Mathematics for Artificial Intelligence

Reading Course

<u>**Elena Agliari**</u>
Dipartimento di Matematica
Sapienza Università di Roma

(TENTATIVE) PLAN OF THE COURSE

Introduction

Chapter 1: Basics of statistical mechanics.
  The Curie-Weiss model

Chapter 2: Neural networks for associative memory and pattern recognition.

Chapter 3: The Hopfield model
  Hopfield model with low-load and solution via log-constrained entropy.
  Hopfield model with high-load and solution via stochastic stability.

Chapter 4: Beyond the Hebbian paradigma

Chapter 5: A gentle introduction to machine learning
  Maximum likelihood
  Rosenblatt and Minsky&Papert perceptrons.

Chapter 6: Neural networks for statistical learning and feature discovery.
  Supervised Boltzmann machines.
  Bayesian equivalence between Hopfield retrieval and Boltzmann learning.

Chapter 7: A few remarks on unsupervised learning, “complex” patterns, deep learning
  Unsupervised Boltzmann machines.
  Non-Gaussian priors.
  Multilayered Boltzmann machines and deep learning.

Seminars: Numerical tools for machine learning; Non-mean-field neural networks; (Bio-)Logic gates; Maximum entropy approach, Hamilton-Jacobi techniques for mean-field models, …

Chapter II
Neural networks for associative memory and pattern recognition

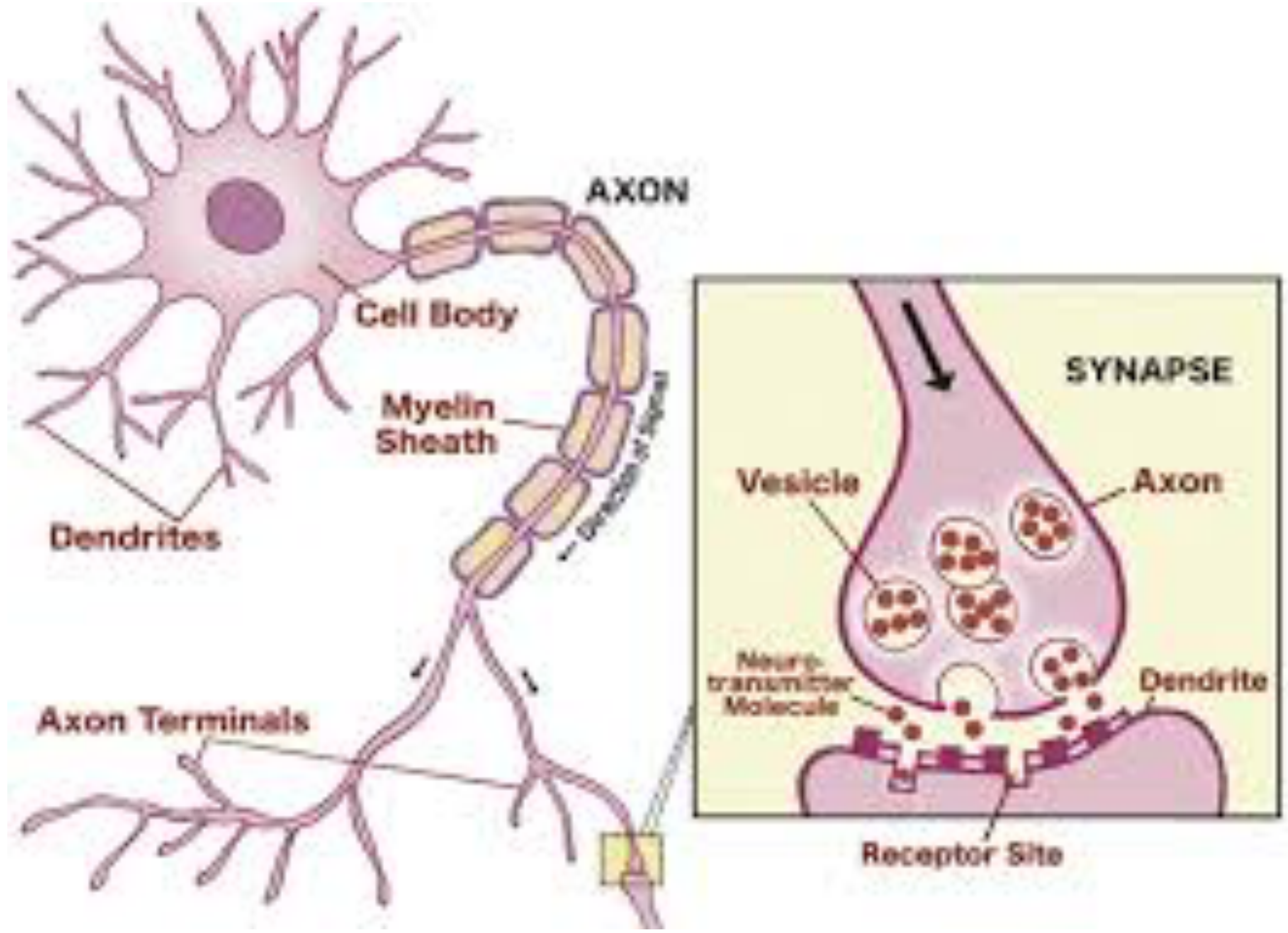
A few biological facts

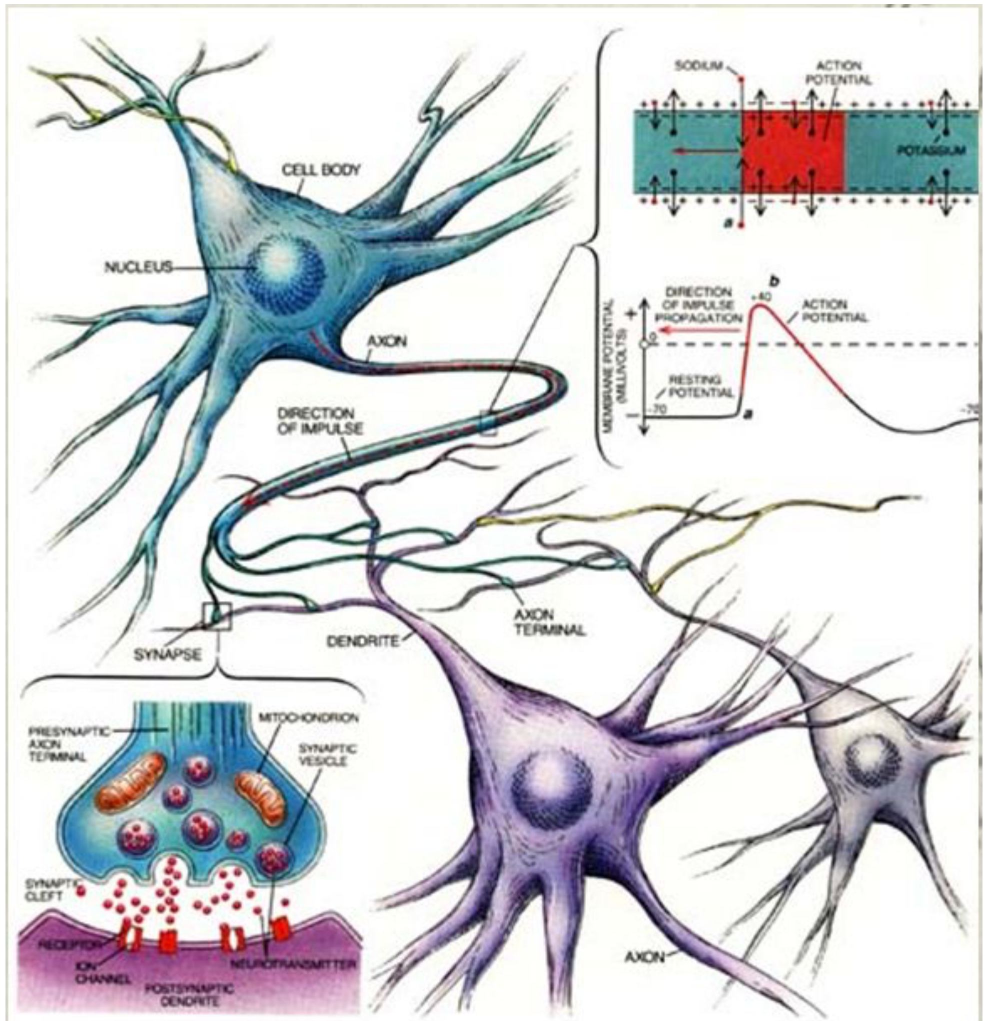
Neurons are big cells covered by a membrane to which are attached different fibers emitting electrical spikes generated by the neuron itself.

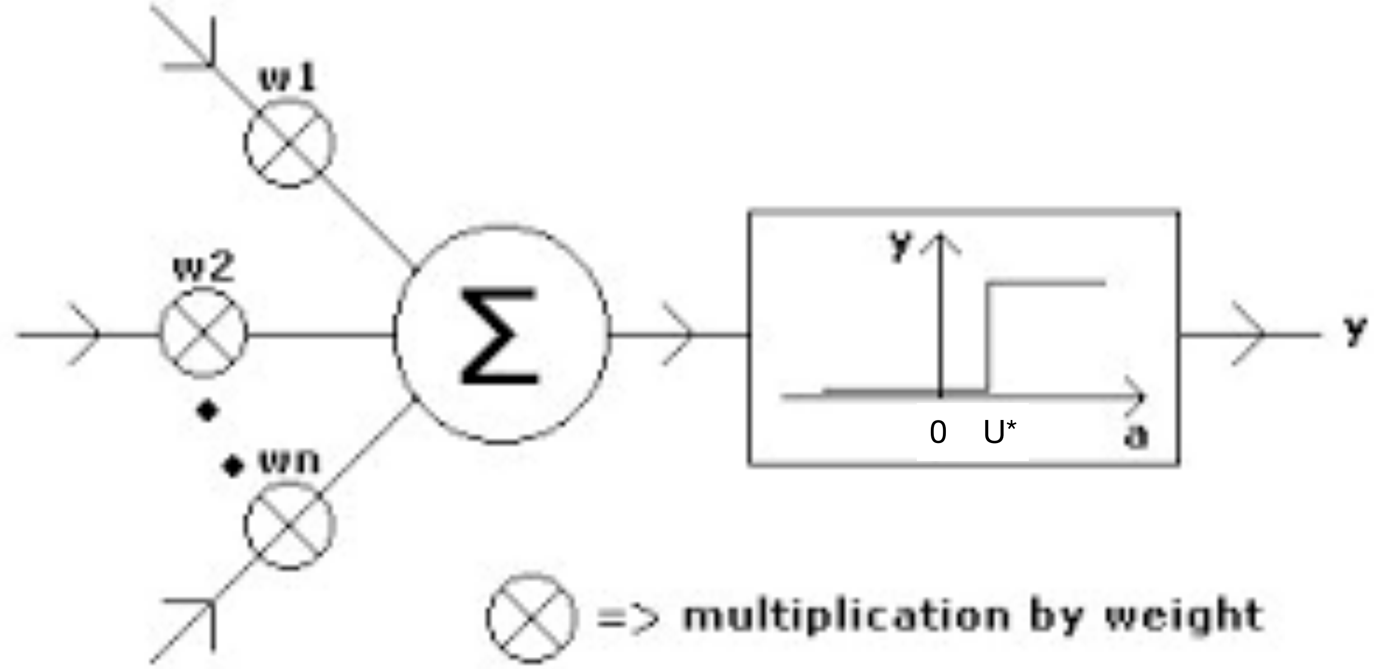
The outgoing signal passes through a bigger fiber conduct called axon.

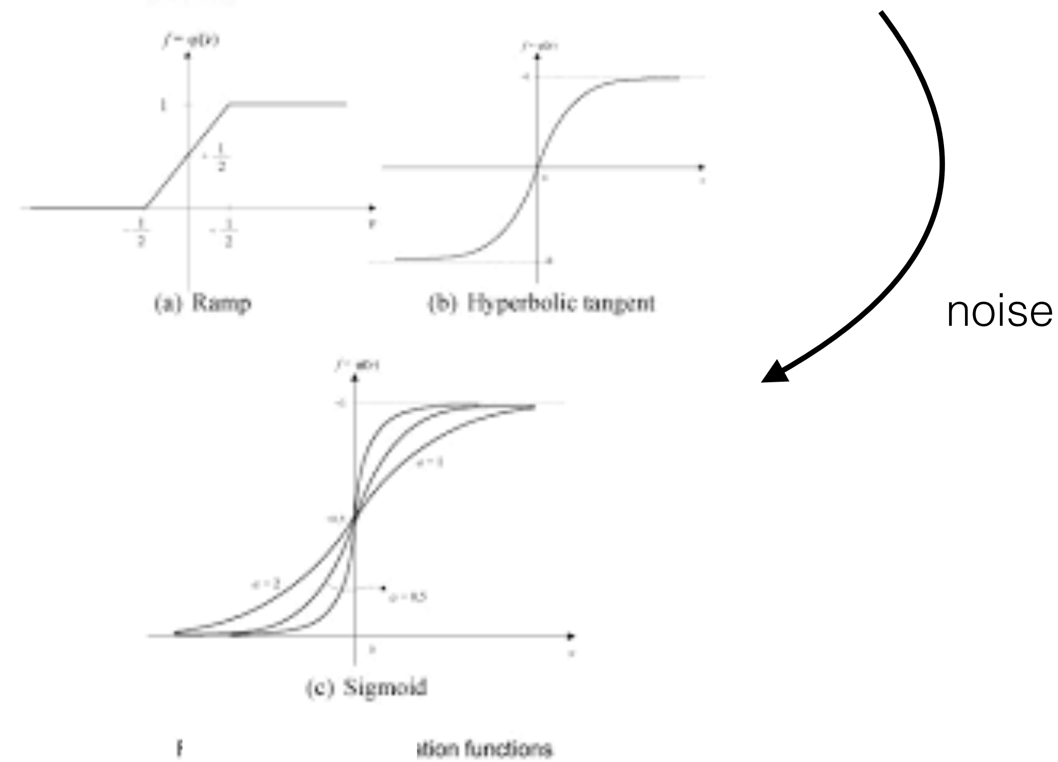
The axon splits into a smaller fibers that are attached, through the dendrites, to the external membrane of other neurons.

The conjunction point of the dendrites with the recipient neuron is called synapse.









Synapses can be either <u>excitatory</u> or <u>inhibitory</u>.

An active neuron emits an electrical wave that travels along the axons, through the different dendrites. At the end of this process there is an electrical potential on the synapse of the recipient neurons. The emission of these packs happen when the <u>total synaptic potential</u>, i.e., the sum of the potentials received from other neurons, is higher than a certain <u>threshold</u> U^*.

Density

Total number of neurons in the human brain is between 10^{10} - 10^{11}

Each neuron is generally connected to 10^4 other neurons

1949 Hebb

Neural pathways are strengthened each time they are used: if two nerves fire at the same time, the connection between them is enhanced ("fire together, wire together")

1988 Miyashita

Trained monkeys show natural activity in a well defined region once a picture is presented for the first time; the same group of neurons reactivates when the monkey sees the same typology of images.

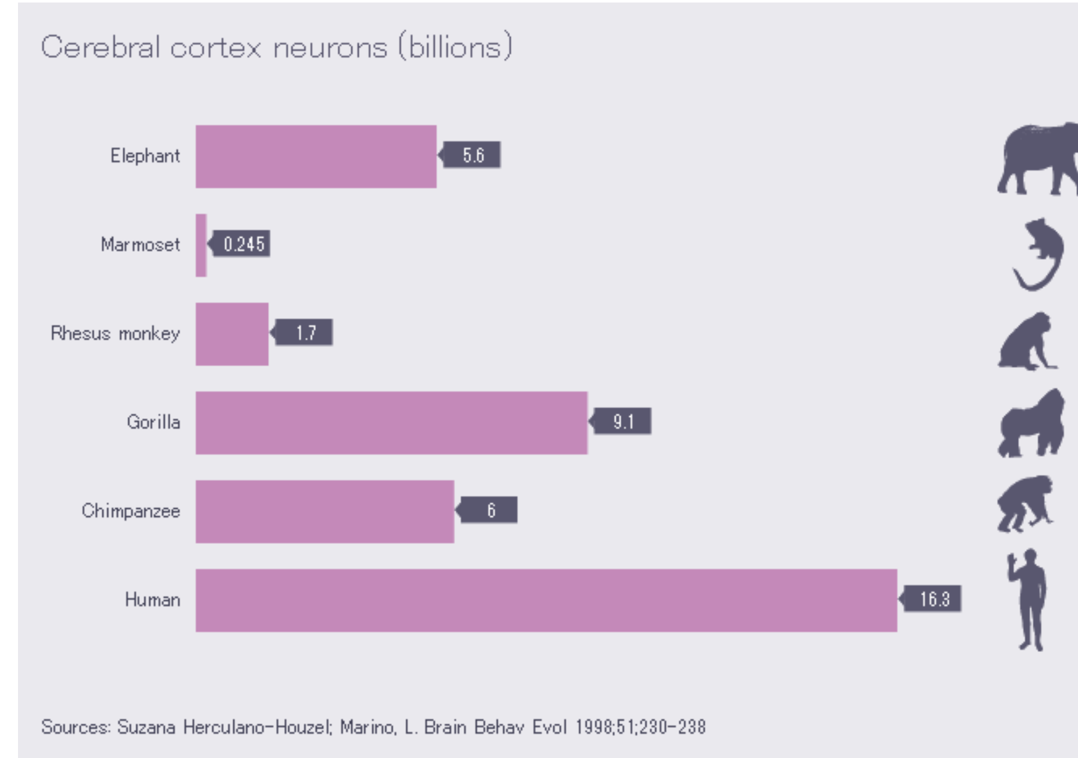
In theoretical developments neural models are built over
the humans brain's modules activity scheme
and preserve the neural network's property of associative memory

stable neural activity ⟷ recognized image or notion

Retrieval is a collective feature of the system

Characteristic time-scales

Duration of action potential:	~1 ms
Refractory period:	~ 3 ms
Synaptic signal transmission:	~ 1 ms
Axonal signal transport:	~ 5 m/s
Short-term synaptic plasticity:	"transient"
Long-term synaptic plasticity:	~hrs-lifetime



L. Sherwood, H. Klandorf, P. Yancey, *Animal Physiology: From Genes to Organisms* Cengage Learning (2012)

F.A.C. Azevedo, et al. "Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain". *The Journal of Comparative Neurology*. **513** (5): 532–541 (2009)

Typical sizes

Neuronal cell body:	~1 ms
Axonal diameter:	~ 3 ms
Synapse size:	~ 1 ms
Synaptic cleft:	~ 5 m/s

Typical densities	#neurons in the brain	#synapses
Caenorhabditis elegans	3.0×10^2	$\sim 7.5 \times 10^3$
Fruit fly	2.5×10^5	$< 10^7$
Honey bee	9.6×10^5	$\sim 10^9$
House mouse	7.1×10^7	$\sim 10^{12}$
Brown rat	2.0×10^8	$\sim 4.5\ 10^{11}$
Cat	7.6×10^8	$\sim 10^{13}$
Human	8.6×10^{10}	$\sim 10^{15}$

McCulloch-Pitts neurons

Warren S. McCulloch (neuroscientist) & Walter Pitts (logician),
“A logical calculus of the ideas immanent in nervous activity” (1943)

$S_i(t) \in \{0, 1\}$ i-th neuron state at time t
$U_i(t)$ post-synaptic potential acting on neuron i at time t
U_i^* threshold noise for neuron i (~ -30 mV)
Δ refractory time (~ 3 ms)

$$U_i(t) = \sum_{k=1}^{N} J_{ik} S_k(t)$$
$$S_i(t + \Delta) = \theta(U_i(t) - U_i^*)$$

Stochastic binary neurons

$$U_i^*(t) = U_i^* - \frac{1}{2} T z_i(t)$$
$$\overline{z_i(t)} = 0, \quad \overline{z_i^2(t)} = 1$$
$$\frac{1}{4} T^2 = \overline{[U_k^*(t) - U_k^*]^2} = \int \mathrm{d}u \; u^2 \; P(u)$$

Stochasticity for thresholds: same variance and write it explicitely

Binary neurons

$$S_i(t) = \frac{1}{2}[\sigma_i(t) + 1]$$
$$U_i^* = \frac{1}{2}\left(\sum_k J_{ik} - \vartheta_i\right)$$

$$S_i(t+\Delta) = \theta\left(\sum_{k=1}^{N} J_{ik} S_k(t) - U_i^* + \frac{1}{2} T z_i(t)\right)$$ stochastic McCulloch-Pitts

$$\Rightarrow \frac{1}{2}[\sigma_i(t+\Delta)+1] = \theta\left(\frac{1}{2}\sum_{k=1}^{N} J_{ik}\sigma_k(t) + \frac{1}{2}\vartheta_i + \frac{1}{2} T z_i(t)\right)$$ binary stochastic

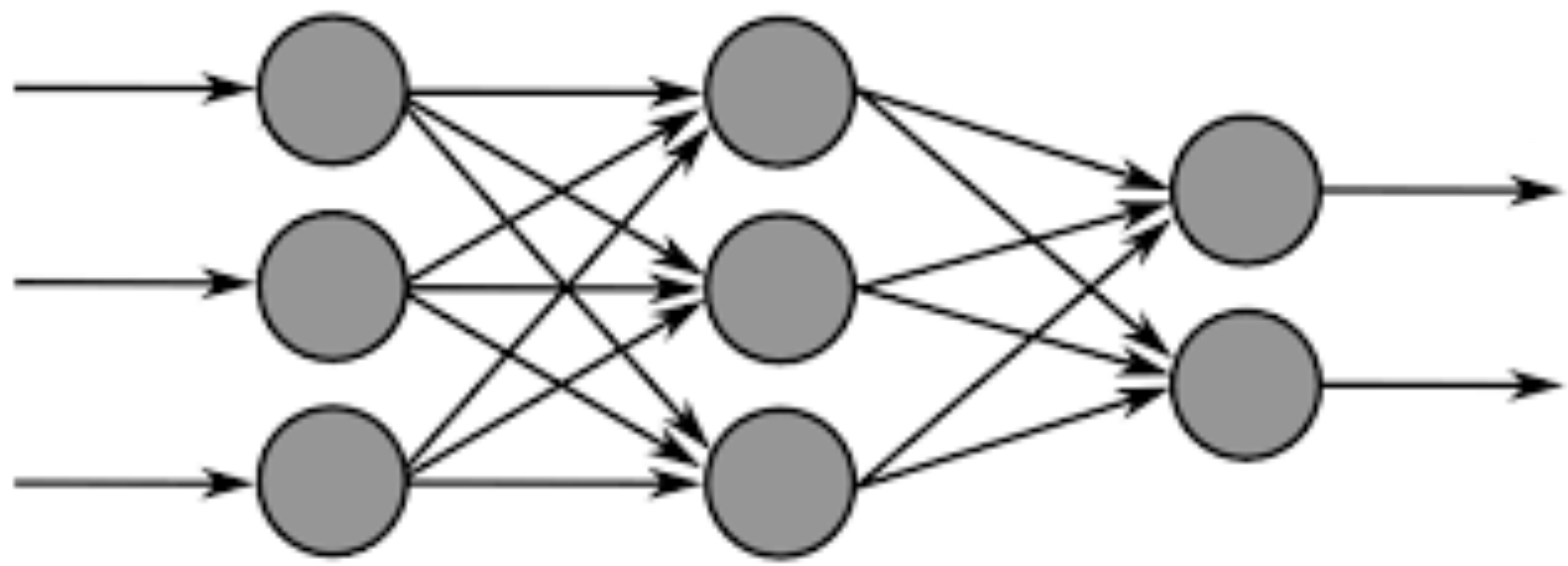
$$\sigma_i(t+\Delta) = \mathrm{sign}\left[h_i(t) + T z_i(t)\right]$$

$$h_i(t) = \sum_{k=1}^{N} J_{ik}\sigma_k(t) + \vartheta_i$$

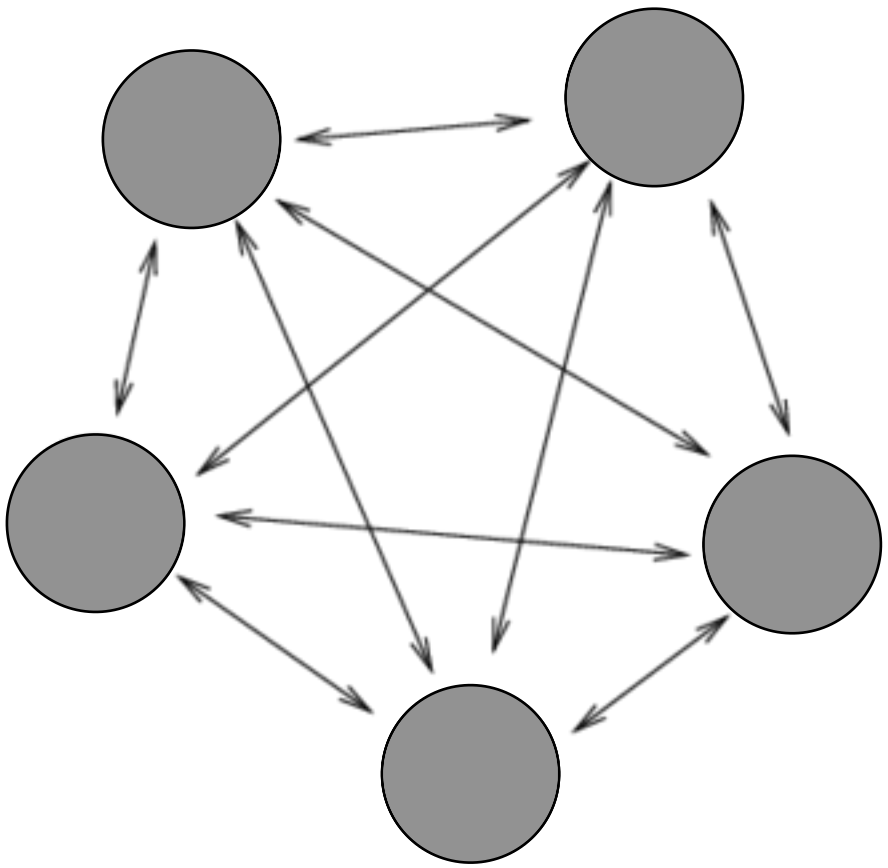
For *symmetric* noise distributions $\mathrm{Prob}[\sigma_i(t+\Delta)] = g(\sigma_i(t+\Delta) h_i(t)/T)$

with $g(x) = \int_{-\infty}^{x} \mathrm{d}z P(z) = \frac{1}{2} + \int_{0}^{x} \mathrm{d}z P(z)$

Recurrent networks



Recurrent networks



feed-back!

Stochastic Local Field Alignment

The microscopic law governing the evolution of a system of N Ising spin neurons $\sigma_i \in \{-1, +1\}$ are defined as a sochastic alignment to local fields $h_i(\boldsymbol{\sigma})$. These fields represent the post-synaptic potentials of the neurons and are assumed to depend linearly on the instantaneous neuron states:

parallel dynamics

$$\mathrm{Prob}[\boldsymbol{\sigma}(t+\Delta)] = \prod_{i=1}^{N} g(\sigma_i(t+\Delta)h_i(t)/T)$$

$$h_i(\boldsymbol{\sigma}(t)) \equiv \sum_{j=1}^{N} J_{ij}\sigma_j(t) + \vartheta_i(t)$$

sequential dynamics

select i randomly from {1, …, N}

$$\mathrm{Prob}[\sigma_i(t+\Delta)] = g(\sigma_i(t+\Delta)h_i(t)/T)$$

The stochasticity is in the iid numbers $z_i(t)$ drawn from $P(u)$ which defines g.
The parameter $\beta=1/T$ controls the impact of this noise on the states $\sigma_i(t+1)$:

- $\beta = \infty \rightarrow$ deterministic process: $\sigma_i(t+1) = \mathrm{sgn}[h_i(\boldsymbol{\sigma}(t))]$
- $\beta = 0 \rightarrow$ fully-random process

The external fields $\vartheta_i(t)$ represent neural thresholds and/or external stimuli.

Stochastic Local Field Alignment

The microscopic law governing the evolution of a system of N Ising spin neurons $\sigma_i \in \{-1, +1\}$ are defined as a sochastic alignment to local fields $h_i(\boldsymbol{\sigma})$. These fields represent the post-synaptic potentials of the neurons and are assumed to depend linearly on the instantaneous neuron states:

parallel dynamics

$$\mathrm{Prob}[\boldsymbol{\sigma}(t+\Delta)] = \prod_{i=1}^{N} g(\sigma_i(t+\Delta)h_i(t)/T)$$

$$h_i(\boldsymbol{\sigma}(t)) \equiv \sum_{j=1}^{N} J_{ij}\sigma_j(t) + \vartheta_i(t)$$

sequential dynamics

select i randomly from {1, …, N}

$$\mathrm{Prob}[\sigma_i(t+\Delta)] = g(\sigma_i(t+\Delta)h_i(t)/T)$$

The stochasticity is in the iid numbers $z_i(t)$ drawn from $P(u)$ which defines g.
The parameter $\beta=1/T$ controls the impact of this noise on the states $\sigma_i(t+1)$:

$\beta = \infty \rightarrow$ deterministic process: $\sigma_i(t+1) = \mathrm{sgn}[h_i(\sigma(t))]$

$\beta = 0 \rightarrow$ fully-random process

The external fields $\vartheta_i(t)$ represent neural thresholds and/or external stimuli.

Interaction Symmetry and Existence of fixed points

Symmetric interactions $J_{ij} = J_{ji}$, for all (i,j)
Non-negative self-interaction $J_{ii} \geq 0$, for all i
Stationary external field ϑ_i
Noiseless system $T=0$

→ Lyapunov function $$L(\boldsymbol{\sigma}) = -\frac{1}{2}\sum_{ij}\sigma_i J_{ij}\sigma_j - \sum_i \sigma_i \vartheta_i$$

for the sequential dynamics
$$\sigma_i(t+1) = \text{sign}\left[\sum_{k=1}^{N} J_{ik}\sigma_k(t) + \vartheta_i\right]$$
$$\sigma_k(t+1) = \sigma_k(t) \quad \forall k \neq i$$

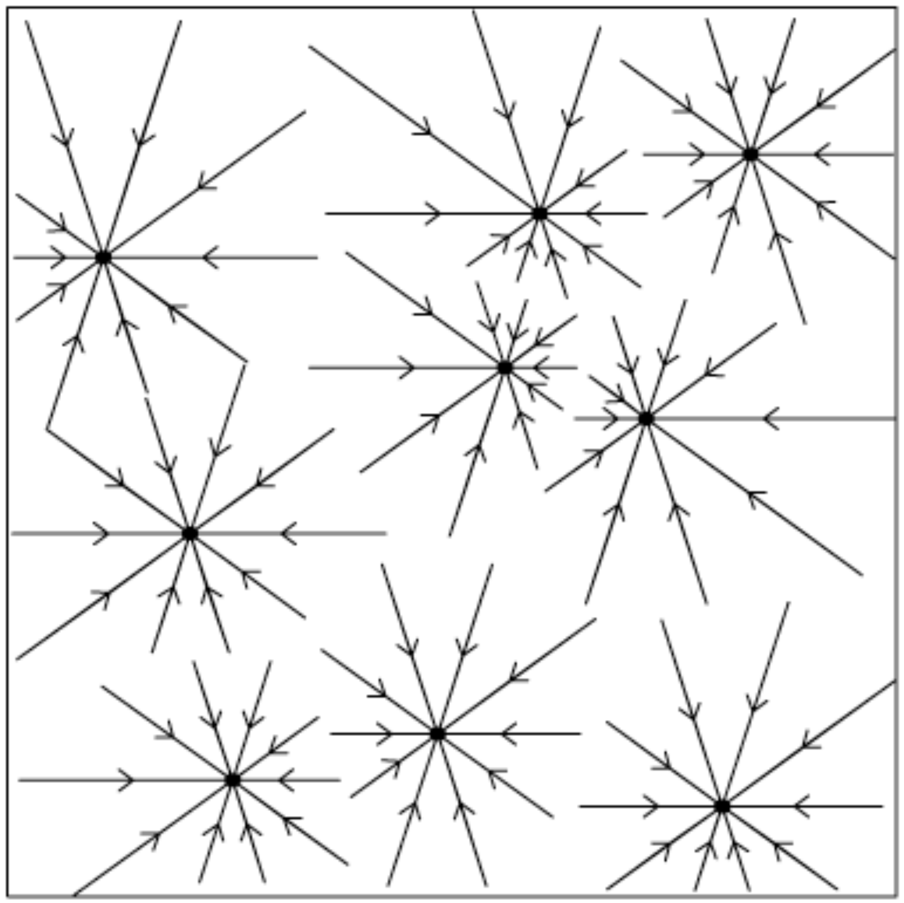
During the iteration of the map the quantity $L(t)$ evolves according to

$$\Delta L = L(\boldsymbol{\sigma'}) - L(\boldsymbol{\sigma}) = -2\left|\sum_j J_{ij}\sigma_j + \vartheta_i\right| - 2J_{ii} < 0$$

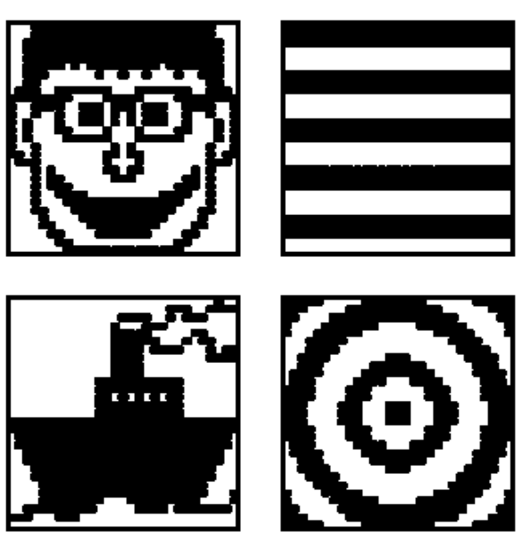
$L(t)$ decreases monotonically until a stage is reached where $\sigma_i(t+1) = \sigma_i(t)$ for all i
⇒ in the deterministic limit, sequential (parallel) systems with symmetric interactions,

non-negative self-interaction (not need for parallel dynamics) and stationary external fields will in always end up with a fixed point (limit cycle with period ≤ 2).

Recurrent networks can be used as associative memories for storing patterns (words, pictures, abstract relations, whatever) through the creation of fixed point attractors



- Represent each of the p items or pattern to be stored as an N-bit vector $\boldsymbol{\xi}^\mu=(\xi_1^\mu, \ldots, \xi_N^\mu) \in \{-1, +1\}^N$, $\mu=1, \ldots, p$
- Construct synaptic interactions $\{J_{ij}\}$ and thresholds $\{\vartheta_i\}$ such that fixed point attractors are created at the p locations of the pattern vectors $\boldsymbol{\xi}^\mu$ in state space
- A given initial configuration $\boldsymbol{\sigma}(0)$ is allowed to evolve in time autonomously, driven by network dynamics, and will end up to the nearest attractor (in some topological sense)
- Final state reached $\boldsymbol{\sigma}(\infty)$ interpreted as the pattern recognized by the network from the input $\boldsymbol{\sigma}(0)$



$t=0$ $t=1$ $t=2$ $t=3$ $t=4$

Inlcude noise and move to a macroscopic setting

particular choice for *P*(*z*): $$P(z) = \frac{1}{2}[1 - \tanh^2(z)] \rightarrow g(x) = \frac{1}{2}[1 + \tanh(x)]$$

parallel: $$\mathrm{Prob}[\boldsymbol{\sigma}(t + \Delta)] = \prod_{i=1}^{N} \left[\frac{1}{2} + \frac{1}{2}\sigma_i(t + \Delta t) \tanh(\beta h_i(\boldsymbol{\sigma}(t))) \right]$$

sequential: choose *i* randomly from {1, …, *N*}
$$\mathrm{Prob}[\sigma_i(t + \Delta)] = \frac{1}{2} + \frac{1}{2}\sigma_i(t + \Delta t) \tanh(\beta h_i(\boldsymbol{\sigma}(t)))$$

with $\beta = T^{-1}$ and $h_i(\boldsymbol{\sigma}) = \sum_{j=1}^{N} J_{ij}\sigma_j + \vartheta_i$

The microscopic equation $$\mathrm{Prob}[\sigma_i(t + \Delta)] = g(\sigma_i(t + \Delta)h_i(t)/T)$$

can be transformed directly into equations for the evolution of the macroscopic state probability $p_t(\boldsymbol{\sigma})$.

Parallel dynamics

If $\boldsymbol{\sigma}(t)$ is given, we find ($\Delta=1$)

$$p_{t+1}(\boldsymbol{\sigma}) = \prod_{i=1}^{N} \frac{1}{2}\{1+\sigma_i \tanh[\beta h_i(\boldsymbol{\sigma}(t))]\} = \prod_{i=1}^{N} \frac{e^{\beta\sigma_i h_i(\boldsymbol{\sigma}(t))}}{2\cosh[\beta h_i(\boldsymbol{\sigma}(t))]}$$

If, instead of $\boldsymbol{\sigma}(t)$, the probability $p_t(\boldsymbol{\sigma})$ is given, the above expression generalises to the corresponding average over the states at time t:

$$p_{t+1}(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\sigma}'} W[\boldsymbol{\sigma};\boldsymbol{\sigma}']p_t(\boldsymbol{\sigma}')$$

W: 2^Nx2^N transition matrix

with

$$W[\boldsymbol{\sigma};\boldsymbol{\sigma}'] \equiv \prod_{i=1}^{N} \frac{e^{\beta\sigma_i h_i(\boldsymbol{\sigma}')}}{2\cosh[\beta h_i(\boldsymbol{\sigma}')]}$$

$$W[\boldsymbol{\sigma};\boldsymbol{\sigma}'] \in [0,1]$$

$$\sum_{\boldsymbol{\sigma}} W[\boldsymbol{\sigma};\boldsymbol{\sigma}'] = 1$$

which is the Markov equation corresponding to the parallel process $\boldsymbol{\sigma}(t) \rightarrow \boldsymbol{\sigma}(t+1)$.

Sequential dynamics

If $\boldsymbol{\sigma}(t)$ and the site i to be updated are given, we find ($\Delta=1$)

$$p_{t+1}(\boldsymbol{\sigma}) = \prod_{i=1}^{N} \frac{1}{2}\{1+\sigma_i \tanh[\beta h_i(\boldsymbol{\sigma}(t))]\} = \prod_{i=1}^{N} \frac{e^{\beta\sigma_i h_i(\boldsymbol{\sigma}(t))}}{2\cosh[\beta h_i(\boldsymbol{\sigma}(t))]}$$

If, instead of $\boldsymbol{\sigma}(t)$, the probability $p_t(\boldsymbol{\sigma})$ is given, the above expression generalises to the corresponding average over the states at time t:

$$p_{t+1}(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\sigma}'} W[\boldsymbol{\sigma};\boldsymbol{\sigma}']p_t(\boldsymbol{\sigma}')$$

W: 2^Nx2^N transition matrix

with

$$W(\boldsymbol{\sigma};\boldsymbol{\sigma}') = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{2}[1+\sigma_i\tanh(\beta h_i(\boldsymbol{\sigma}'))]\prod_{j\neq i}\delta_{\sigma_j,\sigma'_j}\right\}$$

$$W[\boldsymbol{\sigma};\boldsymbol{\sigma}'] \in [0,1]$$

$$\sum_{\boldsymbol{\sigma}} W[\boldsymbol{\sigma};\boldsymbol{\sigma}'] = 1$$

which is the Markov equation corresponding to the parallel process $\boldsymbol{\sigma}(t) \rightarrow \boldsymbol{\sigma}(t+1)$.

Prop: the Markov process described by $W(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ is ergodic, namely

whatever the initial condition $p_0(\boldsymbol{\sigma})$, there exists a time τ such that

$$p_t(\boldsymbol{\sigma}) > 0 \quad \text{for all} \ \ \boldsymbol{\sigma} \ \ \text{and all} \ \ t \geq \tau$$

otherwise stated

$$W^t(\boldsymbol{\sigma}, \boldsymbol{\sigma'}) > 0 \quad \text{for all} \ \ \boldsymbol{\sigma}, \boldsymbol{\sigma'} \ \ \text{and all} \ \ t \geq \tau$$

Theorem: Ergodic Markov processes have a unique stationary distribution p_∞ to which they will converge from any initial distributions over states.

This distribution is reached in the limit $t \to \infty$ (at fixed N) and it is determined by the stationary condition

$$\text{for all} \ \ \boldsymbol{\sigma} \in \{-1, +1\}^N : \quad p_\infty(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\sigma'}} W(\boldsymbol{\sigma}, \boldsymbol{\sigma'}) p_\infty(\boldsymbol{\sigma'})$$

To calculate p_∞ we would have to solve this system of 2^N linear equations for the 2^N values of p_∞, subject to positivity and normalization…

Theorem: Ergodic Markov processes have a unique stationary distribution p_∞ to which they will converge from any initial distributions over states.

This distribution is reached in the limit t → ∞ (at fixed N) and it is determined by the stationary condition

$$\text{for all} \quad \boldsymbol{\sigma} \in \{-1,+1\}^N: \quad p_\infty(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\sigma'}} W(\boldsymbol{\sigma},\boldsymbol{\sigma'}) p_\infty(\boldsymbol{\sigma'})$$

To calculate p_∞ we would have to solve this system of *2N* linear equations for the *2N* values of p_∞, subject to positivity and normalization…

Stronger condition: ***detailed balance***

$$W(\boldsymbol{\sigma},\boldsymbol{\sigma'}) p_\infty(\boldsymbol{\sigma'}) = W(\boldsymbol{\sigma'},\boldsymbol{\sigma}) p_\infty(\boldsymbol{\sigma}) \quad \text{for all} \quad \boldsymbol{\sigma},\boldsymbol{\sigma} \in \{-1,+1\}^N$$

⇒ in addition to the probability distribution being stationary, the latter describes equilibrium ⇒ greatly simplifies the calculation of p_∞

Theorem: for sequential dynamics without self-interactions, interaction symmetry implies detailed balance and vice versa.
If detailed balance holds, the equilibrium distribution is given by

$$p_\infty(\boldsymbol{\sigma}) \propto e^{-H(\boldsymbol{\sigma})/T}$$

$$H(\boldsymbol{\sigma}) = -\frac{1}{2}\sum_{ij}^{N} \sigma_i J_{ij} \sigma_j - \sum_{i=1}^{N} \vartheta_i \sigma_i$$

$H(\boldsymbol{\sigma})$ is called Hamiltonian

In stat-mech this corresponds to a Boltzmann distribution.

Remark: there exists an analogous version for the case of parallel dynamics, where constraint on self-interaction is not necessary and the resulting cost function is slightly different.

Remark: sequential dynamics with self-interactions is considerably more complicated. In principle, d.b. may now hold for both symmetric and non-symmetric systems.

Chapter III
The Hopfield model

Hopfield neural network

prototype example for a vast class of associative memory models

[J.J. Hopfield (1982) *Neural networks and physical systems with emergent collective computational abilities*, PNAS]

Data is stored in the form of <u>patterns of information</u>,
i.e., vectors where each component codifies a particular feature of the image

Original pattern

$\xi = (1,1,\ldots,-1,\ldots -1)$

$\xi_i = +1$ if the pixel labelled as *i* is white

$\xi_i = -1$ if the pixel labelled as *i* is black

Hopfield neural network

prototype example for a vast class of associative memory models

[J.J. Hopfield (1982) *Neural networks and physical systems with emergent collective computational abilities*, PNAS]

Data is stored in the form of <u>patterns of information</u>,
i.e., vectors where each component codifies a particular feature of the image

Original pattern

ξ = (1,1,…,-1,…-1)

ξ_i = +1 if the pixel labelled as *i* is white

ξ_i = -1 if the pixel labelled as *i* is black

Corrupted

Reconstruction

Hopfield neural network

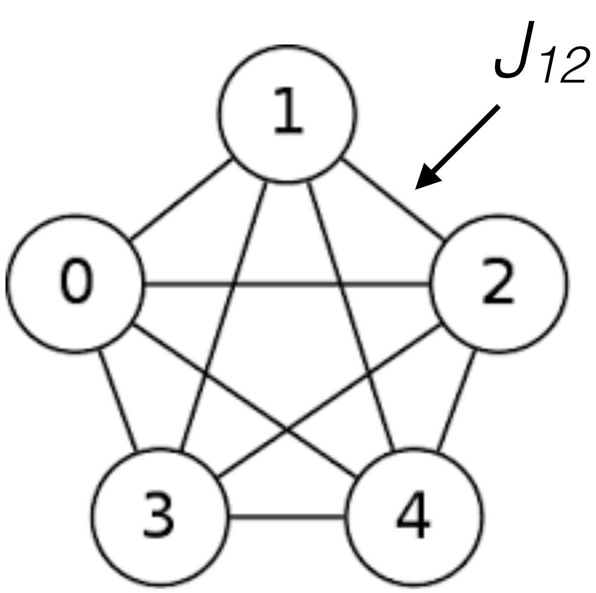
prototype example for a vast class of associative memory models

[J.J. Hopfield (1982) *Neural networks and physical systems with emergent collective computational abilities*, PNAS]

Network of N binary neural units $\sigma_i \in \{-1, +1\}$

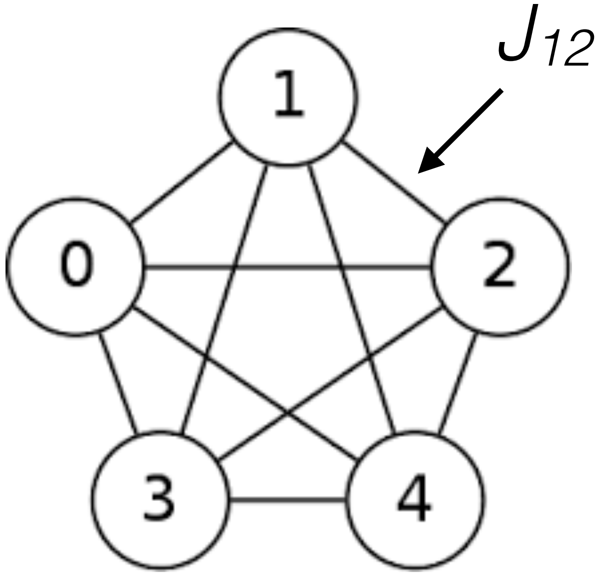
Mean-field interaction

Couplings among neurons quenched (encode for stored patterns)



N=5

$$H_N(\sigma|\mathbf{J},\boldsymbol{\vartheta}) = -\sum_{1\leq i<j\leq N} \sigma_i J_{ij} \sigma_j - \cancel{\sum_{i=1}^{N} \sigma_i \vartheta_i}$$

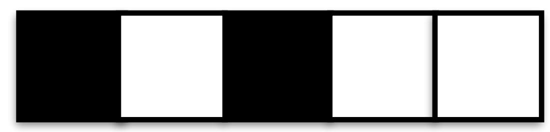


N=5

$$H_N(\sigma|\mathbf{J},\boldsymbol{\vartheta}) = -\sum_{1\leq i<j\leq N} \sigma_i J_{ij}\sigma_j$$

The configuration $\{\sigma\}_{i=1,\dots,N}$ is associated to a pattern

σ ={-1,+1,-1,+1,+1}



Store *P* patterns: $\boldsymbol{\xi}^1, \boldsymbol{\xi}^2, \dots, \boldsymbol{\xi}^P$

Pattern $\boldsymbol{\xi}^\mu = (\xi_1^\mu, \dots, \xi_N^\mu)$ with Boolean entries

Retrieval of *μ*-th pattern $\boldsymbol{\xi}^\mu \longleftrightarrow \sigma_i = \xi_i^\mu$, $\forall i$ equilibrium

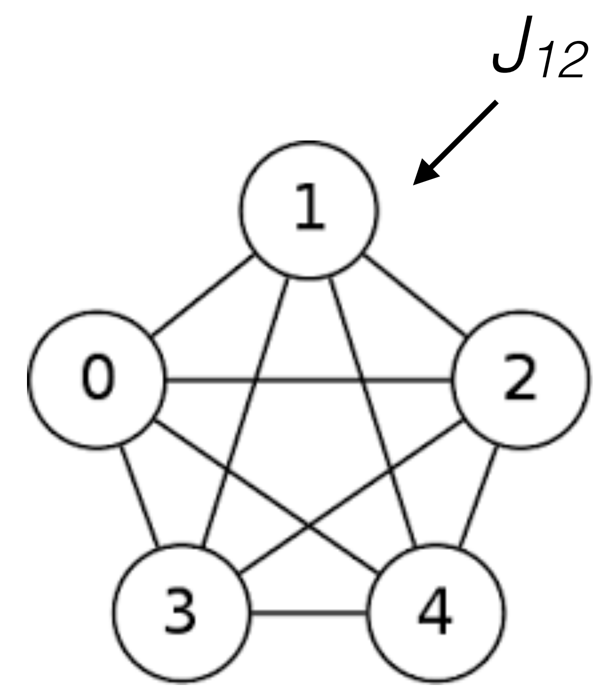
Each entries drawn from $P(\xi_i^\mu = +1) = 1 - P(\xi_i^\mu = -1) = 1/2$, $\forall i, \mu$

➢ ξ_i^μ is the *i*-th entry of the *μ*-th pattern, $\mu = 1, \dots, P$

J_{ij} intensity of the synaptic action of the neuron j over neuron i

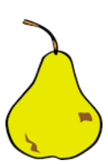
Hebb's learning rule for the synaptic coupling

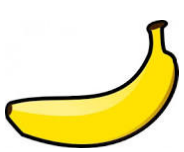
$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu$$



While learning ξ^1 ➛ $\sigma_i = \xi^1_i$ for all i

 $\xi^1=(+1,+1,-1,-1)$ ➛ $\sigma=(+1,+1,-1,-1)$ ➛ J_{12} and J_{34} reinforced

 $\xi^2=(+1,-1,-1,+1)$ ➛ $\sigma=(+1,-1,-1,+1)$ ➛ J_{14} and J_{23} reinforced

 $\xi^3=(+1,-1,+1,-1)$ if not memorized J_{13} and J_{24} not reinforced

Gauge symmetry
Quenched couplings

N neurons

$\sigma_i \in \{-1, +1\}$

$\{\sigma_i\}_{i=1, .., N}$

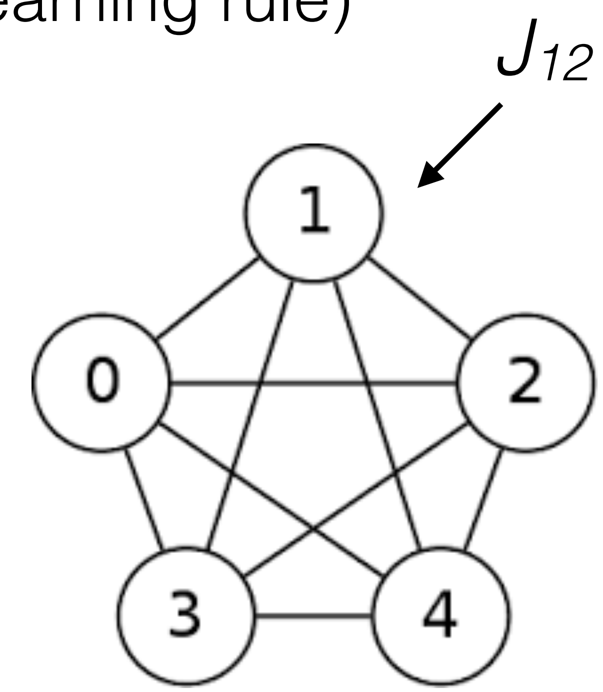
Fully connected network (mathematically convenient and biologically reasonable)

$P(\xi_i = +1) = 1 - P(\xi_i = -1) = 1/2$ (most informative by Shannon theorem)

$J_{ij} = \xi_i \cdot \xi_j / N$ intensity of the synaptic action (mimicking Hebb's learning rule)

Synaptic potential that the *i*-th neuron receives from the other *N*-1

$$h_i = \sum_{i \neq j = 1}^{N} J_{ij} \sigma_j$$



Hamiltonian

$$H_N(\sigma|\mathbf{J}, \boldsymbol{\vartheta}) = -\sum_{1 \leq i < j \leq N} \sigma_i J_{ij} \sigma_j - \sum_i \sigma_i \vartheta_i = -\frac{1}{2N} \sum_{i,j}^{N} \sum_{\mu=1}^{P} \sigma_i \xi_i^\mu \xi_j^\mu \sigma_j + \frac{P}{2}$$

$\vartheta_i = 0$
Hebb

constant term
to be neglected

<u>Overlap</u> between the patterns and the neurons (a.k.a. Mattis magnetization) is an order parameter

$$m_\mu(\sigma) = \frac{1}{N}\sum_{i=1}^{N} \xi_i^\mu \sigma_i \qquad \in [-1,1]$$

$$J_{ij} = \frac{1}{N}\sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu$$

$$H_N(\sigma|\mathbf{J}) = -\frac{1}{2N}\sum_{i,j}^{N}\sum_{\mu=1}^{P} \sigma_i \xi_i^\mu \xi_j^\mu \sigma_j = -\frac{N}{2}\sum_{\mu=1}^{P} m_\mu^2 \qquad \to H_N(\sigma|\mathbf{J}) = -\frac{N}{2}\mathbf{m}^2$$

J symmetric & $J_{ii} \geq 0$, $\forall i$

Noiseless dynamics $\sigma_i(t+1) = \mathrm{sgn}(h_i(t))$

H_N is a Lyapunov function

In fact, when T=0, min(F) = min(U)

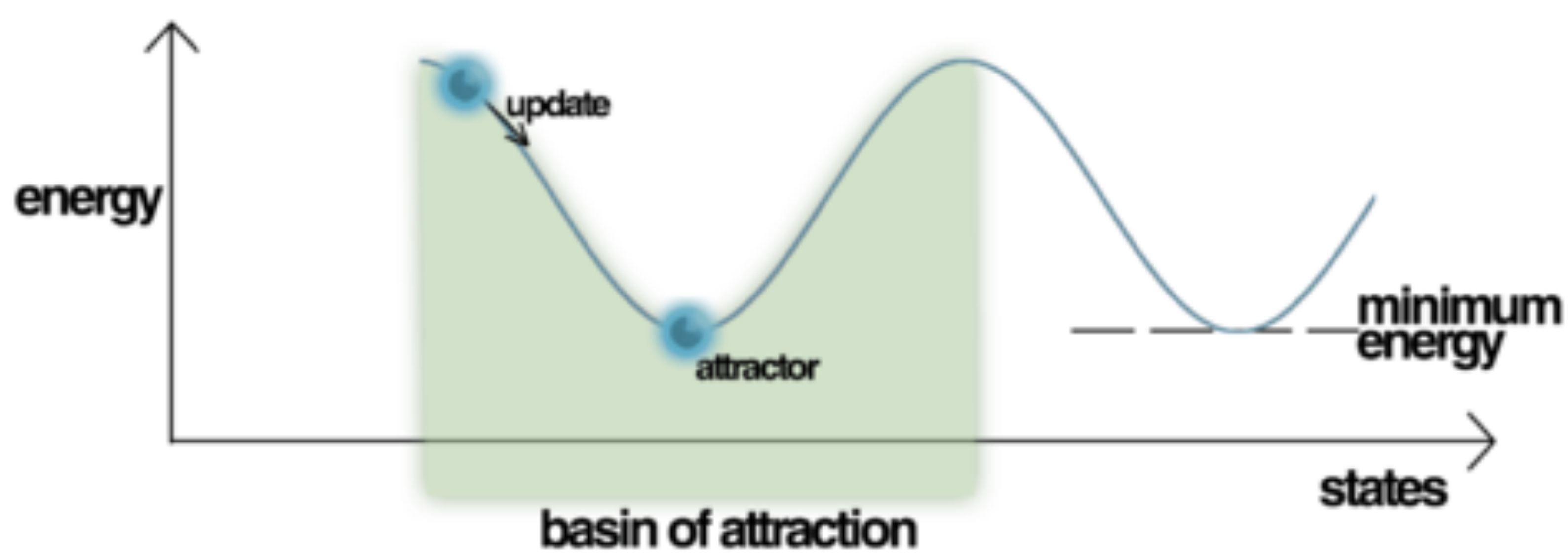
$$m_\mu(\sigma) = \frac{1}{N}\sum_{i=1}^{N} \xi_i^\mu \sigma_i \qquad \in [-1, 1]$$

$$H_N(\sigma|\mathbf{J}) = -\frac{1}{2N}\sum_{i,j}^{N}\sum_{\mu=1}^{P} \sigma_i \xi_i^\mu \xi_j^\mu \sigma_j = -\frac{N}{2}\sum_{\mu=1}^{P} m_\mu^2 \qquad \rightarrow H_N(\sigma|\mathbf{J}) = -\frac{N}{2}\mathbf{m}^2$$

Noiseless dynamics $\quad \sigma_i(t+1) = \mathrm{sgn}(h_i(t) + \vartheta_i)$

H_N is a Lyapunov function

Energy minimization ⇔ Neurons are all parallel to one pattern ⇔ Retrieving

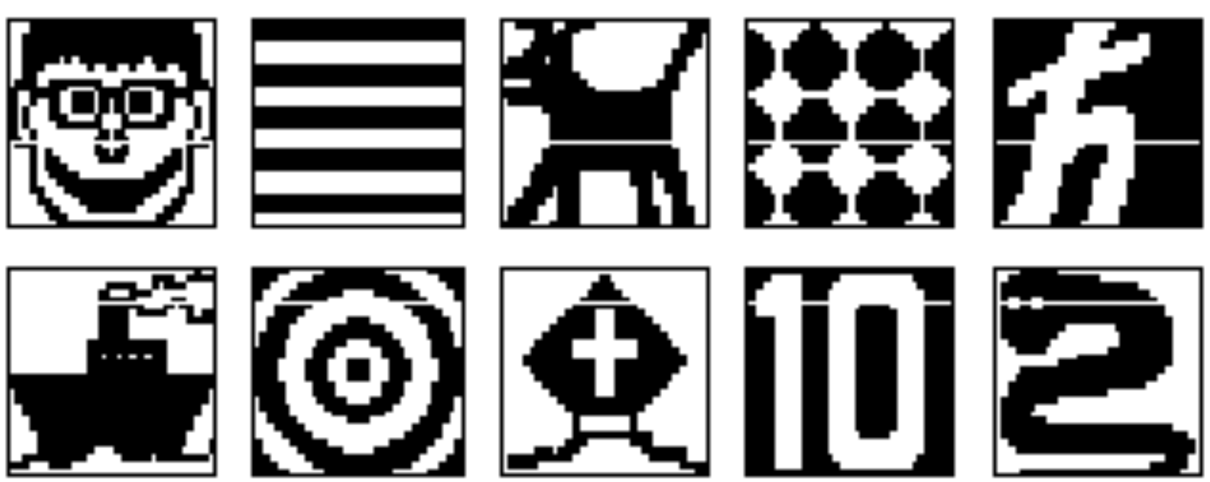


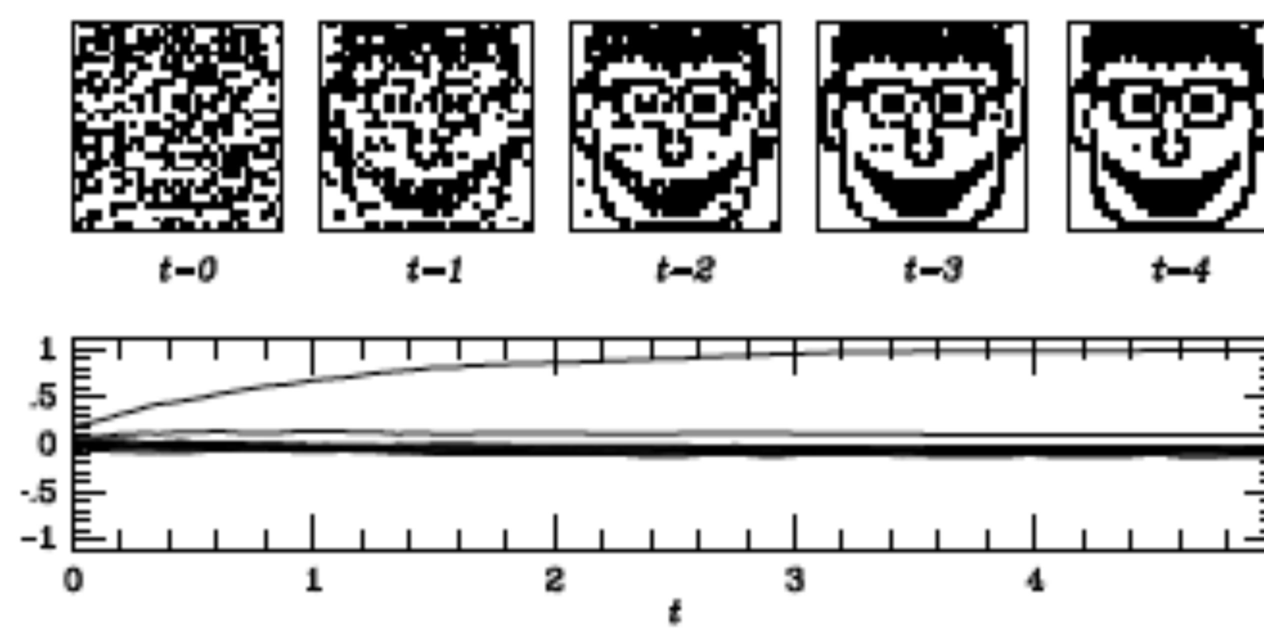
$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu$$

$$H_N(\sigma, \xi) = -\frac{1}{N} \sum_{\mu=1}^{p} \sum_{1 \le i < j \le N} \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j \qquad\qquad H_N(\sigma, \xi) \sim -\frac{N}{2} \sum_{\mu=1}^{p} m_\mu^2$$

$$m_\mu(\sigma) = \frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu \sigma_i \qquad \in [-1, 1]$$

Overlap m_μ says whether the pattern μ has been retrieved or not



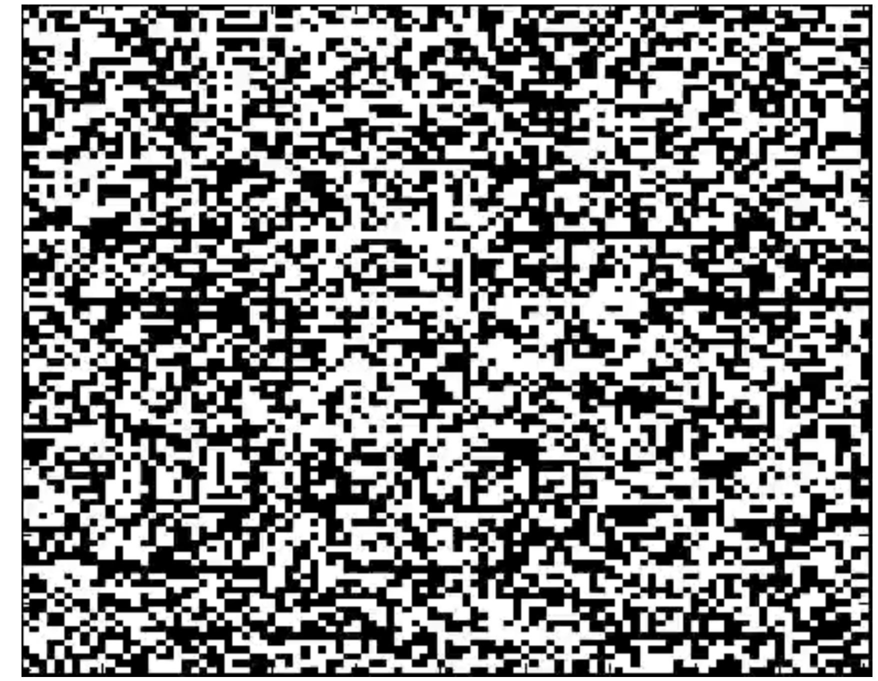


(Fast) Noise β

In real systems and devices noise is unavoidable

E.g., in the brain spikes occur at random times and spiking thresholds are random, data corruption (FN)

E.g., interference among stored patterns (SN)



the system does not move deterministically towards its ground state and does not stays permanently there due to intrinsic randomness

$$\sigma_i(t+\Delta) = \mathrm{sgn}(h_i(t) + T z_i(t))$$

$z_i(t)$, iid variables with central distribution

Neuronal configuration evolves according to a Markov chain fulfilling detailed balance ➛ *p(**σ**)* converges to equilibrium distribution

$$p_\infty(\sigma) \propto e^{-\beta H(\sigma)} \quad \text{with} \quad H(\sigma) = \frac{1}{2}\sum_{i,j=1}^{N} \sigma_i J_{ij} \sigma_j - \sum_{i=1}^{N} \vartheta_i \sigma_i$$

Recover the Lyapunov function of the noiseless case

β=1/T Parameter tuning the (fast) noise

(Slow) Noise α

The number p of stored patterns (encoded in $\boldsymbol{J}$) constitutes a source of noise

If I try to store too much information in the network, the network retrieval capabilities are impaired
➛ black-out scenario

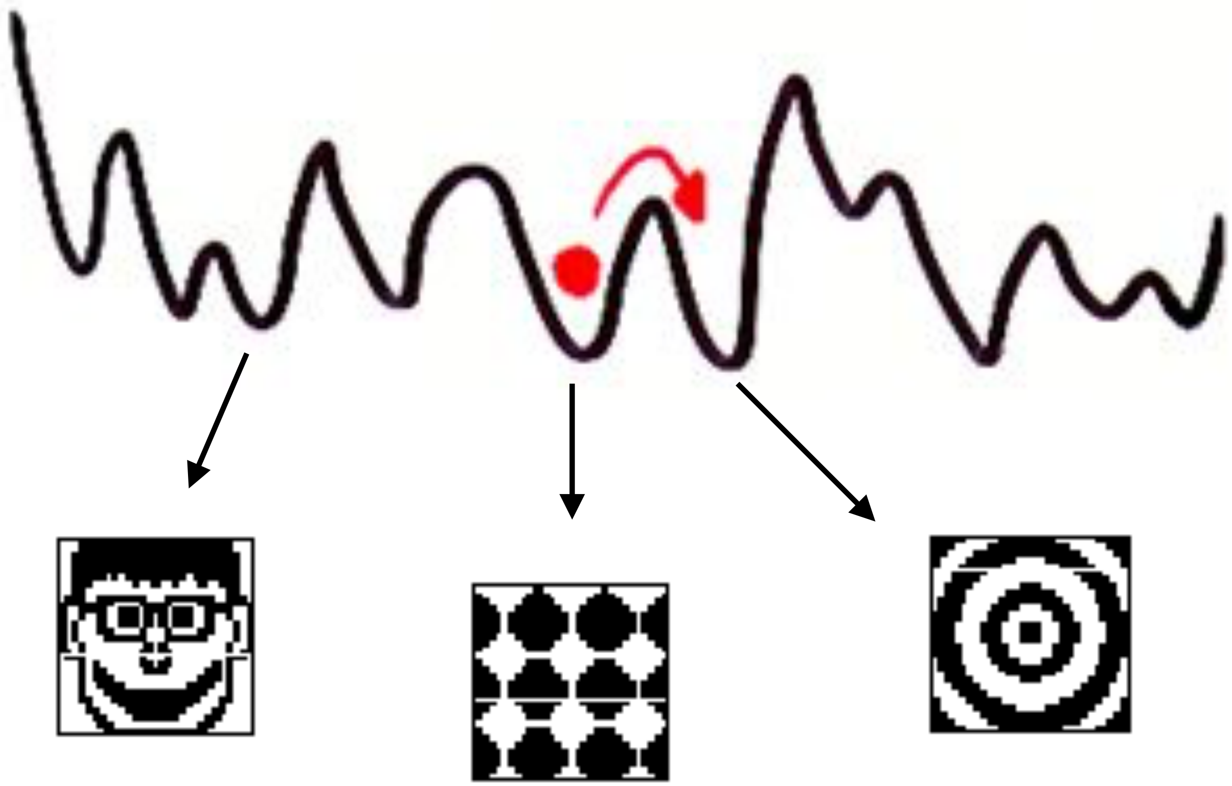
$$\alpha = \lim_{N\to\infty} \frac{p}{N}$$

Parameter tuning the slow noise

Low storage: p scales sub-linearly with N → $\alpha=0$

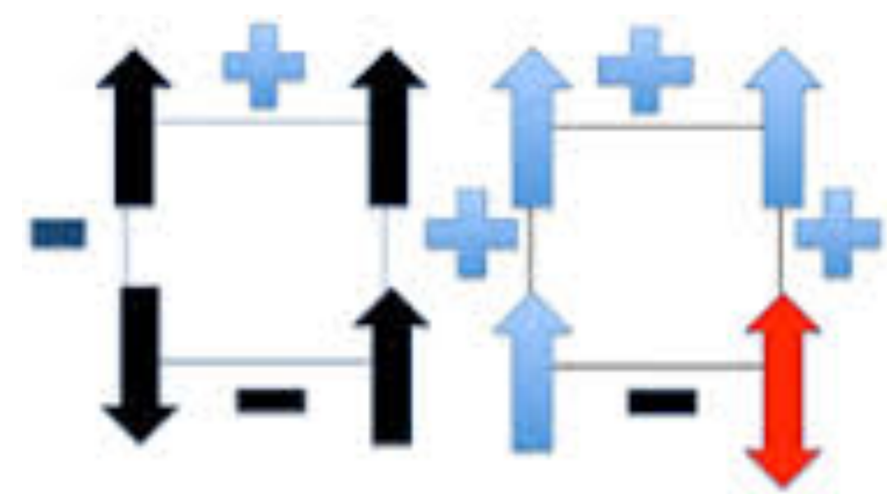
High storage: p scales linearly with N → $\alpha>0$

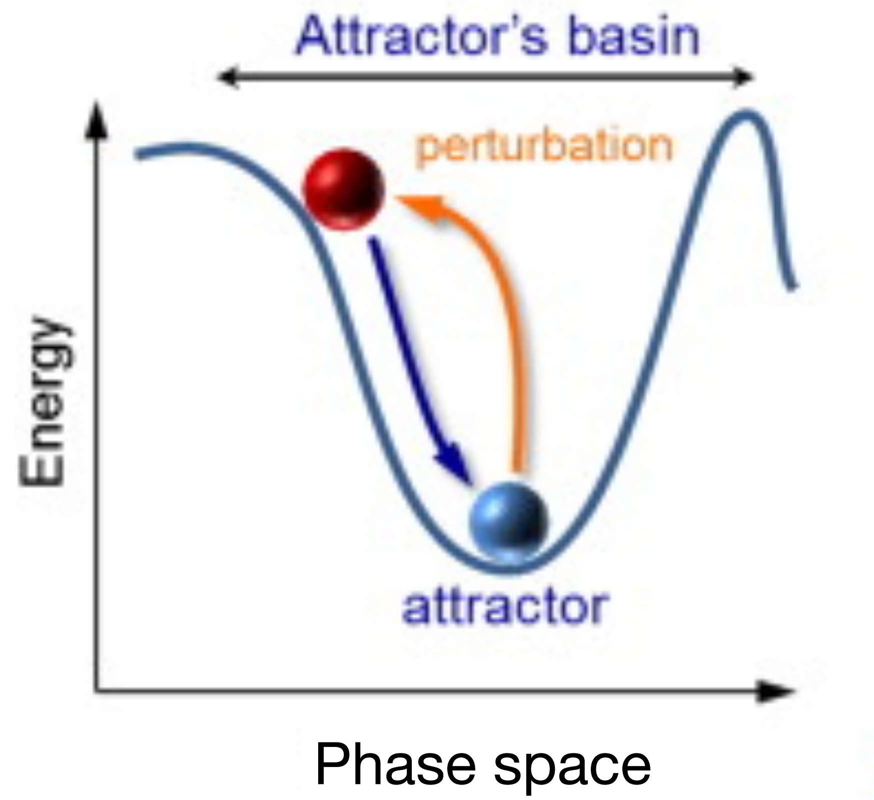
Following statistical mechanics prescription we are interested in the minima of the free energy. They correspond to a state $<m>$ which is the one where the system spontaneously relaxes once the parameters T and p are fixed.



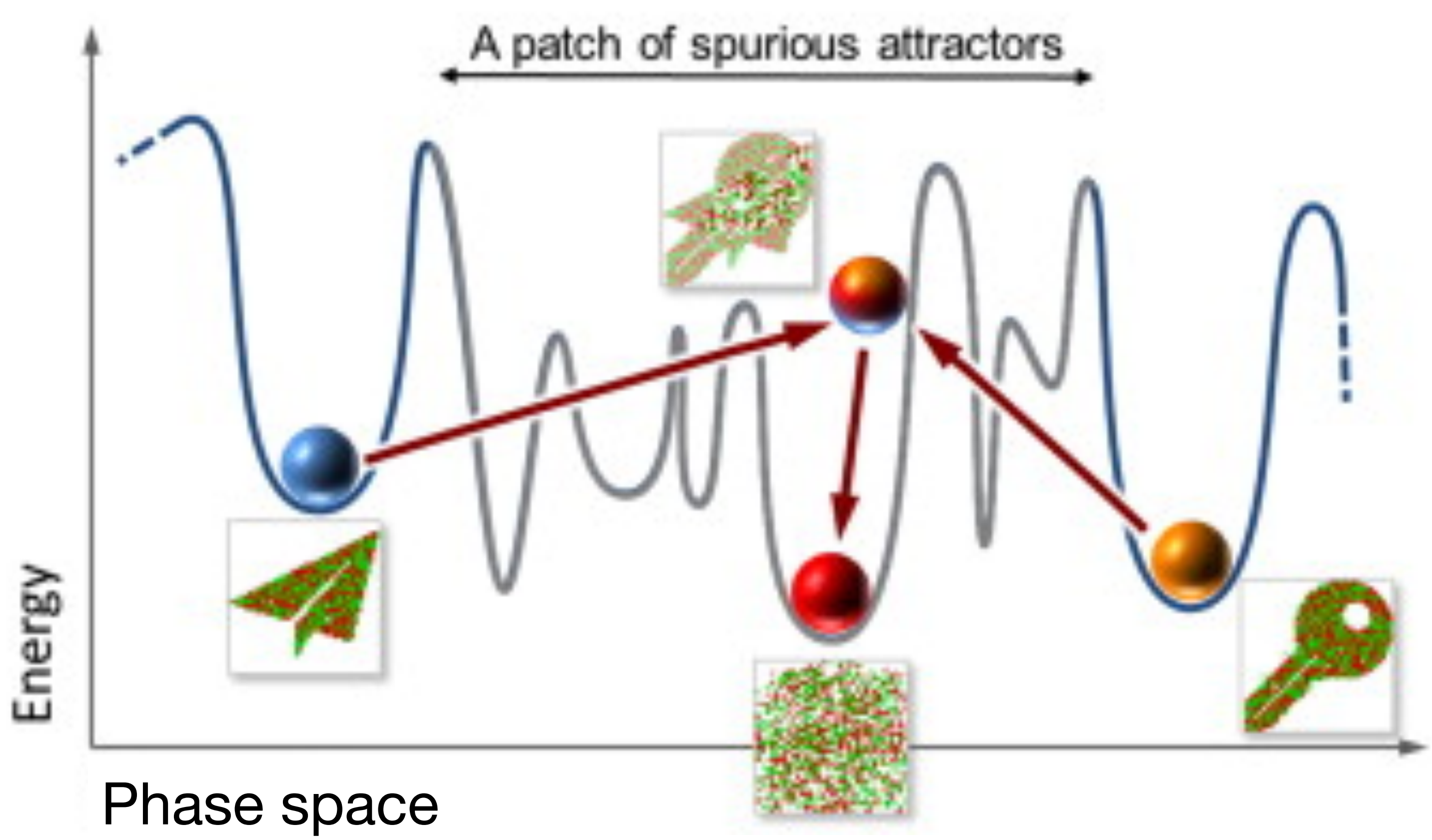
Rugged free energy landscape

The presence of both positive and negative connections makes the system <u>complex</u>
The larger the number of patterns and the more frustrated the system, i.e., the more rugged the landscape



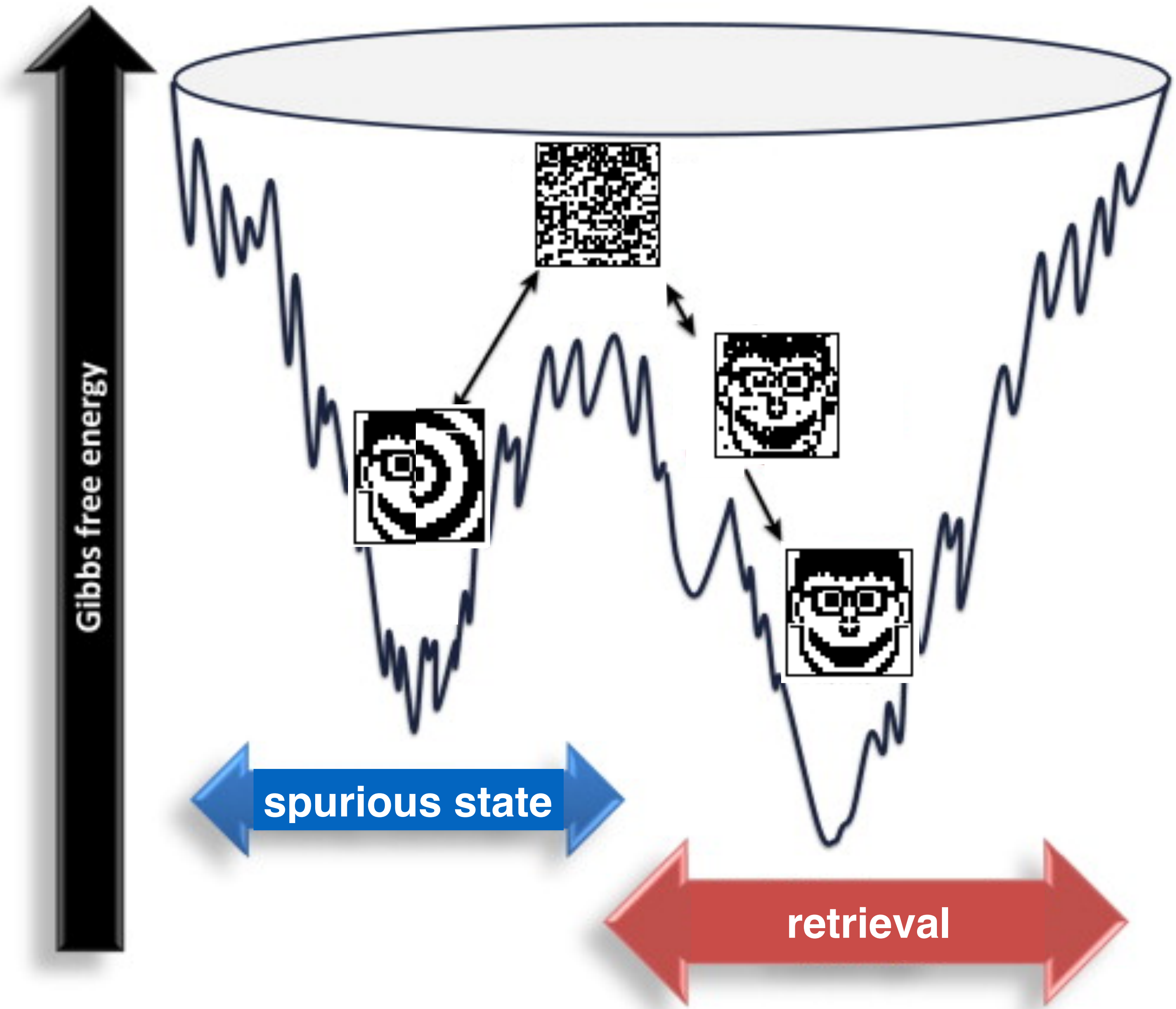


(Fast) noise yield perturbations (thermal excitations) which may impair retrieval.



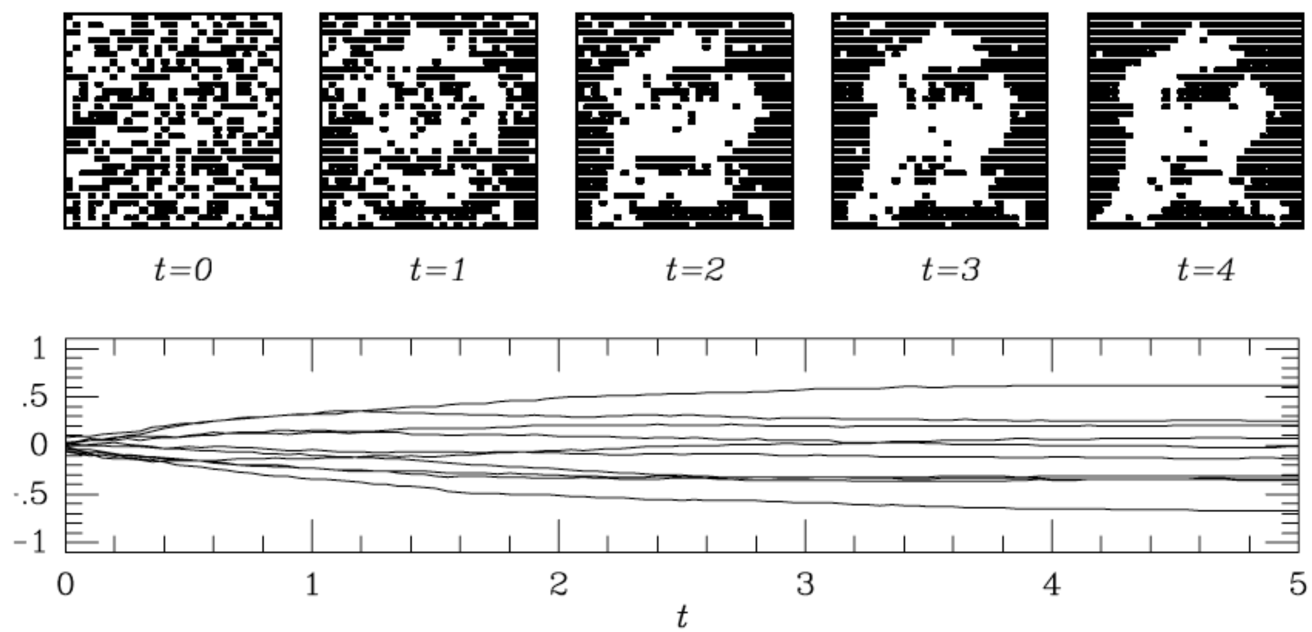
(Slow) noise makes the energy landscape rugged and spurious states may get global minima.

Actually things are a bit more “complex" because the system is “complex”

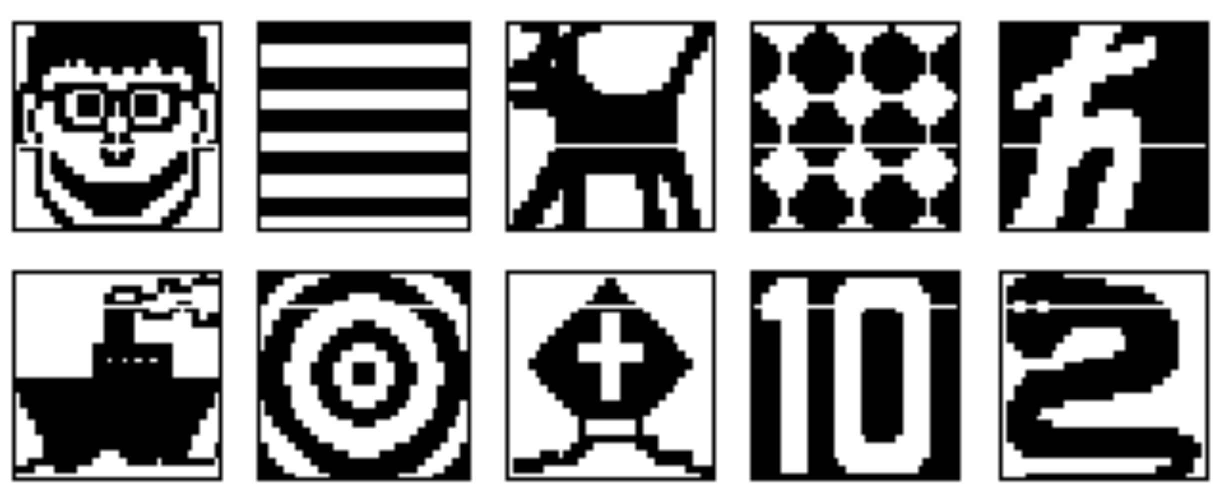


S. Albeverio, B. Tirozzi, B. Zegarlinski (1992) Rigorous results for the free energy in the Hopfield model, Comm. Math. Phys.
L. Pastur, M. Shcherbina, B. Tirozzi (1994) *The replica-symmetric solution without replica trick for the Hopfield model*, J. Stat. Phys.
Mathematical aspects of Spin Glasses and Neural Networks, (1998) A. Bovier and P. Picco Eds., Birkhäuser
A Bovier, V Gayrard, P Picco (1995) Gibbs states of the Hopfield model in the regime of perfect memory, Prob. Theor. and Rel. Fields
A Bovier, V Gayrard (1997) *The retrieval phase of the Hopfield model: A rigorous analysis of the overlap distribution*, Prob. Theor. and Rel. Fields
M. Talagrand (1998) *Rigorous results for the Hopfield model with Many Patterns*, Prob. Theor. Rel. Fields

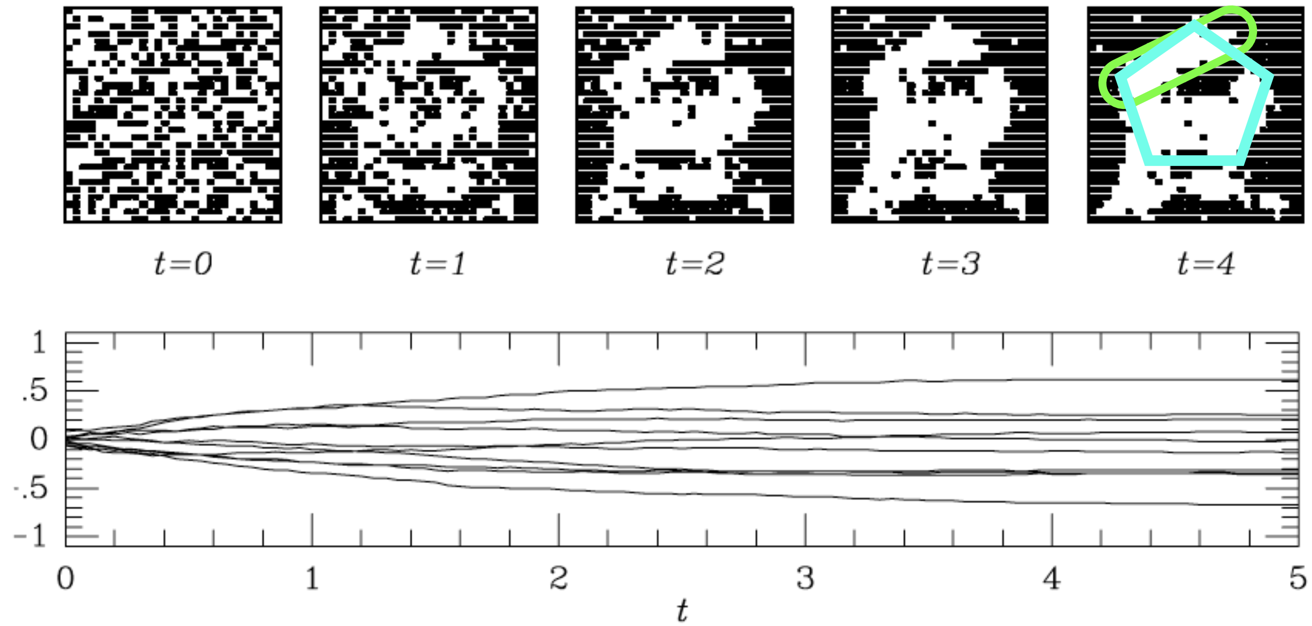
Thermalization in a spurious state: wrong information retrieval!



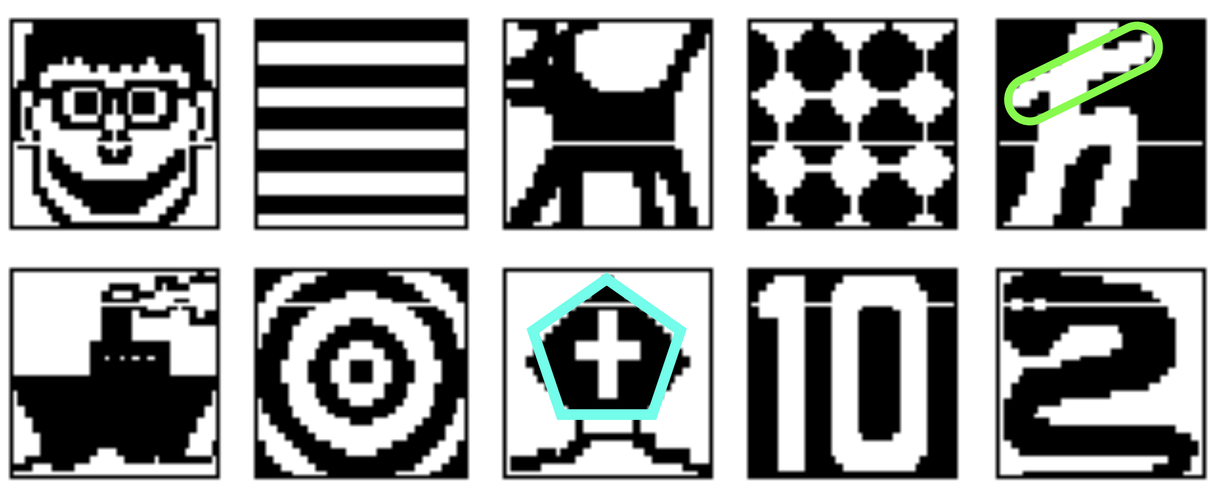
Evolution towards a spurious state at T=0.1 from a randomly drawn initial state. Top row: snapshots of the microscopic system state at times t=0, 1, 2, 3, 4 interactions/spin. Bottom: the corresponding values of the P=10 overlap order parameters as functions of time.



Thermalization in a spurious state: wrong information retrieval!



Evolution towards a spurious state at T=0.1 from a randomly drawn initial state. Top row: snapshots of the microscopic system state at times t=0, 1, 2, 3, 4 interactions/spin. Bottom: the corresponding values of the P=10 overlap order parameters as functions of time.



Phase Diagram

DJ Amit, H Gutfreund, H Sompolinsky (1985) *Storing infinite numbers of patterns in a spin-glass model of neural networks*, Phys. Rev. Lett.