern with each of the stored patterns. We call this the ***quadratic energy function***:

$$E(\boldsymbol{\sigma}) = -\frac{1}{2N}\sum_{\mu=1}^{M}(\boldsymbol{\xi}^{\mu}\cdot\boldsymbol{\sigma})^2.$$

- This could be rewritten using the Hopfield weight matrix:

$$E(\boldsymbol{\sigma}) = -\frac{1}{2}\boldsymbol{\sigma}^T\mathbf{W}\boldsymbol{\sigma}, \quad \mathbf{W} := \frac{1}{N}\sum_{\mu=1}^{M}\boldsymbol{\xi}^{\mu}(\boldsymbol{\xi}^{\mu})^T.$$

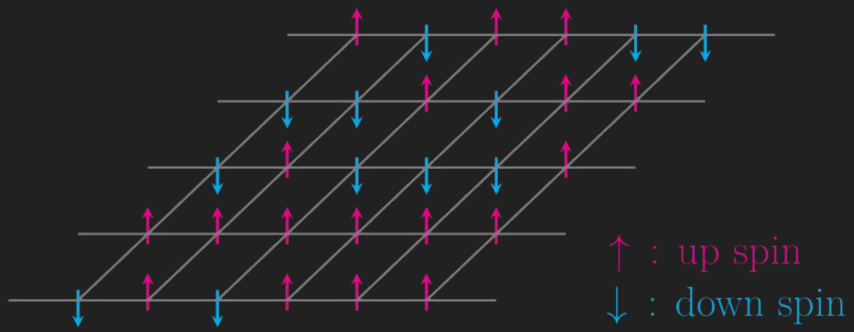
Successful retrieval rate vs. num patterns (100 Neurons)



Student generated

The Ising model

Want to model a macroscopic magnet as a lattice of little magnets ("spins") that can be in either of two orientations ("up" and "down").



Energy associated with a state $\boldsymbol{x} \in \{1,-1\}^N$ is (assuming no applied field)

$$E = -\frac{1}{2}\sum_{i,j} J_{ij} x_i x_j$$

Ising model continued

If the magnet is held at temperature *T*, the probability of finding its spins in configuration ***x*** is

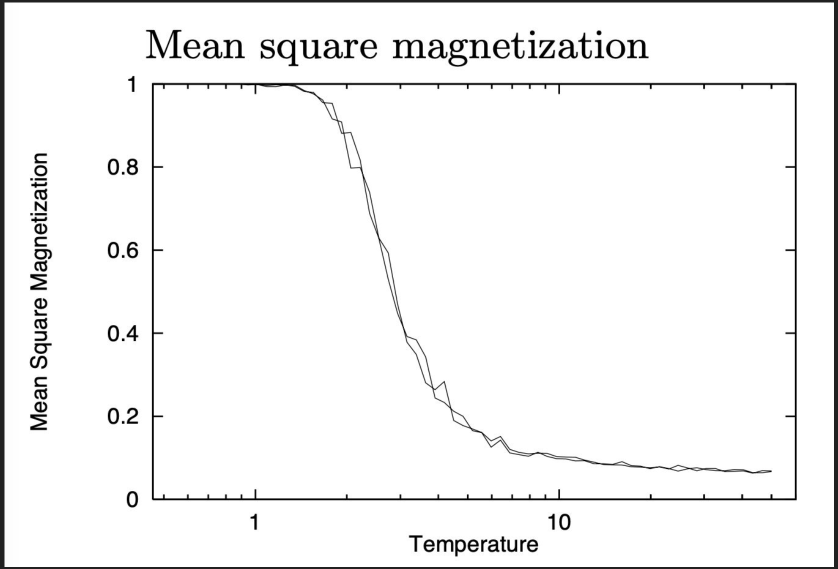
$$P(\mathbf{x}) = \frac{\exp\left(-E(\mathbf{x})/k_{\mathrm{B}}T\right)}{Z(T)}$$

This is known to physicists as the Boltzmann distribution

Criticality in the Ising model

How does the behavior of the magnet vary with temperature?

Intuitively, it should demagnetize, and the Ising model delivers this intuition.



From MacKay, *Information Theory, Inference, and Learning Algorithms*

How does Ising relate to Hopfield?

The Hopfield net's energy function is identical to that of the Ising magnet without applied field.

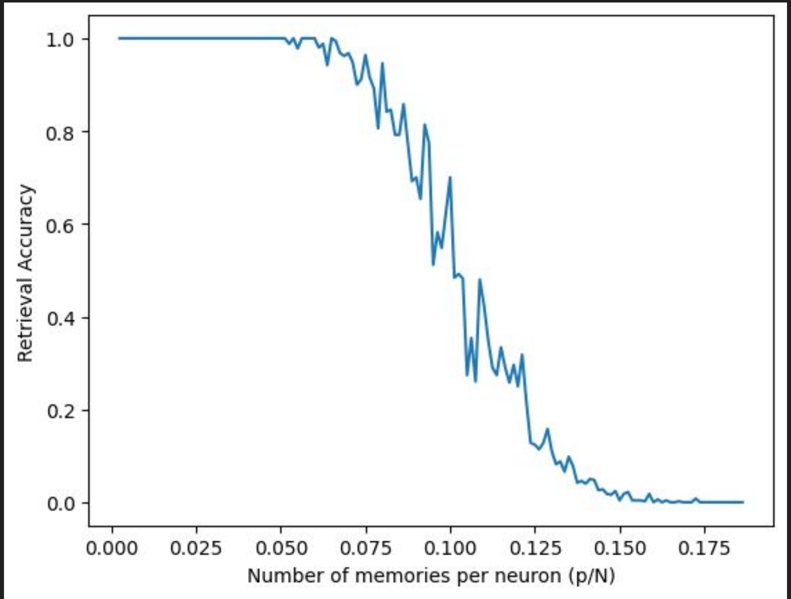
What's the temperature? Suppose ***x*** is close to memory ξ^1. Then,

$$\sum_j W_{ij} x_j = \frac{1}{N} \sum_j \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu x_j$$

$$\approx \xi_i^1 + \underbrace{\frac{1}{N} \sum_j \sum_{\mu=2}^{p} \xi_i^\mu \xi_j^\mu x_j}_{\text{Noise of order } p/N}$$

Criticality in the Hopfield net

As we might expect, the Hopfield net “demagnetizes” when the number of memories per neuron reaches a critical level.



Student-generated

Physical Realization

PHYSICAL REVIEW LETTERS **135,** 160403 (2025)

Editors' Suggestion

Multimode Cavity QED Ising Spin Glass

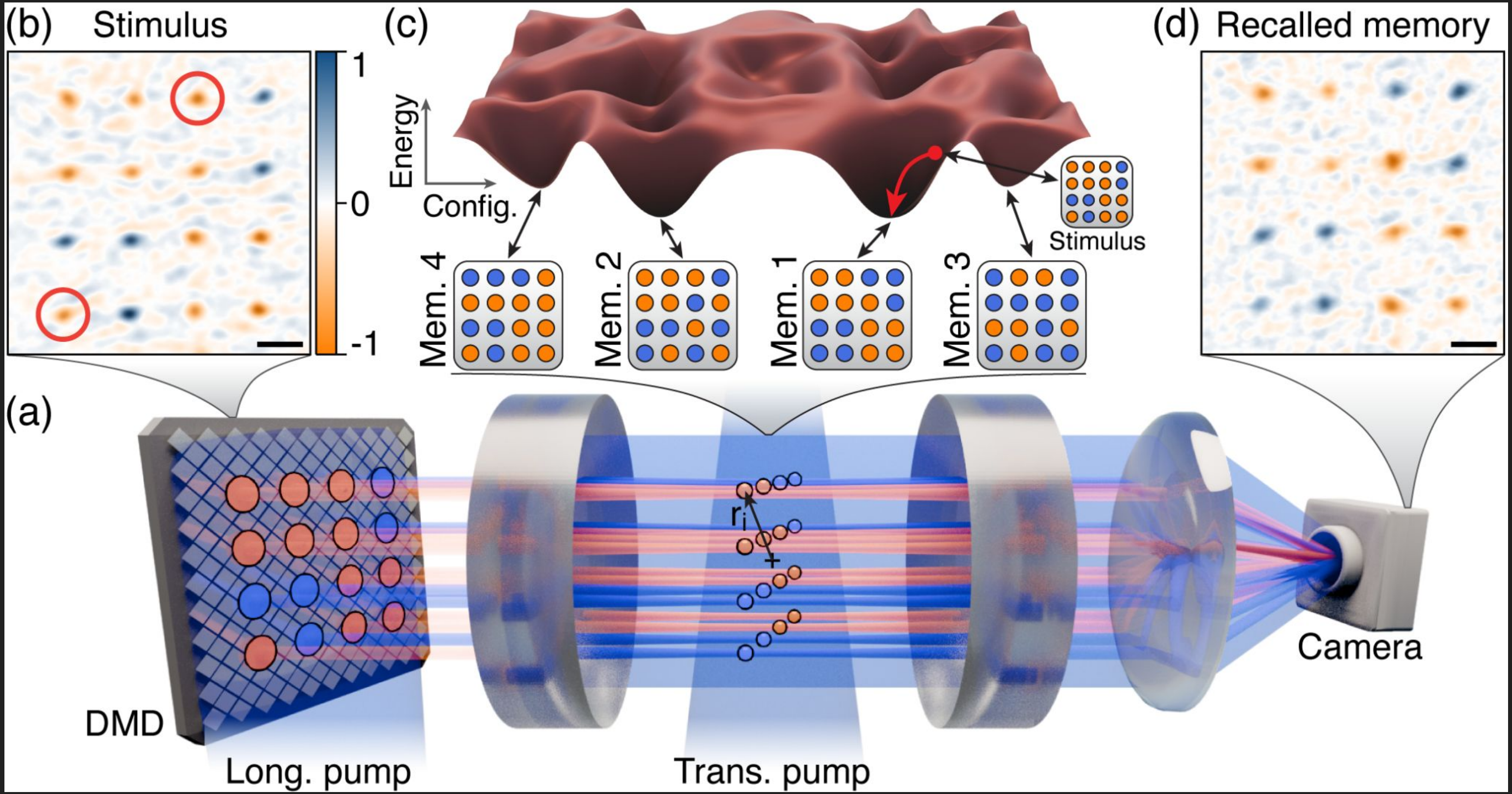
Brendan P. Marsh,[1,2] David Atri Schuller,[1,2] Yunpeng Ji,[1,2,3] Henry S. Hunt,[2,3] Giulia Z. Socolof,[1,2] Deven P. Bowman,[2,3] Jonathan Keeling,[4] and Benjamin L. Lev[1,2,3]

[1]*Department of Applied Physics, Stanford University, Stanford, California 94305, USA*
[2]*E. L. Ginzton Laboratory, Stanford University, Stanford, California 94305, USA*
[3]*Department of Physics, Stanford University, Stanford, California 94305, USA*
[4]*SUPA, School of Physics and Astronomy, University of St. Andrews, St. Andrews KY16 9SS, United Kingdom*

High-capacity associative memory in a quantum-optical spin glass

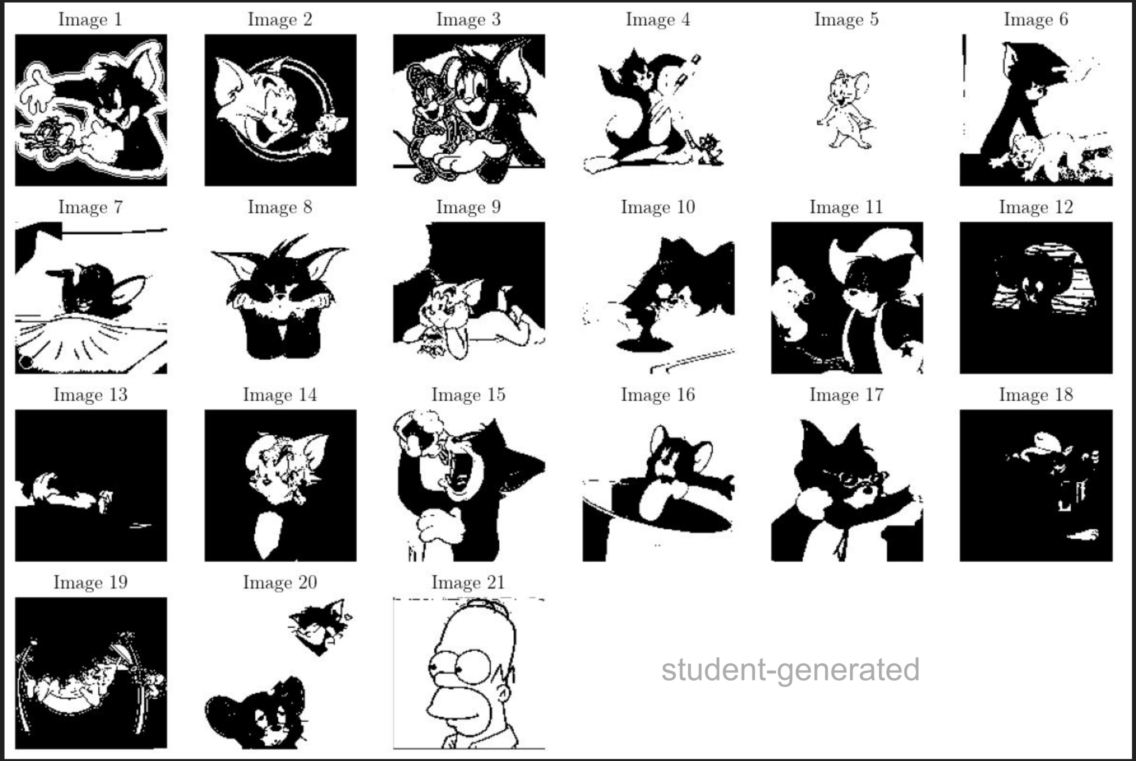
Brendan P. Marsh,[1,2] David Atri Schuller,[1,2] Yunpeng Ji,[1,2,3] Henry S. Hunt,[2,3] Surya Ganguli,[1] Sarang Gopalakrishnan,[4] Jonathan Keeling,[5] and Benjamin L. Lev[1,2,3]

Each “spin” or “neuron” is a Bose-Einstein condensate of 60000 rubidium atoms!



2. Problems with the simple quadratic-*E* network

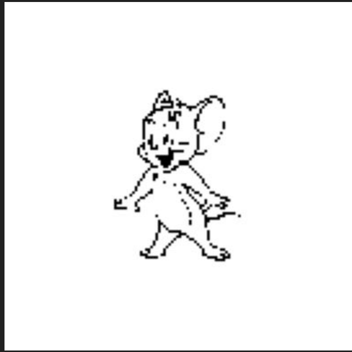
- Images from internet, stored in 128 x 128 pixel arrays with black = 1
  And white = -1



- N = *128* x *128* Neurons
- Theoretically, network should be able to store 0.138 N = 2261 patterns
- In reality, the network gets confused storing just 10 patterns! Why?

starting

Network output



starting

Network output

student-generated

- Spurious minima resulting from similar patterns.

About 0.15 N states can be simultaneously remembered before error in recall is severe. Computer modeling of memory storage according to Eq. 2 was carried out for $N = 30$ and $N = 100$. n random memory states were chosen and the corresponding T_{ij} was generated. If a nervous system preprocessed

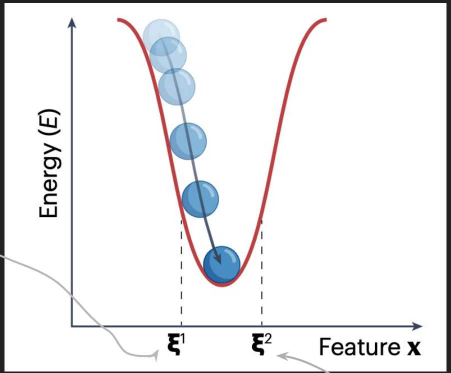
Memories too close to each other are confused and tend to merge. For $N = 100$, a pair of random memories should be separated by 50 ± 5 Hamming units. The case $N = 100$, $n = 8$, was studied with seven random memories and the eighth made up a Hamming distance of only 30, 20, or 10 from one of the other seven memories. At a distance of 30, both similar memories were usually stable. At a distance of 20, the minima were usually distinct but displaced. At a distance of 10, the minima were often fused.

(Hopfield 1982)

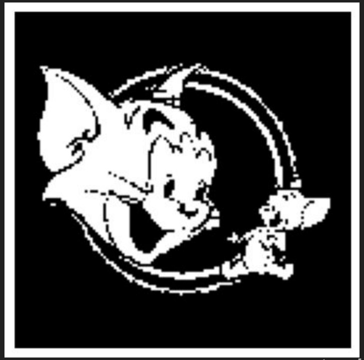
- These errors correspond to exactly the same spurious minimum!

$\boldsymbol{\xi}^1$

$\boldsymbol{\xi}^2$



Krotov 2023



student-generated

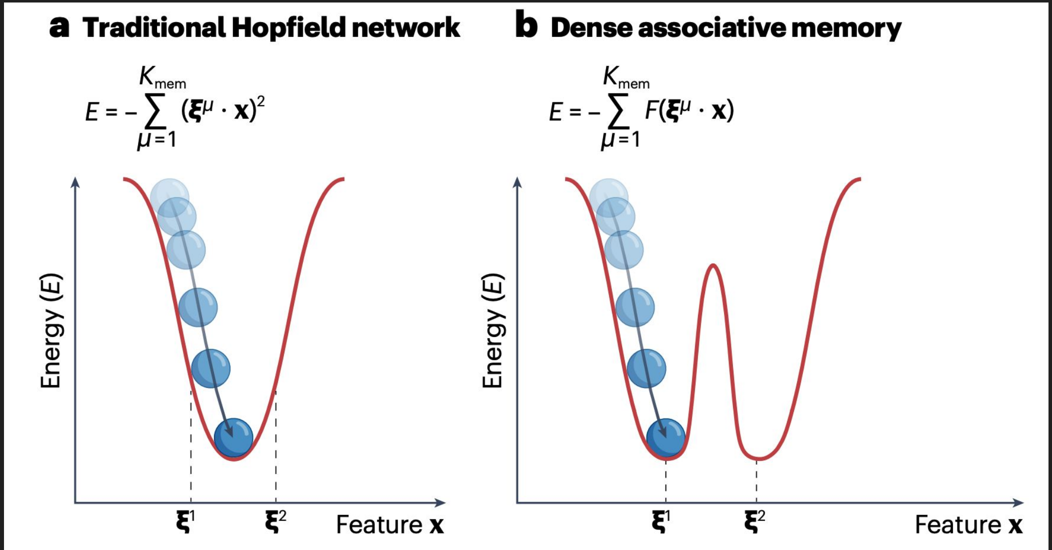
3. Solution: Generalized Energy Functions

Comment | Published: 18 May 2023

A new frontier for Hopfield networks

Dmitry Krotov

Nature Reviews Physics **5**, 366–367 (2023) | Cite this article



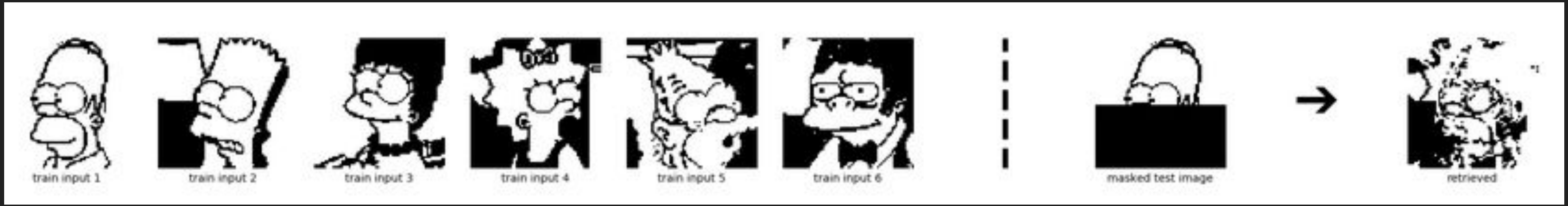
New Energy Function: Dense associative memory

- Quadratic case:

$$E(\boldsymbol{\sigma}) = -\frac{1}{2N}\sum_{\mu=1}^{M}(\boldsymbol{\xi}^{\mu}\cdot\boldsymbol{\sigma})^2.$$

- Generalized energy:

$$E(\boldsymbol{\sigma}) = -\frac{1}{2N}\sum_{\mu=1}^{M}F(\boldsymbol{\xi}^{\mu}\cdot\boldsymbol{\sigma}).$$

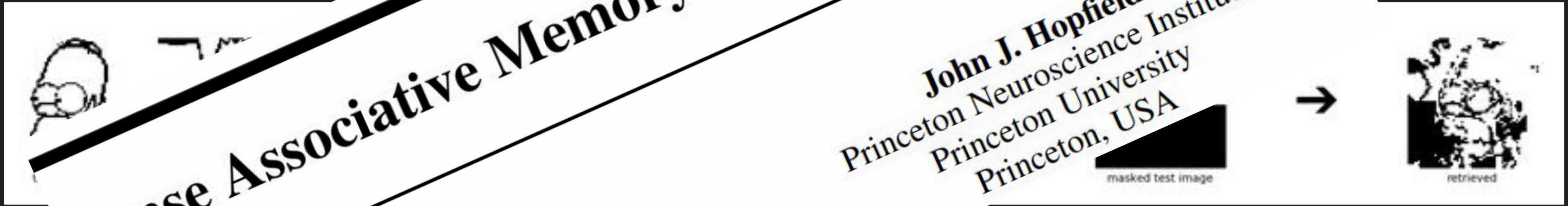


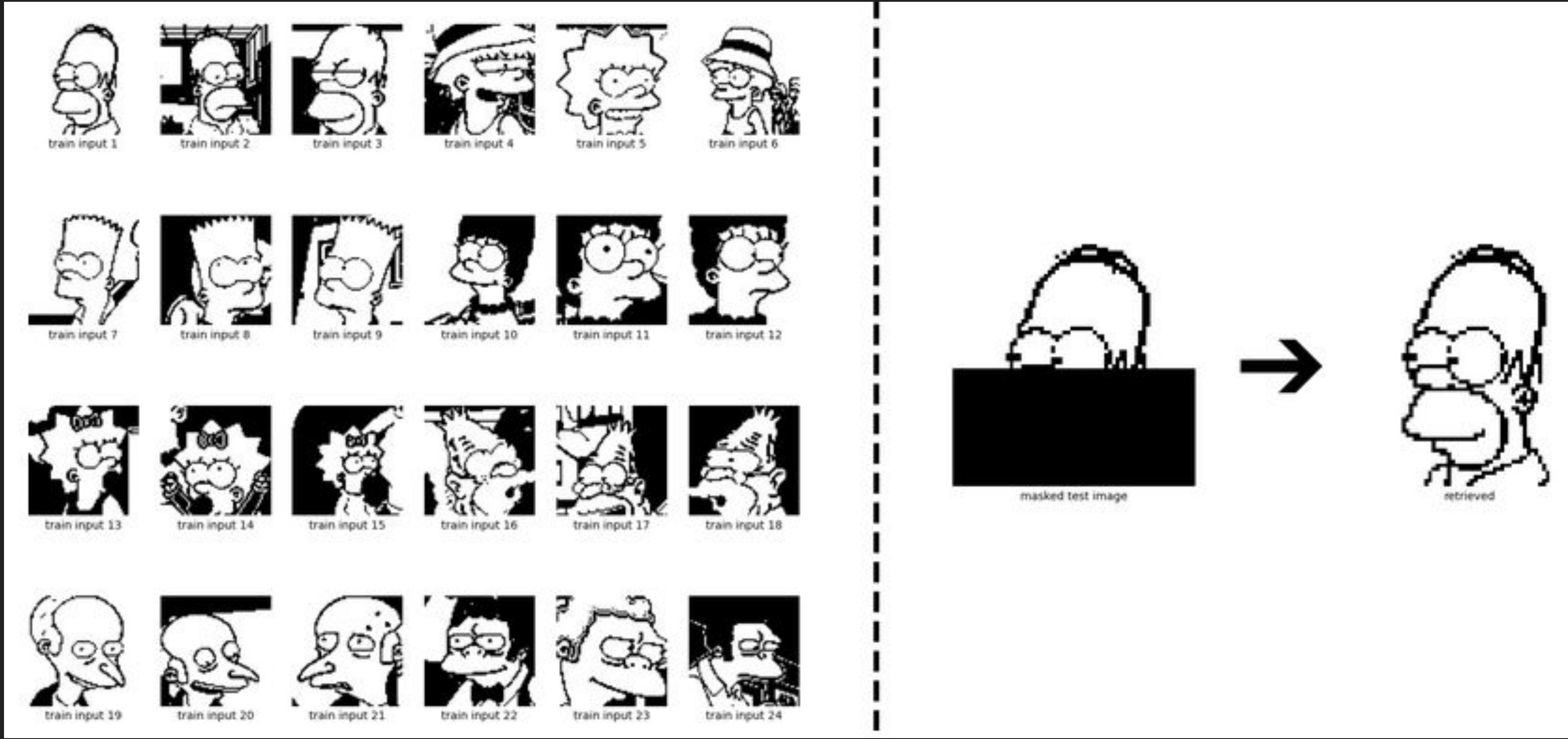
https://ml-jku.github.io/hopfield-layers/#hfnetworks

Dense Associative Memory for Pattern Recognition

Dmitry Krotov
Simons Center for Systems Biology
Institute for Advanced Study
Princeton, USA

John J. Hopfield
Princeton Neuroscience Institute
Princeton University
Princeton, USA





https://ml-jku.github.io/hopfield-layers/#hfnetworks

Dense Associative Memory

$$E = -\sum_{\mu=1}^{K} F(\xi_i^\mu \sigma_i)$$

$$F(x) = \begin{cases} x^n, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

next neuron i state

example μ's neuron i state

current state of neuron j

$$\sigma_i^{(t+1)} = Sign\Bigg[\sum_{\mu=1}^{K} \Bigg(\underbrace{F\Big(\xi_i^\mu + \sum_{j \neq i} \xi_j^\mu \sigma_j^{(t)}\Big) - F\Big(-\xi_i^\mu + \sum_{j \neq i} \xi_j^\mu \sigma_j^{(t)}\Big)}_{\text{positive if } \xi_i^\mu = 1 \text{ and input to } F > 0.} \Bigg)\Bigg],$$

sum over all K memories

https://arxiv.org/abs/1606.01164

Hopfield Demo

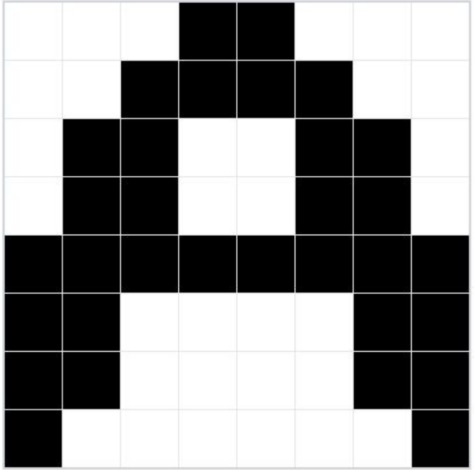
Original Hopfield Stochastic Synchronous Updates

Click and drag to draw.

Draw a letter and converge to my drawing (or not)!

Cells with value 1: 32

Current Energy of State: -64.0625



Clear Grid | Log to Console | Export to JSON | Update (whole state)

Steps

1. Draw an A, B, or C on the left grid and hit the update button to converge to the closest letter.
2. Switch the range to A-G and try the same process with a different letter.
3. Try drawing that letter on the right grid
4. See if you get better convergence on the right.

Choose range:

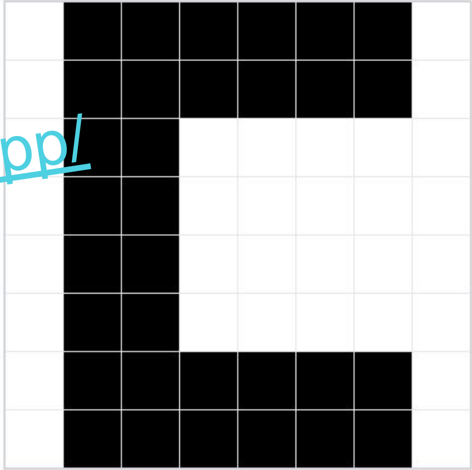
A-C

https://drawfield.vercel.app/

Dense Associative Memory Asynchronous Updates

Cells with value 1: 32

Current Energy of State (cubic): -524288



Clear Grid | Log to Console | Export to JSON | Update (Row at a time)

F = exp() yields interesting results

- β is the *inverse temperature* parameter (1/*kT* in thermodynamics)
- Large β = always converge deterministically to closest pattern
- Small β = possibility of going to unexpected patterns. Enables creativity

$$E(\boldsymbol{\sigma}) = -\frac{1}{2N}\sum_{\mu=1}^{M}\exp(\beta\boldsymbol{\xi}^{\mu}\cdot\boldsymbol{\sigma}).$$

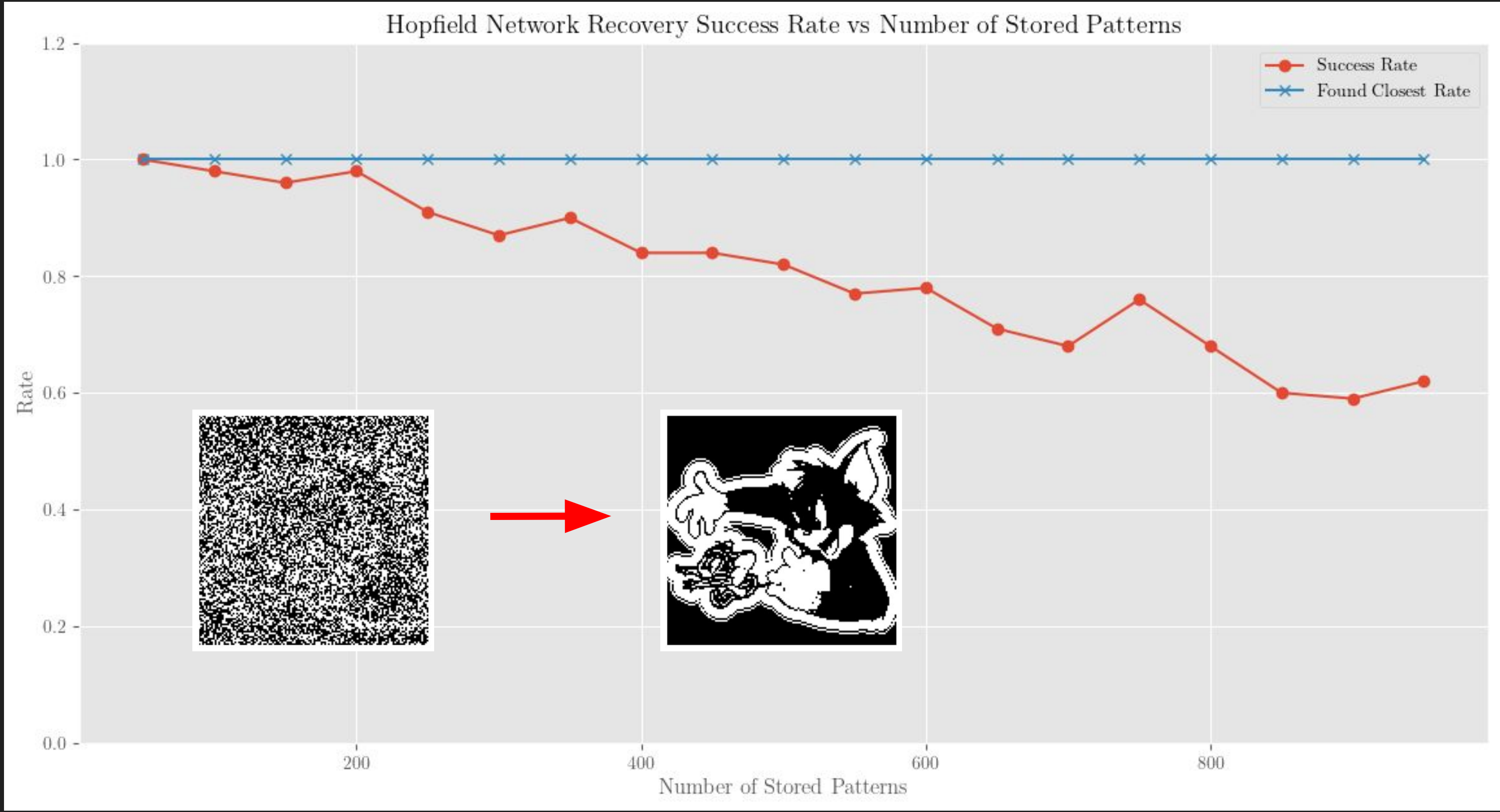
F = exp yields interesting results

$$E(\boldsymbol{\sigma}) = -\frac{1}{2N}\sum_{\mu=1}^{M}\exp(\beta\boldsymbol{\xi}^{\mu}\cdot\boldsymbol{\sigma}).$$

$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$

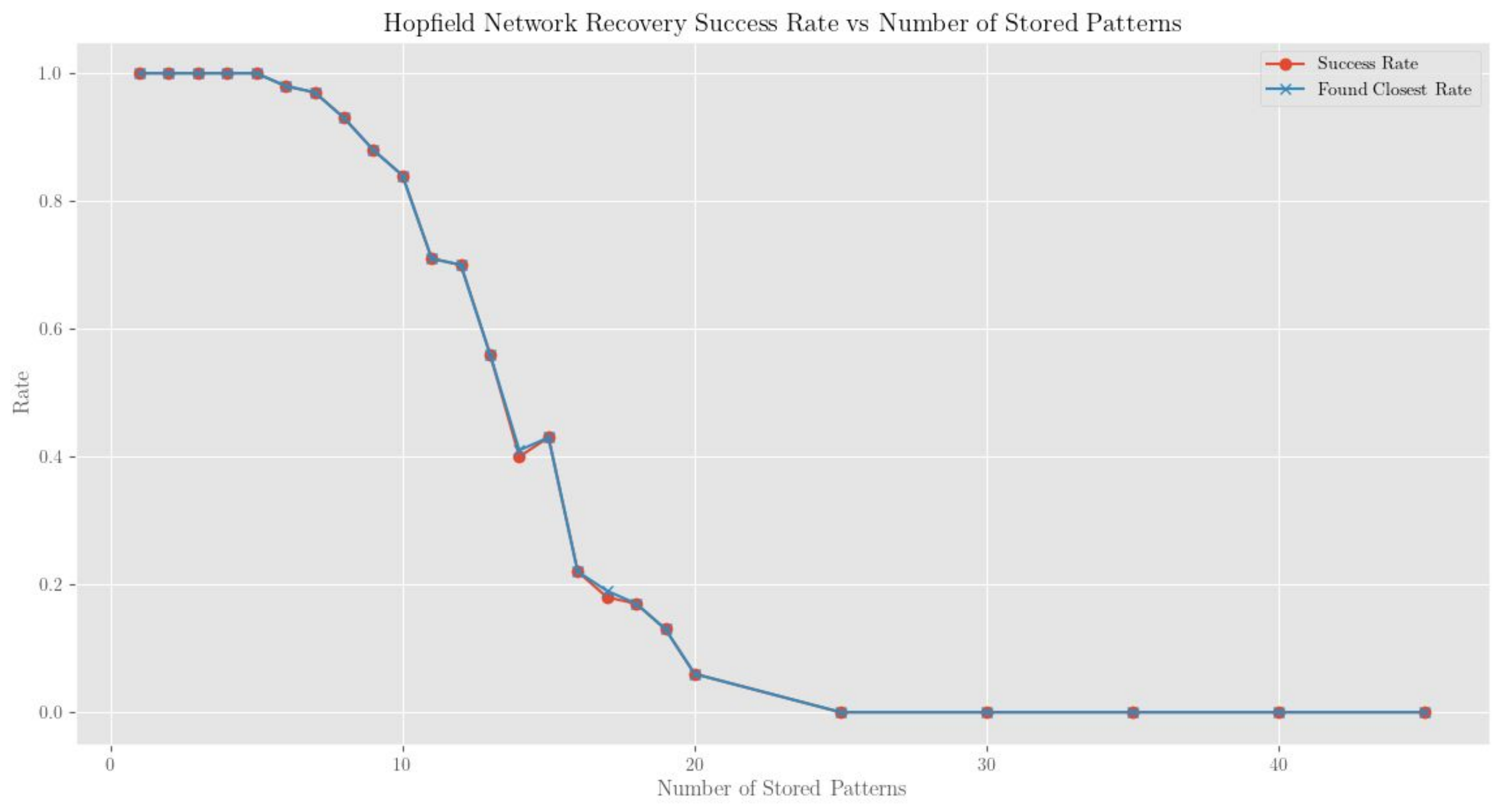
- With suitable modifications, a Hopfield network with this energy function is identical to the self-attention in transformer networks!
- Intuition: user provides text input. LLM predicts most probable sentences that follow user input. This is light retrieving a memory from an incomplete or fuzzy input in a Hopfield network.
- β enables some “creative thinking” by the LLM; if inverse temperature is set to infinity, then the dynamics of the network would be deterministic, and GPT would produce the same answer every time you ask it some question.

Dense associative memory: storage capacity with 100 neurons



Student generated

Successful retrieval rate (classical quadratic function)

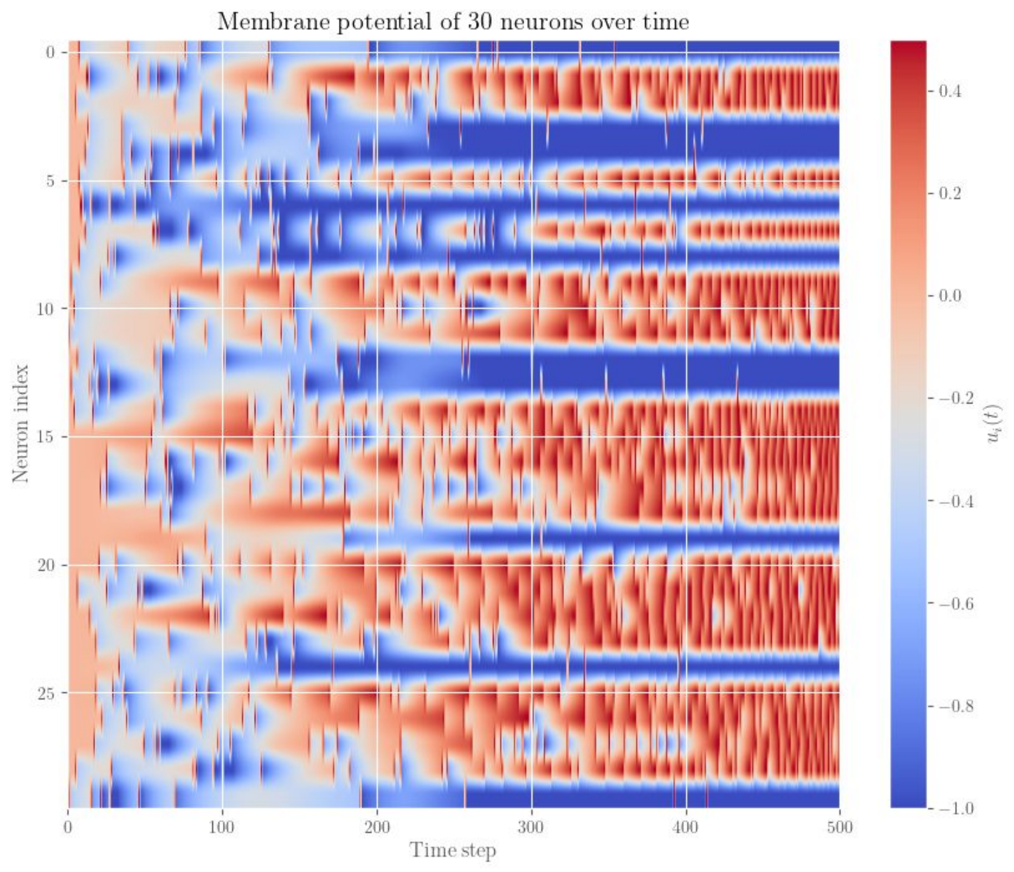


Student generated

4. Biological models

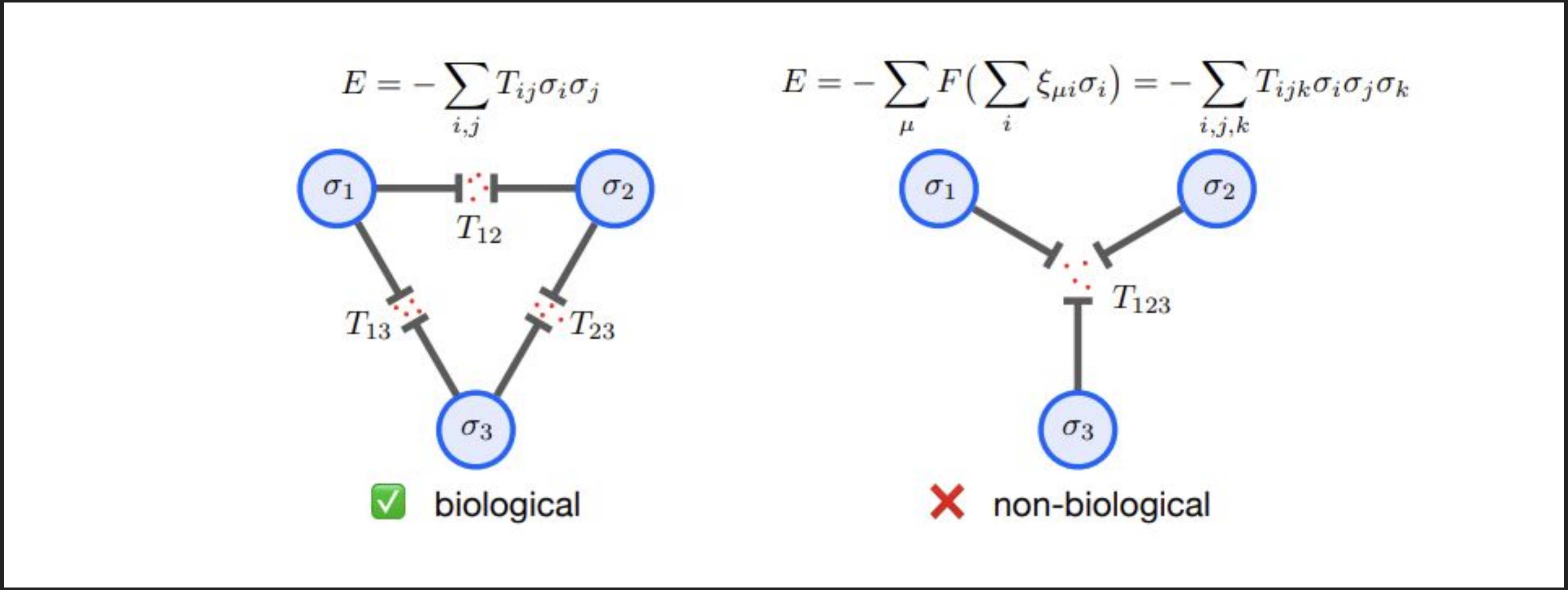
$$u_i(t) = \sum_f \eta(t - t_i^f) + h_i(t) + u_{\text{rest}},$$

$$h_i(t) = \sum_j w_{ij} \varepsilon(t - t_j^f)$$



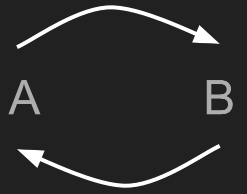
Student generated

https://arxiv.org/abs/2008.06996

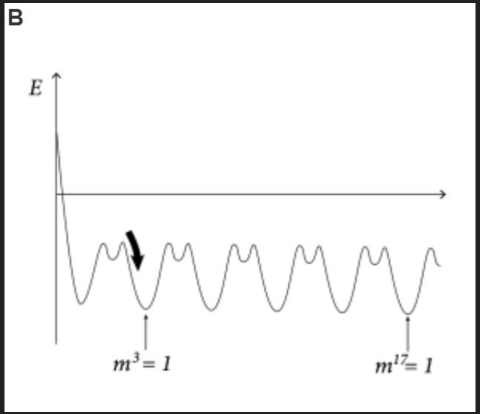


(More on this from Neel)

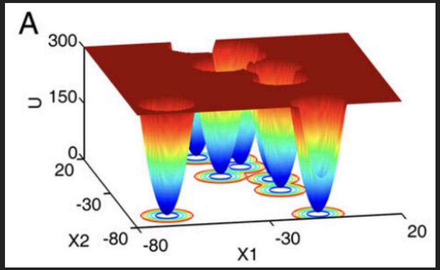
A Hopfield Network For Everything

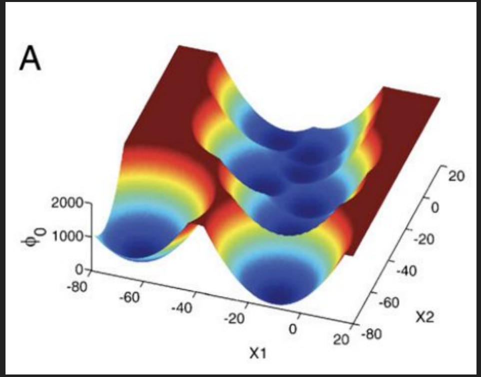


$$\begin{bmatrix} 0 & 0 & -1 \\ 1 & 0 & -1 \\ -1 & -1 & 0 \end{bmatrix}$$



https://neuronaldynamics.epfl.ch/online/x550.png

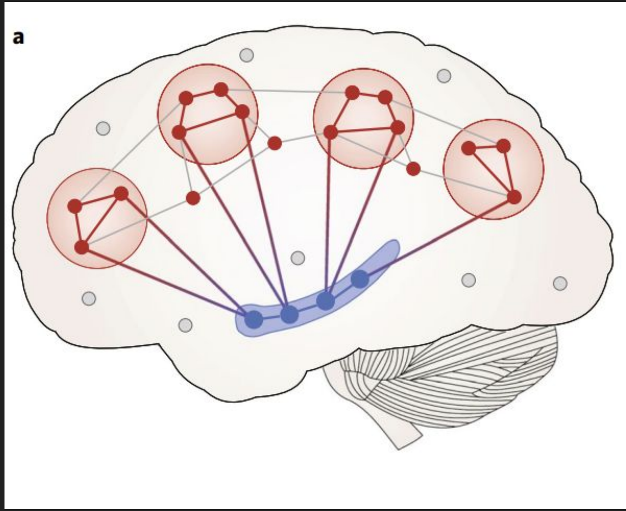




http://pnas.org/doi/pdf/10.1073/pnas.1310692110

Sequential Memory

$p^A \rightarrow p^B \rightarrow p^C \rightarrow p^D$



https://www.nature.com/articles/s41593-019-0467-3.pdf

nature human behaviour

Article https://doi.org/10.1038/s41562-023-01799-z

A generative model of memory construction and consolidation

Memory & Imagination

Neuron

Predictive sequence learning in the hippocampal formation

Hippocampal Replay

Sequential Memory with Temporal Predictive Coding

Mathematical model matches behavior

nature human behaviour

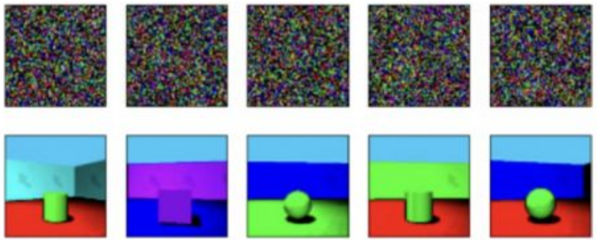
Article https://doi.org/10.1038/s41562-023-01799-z

A generative model of memory construction and consolidation

Hippocampus:

Encodes “snapshots” as episodic memory using associative networks

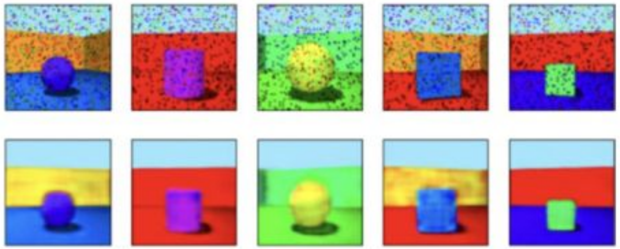
c



Neo-cortex:

Generative model with student-teacher learning from hippocampal neural network

d



Neuron

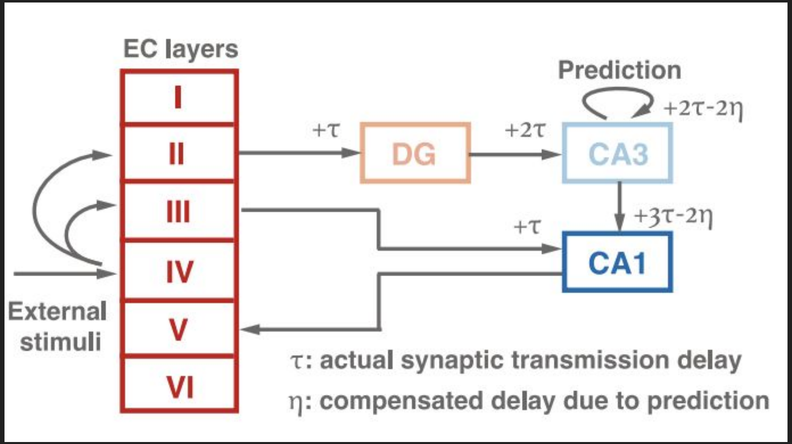
Arti

Predictive sequence learning in the hippocampal formation

CA3 “predicts” patterns sequentially

CA1 error corrects to reality

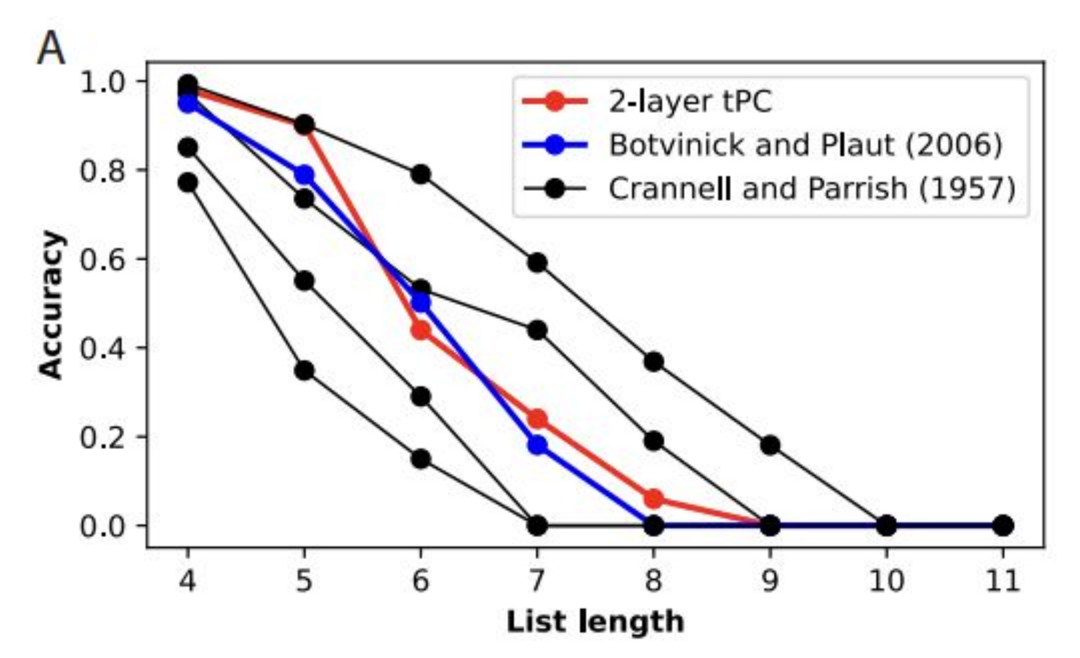
Dreaming helps export “whitened” memories!



Sequential Memory with Temporal Predictive Coding

Asymmetric associative network model

Matched recall of sequences of words in humans



References

[1] Johannes Brandstetter. *Hopfield Networks is All You Need*. Aug. 2020. URL: https://ml-jku.github.io/hopfield-layers/ (visited on 11/30/2025).

[2] Yusi Chen et al. "Predictive Sequence Learning in the Hippocampal Formation". In: *Neuron* 112 (June 2024), 2645–2658.e4. DOI: 10.1016/j.neuron.2024.05.024. (Visited on 12/01/2025).

[3] Mete Demircigil et al. "On a Model of Associative Memory with Huge Storage Capacity". In: *Journal of Statistical Physics* 168 (May 2017), pp. 288–299. DOI: 10.1007/s10955-017-1806-y. (Visited on 12/01/2025).

[4] Wulfram Gerstner et al. *Neuronal Dynamics: from Single Neurons to Networks and Models of Cognition*. Cambridge University Press, 2014.

[5] J. J. Hopfield. "Neural Networks and Physical Systems with Emergent Collective Computational abilities." In: *Proceedings of the National Academy of Sciences* 79 (Apr. 1982), pp. 2554–2558. DOI: 10.1073/pnas.79.8.2554. URL: https://www.pnas.org/content/79/8/2554 (visited on 11/30/2025).

[6] Dmitry Krotov. "A New Frontier for Hopfield Networks". In: *Nature Reviews Physics* 5 (July 2023), pp. 366–367. DOI: 10.1038/s42254-023-00595-y. URL: https://www.nature.com/articles/s42254-023-00595-y (visited on 11/30/2025).

[7] Dmitry Krotov and John Hopfield. "Dense Associative Memory for Pattern Recognition". In: *Neural Information Processing Systems* 29 (June 2016), pp. 1172–1180.

[8] Dmitry Krotov and John Hopfield. "Large Associative Memory Problem in Neurobiology and Machine Learning". In: *ICLR 2021 Conference Paper* (Apr. 2021). URL: https://arxiv.org/abs/2008.06996 (visited on 11/30/2025).

[9] Brendan P Marsh et al. "High-capacity Associative Memory in a quantum-optical Spin Glass". In: *arXiv preprint* (Sept. 2025). DOI: 10.48550/arxiv.2509.12202. (Visited on 11/30/2025).

[10] Brendan P Marsh et al. "Multimode Cavity QED Ising Spin Glass". In: *Physical Review Letters* 135 (Oct. 2025). DOI: 10.1103/x19r-pzyb.

[11] Hubert Ramsauer et al. "Hopfield Networks Is All You Need". In: *ICLR 2021 Conference Paper* (Jan. 2020). DOI: 10.48550/arxiv.2008.02217. URL: https://iclr.cc/virtual/2021/poster/2657 (visited on 11/30/2025).

[12] Eleanor Spens and Neil Burgess. "A Generative Model of Memory Construction and Consolidation". In: *Nature Human Behaviour* 8 (Jan. 2024), pp. 1–18. DOI: 10.1038/s41562-023-01799-z. URL: https://www.nature.com/articles/s41562-023-01799-z.

[13] Mufeng Tang, Helen Barron, and Rafal Bogacz. "Sequential Memory with Temporal Predictive Coding". In: *PubMed* 36 (May 2023), pp. 44341–44355. DOI: 10.48550/arxiv.2305.11982. (Visited on 12/01/2025).

Problem: Memory as Storing Patterns

- We have *N* binary neurons (+1 if firing, -1 if not firing) and want to remember a number of patterns. Each pattern is an array of *N* bits (a bit vector). For example, a 6-bit pattern would be (1,-1,1,1,-1,1)

- Patterns are stored via a Weight matrix W, where W_{ij} is the strength of the synaptic connection between neurons *i* and *j*

- The network is presented with a starting bit vector and want to retrieve the stored pattern that most resembles the starting pattern.

- How should we choose the weights W_{ij} and the update rule?

Energy Minimization

- Idea: frame problem as the minimization of some objective function, which we will call the energy E. E is a function of configuration of the N neurons.

- The update rule is designed so that given a starting pattern $\boldsymbol{x}$, we successively modify $\boldsymbol{x}$ so that the energy decreases after each modification.

Such matrices T_{ij} have been used in theories of linear associative nets (15–19) to produce an output pattern from a paired input stimulus, $S_1 \rightarrow O_1$. A second association $S_2 \rightarrow O_2$ can be simultaneously stored in the same network. But the confusing simulus $0.6\ S_1 + 0.4\ S_2$ will produce a generally meaningless mixed output $0.6\ O_1 + 0.4\ O_2$. Our model, in contrast, will use its strong nonlinearity to make choices, produce categories, and regenerate information and, with high probability, will generate the output O_1 from such a confusing mixed stimulus.

Now suppose that we flip the i^{th} component of $\boldsymbol{\sigma}$ from σ_i to $-\sigma_i = \sigma_i + \Delta\sigma_i$. Then the energy change caused by this flip is

$$\Delta E = -\Delta\sigma_i \sum_{j \neq i} W_{ij}\sigma_j.$$

Thus, if we want ΔE to be *negative*, we want $\Delta\sigma_i$ to always have the same sign as $\sum_{j\neq i} W_{ij}\sigma_j$, resulting in the update rule

$$\sigma_i^{(t+1)} = \operatorname{sgn}\left(\sum_{j \neq i} W_{ij}\sigma_j^{(t)}\right).$$

Since it is inconvenient to carry around the restriction $j \neq i$ in the summation, it is conventional to modify the definition (1) for the weight matrix:

$$\mathbf{W} := \frac{1}{N}\sum_{\mu=1}^{M} \boldsymbol{\xi}^{\mu}(\boldsymbol{\xi}^{\mu})^T, \quad W_{ii} = 0. \tag{2}$$

Note that this does not affect the interpretation of the energy as the sum of squares of dot products, since the diagonal terms always contribute a fixed amount to the energy, and since we care about the derivative of the energy and not its actual value, this constant could be subtracted without difficulty.