The Nobel Prize in Physics 2024

John Hopfield

“for foundational discoveries and inventions that enable machine learning with artificial neural networks”

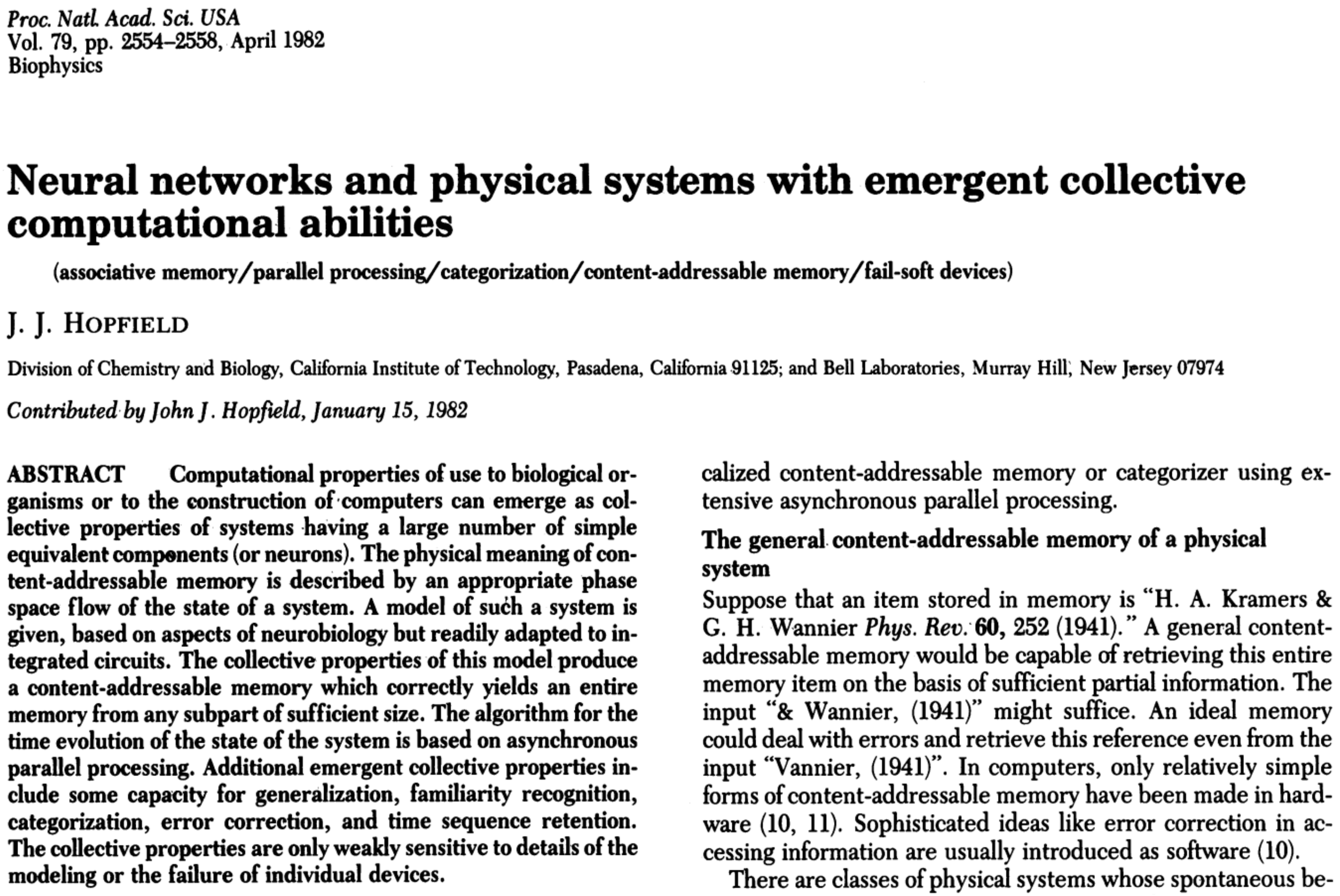
John Hopfield. Ill. Niklas Elmehed © Nobel Prize Outreach

Geoffrey Hinton

“for foundational discoveries and inventions that enable machine learning with artificial neural networks”

Geoffrey Hinton. Ill. Niklas Elmehed © Nobel Prize Outreach

Hopefield 1982 paper
Presented by Xun Shi
2024.10.16



Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 2554–2558, April 1982
Biophysics

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

Contributed by John J. Hopfield, January 15, 1982

ABSTRACT Computational properties of use to biological organisms or to the construction of computers can emerge as collective properties of systems having a large number of simple equivalent components (or neurons). The physical meaning of content-addressable memory is described by an appropriate phase space flow of the state of a system. A model of such a system is given, based on aspects of neurobiology but readily adapted to integrated circuits. The collective properties of this model produce a content-addressable memory which correctly yields an entire memory from any subpart of sufficient size. The algorithm for the time evolution of the state of the system is based on asynchronous parallel processing. Additional emergent collective properties include some capacity for generalization, familiarity recognition, categorization, error correction, and time sequence retention. The collective properties are only weakly sensitive to details of the modeling or the failure of individual devices.

calized content-addressable memory or categorizer using extensive asynchronous parallel processing.

The general content-addressable memory of a physical system

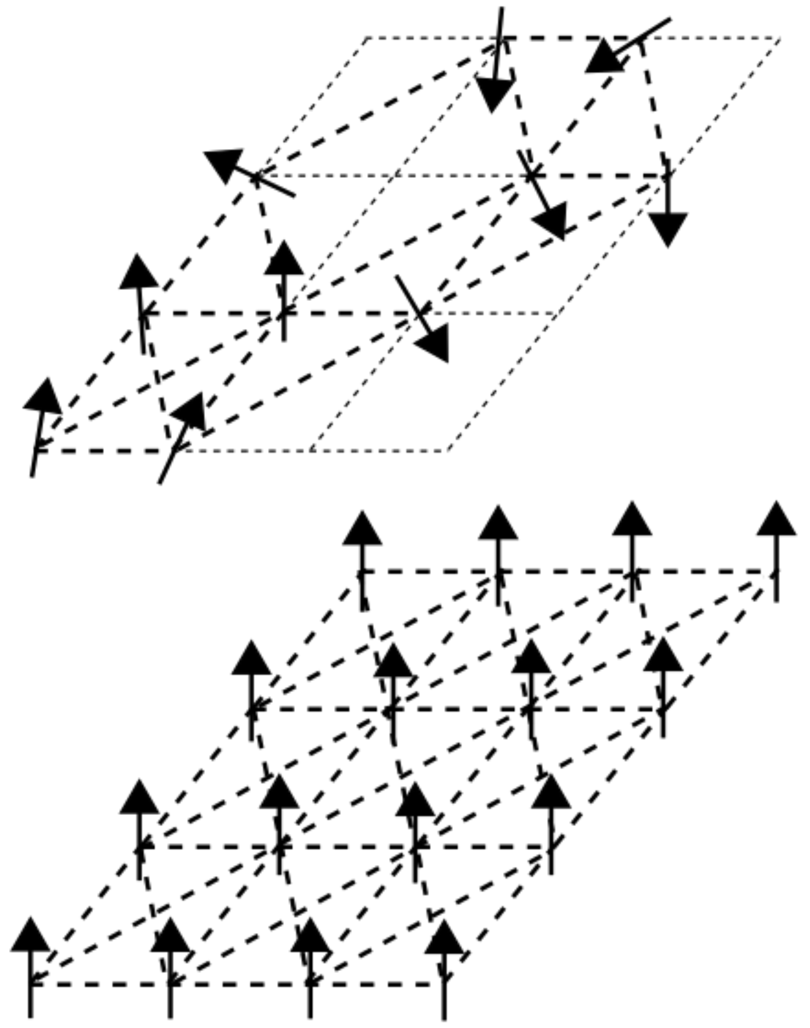
Suppose that an item stored in memory is "H. A. Kramers & G. H. Wannier *Phys. Rev.* **60,** 252 (1941)." A general content-addressable memory would be capable of retrieving this entire memory item on the basis of sufficient partial information. The input "& Wannier, (1941)" might suffice. An ideal memory could deal with errors and retrieve this reference even from the input "Vannier, (1941)". In computers, only relatively simple forms of content-addressable memory have been made in hardware (10, 11). Sophisticated ideas like error correction in accessing information are usually introduced as software (10).

There are classes of physical systems whose spontaneous be-

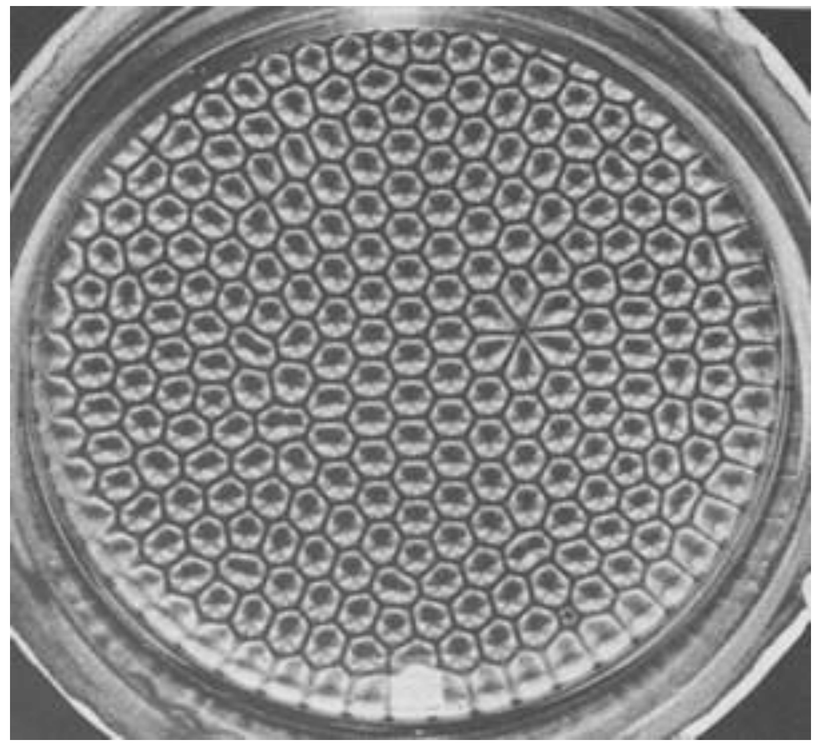
the "Hopfield network"

Collective phenomena in physics:

物理中的集体自组织现象

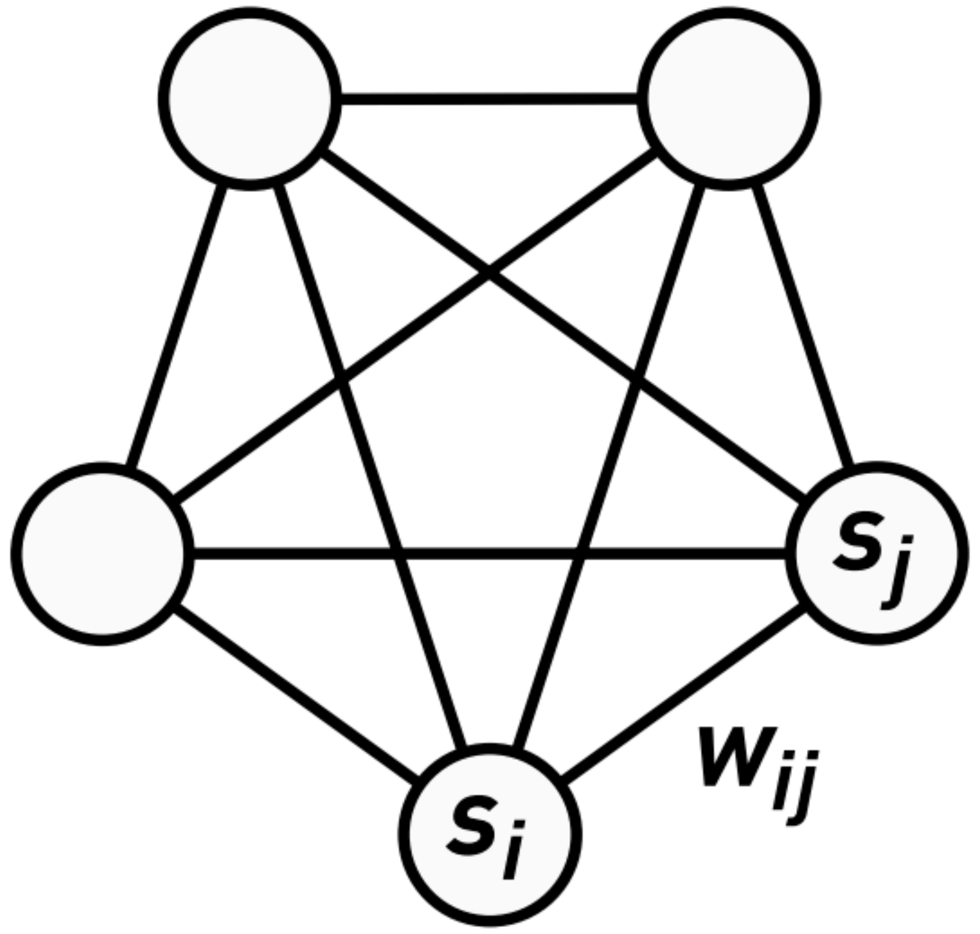


自旋玻璃



贝纳德原胞

A network of neurons (~ our brain) can memorize — could this be an emergent property and collective in origin?



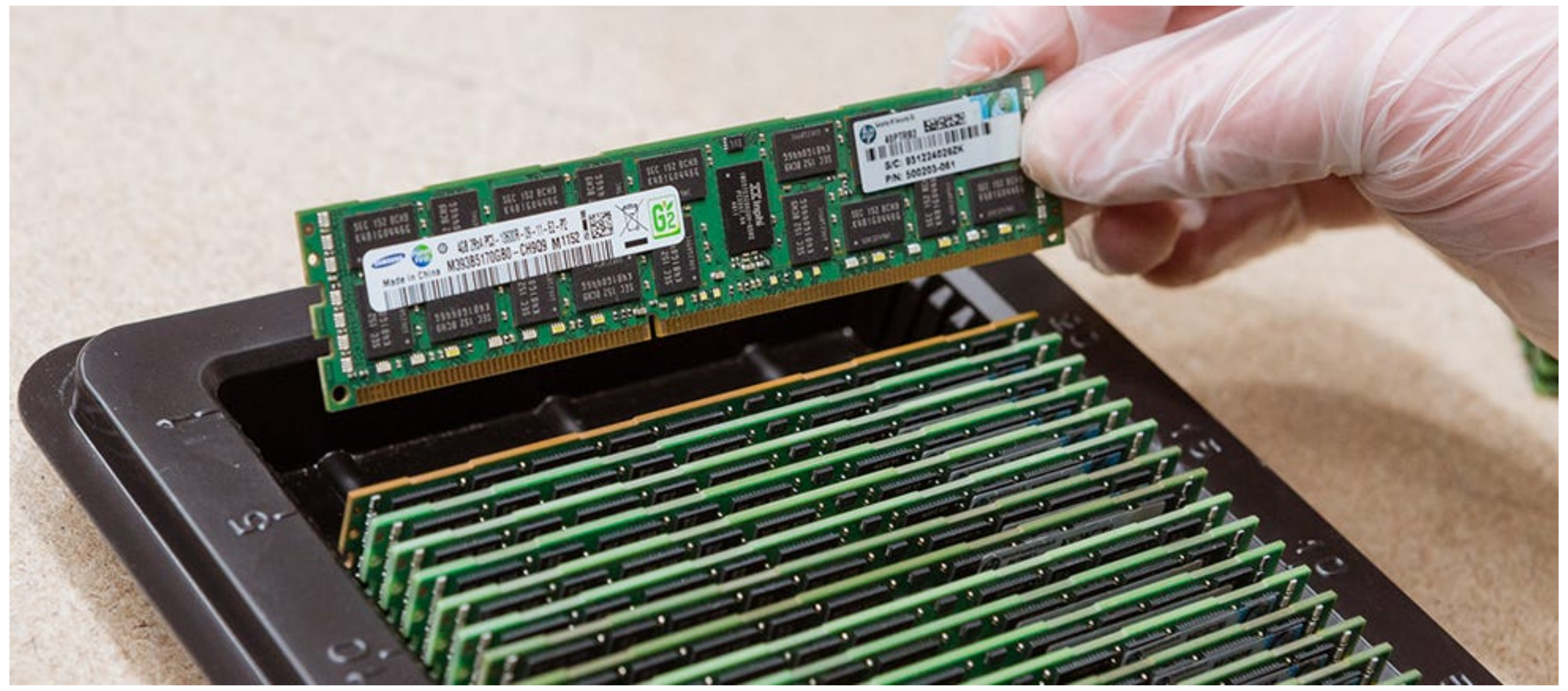
Can this memorize?

一个简单的神经网络可以自发产生记忆的本领吗?

Two types of memories

“random access memory” RAM 随机存储内存

— Input address, output content

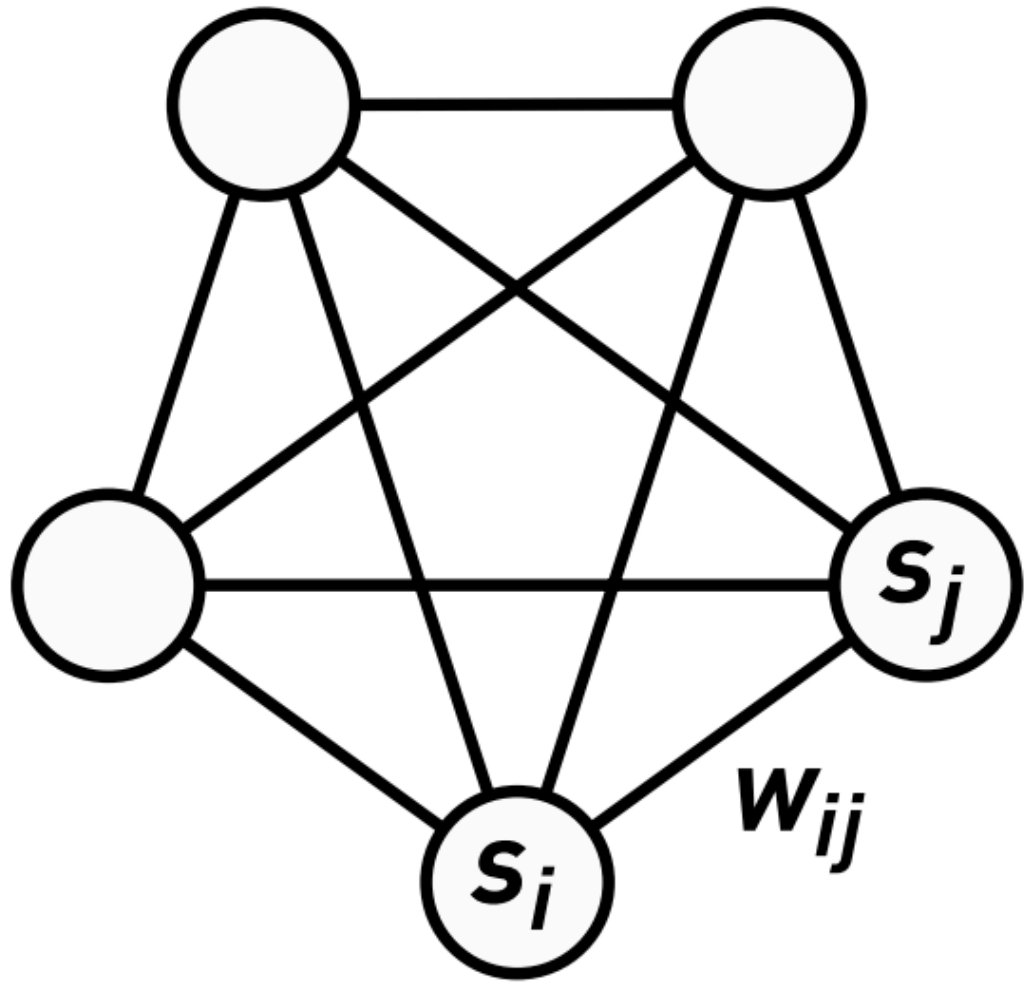


“content-addressable memory CAM 内容可寻址内存” / “associative memory” 联想内存

— Input “stimulus”, output content

A new paradigm of information storage and retrieval

A simple model of this process:



the “Hopfield network”

- A neuron has a state s_i (e.g. 0 or 1)
- Interaction ~ a weight between a neuron pair w_{ij}

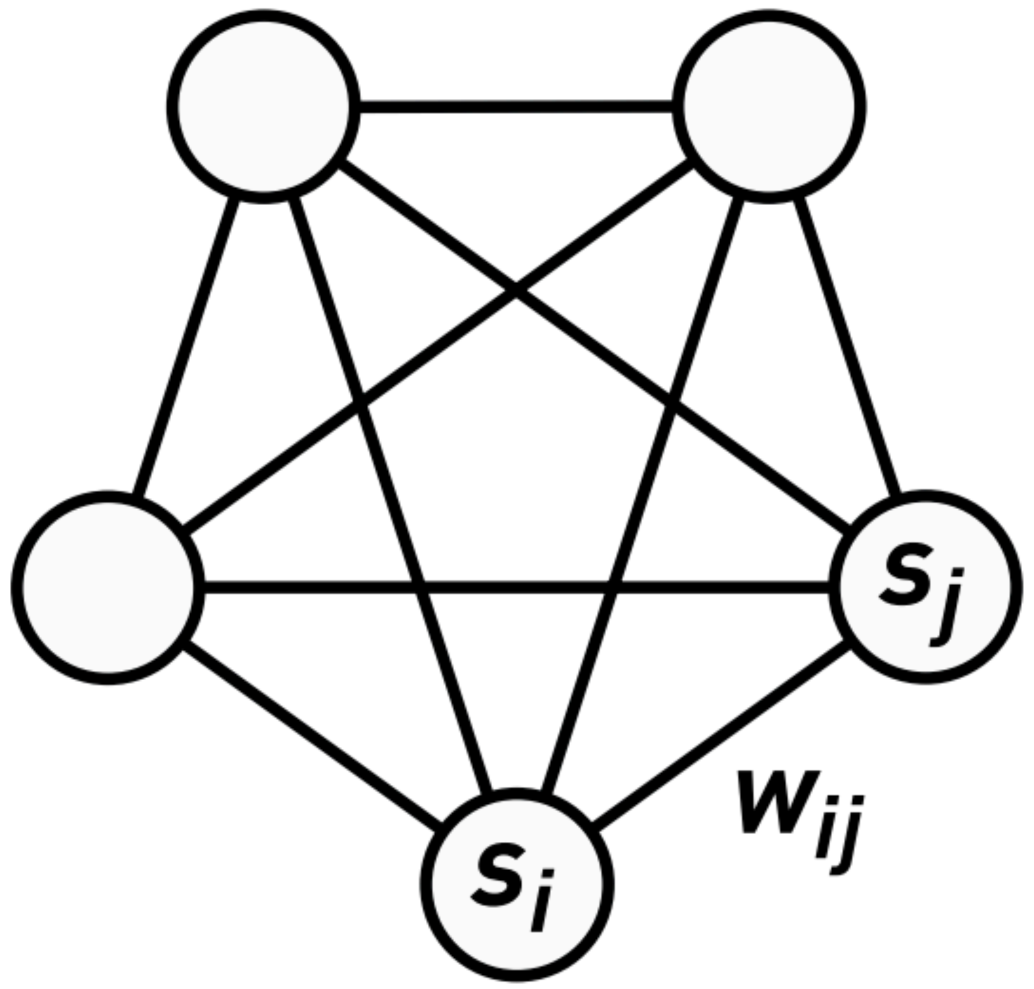
— N-body system

- Neurons are updated one at a time as a result of all interactions $s_{i,t=1} = f\left(\sum_{i \neq j} w_{ij} s_{j,t=0}\right)$
- f(x) is highly nonlinear e.g. a Heaviside function

— non-linear dynamics

- “memory” ~ a steady state of the network
- What the network memorizes is determined by the weights w_{ij}

To memorize and retrieve memory



the “Hopfield network”

- To let the network memorize a pattern $\{s_i^A\}$:

Let the weight be $w_{ij} = (2s_i^A - 1)(2s_j^A - 1)$

~the outer product (the “Hebb rule”)

- To let the network memorize two patterns $\{s_i^A\}$ and $\{s_i^B\}$:

Let the weight be $w_{ij} = (2s_i^A - 1)(2s_j^A - 1) + (2s_i^B - 1)(2s_j^B - 1)$

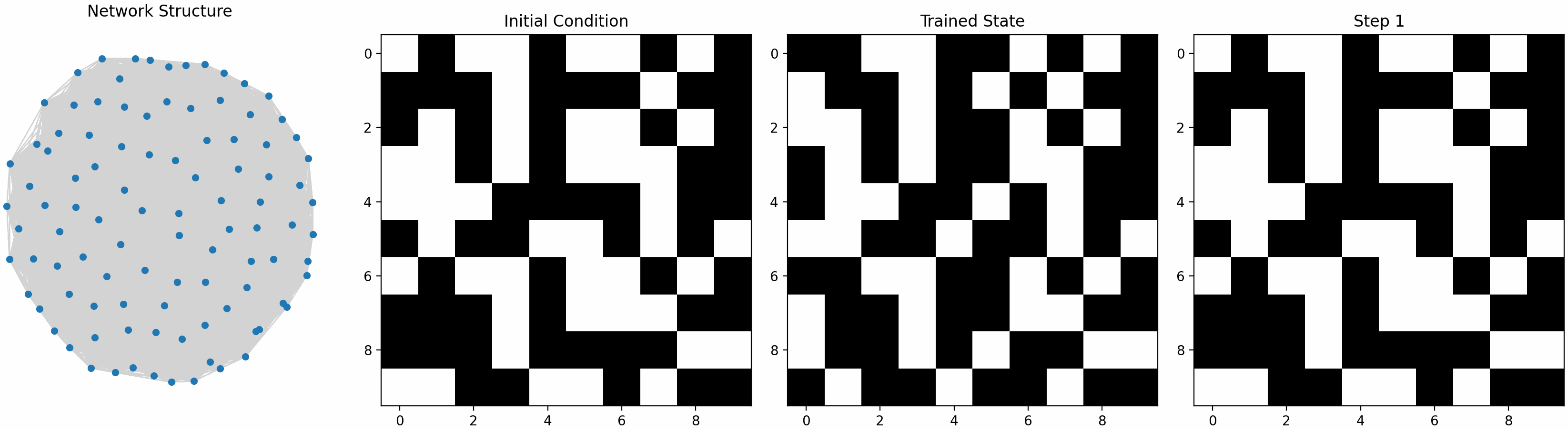
—the network learns the two-point correlations

One can prove that the stored states $\{s_i^A\}$ and $\{s_i^B\}$ are stable under the updating algorithm.

The memories can be retrieved by inputing similar patterns (“stimulus”).

— “content-addressable memory CAM 内容可寻址内存”

An example with N=100 neurons (Hopfield 1982 used N=30)



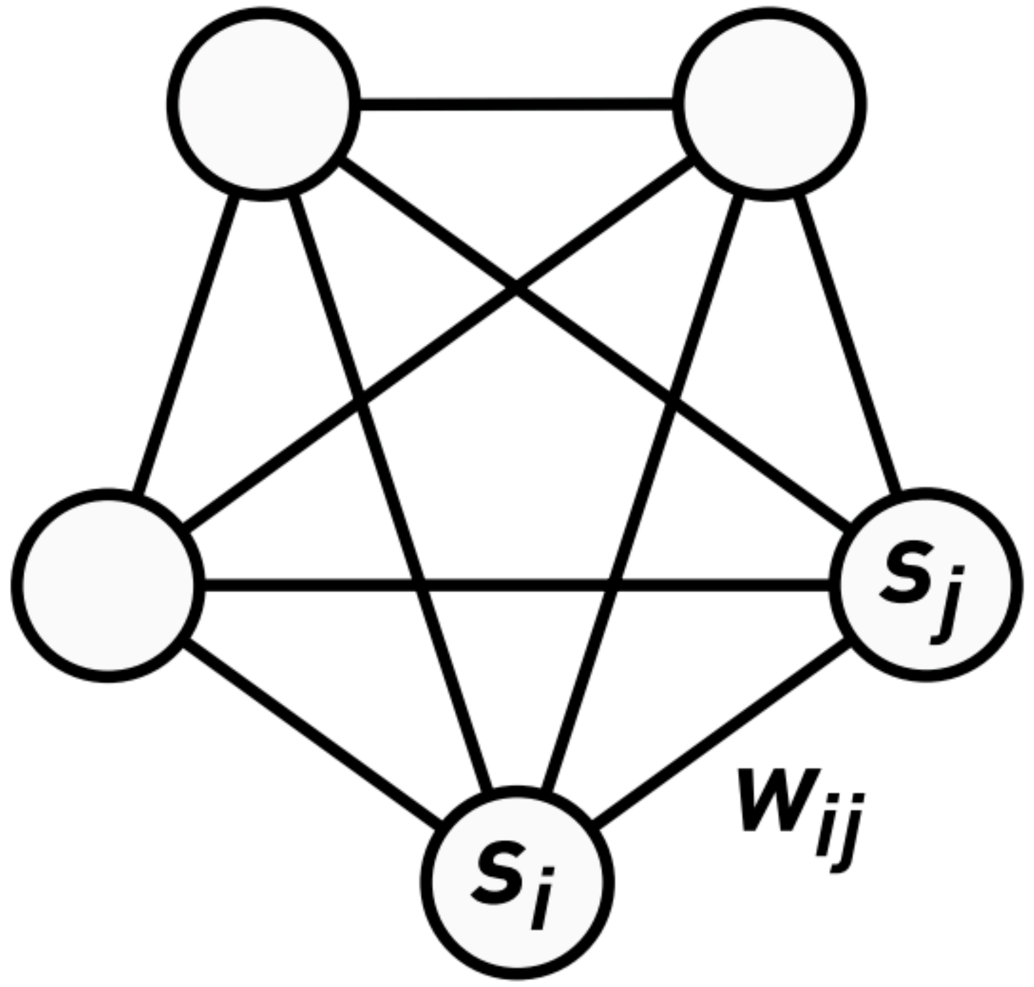
“Stimulus”
刺激

“Memory”
记忆

From stimulus to memory
通过刺激找到记忆

Code credit: GPT-4o

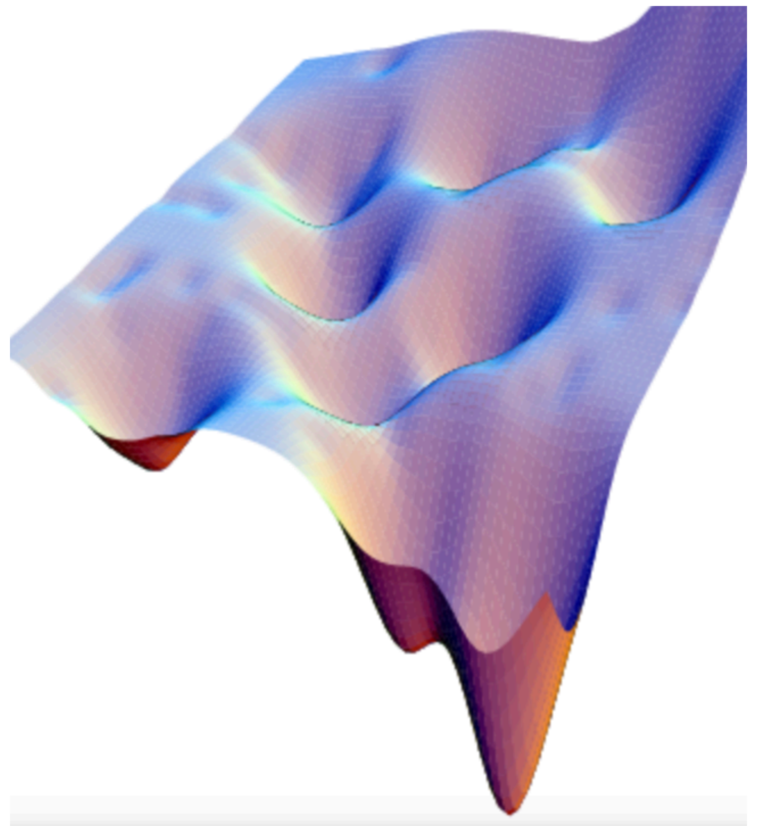
Why can this memorize?



the “Hopfield network”

An energy can be defined of the network

$$E = -\frac{1}{2}\sum_{i \neq j}\sum_{j} w_{ij}s_is_j$$

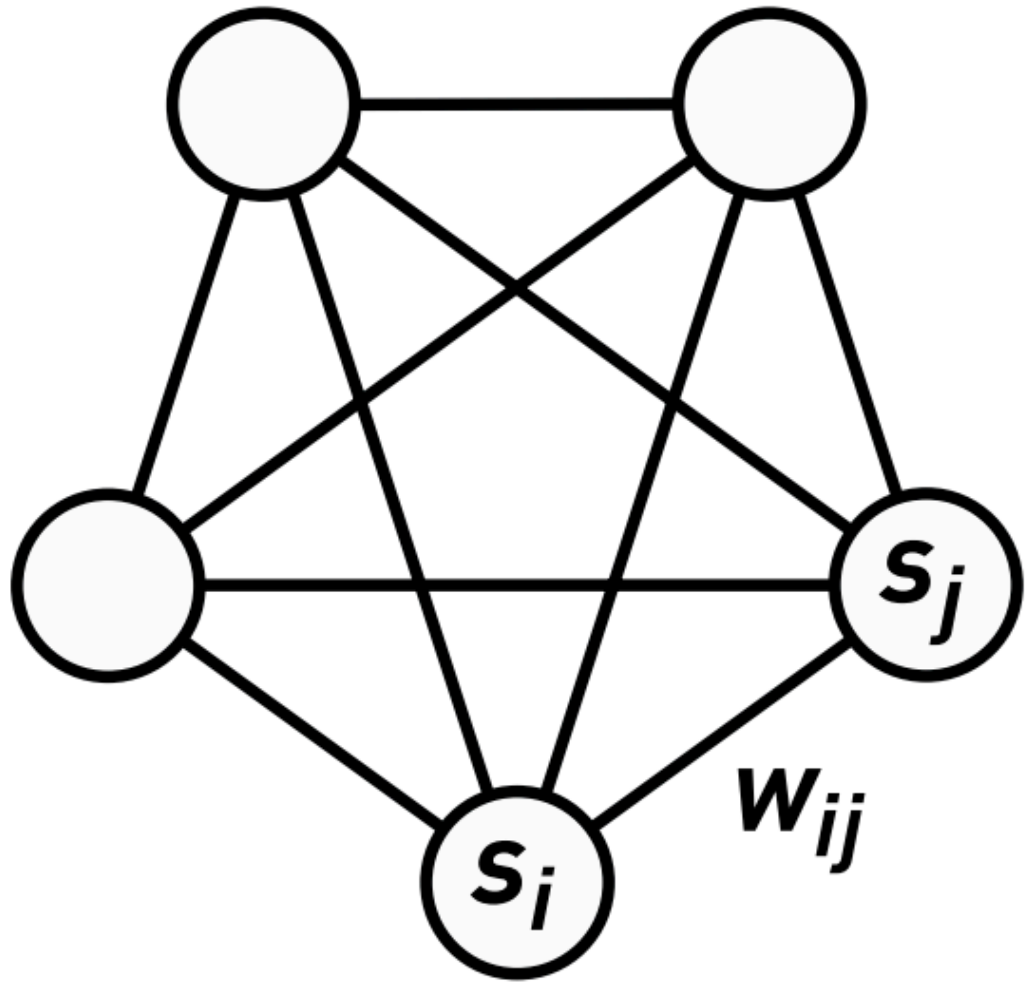


memories ~ potential wells in the state space

- When w_{ij} is symmetric and deterministic: E decreases when the network updates (~ Ising model)
- When w_{ij} is symmetric and random (~ spin glass): many locally stable states exist
- When w_{ij} is asymmetric: richer dynamics (limiting points/cycles, chaotic wandering)

In all cases, the flow in the phase space has the necessary properties for a physical content-addressable memory.

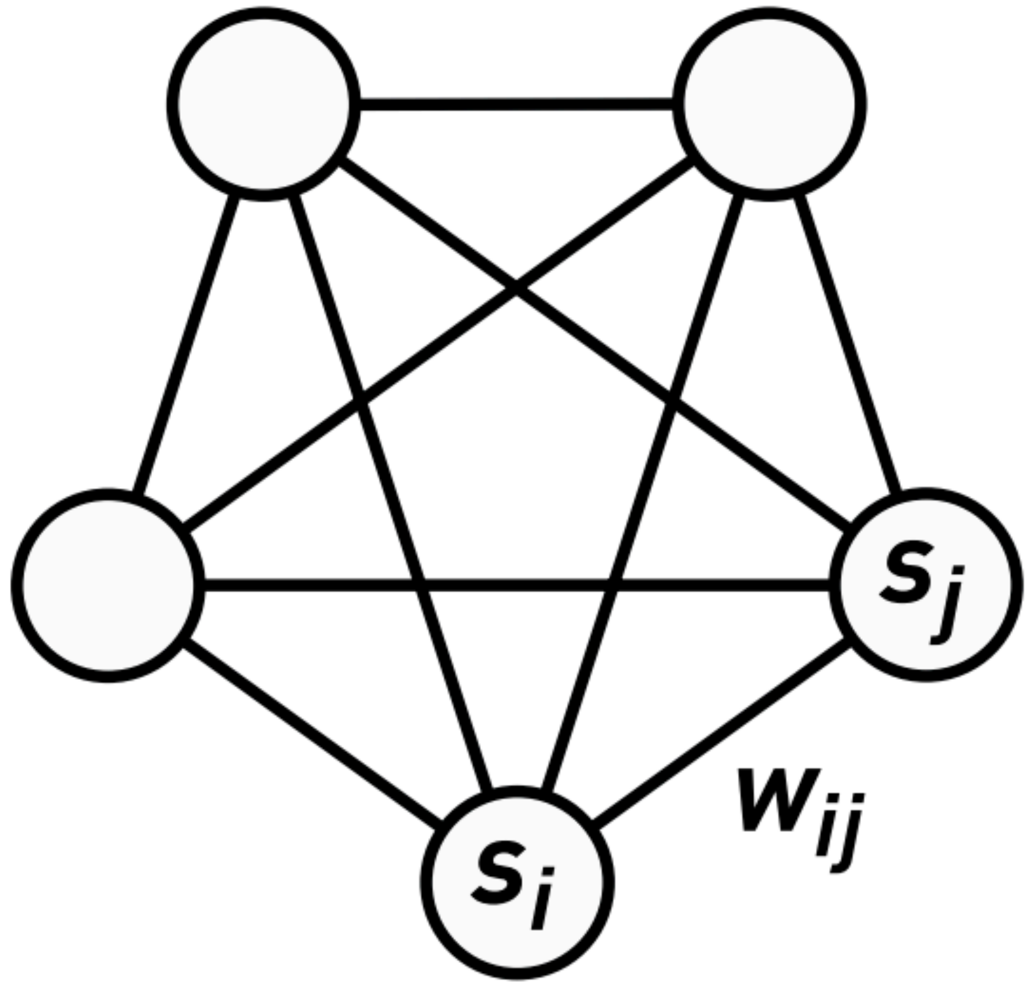
Properties



the "Hopfield network"

- Robust to model details
- New memories can be added
- A network of a certain size can saturate
- Can work with "brain damage"
- Memories too close to each other tend to confuse and merge
- Not entirely deterministic

What can it do



the “Hopfield network”

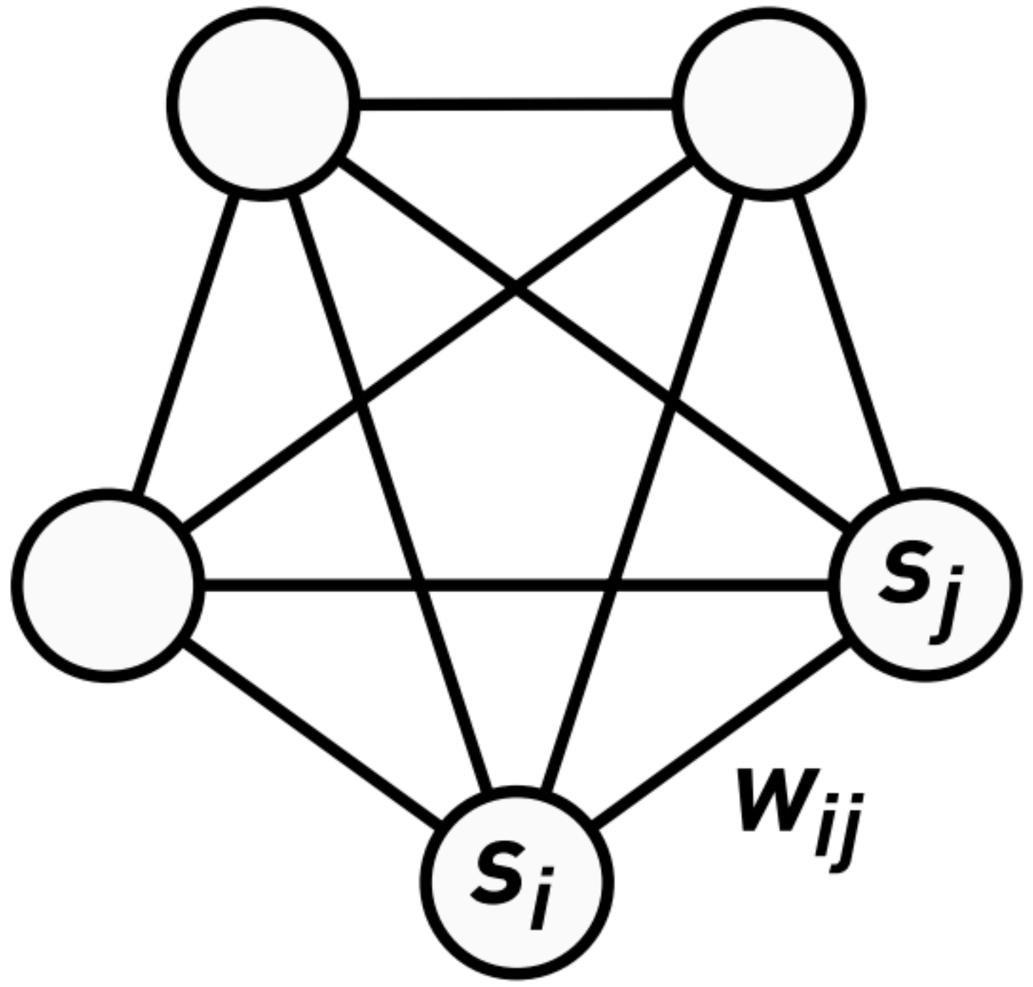
As associative memory 联想内存

- ➤ pattern completion 图案补全
- ➤ familiarity recognition 识别
- ➤ categorization 分类
- ➤ error correction 纠错
- ➤ time sequence retention 时间顺序保留

As a minimizer of an energy function

- ➤ Solve optimization problems 优化问题求解

e.g. Hopfield and Tank (1985) traveling salesman problem



In this model network each “neuron” has elementary properties, and the network has little structure. Nonetheless, collective computational properties spontaneously arose.