Classical Hopfield Networks and Sparse Distributed Memory in Pattern Completion

Henrik Formoe, Anna Mihály, & Morgan Fitzgerald

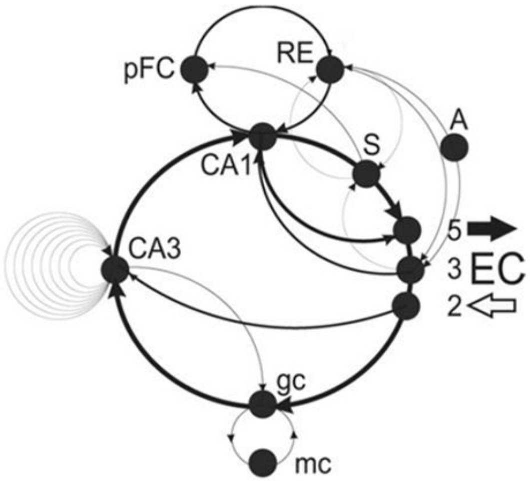
Mentor: Nicolai Haug

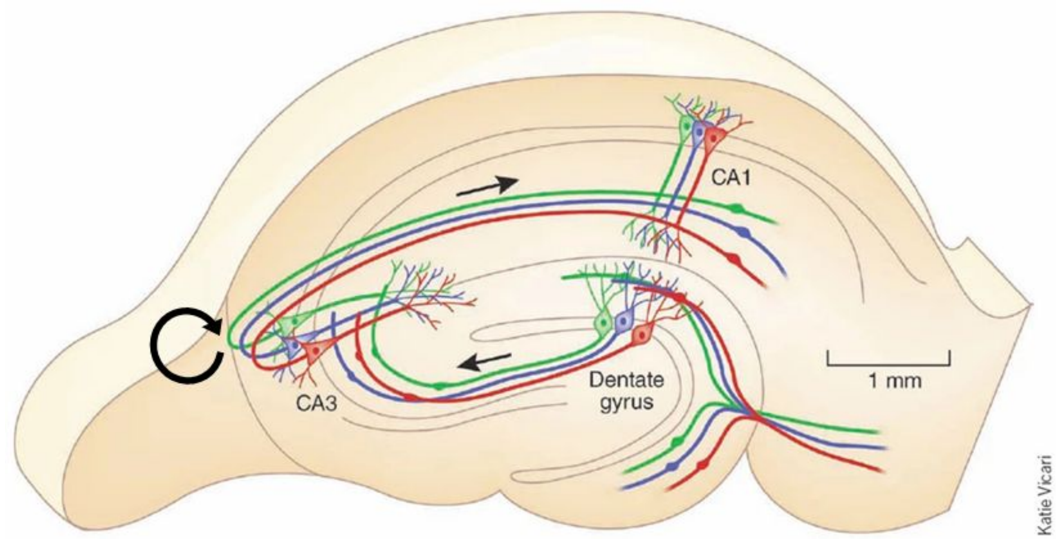
Simula Summer School 2025

sim ? ?

simula

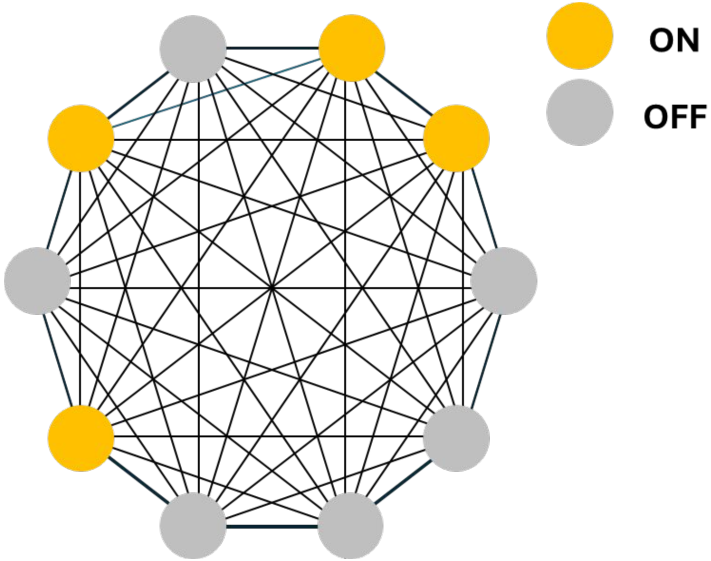
Pattern completion: retrieving and entire memory from a partial or noisy input

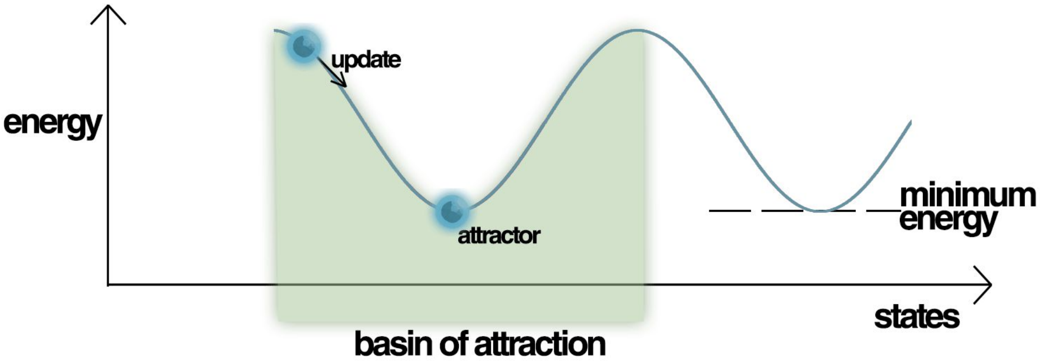


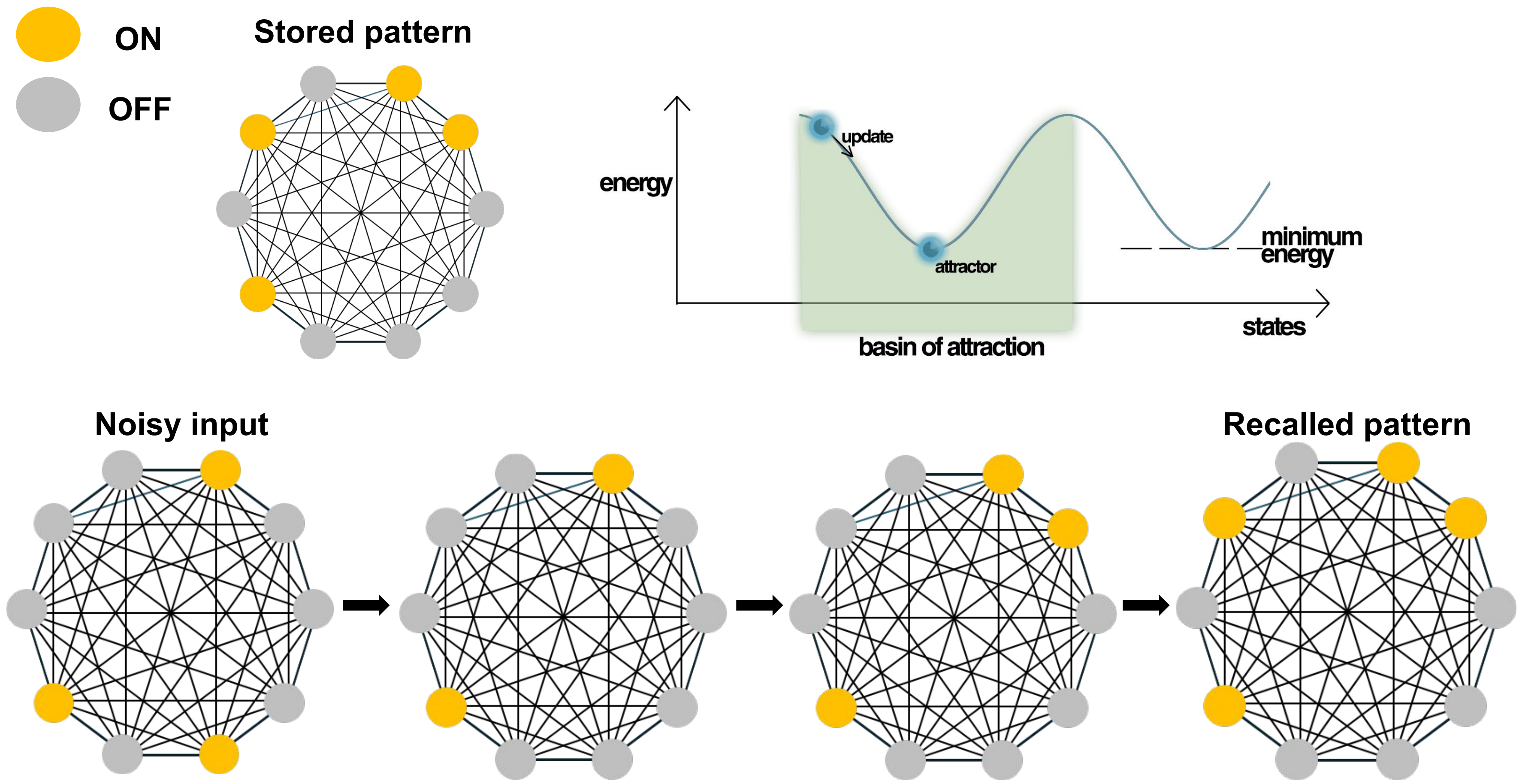


(Buzsáki, 2011)

Classical Hopfield Network: a fully connected recurrent network using Hebbian learning

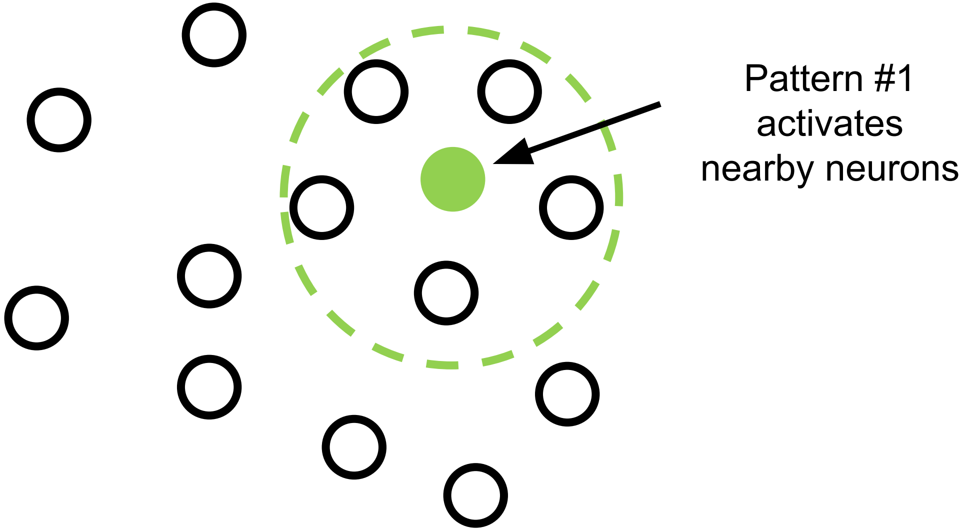




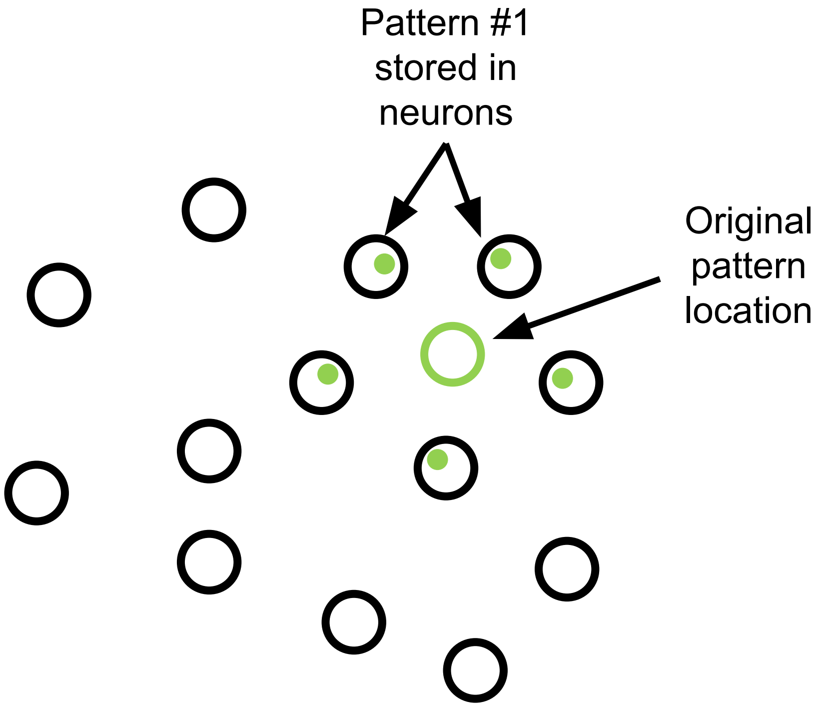

ON
OFF
Stored pattern
energy
update
attractor
minimum energy
states
basin of attraction
Noisy input
Recalled pattern

Kanerva’s spare distributed memory (SDM)

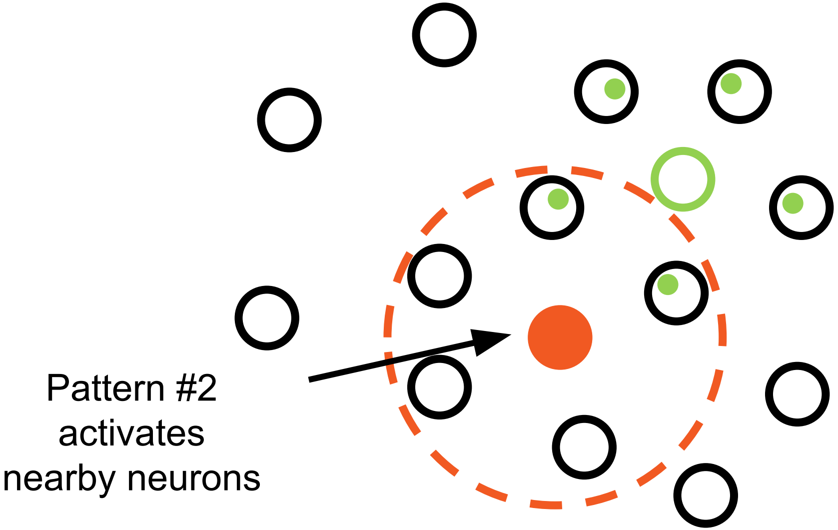
- High dimensional binary vector space
- Uses Hamming distance
- Patterns are stored in nearby neurons and then disappear



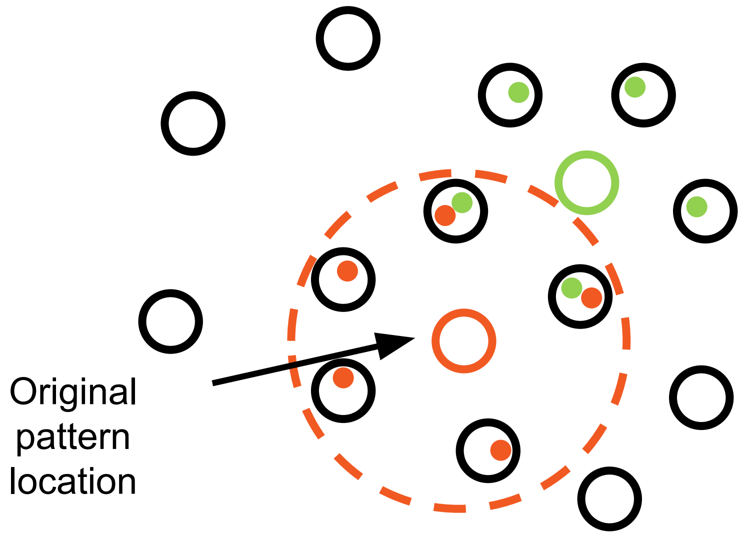
SDM writing operation



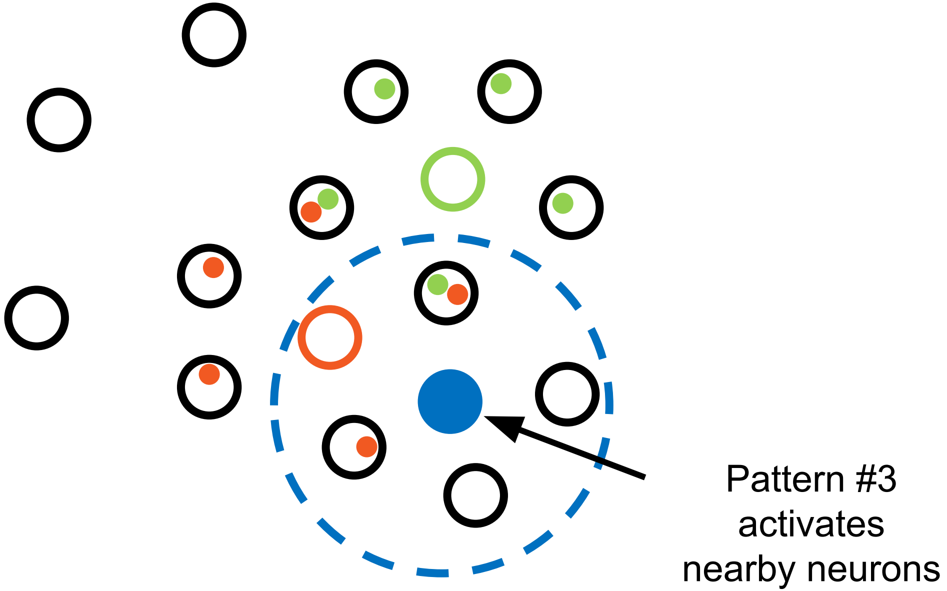
SDM writing operation



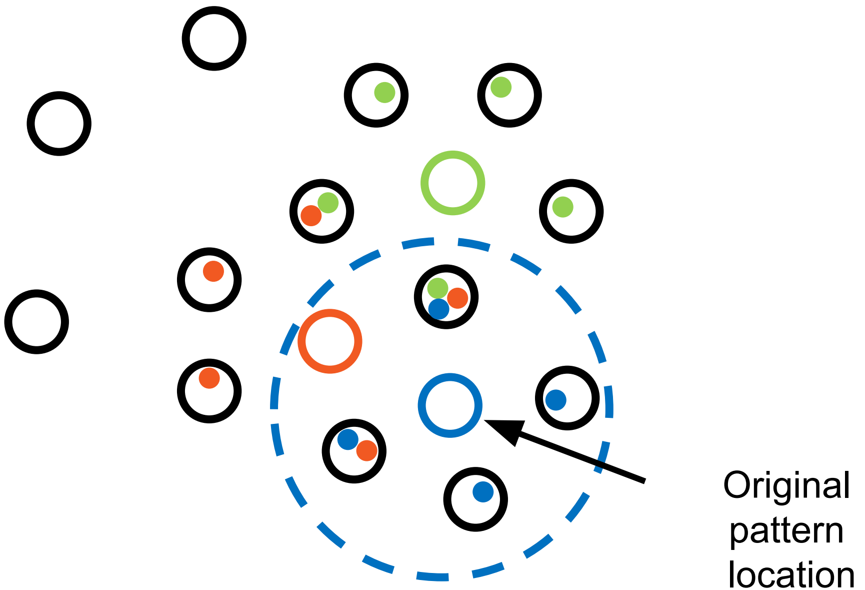
SDM writing operation



SDM writing operation

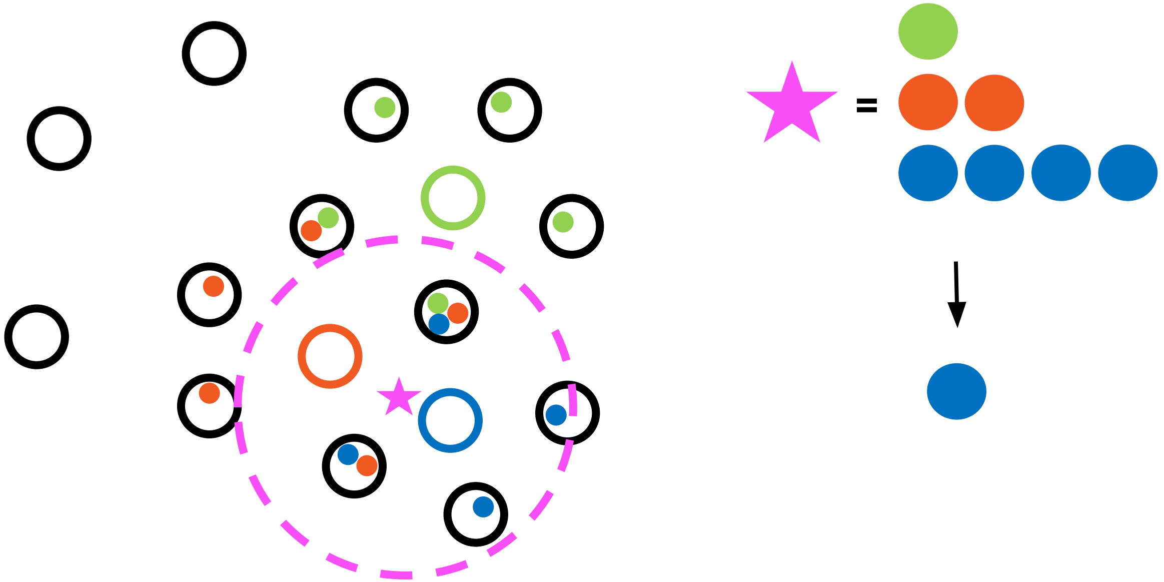


SDM writing operation



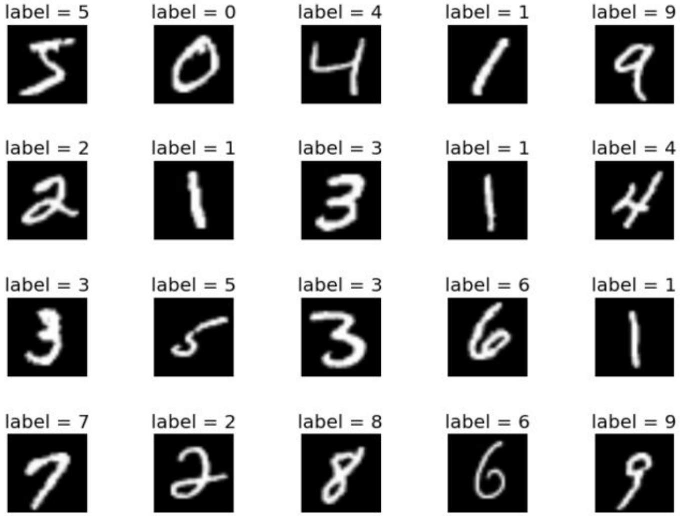
SDM reading operation

- STAR is a vector that activates all nearby neurons within its radius
- Activated neurons' output are the stored patterns
- Majority pattern is blue, so the output will converge to blue



Our experimental goal

- Build the Classical Hopfield Network and Kanerva’s SDM model
- MNIST dataset
- Add different levels of noise
- Reconstruct the clean digit



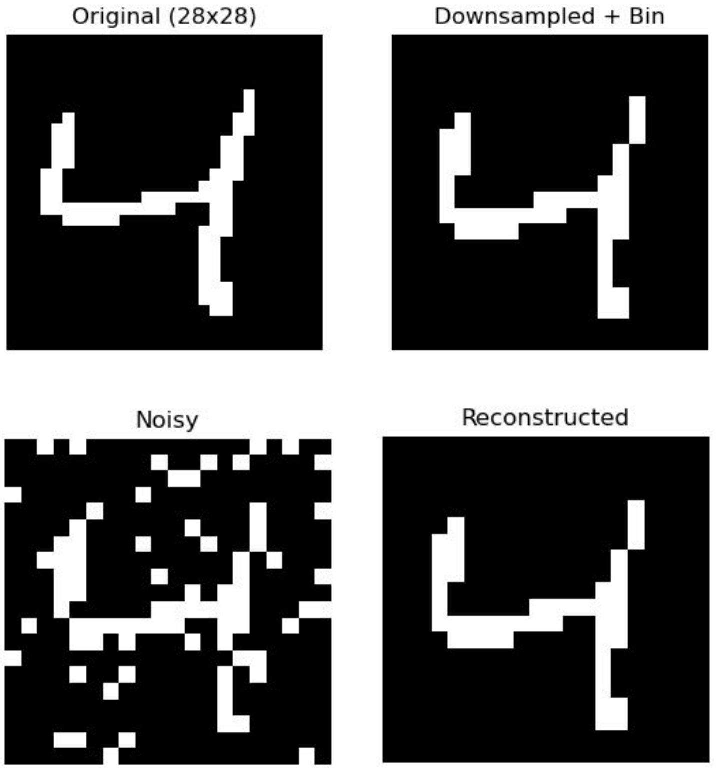
The classical Hopfield Network: model and task

MODEL

- Bipolar units
  - Active (+1)
  - Inactive (-1)
- Hebbian learning
- Synchronous recall

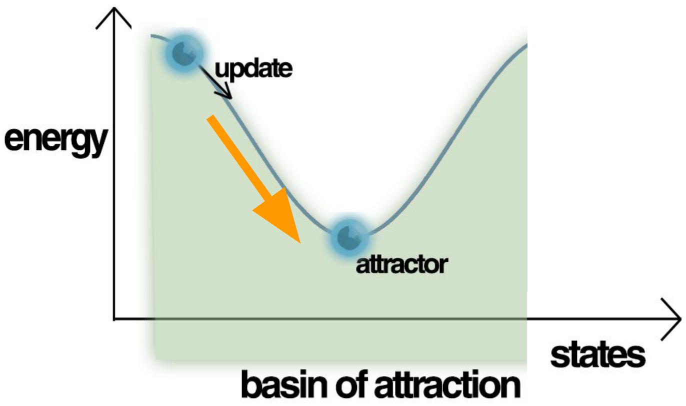
TASK

- **See if the network can recall clean patterns when given corrupted versions.**

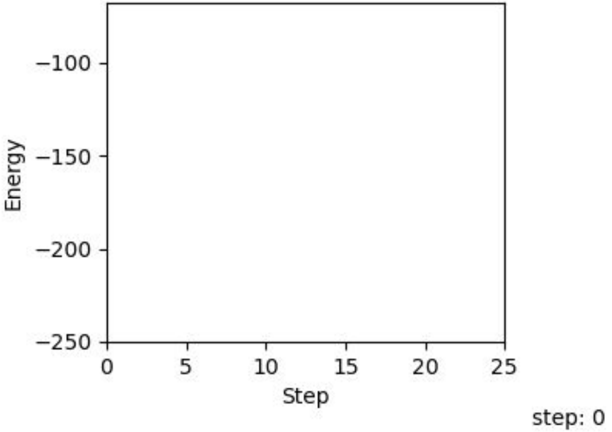


Example Noise Degradation of Digit 4

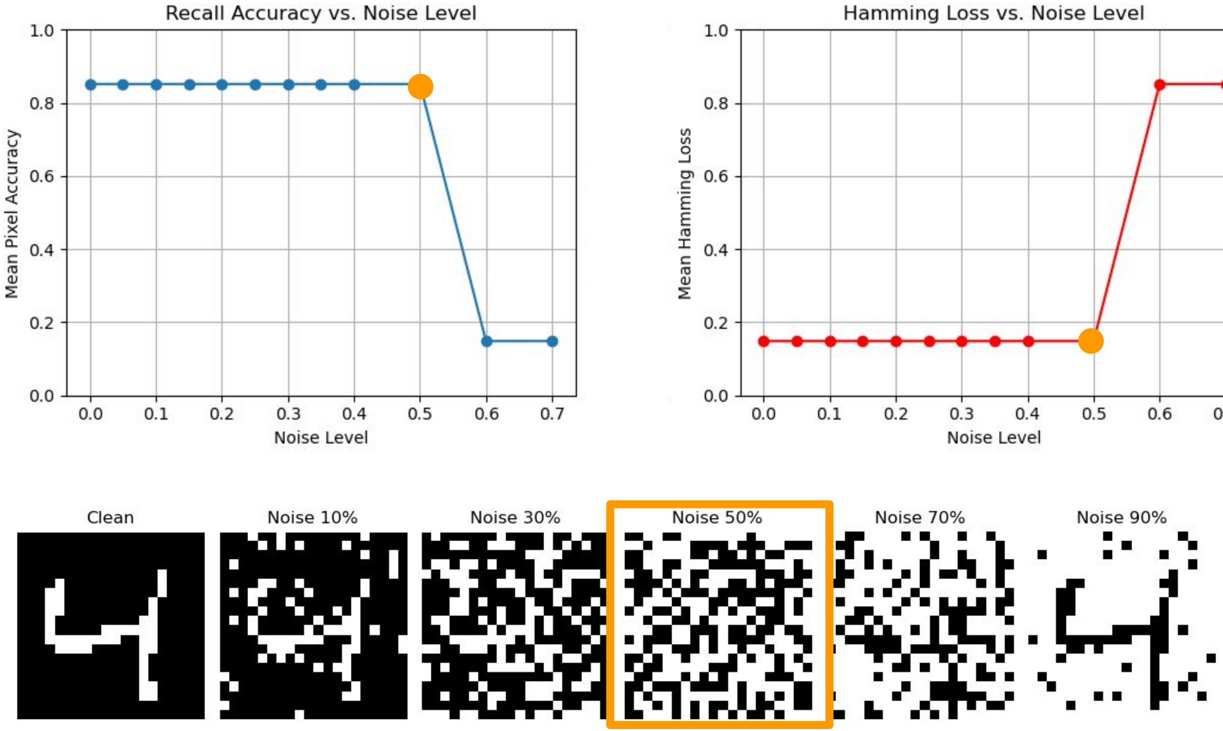
Given a noisy cue, the network retrieves the closest stored pattern



Starting state:
noisy bipolar image



The classical Hopfield Network recall performance degrades with noise



HOPFIELD NETWORK CAPACITY

Bottleneck:
How many patterns a network can store and still recall accurately.

Theoretical Capacity:

$$p\text{max} \approx 0.138 \times N$$

- pmax = is the maximum number of storable patterns (without errors)
- N is the number of neurons (or binary features in each pattern),

HOPFIELD NETWORK CAPACITY

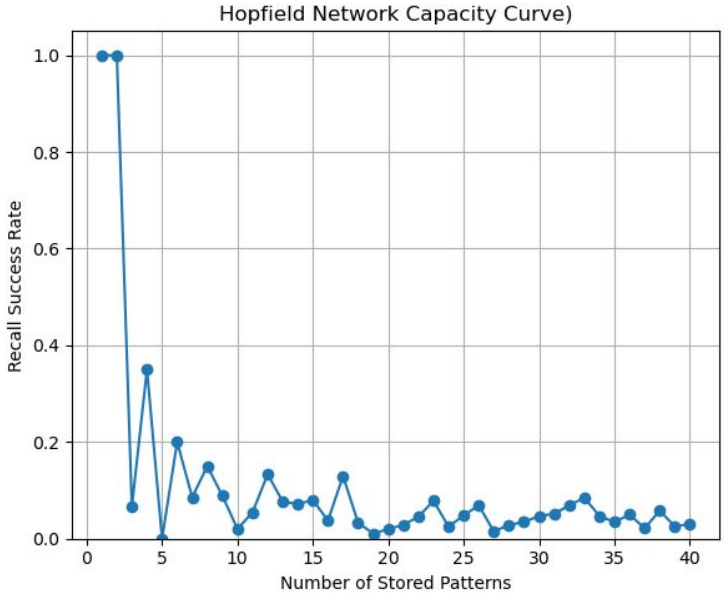
Downsampled 20×20 images

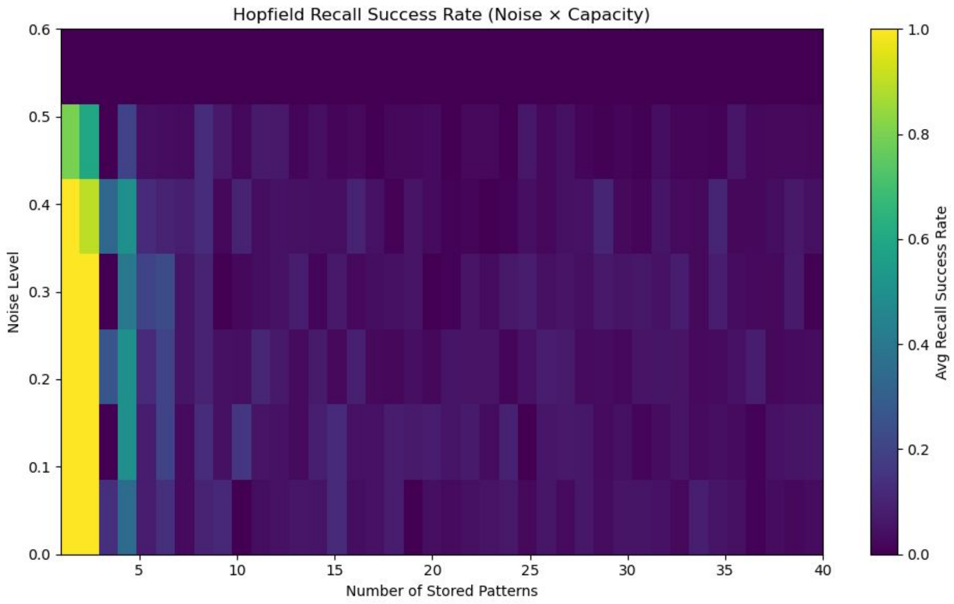
N=400 binary pixels per pattern (aka neurons)

So the theoretical maximum is:

$$\text{pmax} \approx 0.138 \times 400 = \mathbf{55.2}$$

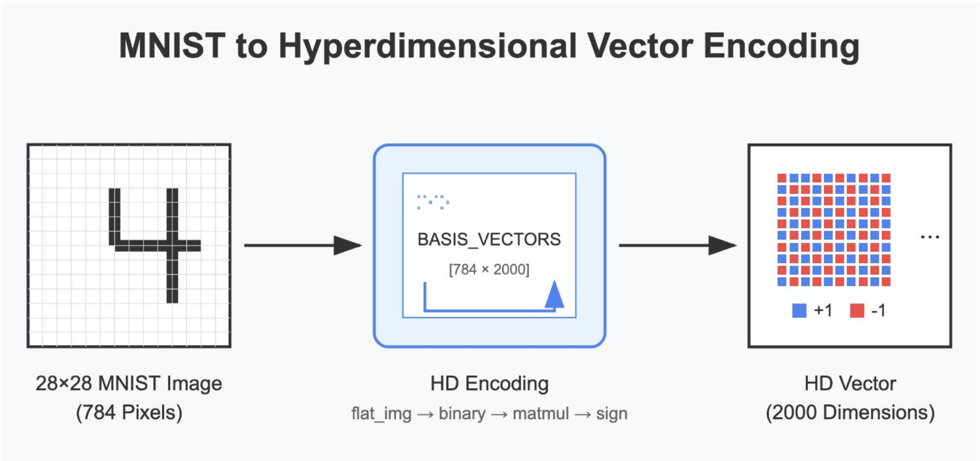
The classical Hopfield Network recall performance degrades with number of stored patterns and noise





Small region of reliable operation, and both noise and pattern load quickly drive the network to failure.

Configuring the Sparse Distributed Memory model



- Dimensionality of Hypervectors:

  2,000

- Number of Memory Locations:

  100,000

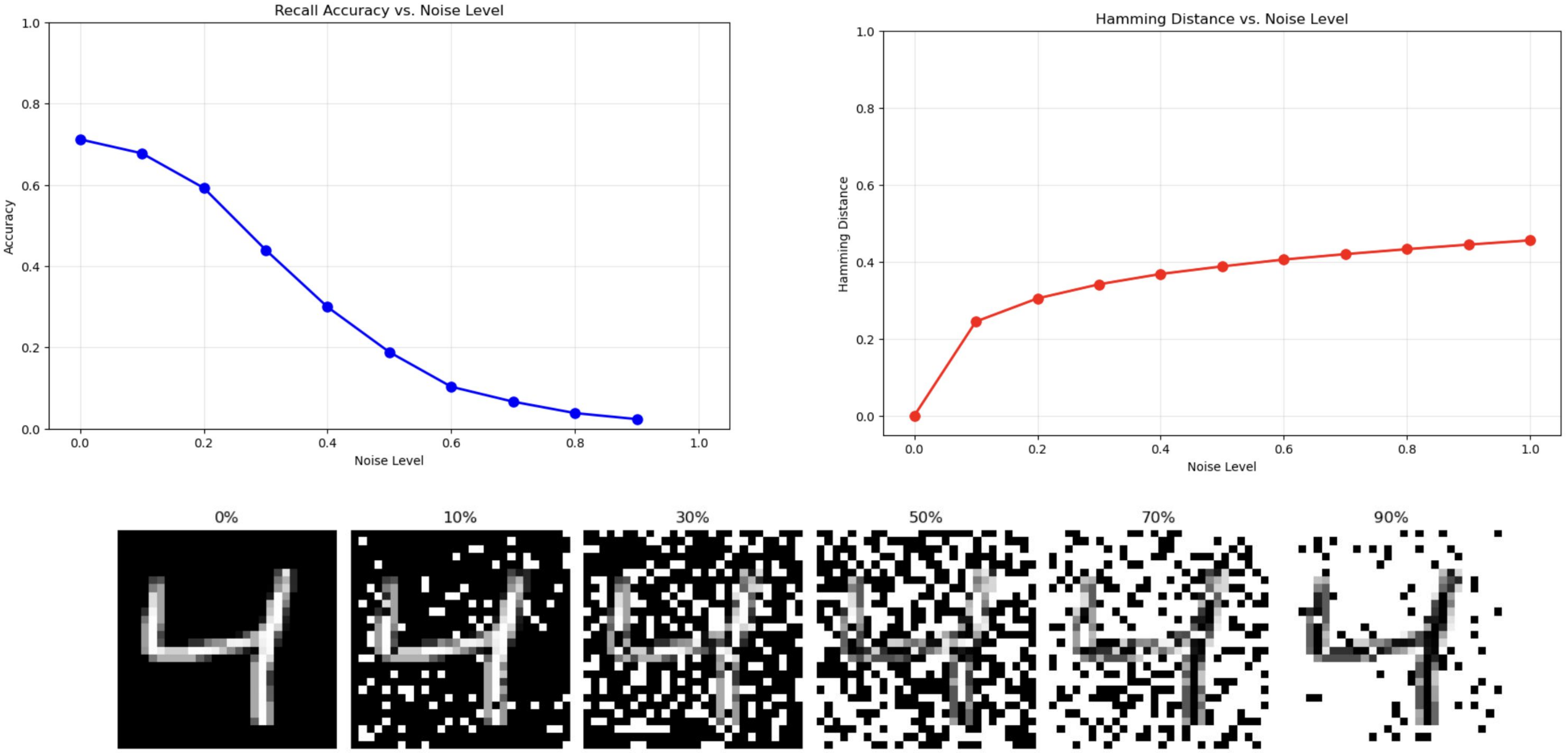
- The expected fraction of memory address that will contain any value.

  $p = 0.005$

- Subset of memory locations likely to be activated

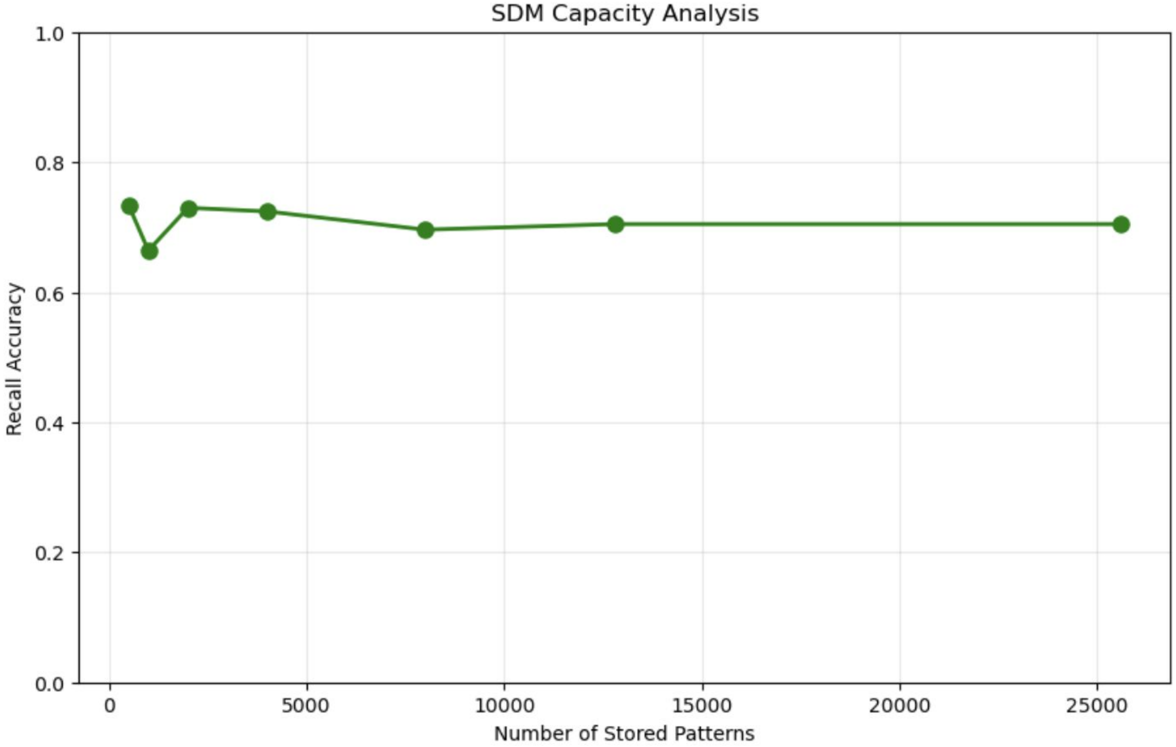
  $0.005 \times 100{,}000 = 500$

Recall accuracy gradually decreases with increasing noise



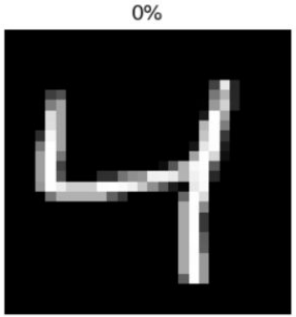
Sparse Distributed Memory Model

Model capacity: initial degradation followed by stable effects of additional stored patterns



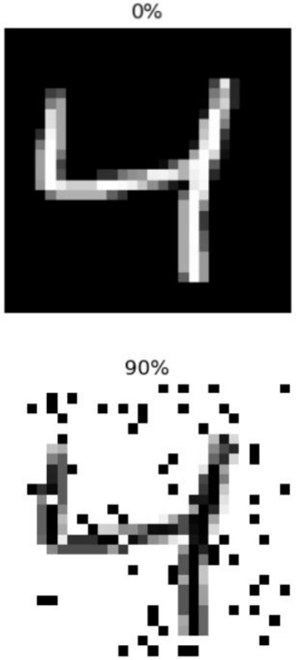
Sparse Distributed Memory Model

Model characteristics and memory recall



- Performance declines gradually as cues become less accurate – compared to the more all-or-nothing retrieval by the Hopfield network
- Failure to recognize inverse representations despite preserved structure

Model characteristics and memory recall

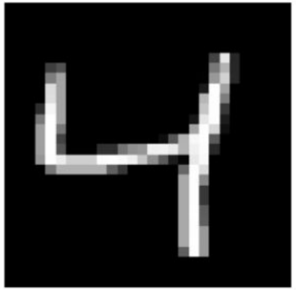


- Performance declines gradually as cues become less accurate – compared to the more all-or-nothing retrieval by the Hopfield network
- Failure to recognize inverse representations despite preserved structure

- Memory recall: the mental process of retrieving information from the past
- Cued recall: retrieving memories when provided with a specific prompt or cue

Model characteristics and memory recall

0%



90%

- Performance declines gradually as cues become less accurate – compared to the more all-or-nothing retrieval by the Hopfield network
- Failure to recognize inverse representations despite preserved structure

- Memory recall: the mental process of retrieving information from the past
- Cued recall: retrieving memories when provided with a specific prompt or cue

Kanerva's SDM is fundamentally a cued recall system:

Input as Cue → Associative Activation → Retrieval Process → Pattern Completion

The classical Hopfield Network and the Sparse Distributed Memory model as Models of Hippocampal Memory

Purpose

- Compare classical Hopfield networks and Sparse Distributed Memory (SDM) as associative memory models of the hippocampus.

Method:

- Tested pattern completion on noisy MNIST digit inputs.

Goal

- Assess robustness and accuracy under varying noise levels and patterns

Models of associative memory

Feature	Classical Hopfield Network	Sparse Distributed Memory

"The brain works with memory based on two contradictory principles. Parts of the brain try to make as much of the information as possible into something that is similar and categorizable to save space, while the hippocampus fights to preserve the uniqueness of events."

Quote by Anders Fjell (University of Oslo) in the book “Adventures in Memory: The Science and Secrets of Remembering and Forgetting" (Østby & Østby, 2018).