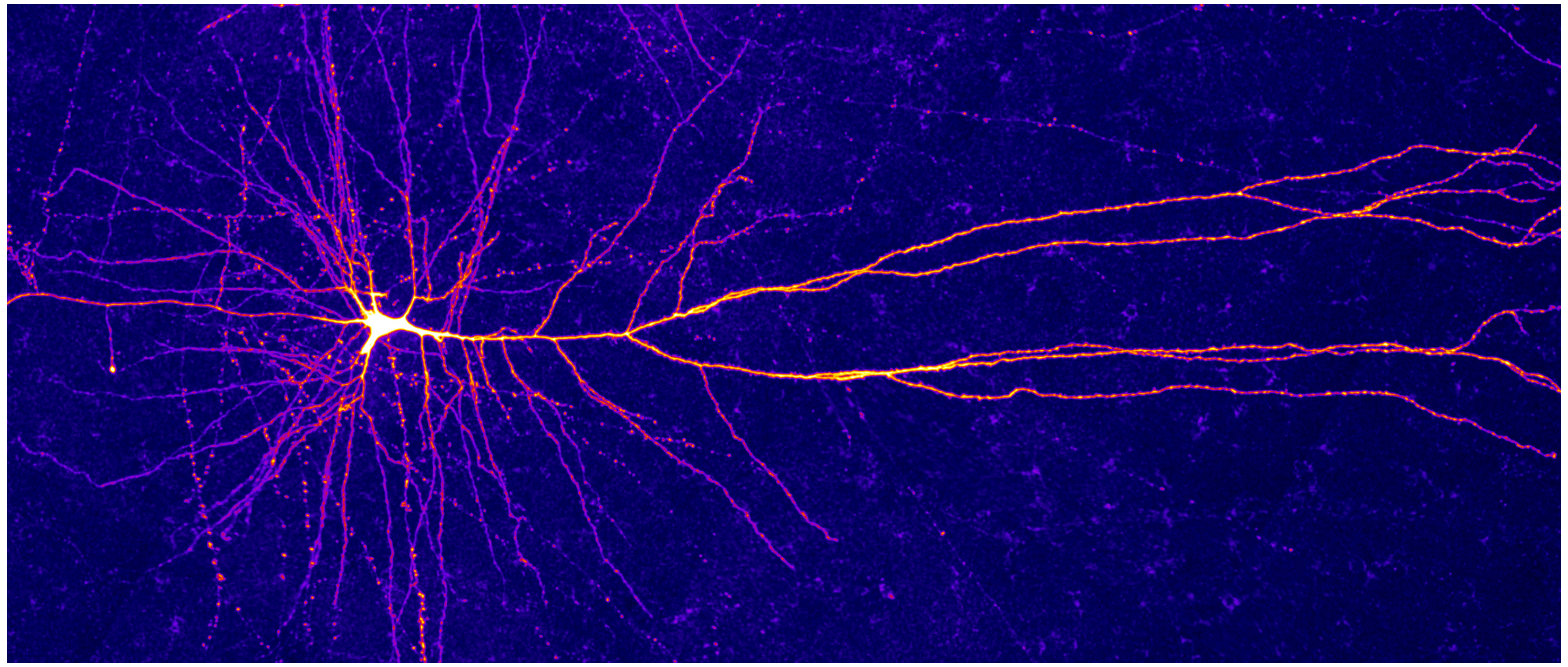
Marco Benedetti

You snooze, you don't loose: the role of dreaming in Neural Networks for associative memory

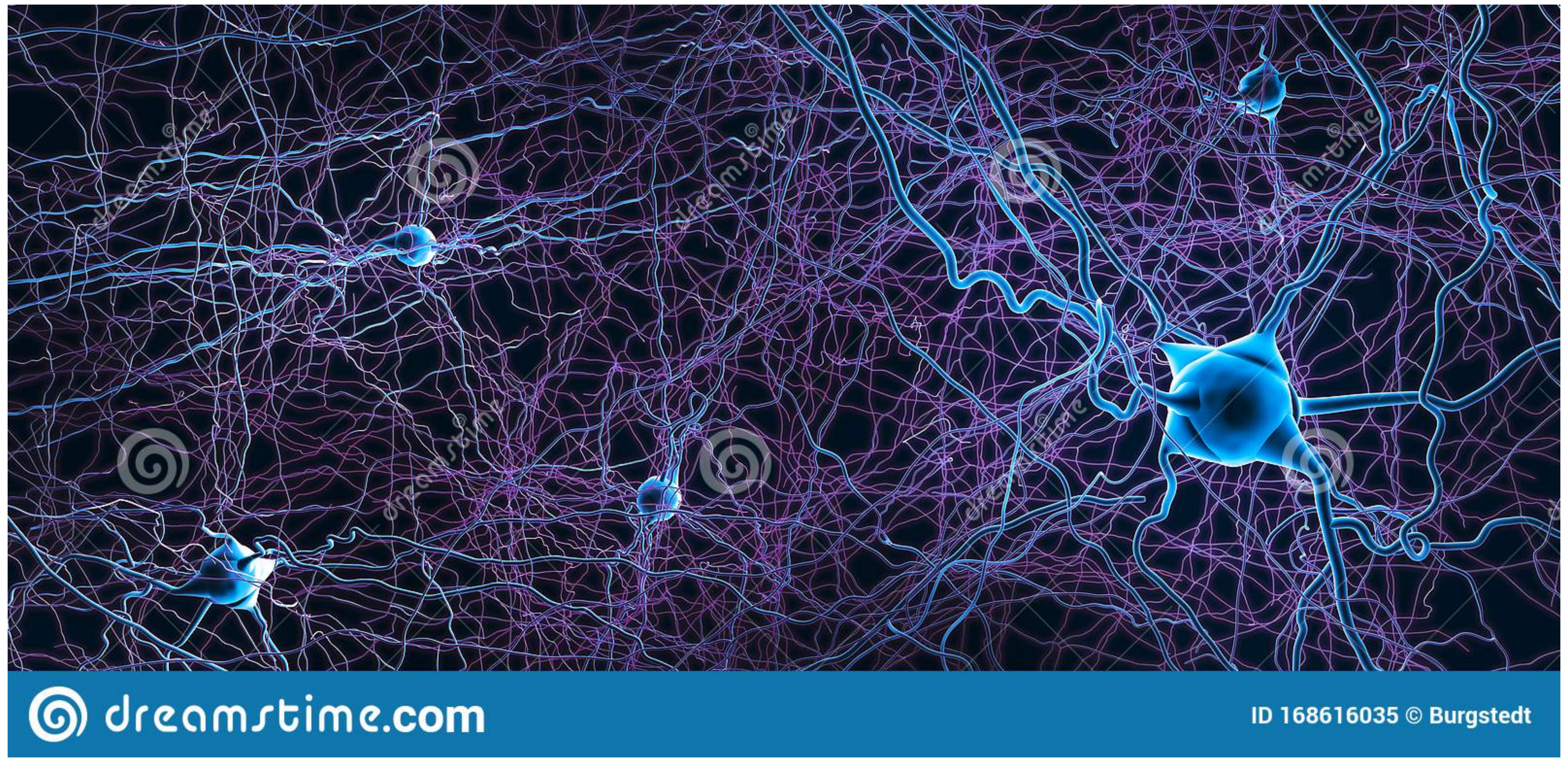
The most modern and successful strategies to build AI are inspired by the structure of the brain itself.



The building block of the brain is the neuron, composed by dendrites, soma and axon. The state of the neuron is characterized by it's firing rate. Neurons are connected by synapses...

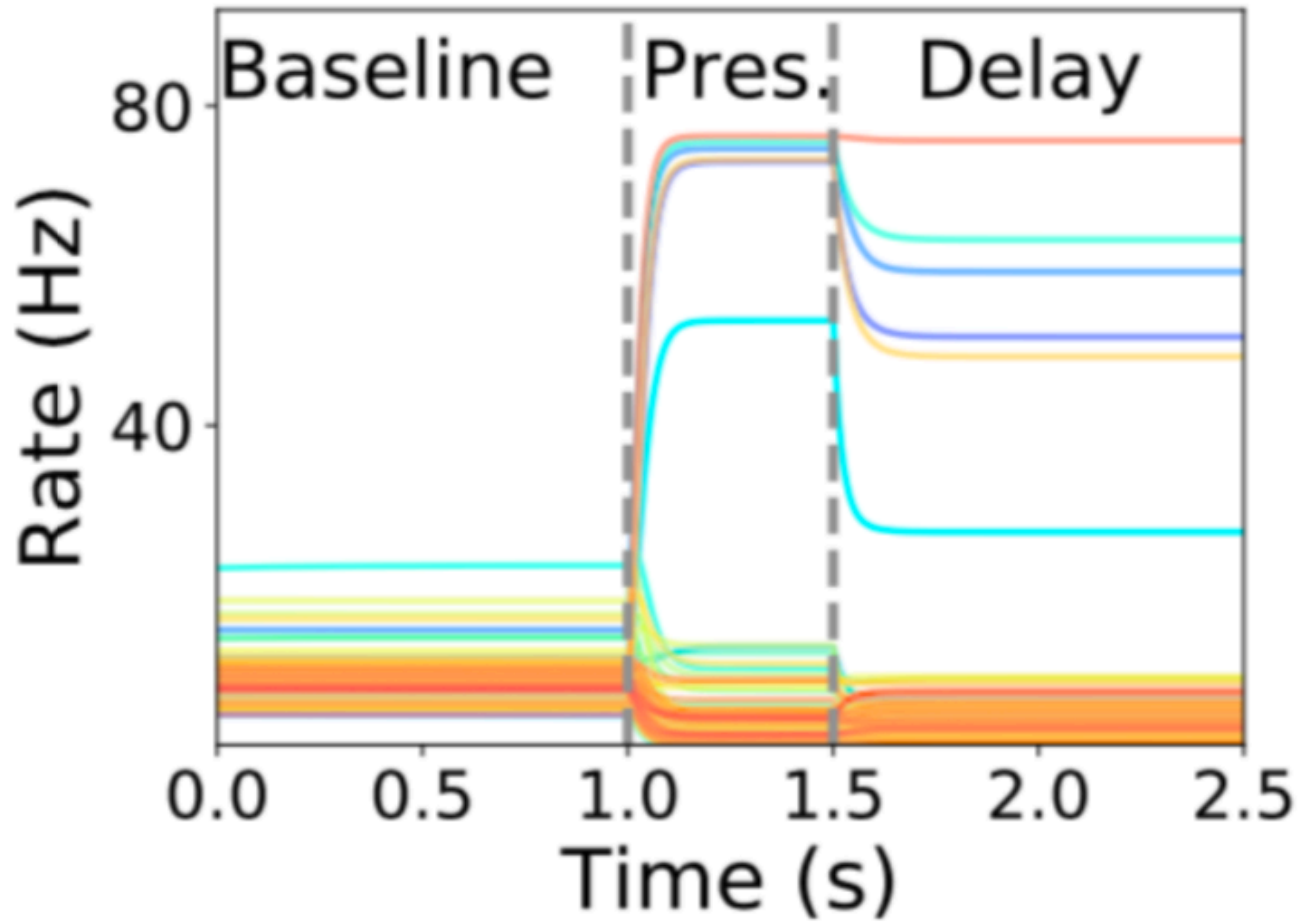
... and there is a lot of them in a human brain:

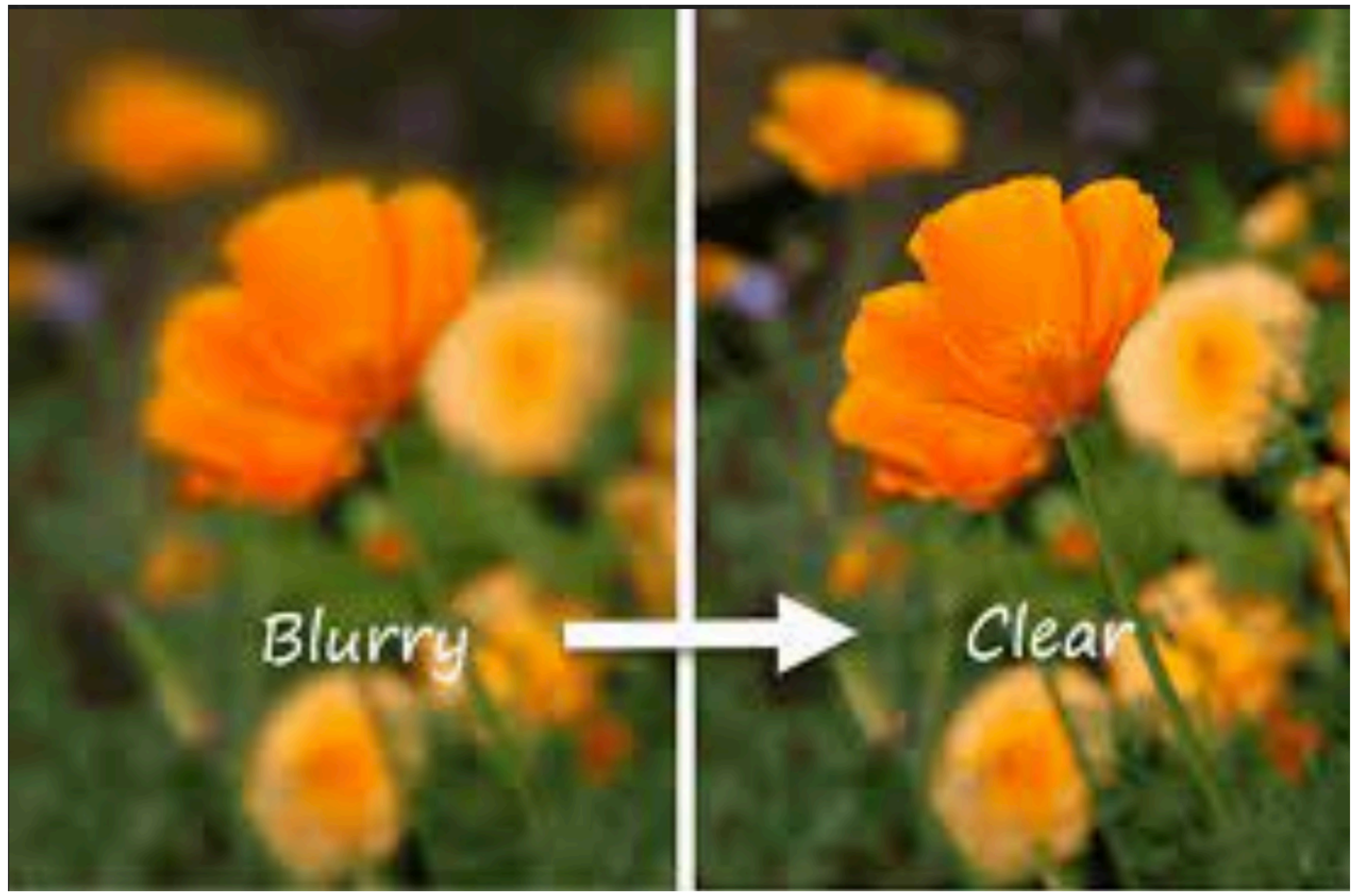
$$\sim 10^5 \frac{\text{neurons}}{\text{mm}^3}, \ \sim 10^9 \frac{\text{synapses}}{\text{mm}^3}$$



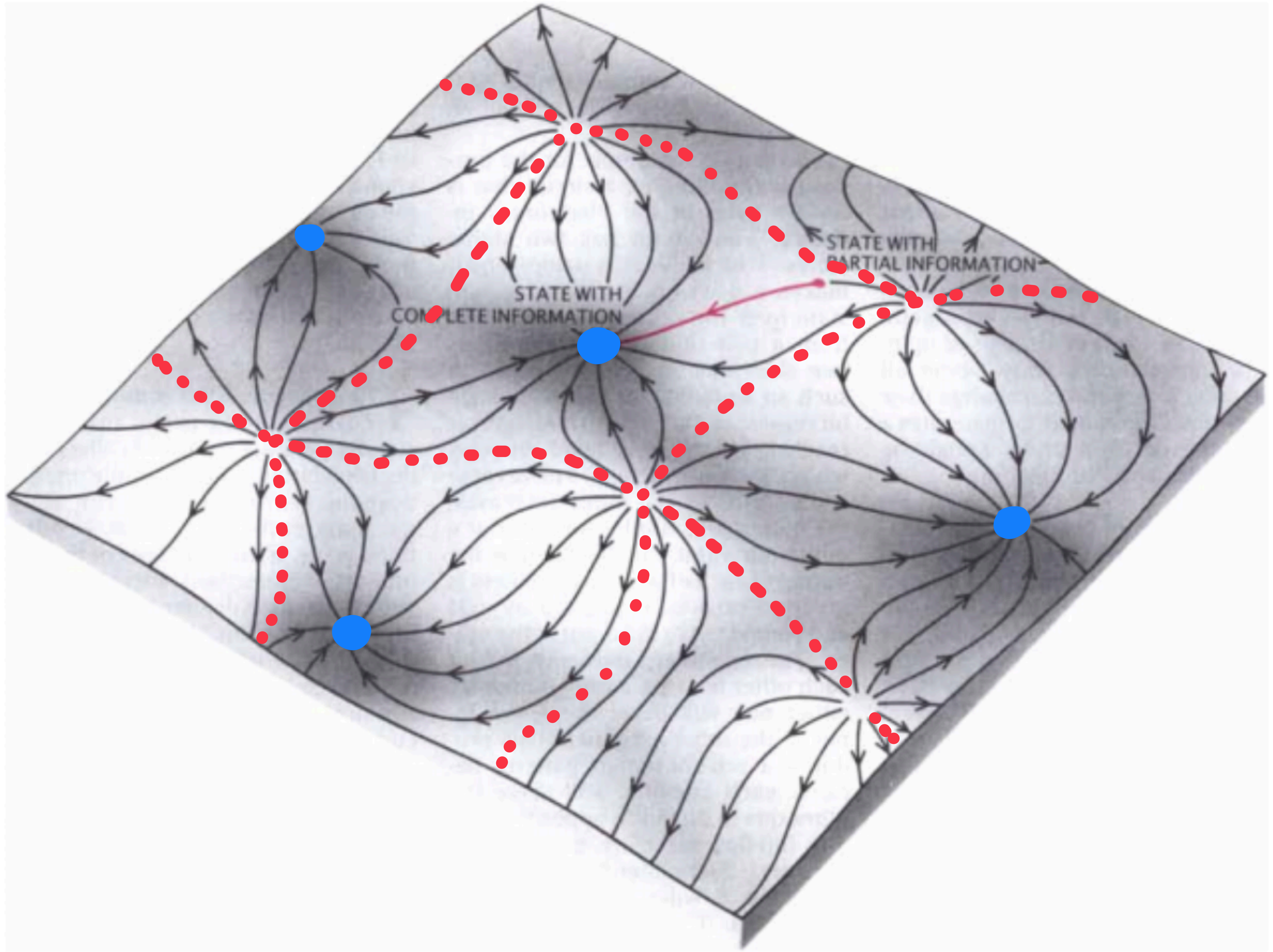
Cognitive abilities are an example of emergent property, very close to the subject of study if statistical mechanics. More specifically, **Neural Networks are certainly complex systems.**

Information is represented in the brain in terms of neural activity patters: sequences of active and quiescent neurons specific to each input.

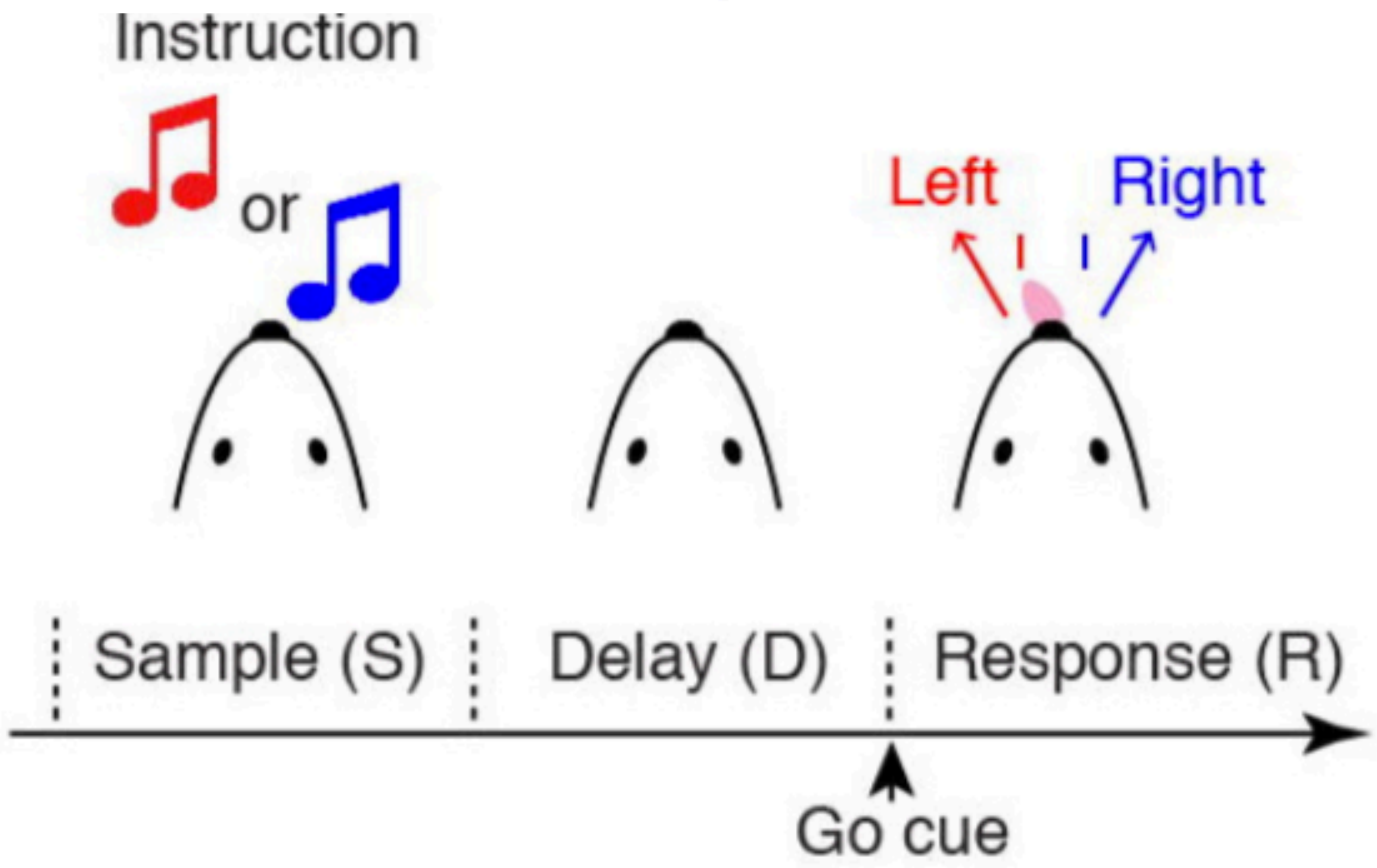


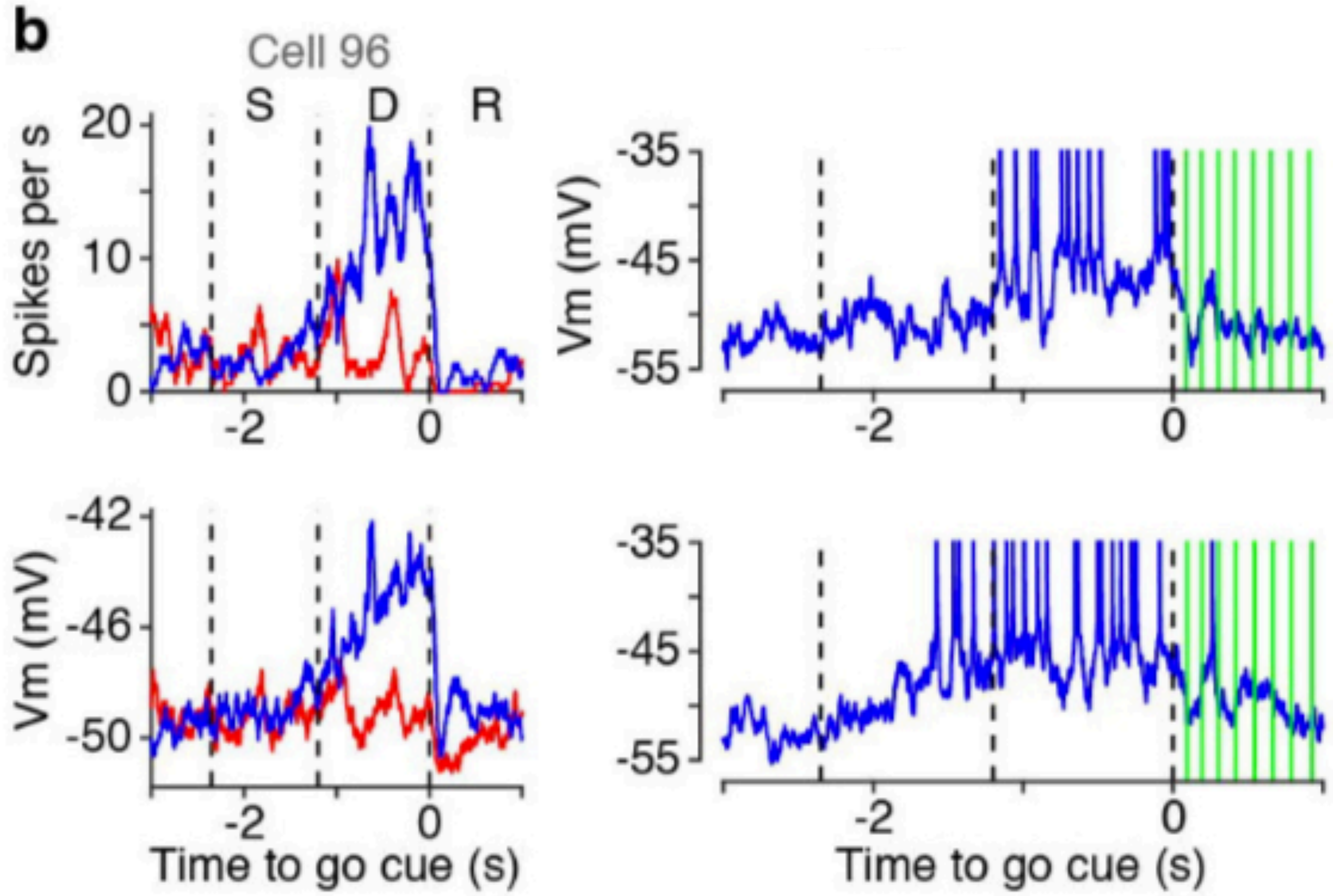


One of the most important functions performed by the brain is **associative memory**: inputs resembling memorized objects should be associated by the network to the objects themselves.



A popular framework to model this ability is that of attractor dynamics: the network is endowed with a specific dynamics, such as to **build attractors in close proximity to the memorized neural activity patterns.** There is convincing evidence that the attractor framework is indeed relevant describing mouse brain activity during Delayed Response Tasks.





One of the earliest successes in brain modeling is Hopfield's model. A system composed of N binary neurons $\{\sigma_i\}_{i=1}^{N}$ is used to store $P = \alpha N$ patterns containing N binary symbols $\xi_i^\mu = \pm 1$, $\mu \in \{1,\ldots,P\},\ i \in \{1,\ldots,N\}$. The dynamics is given by a Markov Process and the asymptotic probability distribution is $P[\sigma] \sim \exp\{-\beta H[\sigma]\}$, where

$$H[\sigma] = -1/2 \sum_{i,j=1}^{N} J_{ij}\sigma_i\sigma_j, \text{ and } J_{ij} = \frac{1}{N}\sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu .$$

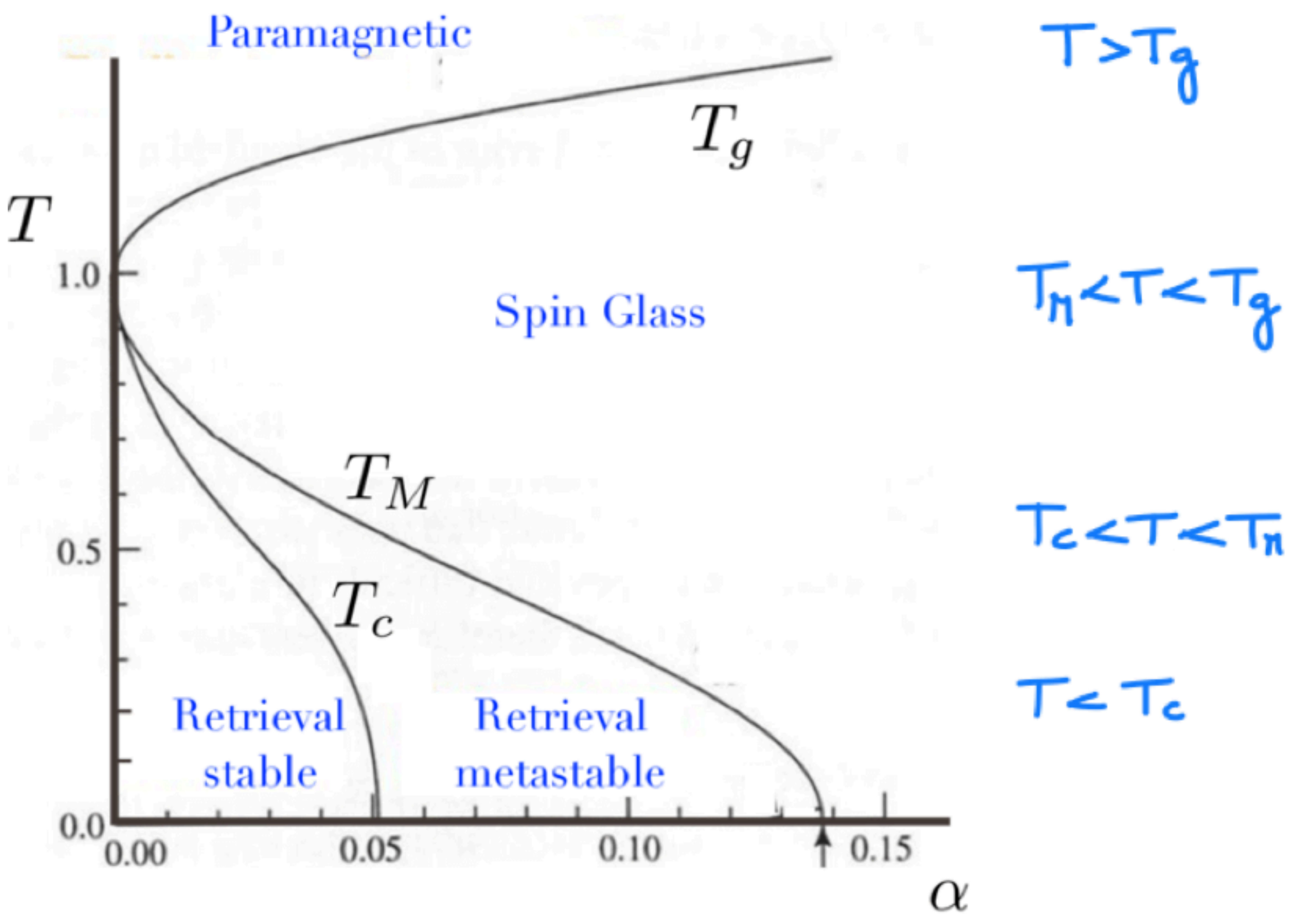
Two words on how to tackle complex systems statistical mechanics

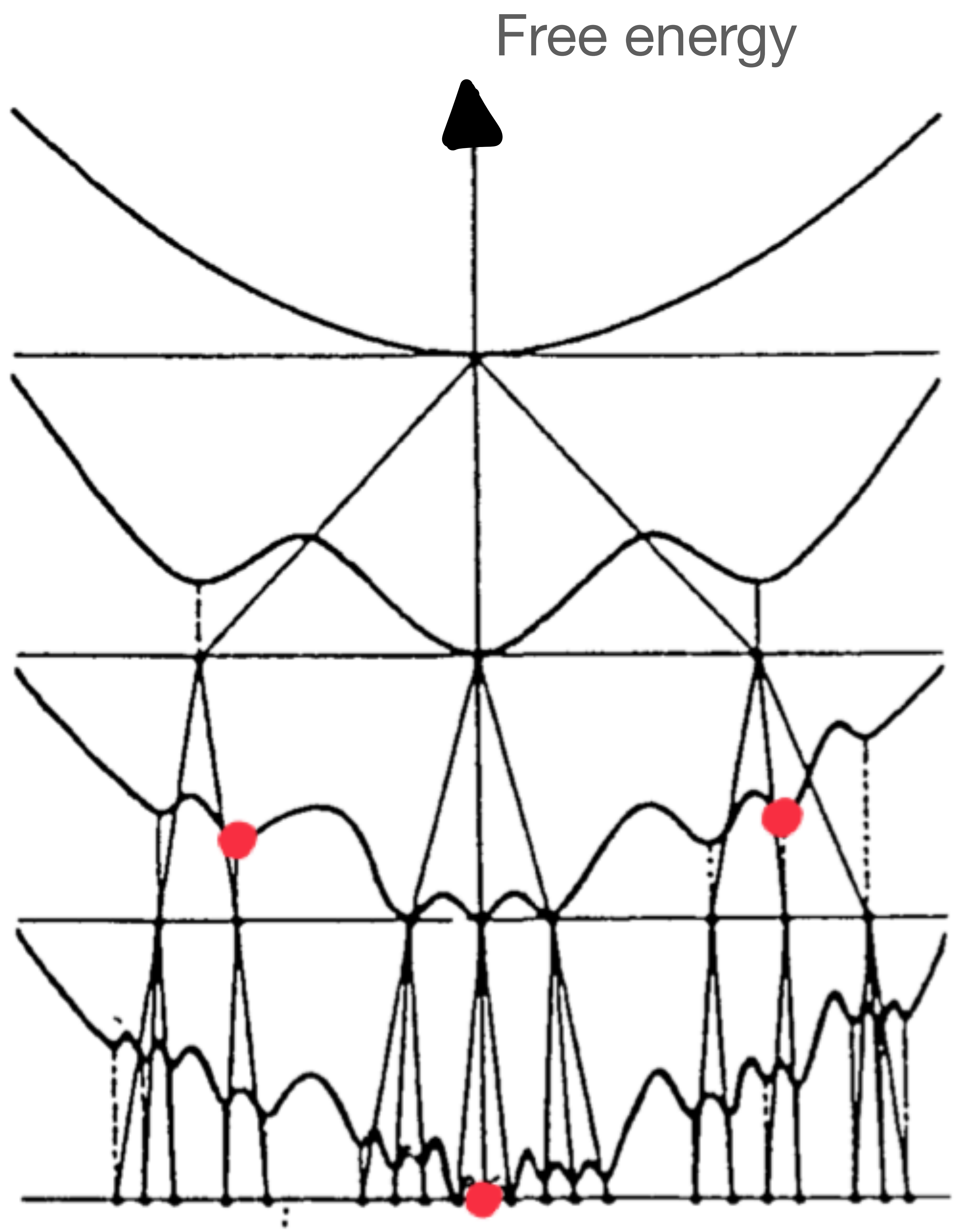
The richness of the free energy landscape of disordered systems poses great challenges to the statistical physicist. **As ergodicity breaks, the Gibbs measure is shattered into an exponential number of probability lumps** (pure states).

Each pure state is characterized by the average magnetization of each spin $\{\langle\sigma_i\rangle\}$ (confront this with an Ising model, where each site has the same average magnetization). Traditional order parameters cannot aid us to sort this mess. On top of this, **we must average over disorder** (the different realizations of the memories)

We can introduce "**replicas**": we will study a system composed of n weakly coupled copies of the original system. The **alignment among those replicas can give us important information on the structure of the pure states**. We don't need to guess the appropriate local field.

The full phase space diagram can be computed with the tools of spin glass physics

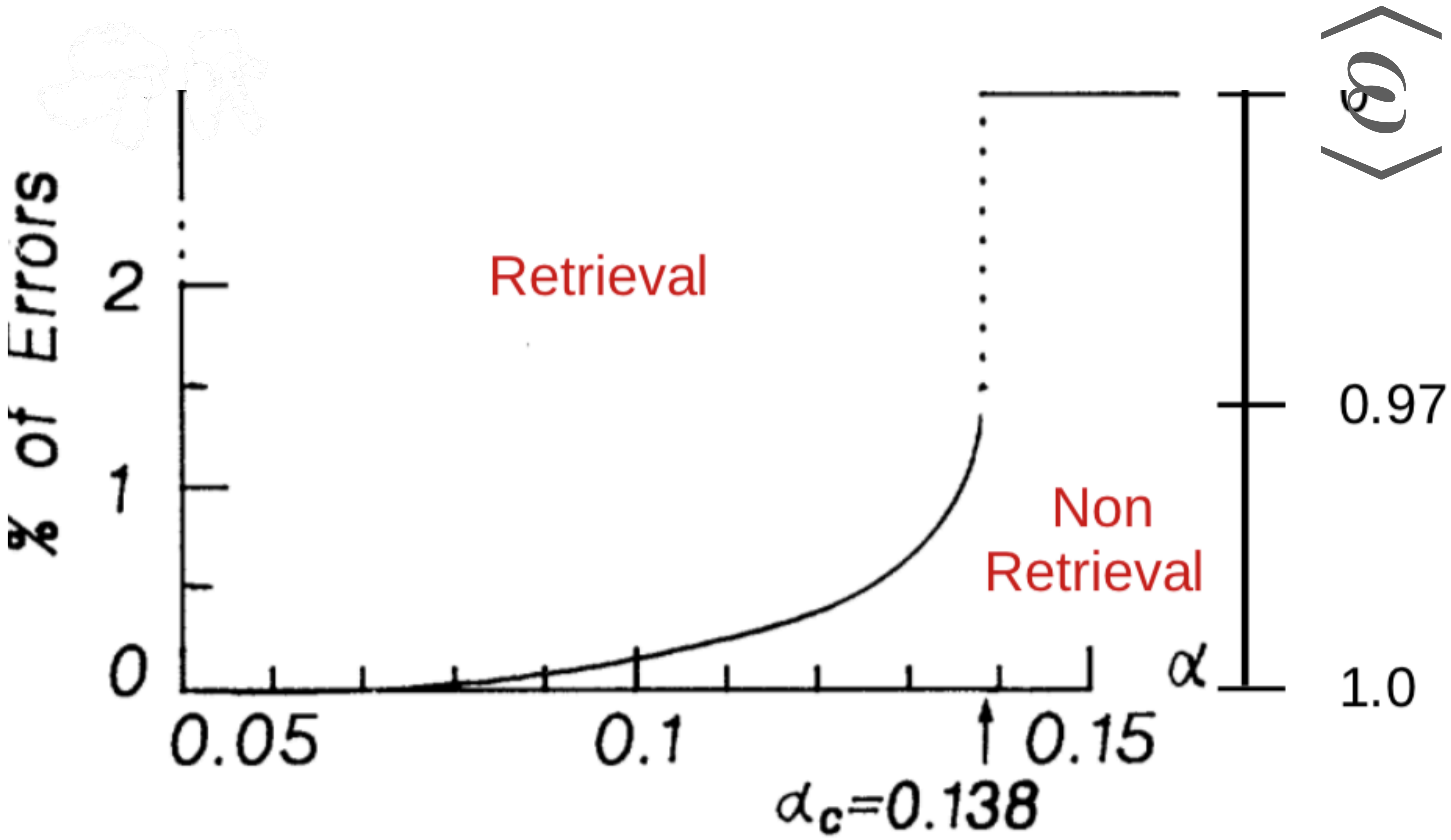




The performance of the network is measured in by the average overlap between stored memories ξ^μ and the configuration σ reached by the dynamics when starting from the exact memories:

$$\langle \omega \rangle = 1/N \sum_{i=1}^{N} \xi_i^\mu \langle \sigma_i \rangle \,.$$

A high value of $\langle \omega \rangle$ means that the random motion in configuration space is limited to a tight region centered on the memory, i.e. memories have finite basins of attraction (error correcting performance).



And now some fun: should an artificial brain sleep?

NATURE VOL. 304 14 JULY 1983

The function of dream sleep

Francis Crick* & Graeme Mitchison*

We propose that the function of dream sleep (more properly rapid-eye movement or REM sleep) is to remove certain undesirable modes of interaction in networks of cells in the cerebral cortex. We postulate that this is done in REM sleep by a reverse learning mechanism (see also p.158), so that the trace in the brain of the unconscious dream is weakened, rather than strengthened, by the dream.

In the same year Hopfield, Feinstein and Palmer independently come up with an inverse learning procedure mimicking the biological dreaming process, as seen by Crick and Mitchison.

- Initalize J according to Hebb's learning rule

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

- Pick at random a starting configuration for the dynamics
- Evolve the initial configuration following zero temperature dynamics

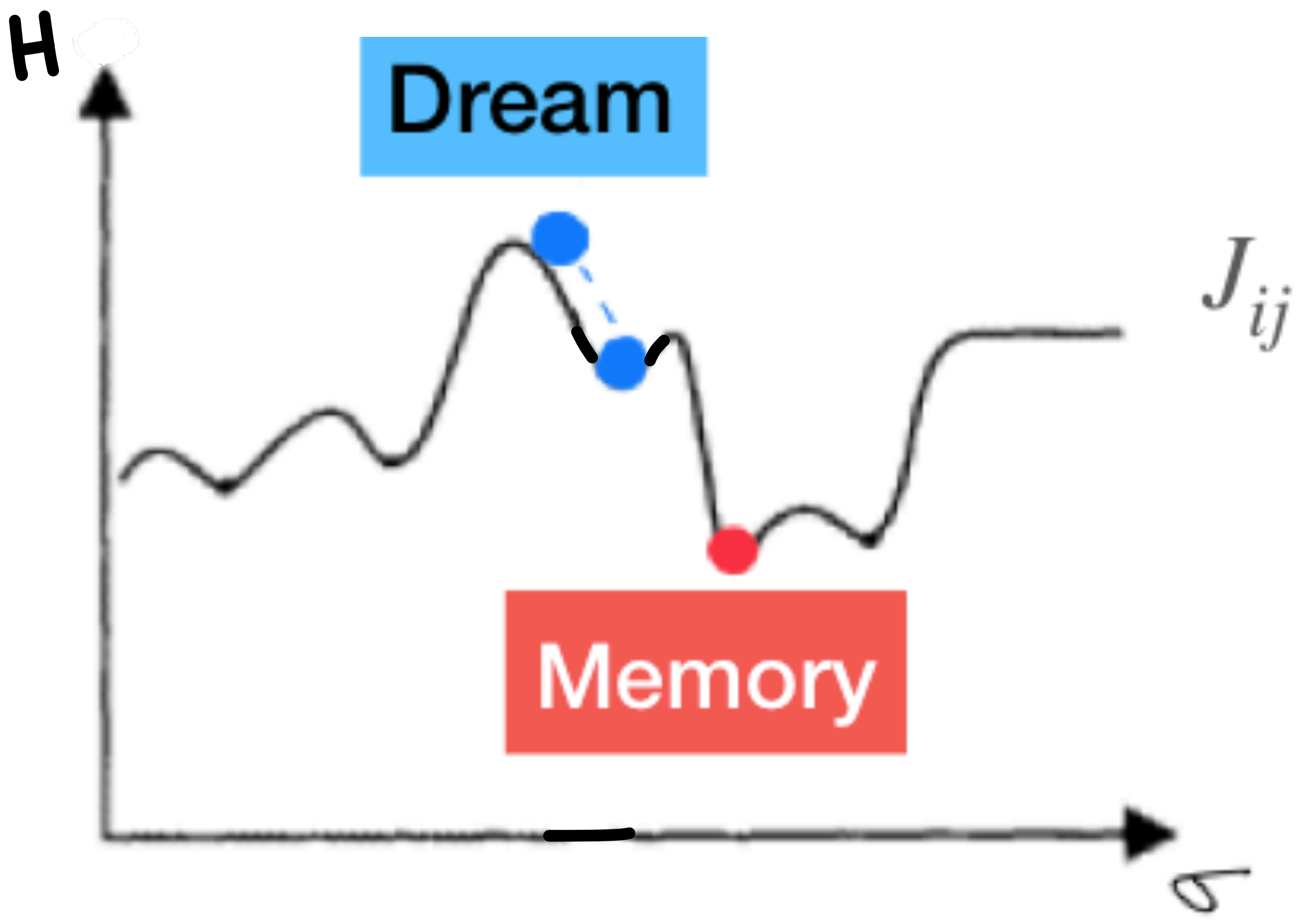
$$\sigma_i^{(t+1)} = \text{sgn}\left(\sum_j J_{ij}\sigma_j^{(t)}\right)$$

  until convergence

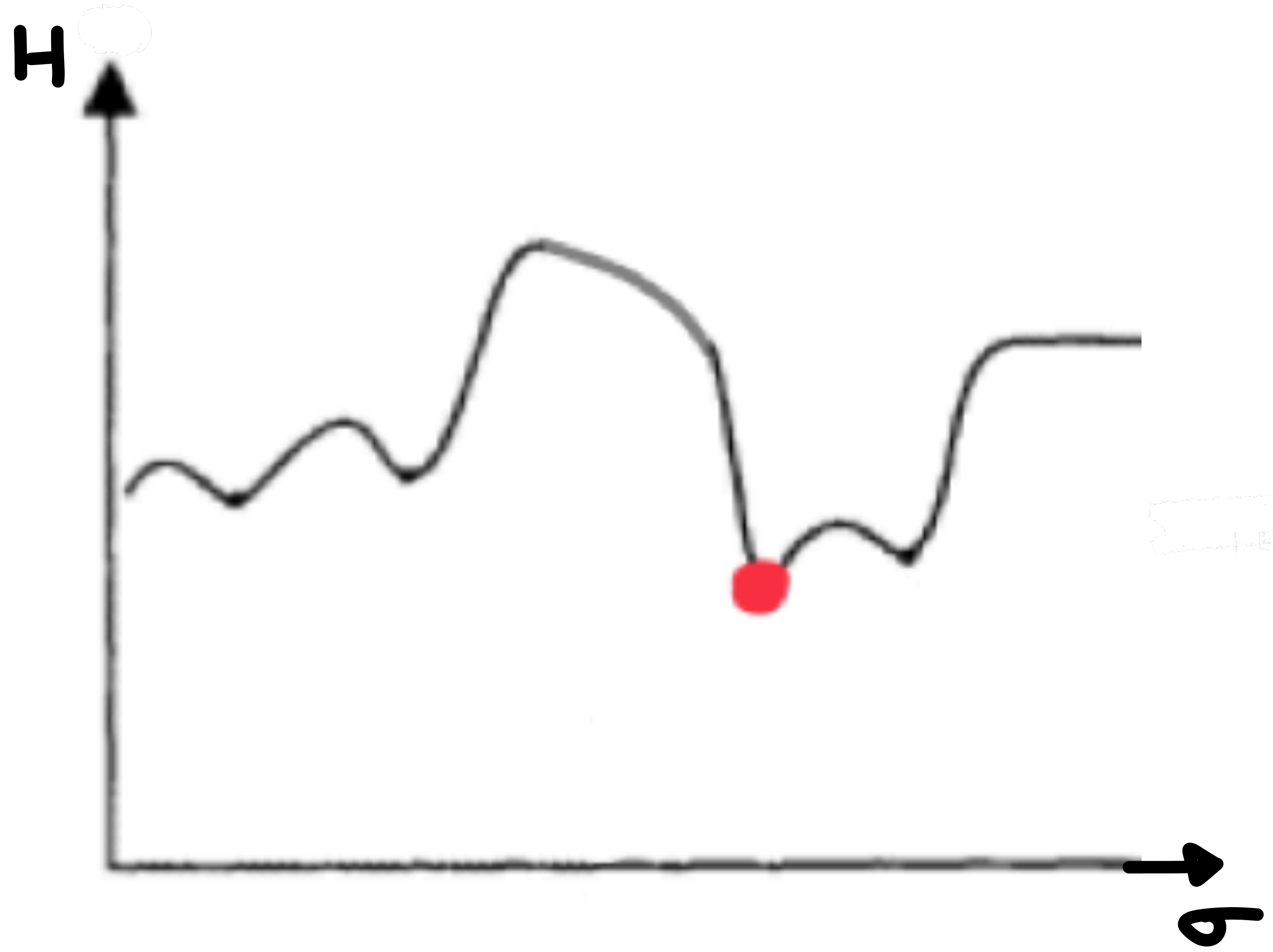
- Update the connectivity matrix

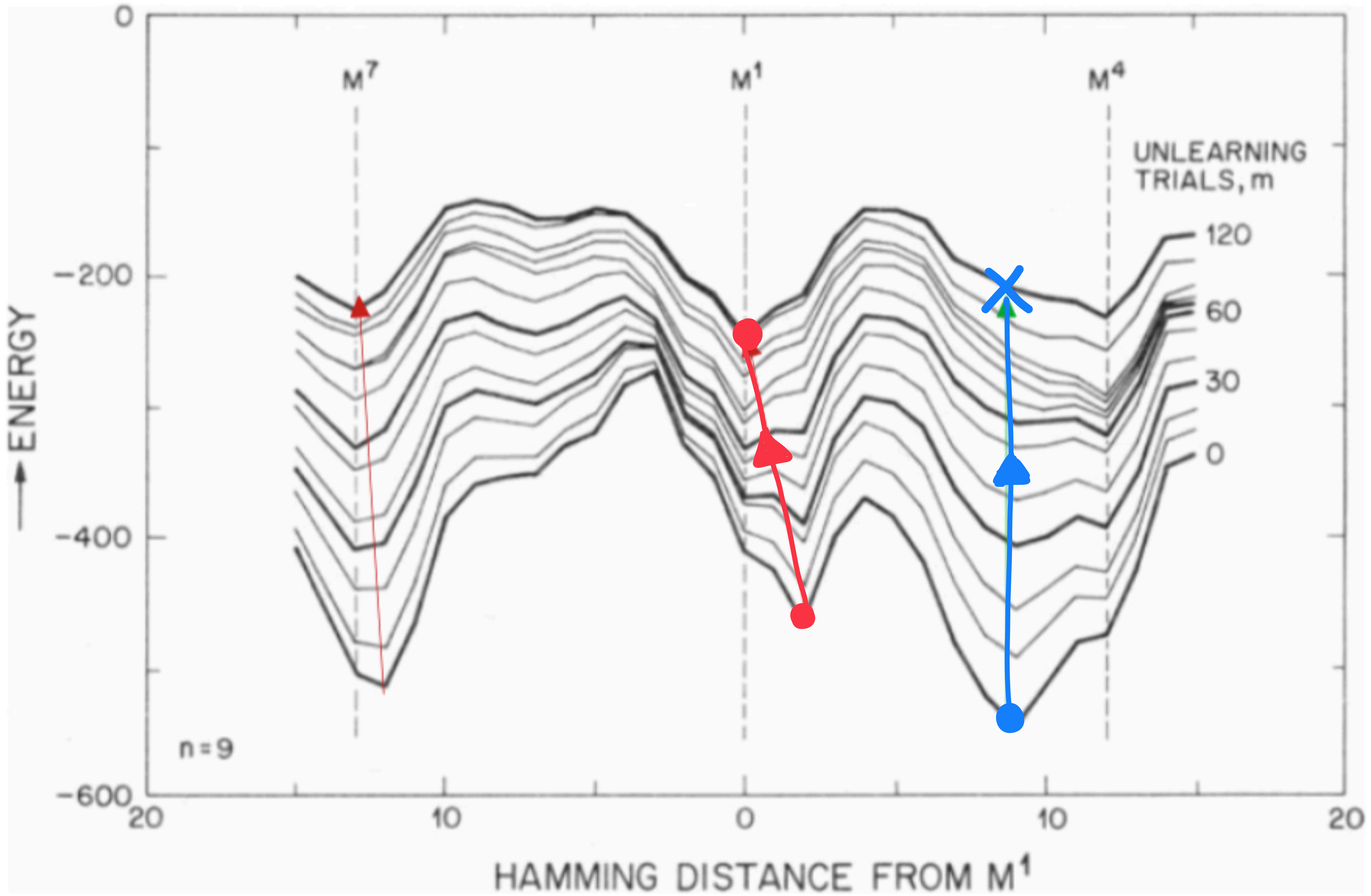
$$J_{ij}^{(t+1)} = J_{ij}^{t} - \epsilon\sigma_i\sigma_j$$

Why does it work?



$$J_{ij} \to J_{ij} - \epsilon\sigma_i\sigma_j$$





Spurious minima disappear

Memories become fixed points

This kind of “**perfect retrival**” is more common among supervised algorithms. **Gardner's algorithm** is a famous example:

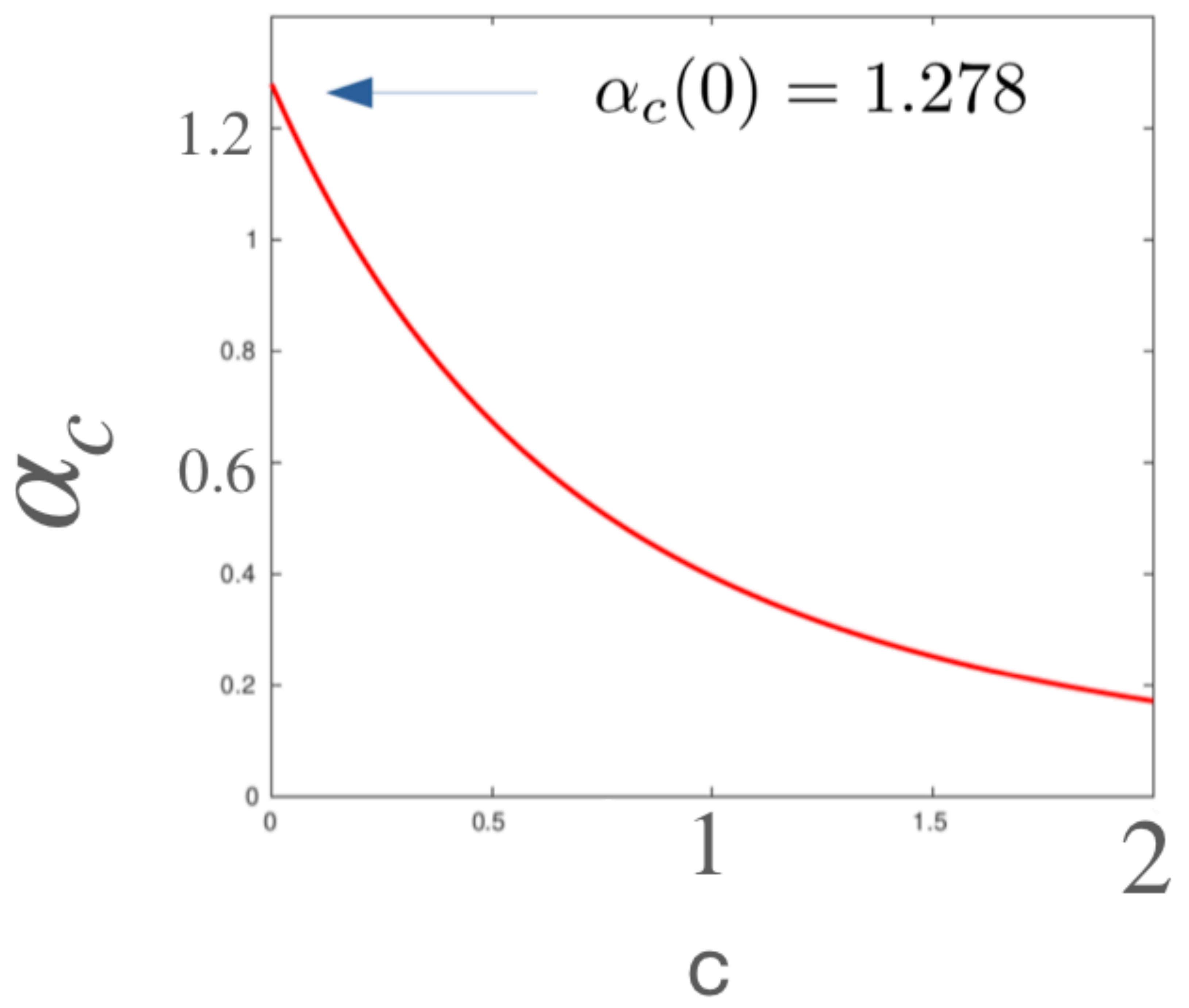
$$J_{ij}^{(d+1)} = J_{ij}^{(d)} + \lambda \sum_{\mu}^{P} (\epsilon_i^{\mu,d} + \epsilon_j^{\mu,d}) \xi_i^\mu \xi_j^\mu \; ; \qquad J_{ii}^{(d)} = 0 \quad \forall i$$

$$\epsilon_i^{\mu,d} = \Theta\left(-\xi_i^\mu \sum_{j=1}^{N} J_{ij}^{(d)} \xi_j^\mu + c \left(\sum_{j=1}^{N} (J_{ij}^{(d)})^2 \right)^{1/2} \right)$$

Mask

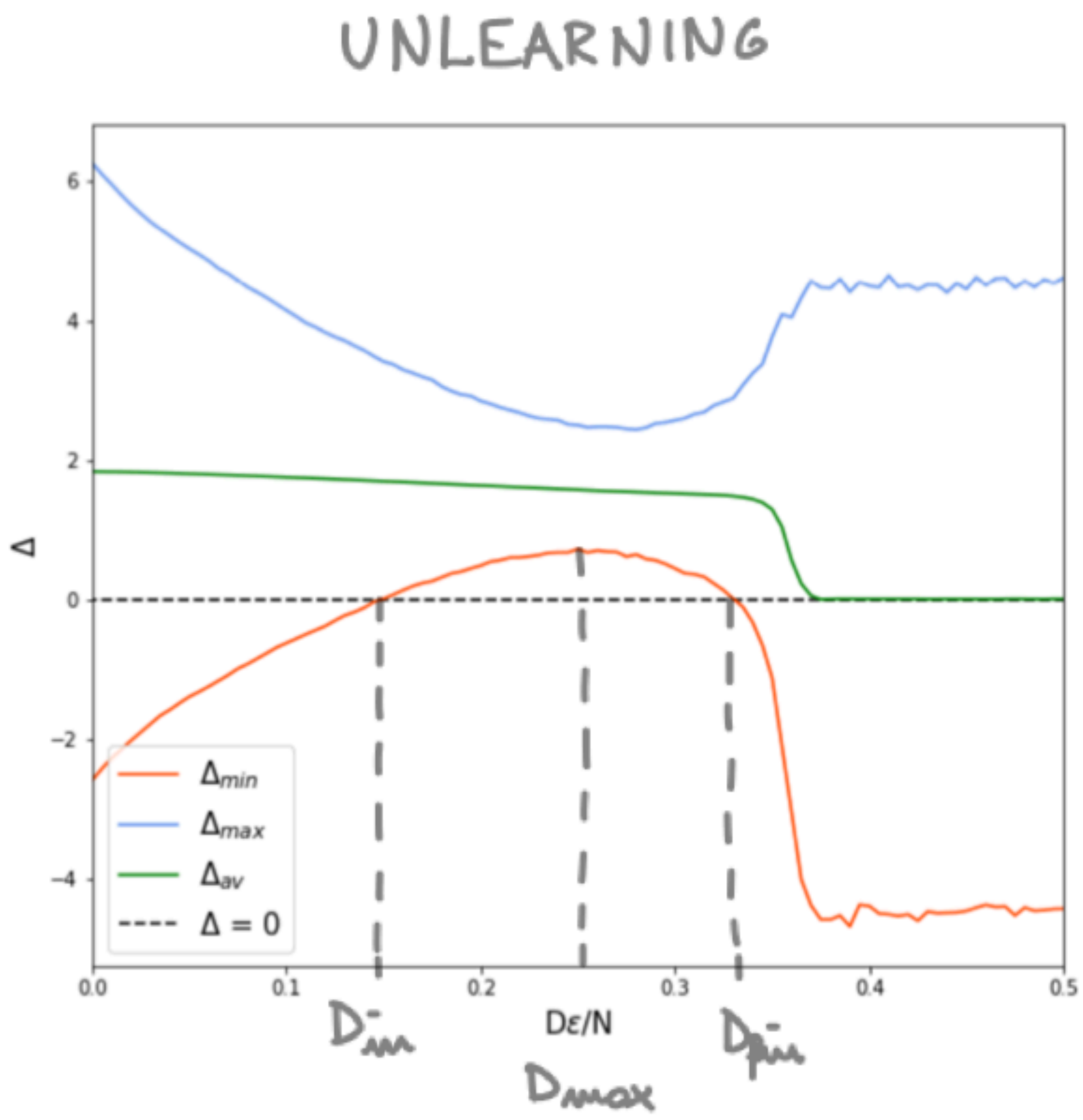
$$\Delta_i^\mu = \xi_i^\mu \sum_{j=1}^{N} \frac{J_{ij}}{|\vec{J}_i|} \xi_j^\mu$$

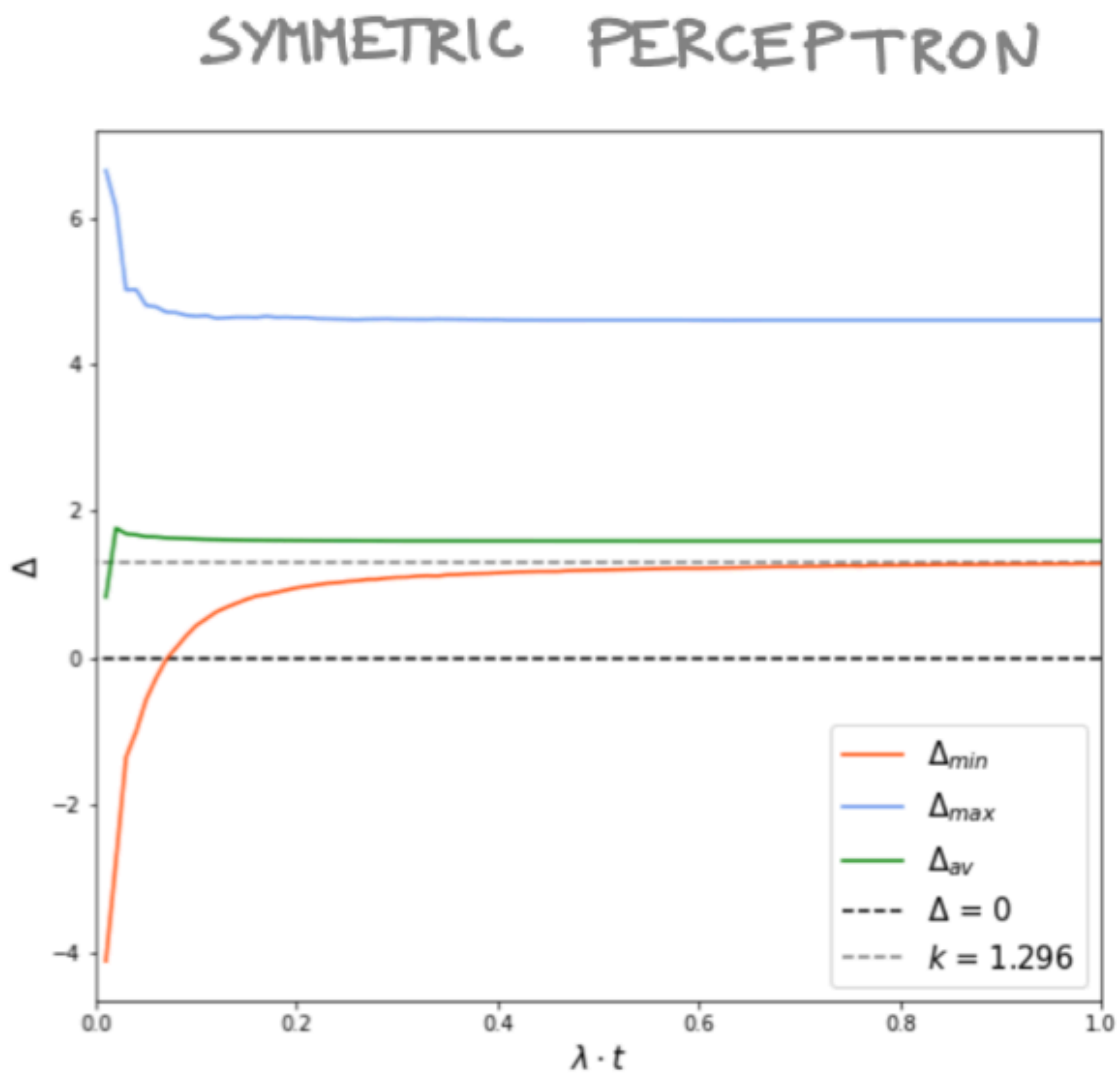
Stability



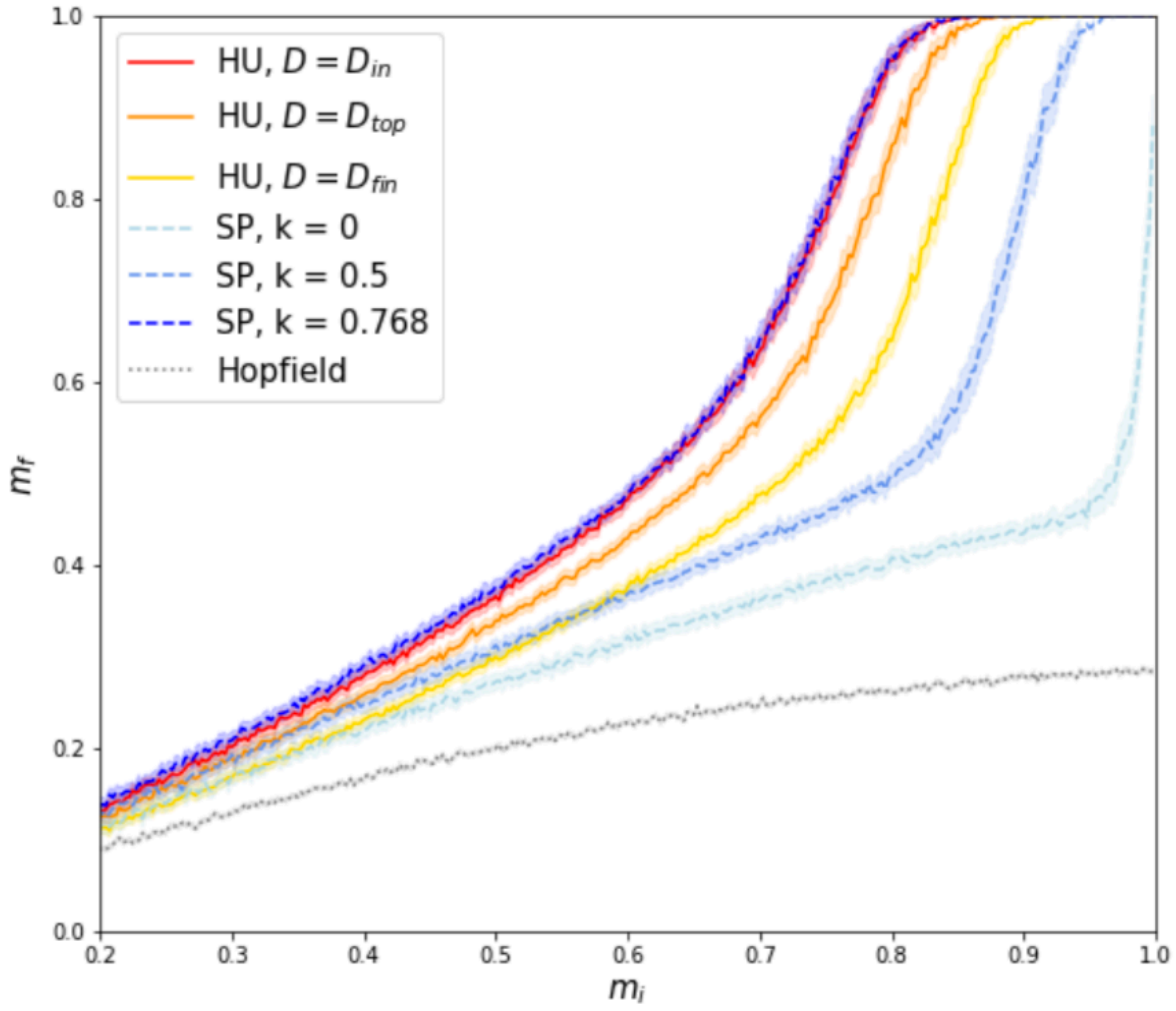
Guaranteed to converge whenever a solution exists!

A intuitive way to visualize the progress over "time" of the algorithm is studying the **evolution of stabilities**...





...and the **attraction basins** obtained.



The two algorithms perform indistinguishably in their respective optimal regimes

Is there a theoretical framework explaining this surprising correspondence between supervised and unsupervised algorithms? If you know one, call me at 320 3141657

Take-home messages

- Even the crudest models of the brain can benefit in surprising ways from educated guesses based on biology.
- Statistical mechanics and in particular disordered systems physics provide a very successful toolbox to tackle analytically models of the brain.


The End