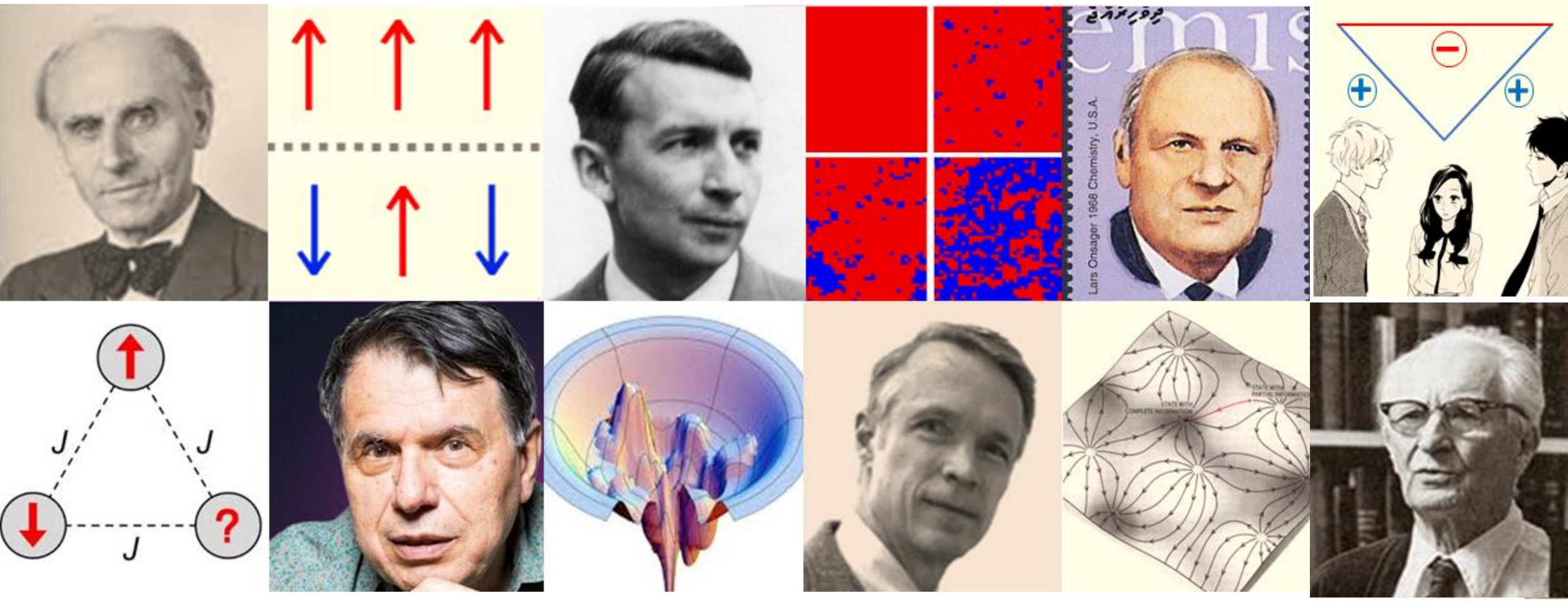
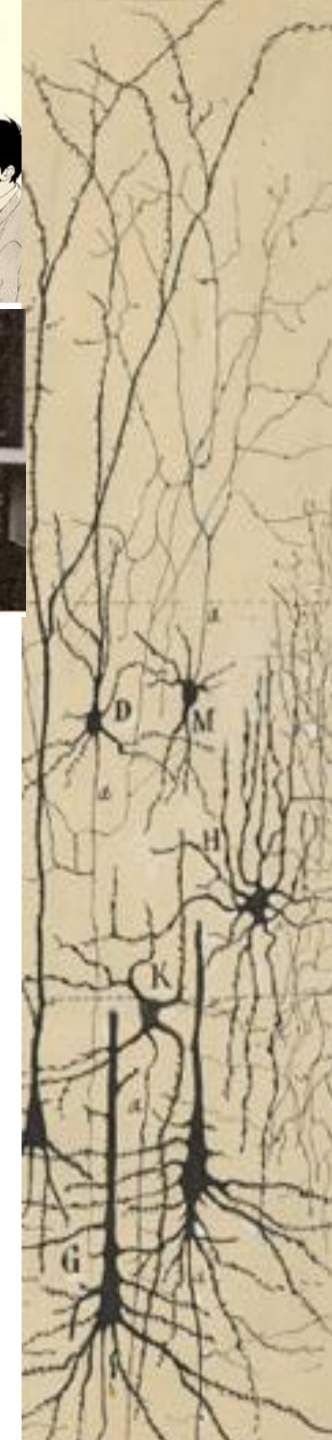




Mind, Memory and Magnets

Introducing Neural Networks

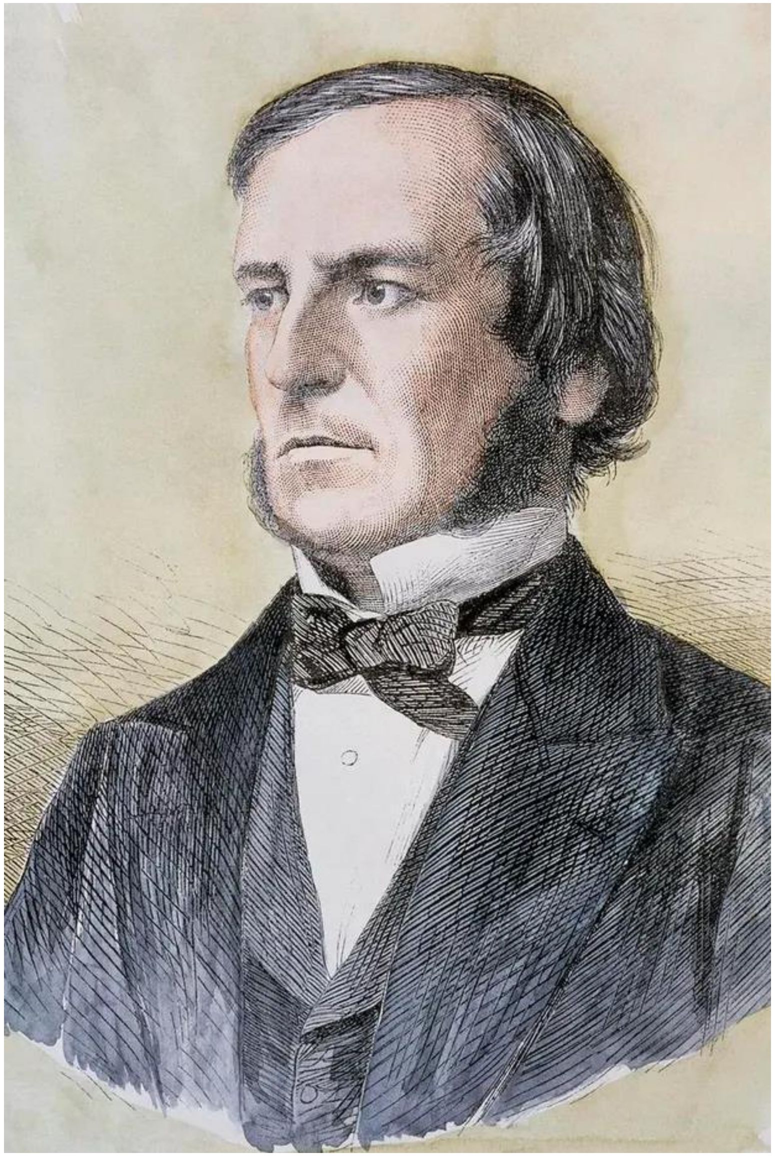
Sitabhra Sinha
IMSc Chennai

The laws of thought

"…to investigate the fundamental laws of those operations of the mind by which reasoning is performed;
to give expression to them in the symbolical language of a Calculus, and upon this foundation to establish the science of Logic and construct its method …
… and, finally, to collect from the various elements of truth brought to view in the course of these inquiries some probable intimations concerning the nature and constitution of the human mind."

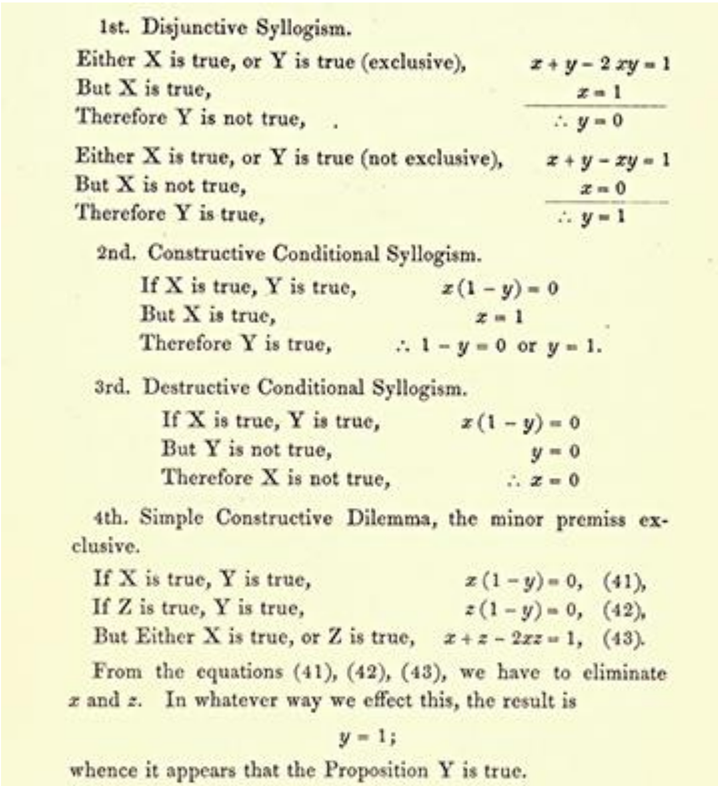
An Investigation of the Laws of Thought (1854)

George Boole
(1815-1864)



https://www.theguardian.com/

The Laws of Thought ⇒ Automated reasoning ?



1st. Disjunctive Syllogism.

Either X is true, or Y is true (exclusive),	$x+y-2xy=1$
But X is true,	$x=1$
Therefore Y is not true,	$\therefore y=0$
Either X is true, or Y is true (not exclusive),	$x+y-xy=1$
But X is not true,	$x=0$
Therefore Y is true,	$\therefore y=1$

2nd. Constructive Conditional Syllogism.

If X is true, Y is true,	$x(1-y)=0$
But X is true,	$x=1$
Therefore Y is true,	$\therefore 1-y=0$ or $y=1$.

3rd. Destructive Conditional Syllogism.

If X is true, Y is true,	$x(1-y)=0$
But Y is not true,	$y=0$
Therefore X is not true,	$\therefore x=0$

4th. Simple Constructive Dilemma, the minor premiss exclusive.

If X is true, Y is true,	$x(1-y)=0$, (41),
If Z is true, Y is true,	$z(1-y)=0$, (42),
But Either X is true, or Z is true,	$x+z-2xz=1$, (43).

From the equations (41), (42), (43), we have to eliminate x and z. In whatever way we effect this, the result is

$$y=1;$$

whence it appears that the Proposition Y is true.

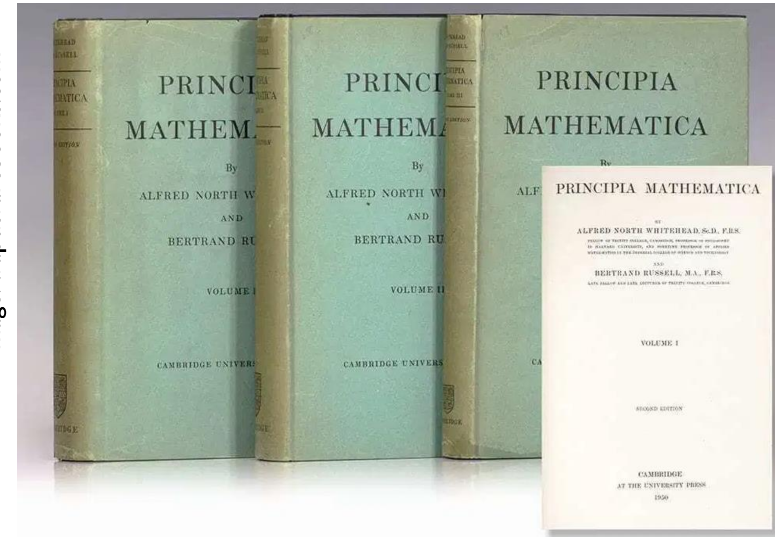
In1937 a 21-year-old Claude Shannon wrote his Master's thesis at MIT demonstrating that electrical applications of boolean algebra could construct and resolve any logical, numerical relationship

⇒ design of digital circuits, digital computers.

Claude Shannon
(1916-2001)

Logical calculus:
The automation of thought

Image: .raptisrarebooks.com

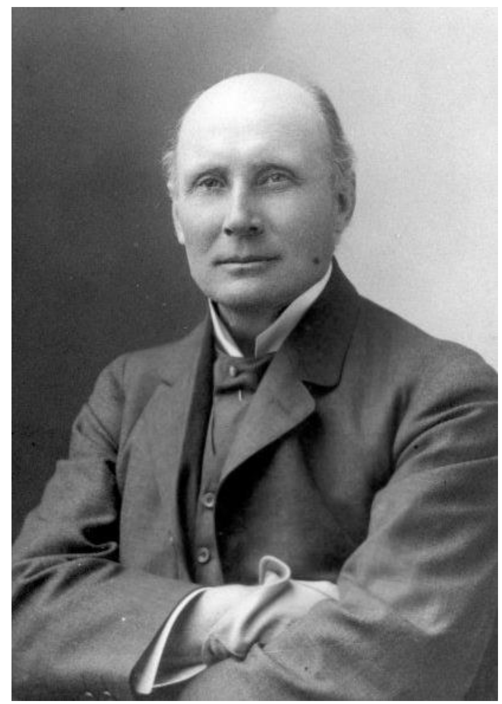


Principia Mathematica (1910-1913)

of Whitehead and Russell provided a model by attempting to derive the entire body of mathematical knowledge by using logical operations such as

- Conjunction (AND)
- Disjunction (OR)
- Negation (NOT)

on a set of simple propositions (either TRUE or FALSE)



Alfred North Whitehead (1861-1947)

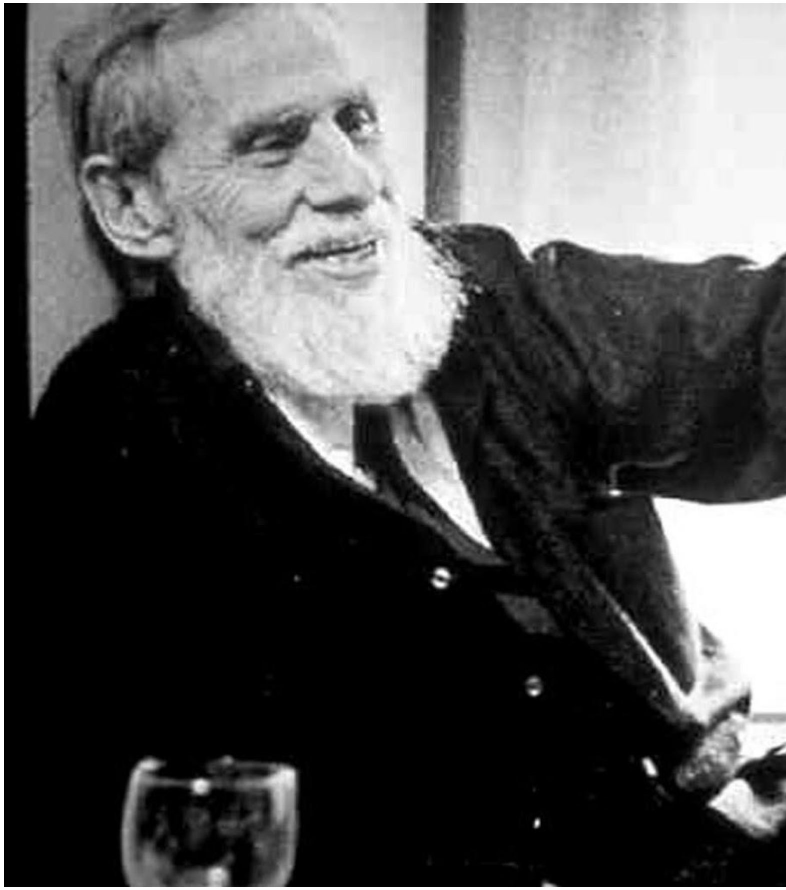
Image: pinterest.co.uk

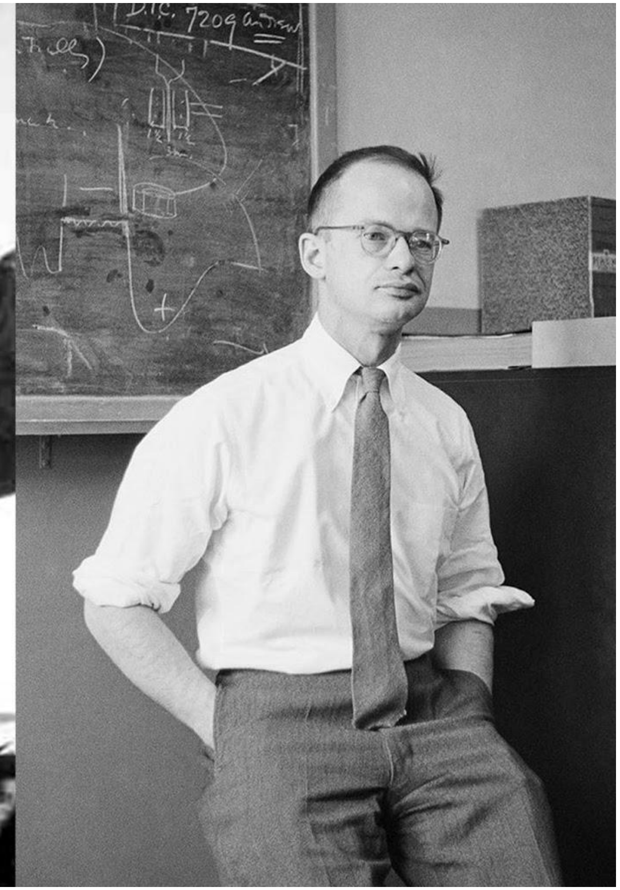
Bertrand Russell (1872-1970)

Image: flickr.com

The logical calculus of nervous activity

http://infonintelli.blogspot.com/2017/11/





https://nautil.us/the-man-who-tried-to-redeem-the-world-with-logic-235253/

Warren S. McCulloch (1898-1969)   Walter H Pitts (1923-1969)

"[They recognized that] the laws governing the embodiment of mind should be sought among the laws governing information rather than energy or matter."

Seymour Papert

The McCulloch-Pitts neuron

Image: Current Biology

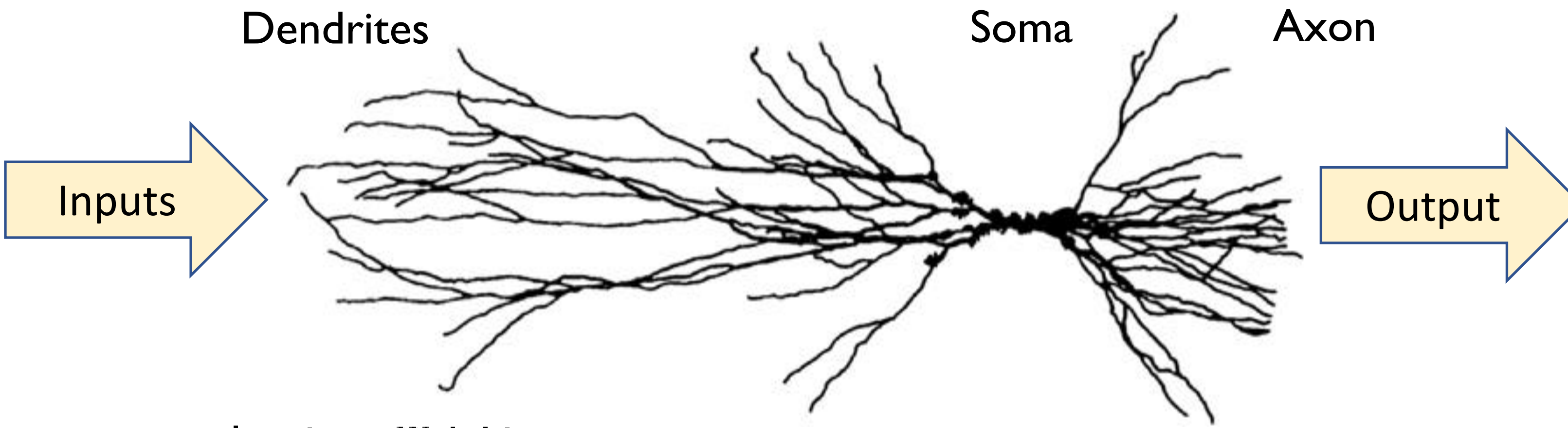
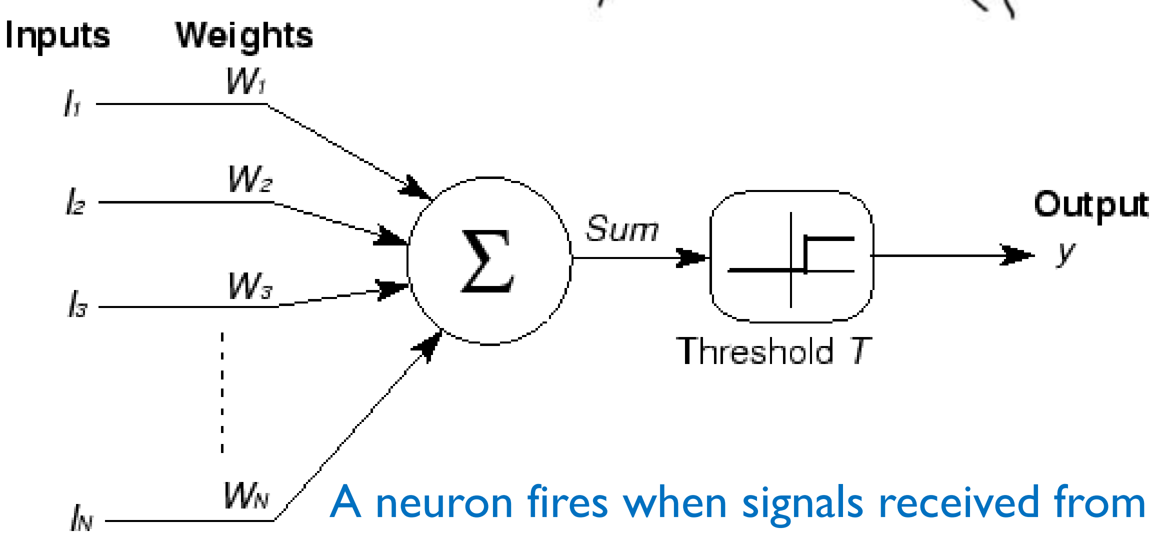


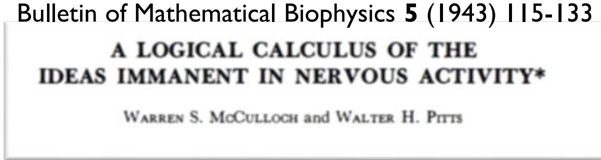
Image: chatbotslife.com/keras-in-a-single-mcculloch-pitts-neuron-317397cccd45



A neuron fires when signals received from neighboring cells exceed a threshold, else it is at rest ⇒ binary state (ON/OFF ≡ TRUE/FALSE)

- Signal: proposition
- Neurons: logic gates (e.g., AND)
- Varying threshold: Different logic gates

Bulletin of Mathematical Biophysics **5** (1943) 115-133

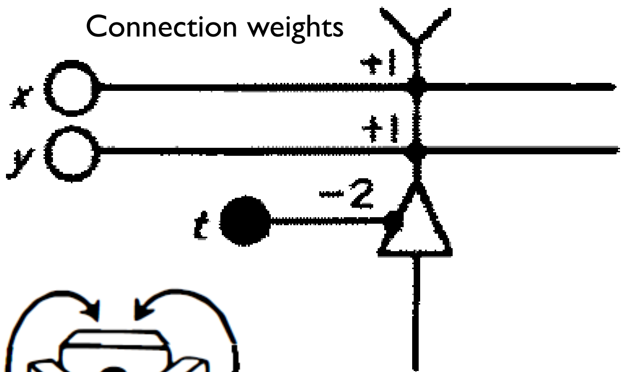

A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY*

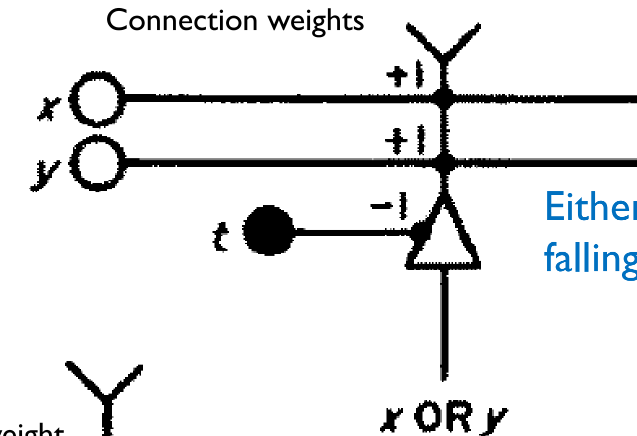
WARREN S. MCCULLOCH and WALTER H. PITTS

The McCulloch-Pitts network

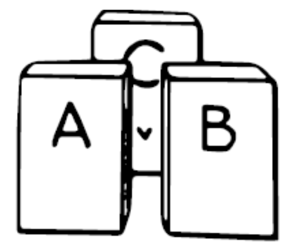
Circuits implementing computational logic

Figure: Cowan, Bull. Math. Biol. (1990)

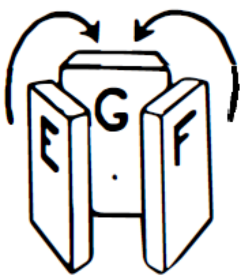




Either A or B blocks falling will make C fall

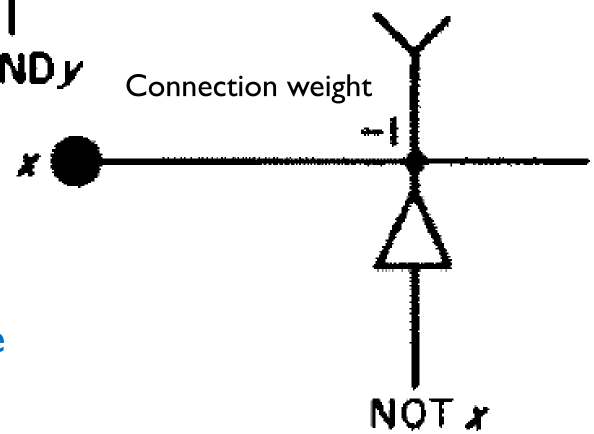


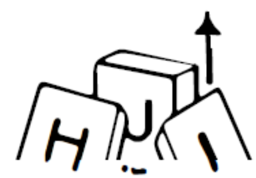
C = A or B



G = E and F

Both E and F blocks have to fall to make G fall





J = H and (not I)

J will be fell by H falling only if I does not fall

Each unit is activated iff its total excitation ≥ 0.

Positive weights: "excitatory"" synapses, negative weights: "inhibitory" synapses

open circles: excitatory neurons; filled circles: inhibitory neurons

Blocks analogy from Warren S. McCulloch, Finality and Form in Nervous Activity,1952

Perceptron

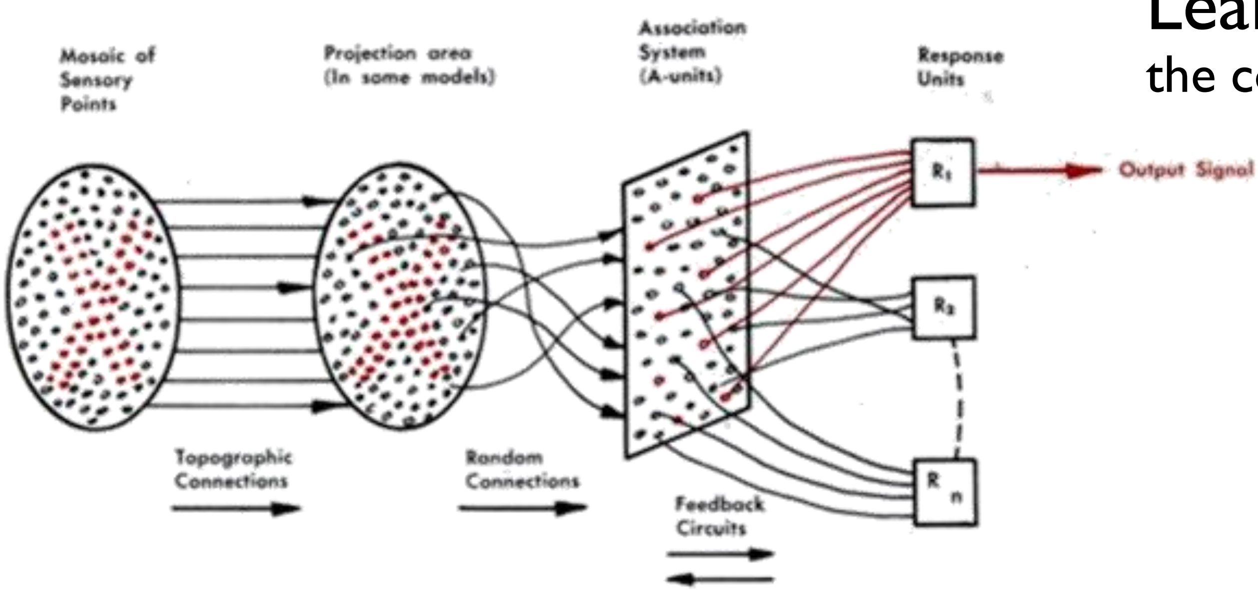
The first neural network

McCulloch-Pitts network +
Learning to adapt the link weights
→ A binary classifier for patterns

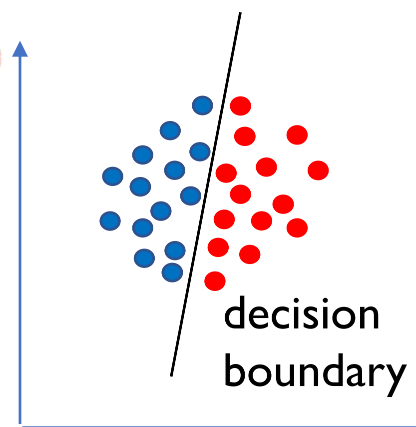
Image: Cornell University

Frank Rosenblatt (1928 –1971)

FIG. 1 — Organization of a biological brain. (Red areas indicate active cells, responding to the letter X.)

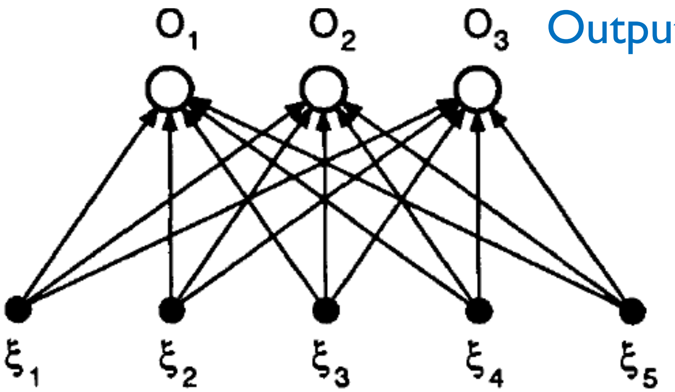


Learning ≡ modifying the connection weights



From: F Rosenblatt, The Design of an Intelligent Automaton (1958)
Image: Cornell University

Single-layer Perceptron



Output $$O_i = g(h_i) = g\left(\sum_k w_{ik}\xi_k\right)$$

g: nonlinear function

Inputs

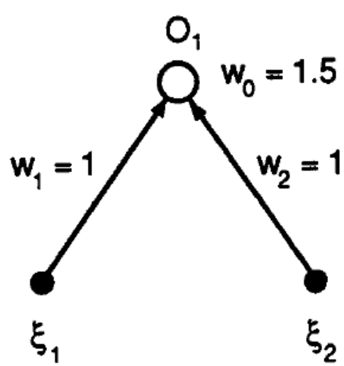
performs hetero-association between input and output that are dissimilar – e.g., as in decision or classification problems

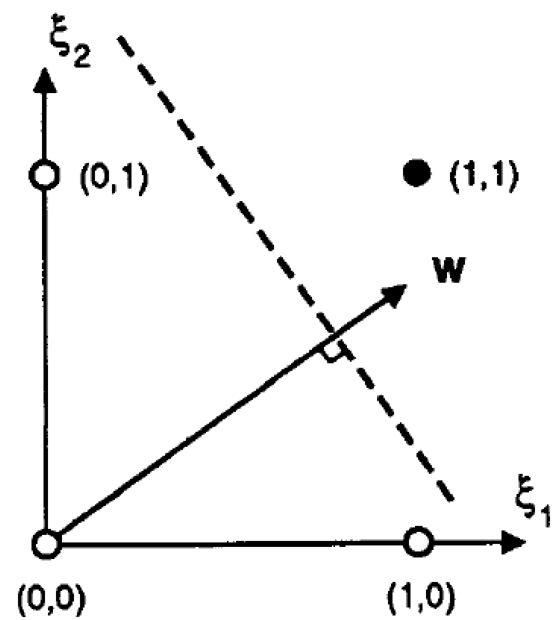
Any classification problem that is linearly separable can be solved by the single-layer perceptron

Example: Learning the AND function

Desired output

ξ_1	ξ_2	ζ
0	0	−1
0	1	−1
1	0	−1
1	1	+1

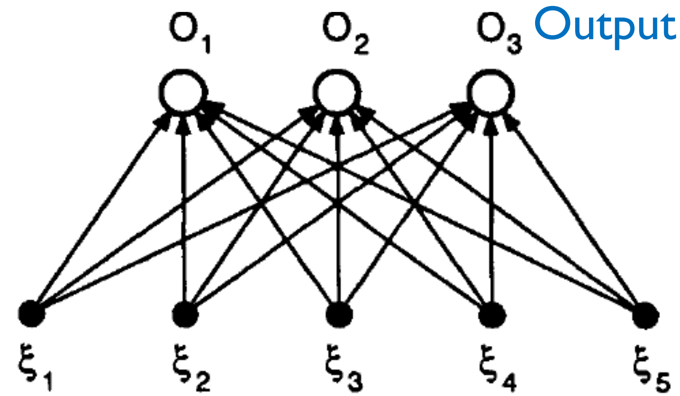




Images: Hertz, Krogh and Palmer, Introduction to the Theory of Neural Computation, Addison-Wesley (1999)

Teaching the single-layer perceptron

Output



Inputs

g: nonlinear function

$$O_i = g(h_i) = g\left(\sum_k w_{ik}\xi_k\right)$$

Scheme:

For each input pattern μ, check if each output unit $O_i^\mu = \zeta_i^\mu$ (desired output)

- If yes, do nothing
- else, correct weight by quantity proportional to product of input & desired output

$\Rightarrow$ $$w_{ik}^{\text{new}} = w_{ik}^{\text{old}} + \Delta w_{ik}$$ where $$\Delta w_{ik} = \eta(1 - \zeta_i^\mu O_i^\mu)\zeta_i^\mu \xi_k^\mu$$

Parameter η : learning rate

requires that the argument *h* of *g*( ) be larger than some margin κ which scales with N

As sum over k scales with N

$\Rightarrow$ $$\zeta_i^\mu h_i^\mu \equiv \zeta_i^\mu \sum_k w_{ik}\xi_k^\mu > N\kappa$$ Desirable!

Implementing this criterion gives

Θ : step function

$\Rightarrow$ Perceptron Learning Rule $$\Delta w_{ik} = \eta\,\Theta(N\kappa - \zeta_i^\mu h_i^\mu)\zeta_i^\mu \xi_k^\mu$$

in single-output vector notation $$\Delta \mathbf{w} = \eta\,\Theta(N\kappa - \mathbf{w}\cdot\mathbf{x}^\mu)\mathbf{x}^\mu$$ where $\mathbf{x}^\mu \equiv \zeta^\mu \xi^\mu$

Images: Hertz, Krogh and Palmer, Introduction to the Theory of Neural Computation, Addison-Wesley (1999)

Perceptron Learning Rule

$$w_i \leftarrow w_i + \Delta w_i$$

where

$$\Delta w_i = \underbrace{\eta}_{\text{learning rate}}(\underbrace{t}_{\text{target value}} - \underbrace{o}_{\text{perceptron output}})\underbrace{x_i}_{\text{input value}}$$

https://dev.to/swyx/supervised-learning-neural-networks-mpo

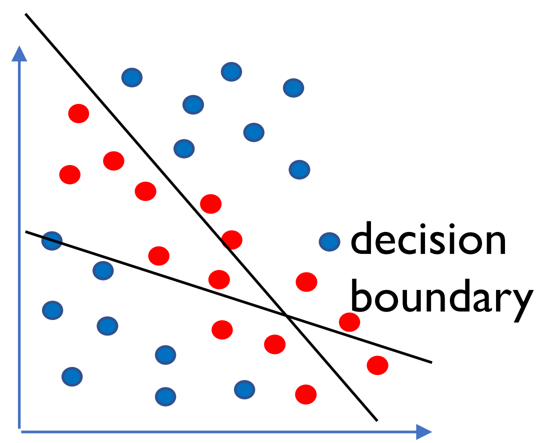
Image: MIT | Cynthia Solomon

Marvin Minsky (1927 –2016) & Seymour Papert (1928-2016)

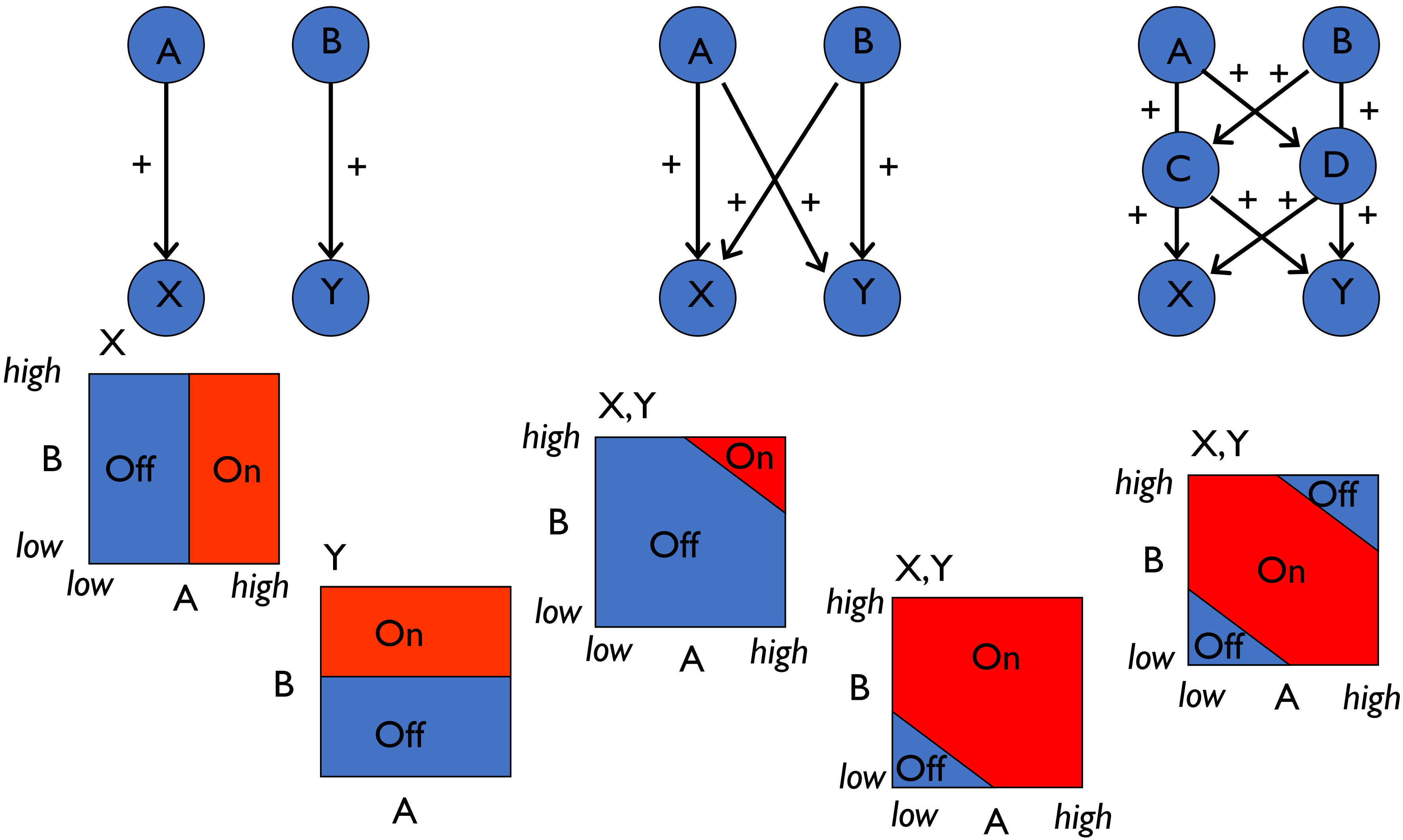
… and the problem of XOR classification

In 1969, Minsky & Papert showed that the perceptron cannot be trained to function as a XOR gate

Only solved once a learning algorithm for multi-layer perceptrons (back-propagation algorithm) was developed in the 1980s



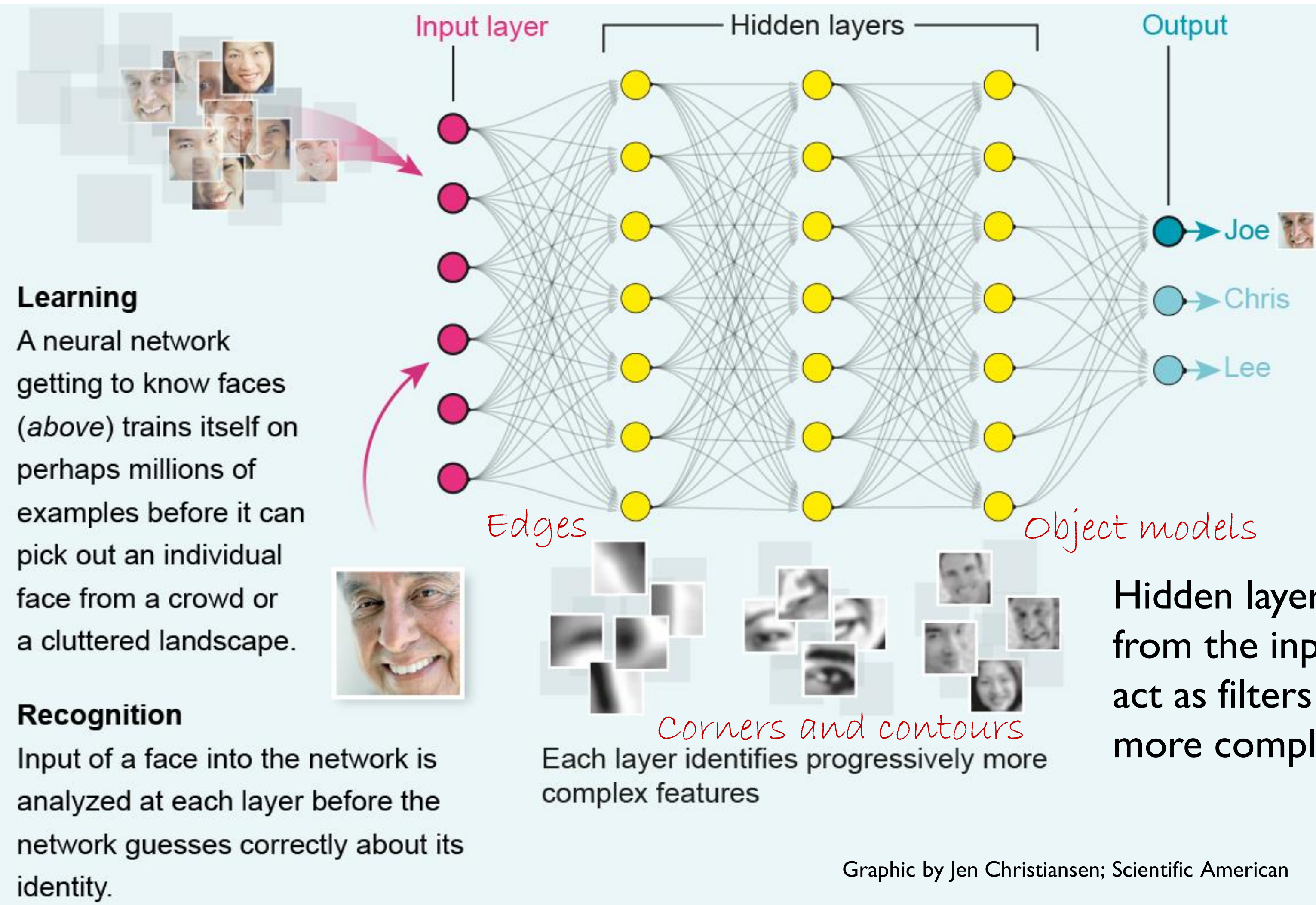
But XOR problem can be solved with an intermediate ("hidden") layer between input and output layers



Problem: How do you train the network ? How do you find the weights for hidden layer ?

What's hidden in the hidden layers?

The activity pattern in the hidden layer(s) represent (encode) significant features of the input space – i.e., they extract features that can be useful for generating correct output



Hidden layers further away from the input layer detect (or act as filters for) progressively more complex features

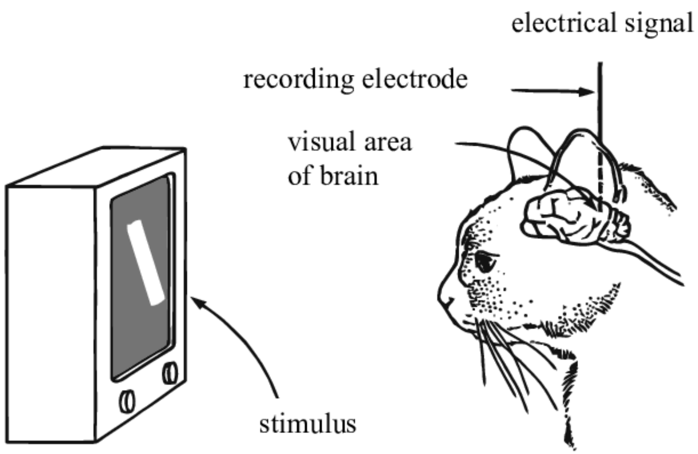
Graphic by Jen Christiansen; Scientific American

Orientation selective cells in Primary Visual Cortex

Image:braintour.harvard.edu

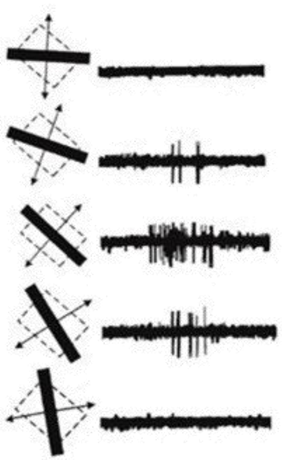
Physiology 1981

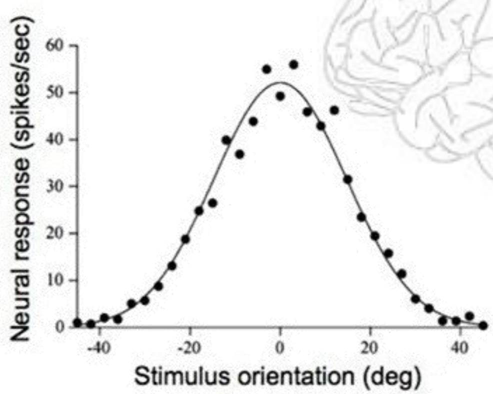
David Hubel and Torsten Wiesel
(1926-2013) (1924-)



Neurons responding to bright stripes against dark background or dark stripes against bright background oriented at specific angles

Image: www.researchgate.net/profile/Anh_Nguyen44/publication/335707980/figure/fig1/





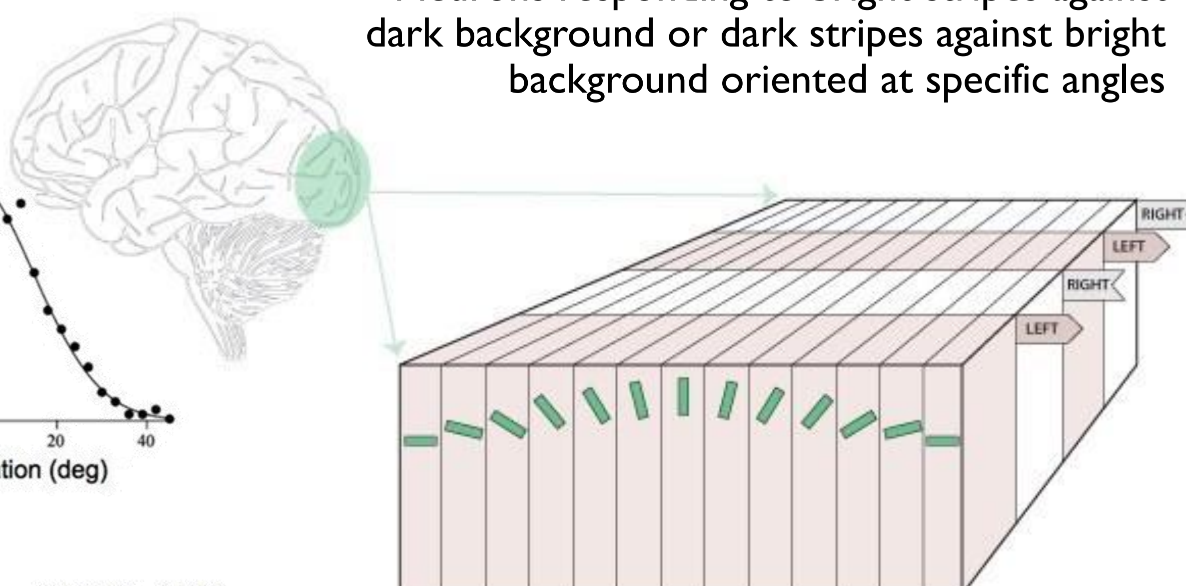
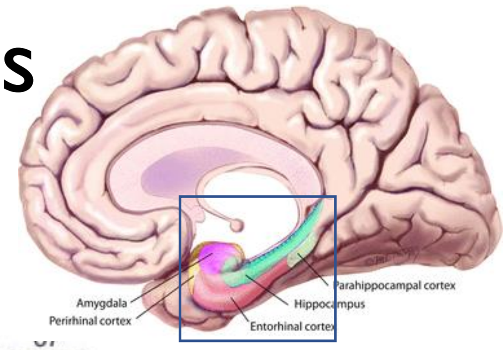


Image: knowingneurons.com

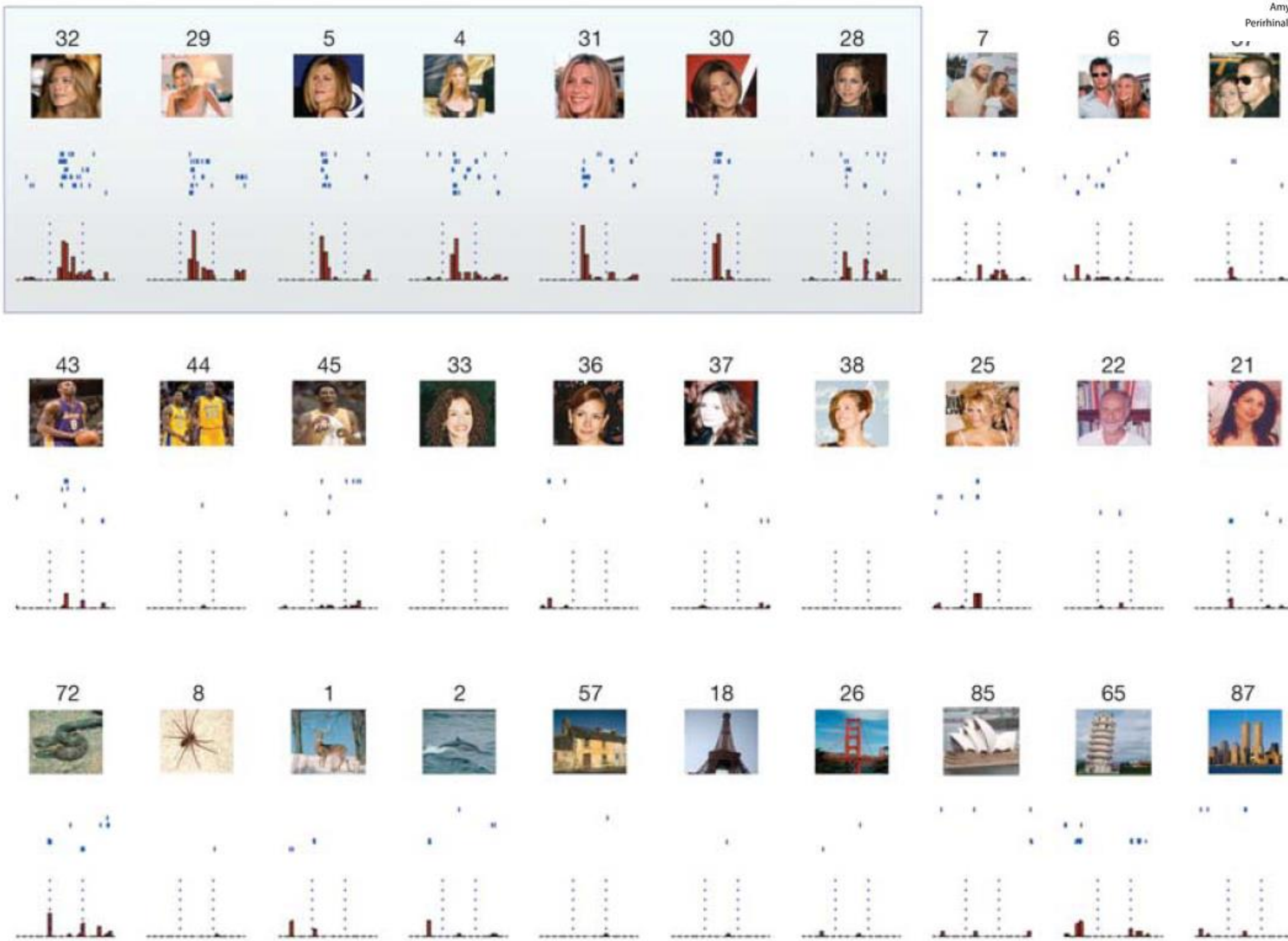
Hubel & Wiesel, 1968

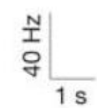
Purves et al, *Principles of Cognitive Neuroscience,* Sinauer (2008)

The "Grandmother cell" hypothesis & the "Jennifer Aniston neuron"



medial temporal lobe





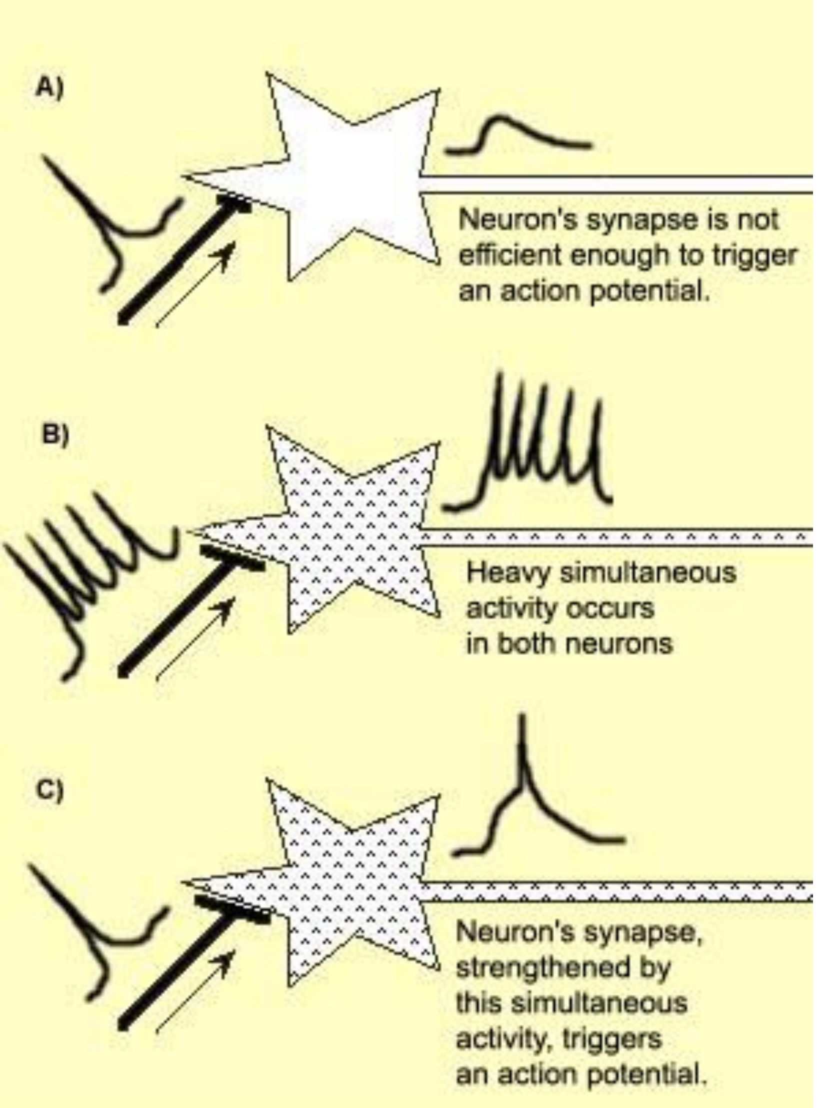
A neuron in left posterior hippocampus is seen to be activated exclusively by images of Jennifer Aniston

For each picture, the corresponding raster plots (the order of trial number is from top to bottom) and post-stimulus time histograms are shown

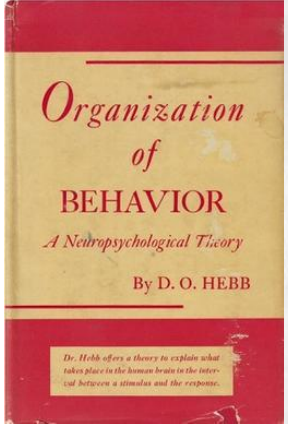
Image: R. Quian Quiroga et al, Nature (2005)

Now, let's backtrack a bit… once again to the 1940s

Hebb's Theory of Learning



1949



Donald Hebb's neuropsychological theory involves the ideas of the Hebbian synapse, cell assembly and the proposal that thinking is the sequential activation of neural assemblies

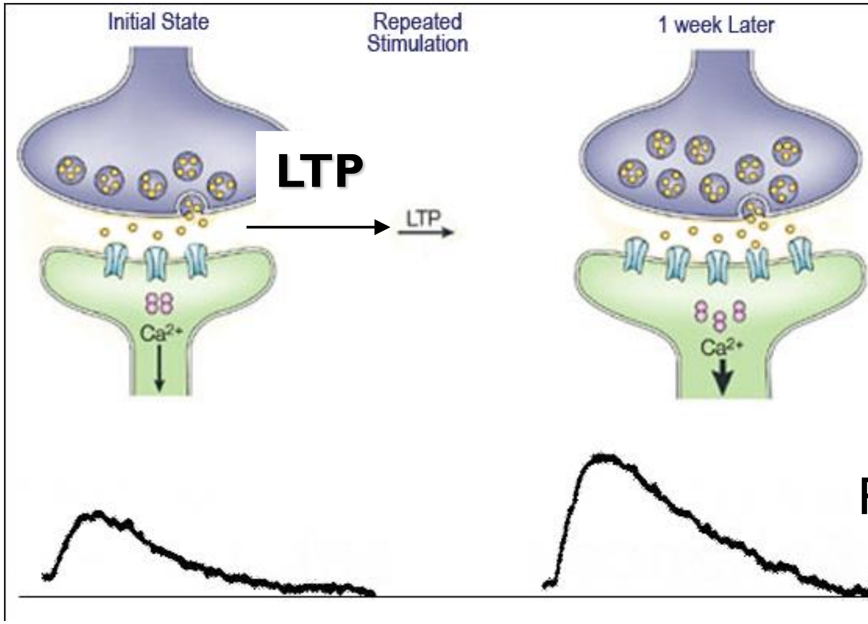
"When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes place in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased."

Neurons that fire together, wire together

http://thebrain.mcgill.ca

Hebb Rule and Biology

Long-term potentiation



https://sites.google.com/site/mcauliffeneur493/home/synaptic-plasticity/Long-Term-Potentiation1.jpg?attredirects=0

First empirical observation (Lomo, 1966) supporting Hebb’s hypothesis

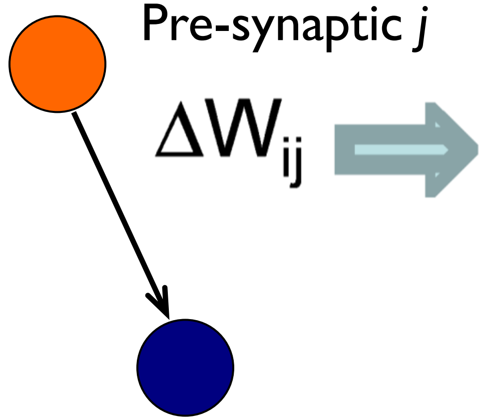
Persistent increase in synaptic strength after brief high-freq stimulation of synapse

Spike-timing dependent plasticity

spike-based formulation of Hebb rule (Markram, 1995)

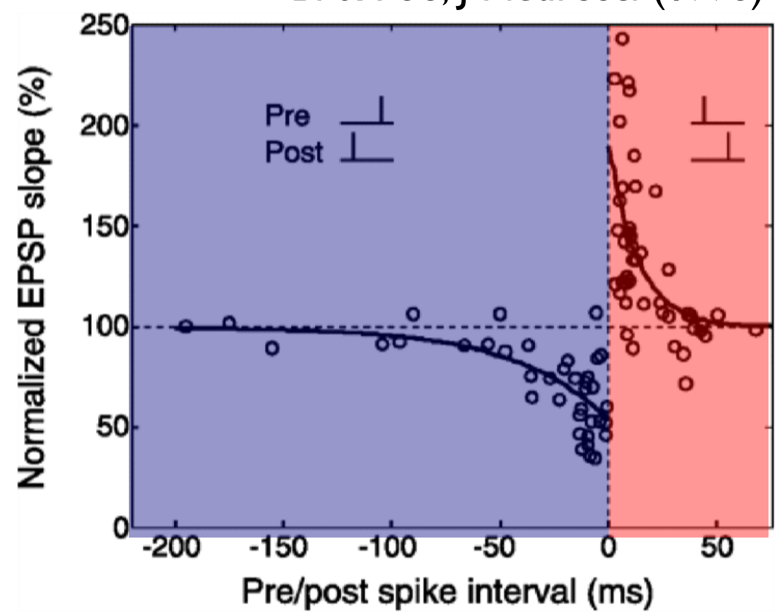
synapse strengthened if presynaptic neuron “repeatedly or persistently takes part in firing” the postsynaptic one (Hebb 1949)

Pre-synaptic *j*



Post-synaptic *i*

Bi & Poo, J Neurosci (1998)



Computers vs Human Memory

Magnetic-core memory: early type of RAM
hysteresis → each core "remembers" its state

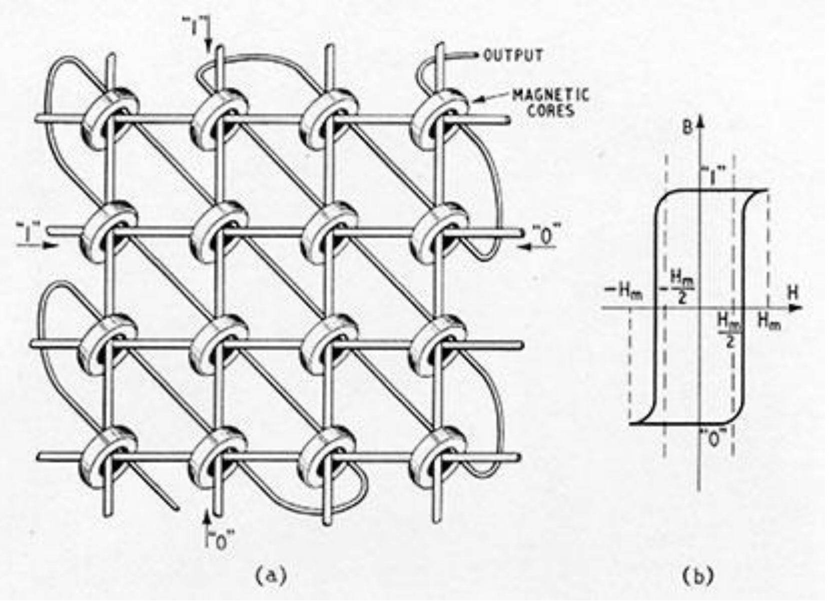


Image: Ivall (Ed), Electronic Computers. Iliffe London 1956

Most commonly used method of storing information in computers is the

random-access memory (RAM)

The process of locating a datum within the storage array involves giving its address. The time needed to retrieve the word remains the same irrespective of the physical location of the word in the array.

The value of the bit stored in a core is 0 or 1 according to direction (clockwise or counter-clockwise) of the core's magnetization

In contrast, humans typically access a memory by partial recall of its content →

Associative or Content Addressable Memory

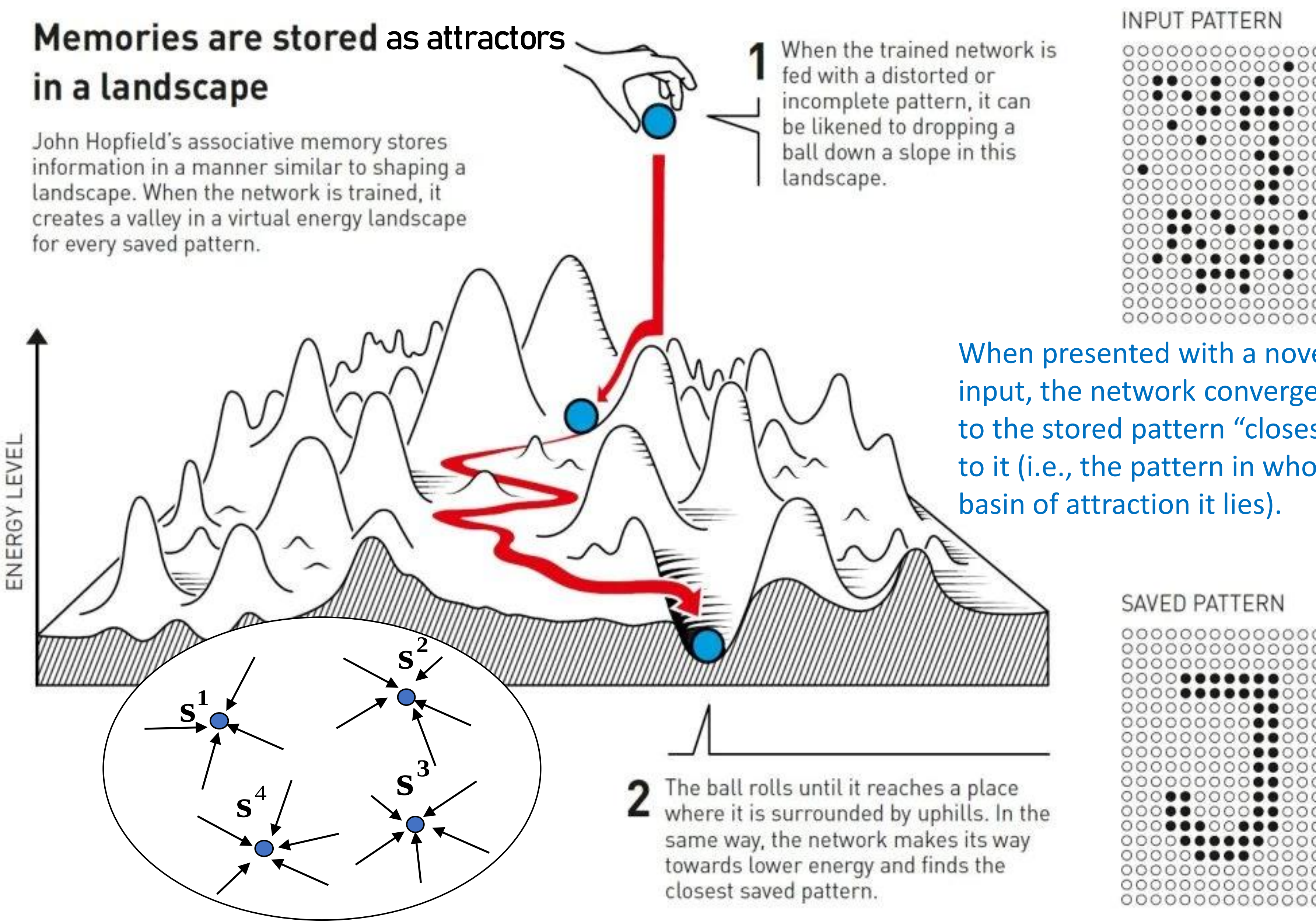
Human memory is associative

…when from a long distant past nothing subsists, after the people are dead, after the things are broken and scattered, still, alone, more fragile, but with more vitality, more unsubstantial, more persistent, more faithful, the smell and taste of things remain poised a long time…

Marcel Proust,
À la recherche du temps perdu (1913-1927)

Photo: Otto Wegener

Associative Memory as Attractor Network



When presented with a novel input, the network converges to the stored pattern "closest" to it (i.e., the pattern in whose basin of attraction it lies).

Image: Johan Jarnestad/The Royal Swedish Academy of Sciences

Hopfield Model

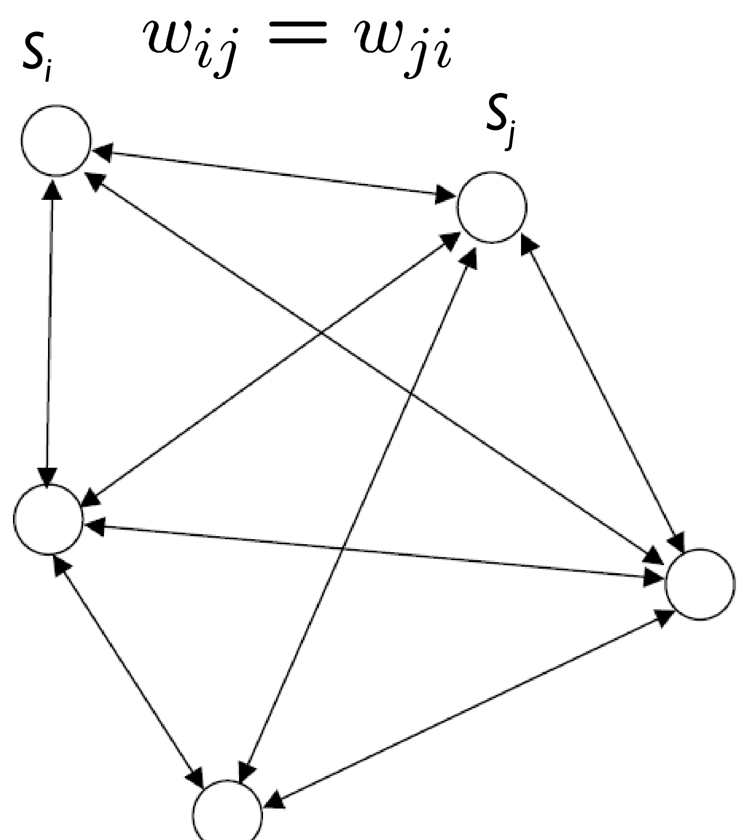
Attractor Network Model for Associative Memory

Image: bulletin.swarthmore.edu

John J Hopfield

Globally connected system of "neurons" (spins)

$$\text{State } S_i = \begin{cases} -1 & \text{Resting} \\ +1 & \text{Firing} \end{cases}$$



T=0 or deterministic dynamics

Time-evolution $S_i = \mathrm{sgn}(\sum_j w_{ij} S_j)$

- $\mathrm{sgn}(q) = -1$, if $q < 0$;
- $\mathrm{sgn}(q) = +1$ otherwise

- ❑ Symmetric connection weights $w_{ij} = w_{ji}$

  "A brilliant step backwards" (Amit)

- ❑ $w_{ii}=0$ (No self connections)

Learning in Hopfield Network

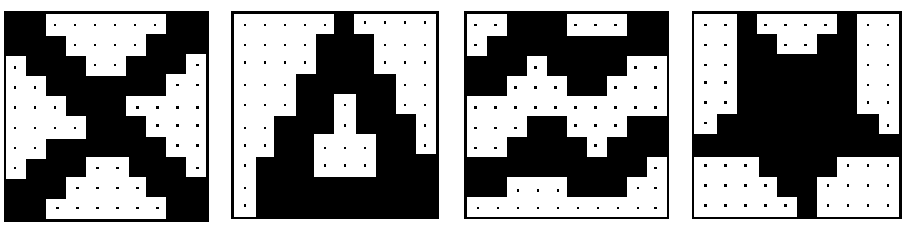
Implementing Hebb rule in synaptic weight determination

"One-shot" learning

$$w_{ij} = \frac{1}{N}\sum_{p=1}^{M} \xi_i^p \xi_j^p$$

ξ_i^p : state of i^{th} neuron in the p^{th} pattern

Four stored patterns in simulation



"■" i^{th} neuron is ON $\xi_i = +1$

"·" i^{th} neuron is OFF $\xi_i = -1$

Convergence to stored pattern

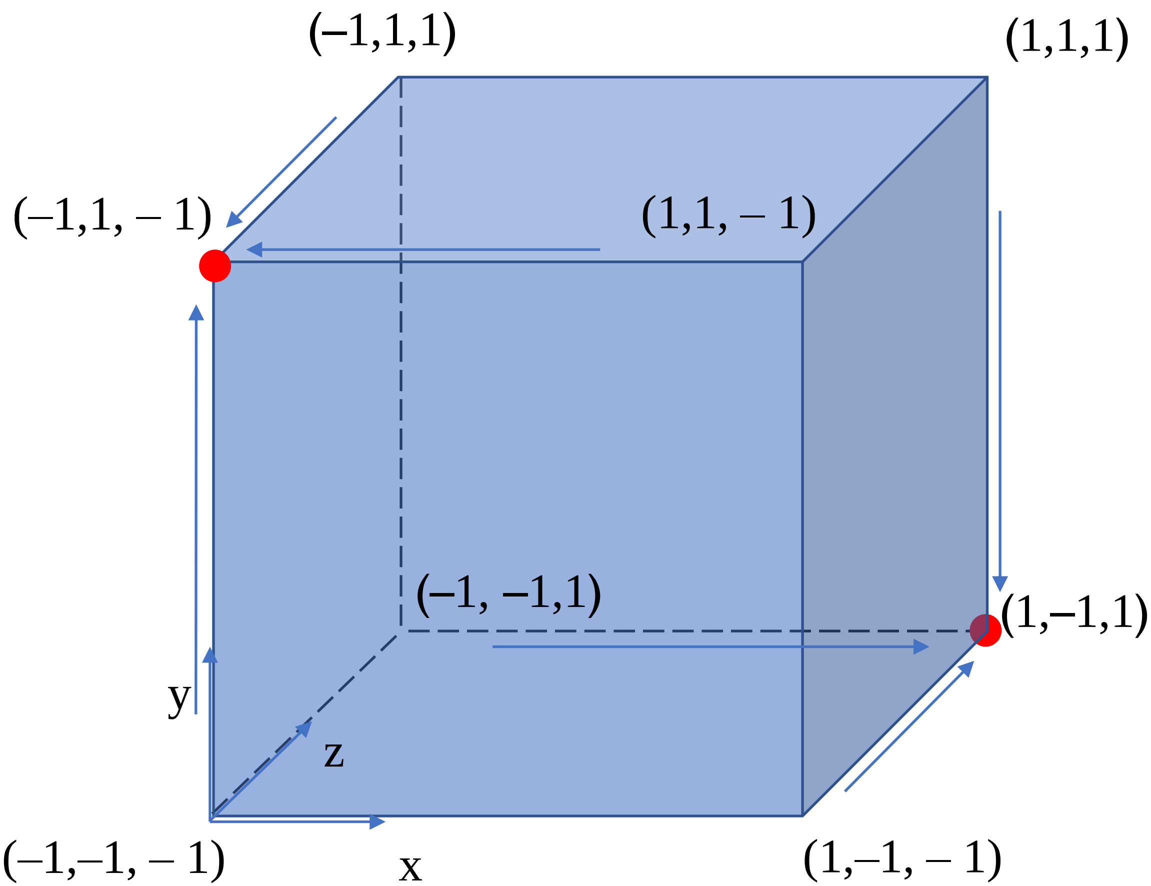
Example: Hopfield Model with N=3, p=2

The strings $(1,-1,1)$ and $(-1,1,-1)$ are the stored patterns have to be made attractors of the network dynamics

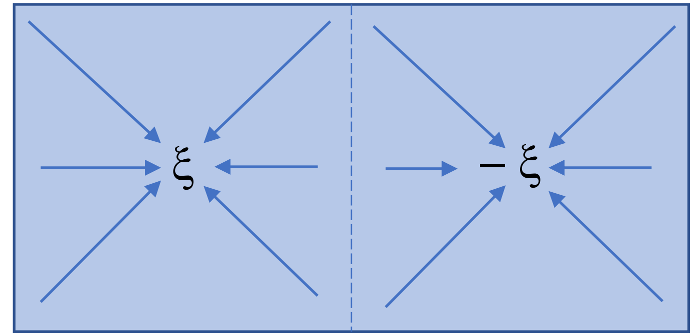
$$w_{ij} = \frac{1}{N}\sum_{p=1}^{M} \xi_i^p \xi_j^p$$

$$w_{ii} = 0$$

$$W = \frac{1}{3}\begin{bmatrix} 0 & -2 & 2 \\ -2 & 0 & -2 \\ 2 & -2 & 0 \end{bmatrix}$$



One pattern (p=1)



ξ_i : pattern memorized

For the pattern to be stable, $\mathrm{sgn}(\sum_j w_{ij}\, \xi_j) = \xi_i$ for all i

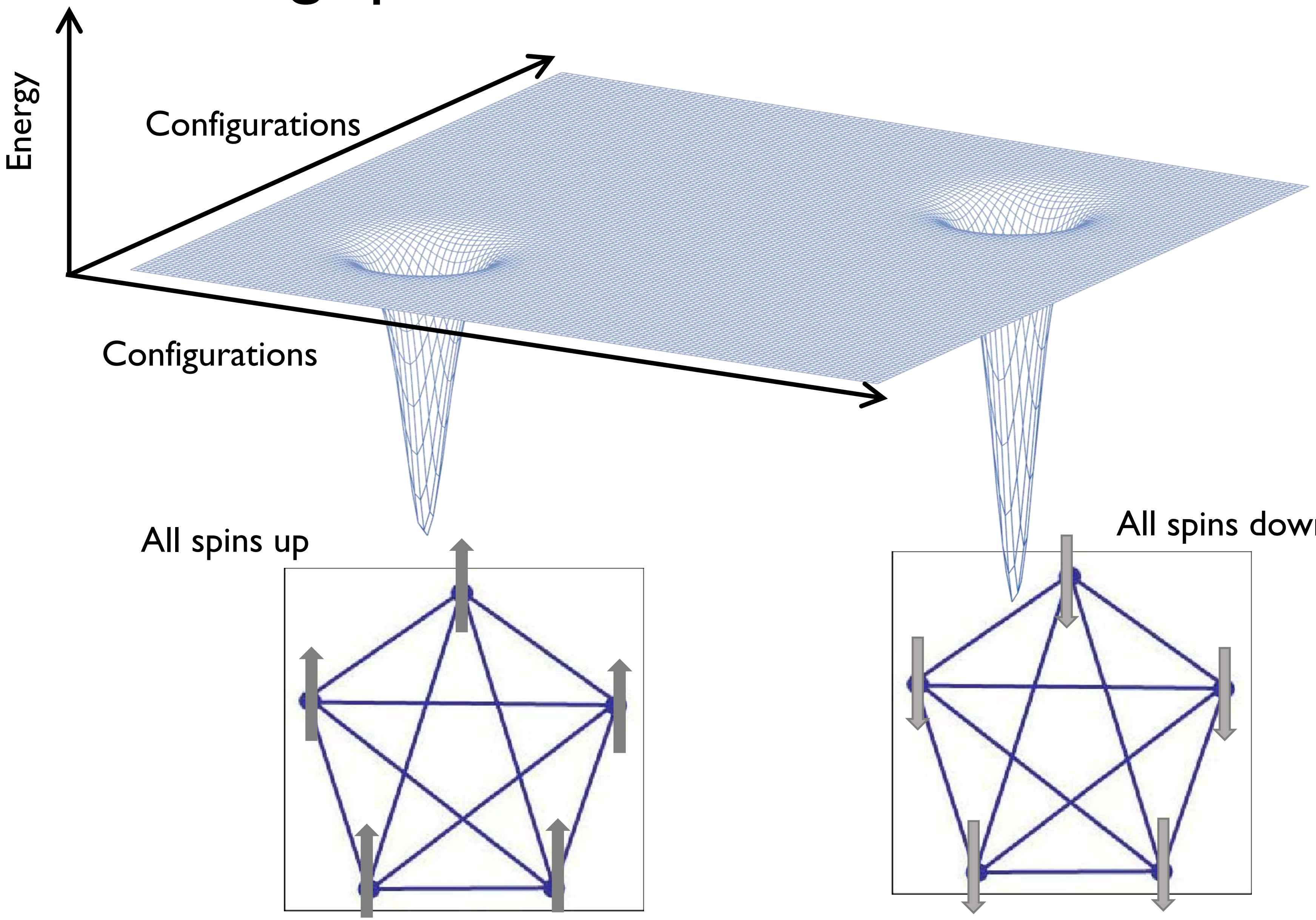
This is true if $w_{ij} \propto \xi_i\, \xi_j$ as $\xi_i^2 = 1$
(the proportionality constant being 1/N)

If M out of N components of the initial state S_i are wrong (opposite to ξ_i)
the input $h_i \equiv \mathrm{sgn}(\sum_j w_{ij}\, S_j) = \mathrm{sgn}(\underbrace{\sum_k w_{ik}\, \xi_k}_{\text{Same sign as } \xi} - \underbrace{\sum_m w_{im}\, \xi_m}_{\text{Opposite sign to } \xi})$

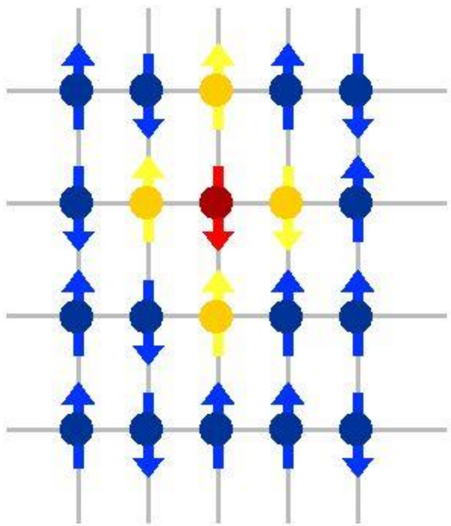
will converge to output same as the stored pattern ξ if $M < N/2$
$\Rightarrow$ Network will correct errors in the initial pattern and converge to ξ, the attractor of the recall dynamics

Identical (under a gauge transformation) to the mean-field

Lenz-Ising spin model



Spin models as a paradigm for Complex Systems



•Spin orientation: mutually exclusive choices

•Choice dynamics: decision based on information about choice of majority in local neighborhood

Simplest case: 2 possible choices

Ising model with Ferromagnetic interactions: each agent can be in one of 2 states (Yes/No , +/-)

Spin models as a paradigm for Complex Systems

The McCulloch-Pitts neuron

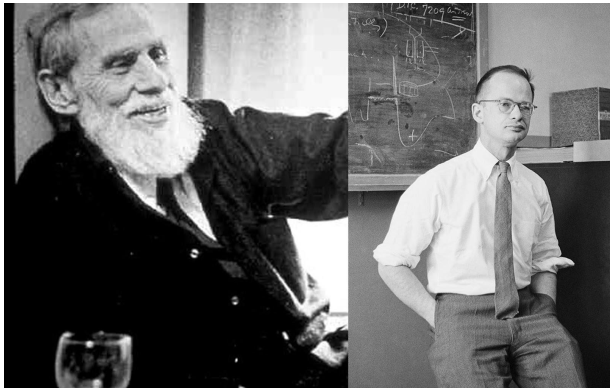
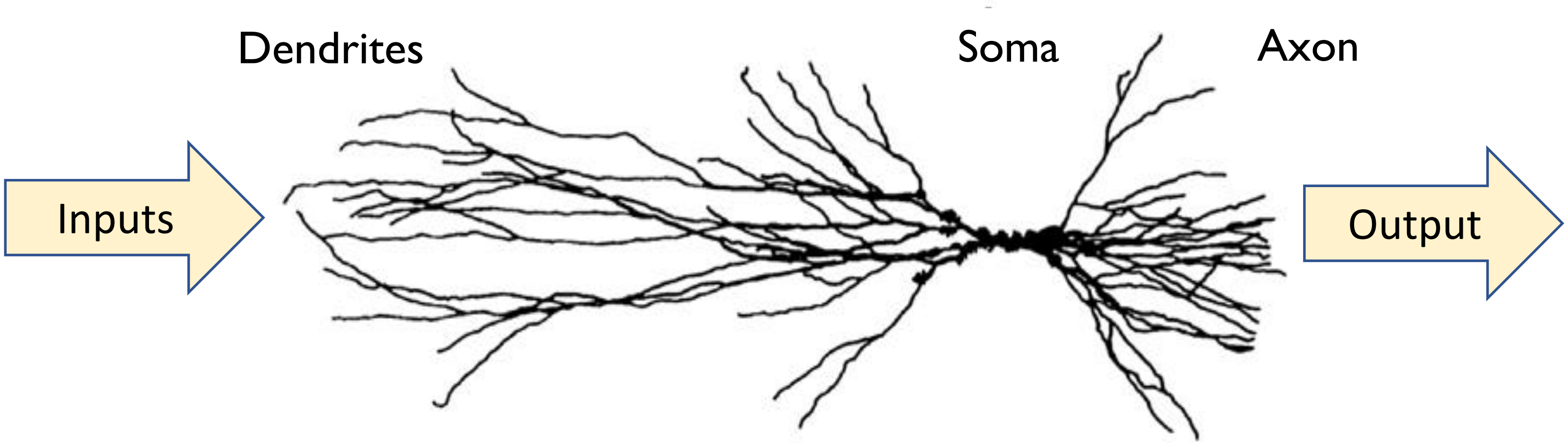


Image:infonintelli.blogspot.com

Image: Current Biology



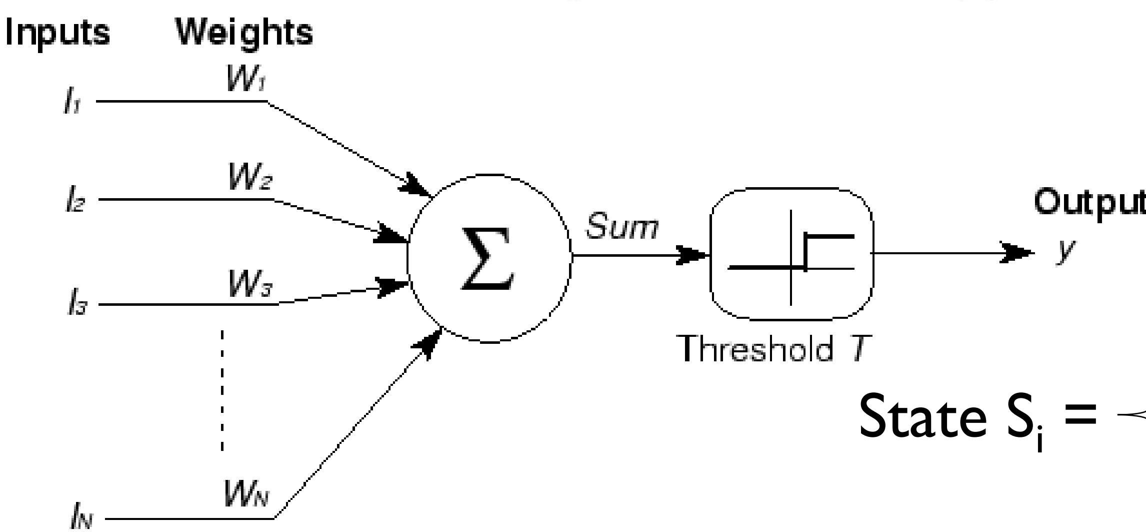
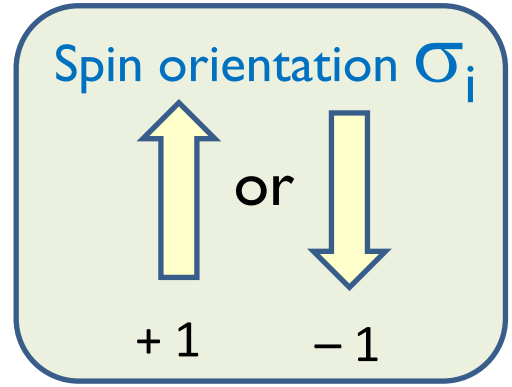


Image: chatbotslife.com/keras-in-a-single-mcculloch-pitts-neuron-317397cccd45

$$\text{State } S_i = \begin{cases} -1 & \text{Resting} \\ +1 & \text{Firing} \end{cases}$$

The Ising model is described by



$$H = -\underbrace{\sum_{ij} \overset{\text{spin-orientation coupling}}{\downarrow \atop J_{ij}}\;\sigma_i\;\sigma_j}_{\text{interaction}} - \underbrace{\sum_i \overset{\text{external field}}{\downarrow \atop h}\;\sigma_i}_{\text{environment}}$$

For spontaneous ordering in a ferromagnet, $J_{ij} = J > 0$ and $h = 0$

Once we introduce thermal fluctuations (at finite temperature $T>0$) system behavior is governed by

$$\text{Free energy } F = U - T\,.\,S$$

Ising model and Maximum Entropy Distribution

For a large system of coupled binary elements (equivalent to spins), impossible to measure the probability distribution of all the network states, $P(\sigma)$

However, by individually recording the values of every element (spin) σ_i , one can measure the **mean order parameter** (e.g., system activity), the expectation value $\langle \sigma_i \rangle$

One can also simultaneously record from a pair of elements (spins), to obtain the **correlations** $C_{ij} = \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle$

Question: what do these measurements say about the system distribution $P(\sigma)$?

Problem! In general, there are **infinitely many distributions** (over the 2^N states of N elements) consistent with the $N(N + 1)/2$ measurements.

However, of all these possible distributions, one reproduces the measurements but otherwise describes a system that is as random as possible (i.e., having the fewest additional assumptions) $\rightarrow$ the maximum entropy distribution

Knowing the expectation values of some functions on the state, $\langle f_\mu(\sigma)\rangle = \overline{f}_\mu(\sigma)$, how to obtain the probability distribution $P(\sigma)$?

Maximize the entropy of the distribution subject to the constraints imposed by the observations using Lagrange multipliers λ_μ

$\Rightarrow$ Maximize $F = -\sum_\sigma P(\sigma)\ln P(\sigma)$ Entropy

$- \sum_\mu \lambda_\mu \left[\sum_\sigma P(\sigma)\, f_\mu(\sigma) - \overline{f}_\mu(\sigma)\right]$ Observations

$- \Lambda \left[\sum_\sigma P(\sigma) - 1\right]$ Normalization constraint

$\Rightarrow$ The optimum is given by $\dfrac{\delta F}{\delta P(\sigma)} = 0 = -\left[\ln P(\sigma) + 1\right] - \sum_\mu \lambda_\mu f_\mu(\sigma) - \Lambda$

$$\Rightarrow \ln P(\sigma) = -\sum_\mu \lambda_\mu f_\mu(\sigma) - (\Lambda+1)$$

$$\Rightarrow P(\sigma) = \frac{1}{Z}\exp\left[-\sum_\mu \lambda_\mu f_\mu(\sigma)\right] \quad \text{where } Z = \exp\left[-(\Lambda+1)\right]$$

λ_μ determined by matching the expectation values in the distribution to observed ones

Derivatives of ln Z (the free energy) give the various averages to be matched to observation

$$\langle f_\nu(\sigma)\rangle = -\frac{\partial \ln Z(\{\langle\lambda_\mu\})}{\partial \lambda_\nu}$$

$\Rightarrow$ need to solve the equations $$-\frac{\partial \ln Z(\{\langle\lambda_\mu\})}{\partial \lambda_\nu} = \bar{f}_\nu(\sigma)$$

In general hard to solve.

Special case: If the expectation values that are measured are $\langle\sigma_i\rangle$ and $\langle\sigma_l\, \sigma_j\rangle$, the maximum entropy distribution yields the…. Ising model

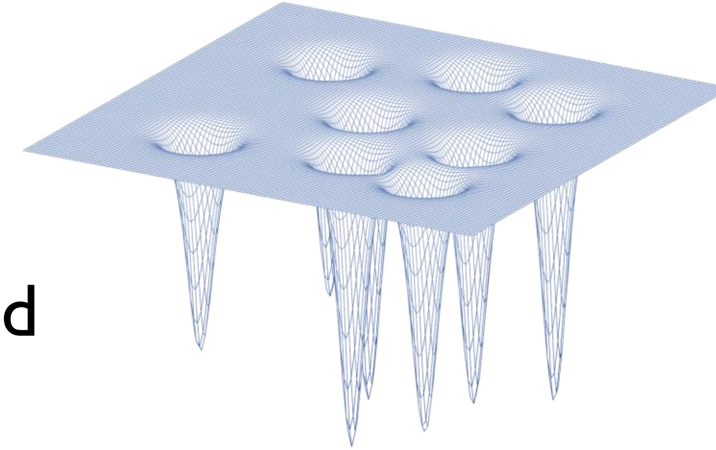
$$P(\sigma) = \frac{1}{Z}\exp\Big[\sum_{i=1,\dots,N} \underset{\substack{\uparrow\\ \text{magnetic fields}}}{h_i}\, \sigma_i + \frac{1}{2}\sum_{i\neq j=1,\dots,N} \underset{\substack{\uparrow\\ \text{exchange couplings}}}{J_{ij}}\, \sigma_i\, \sigma_j\Big]$$

The fields $\{h_i\}$ and interactions $\{J_{ij}\}$ are chosen so as to reproduce the measured values of the order parameter $\{\langle\sigma_i\rangle\}$ and correlations $\{C_{ij}\} \sim \{\langle\sigma_i\, \sigma_j\rangle\}$

$\rightarrow$ Ising model with pairwise interactions among spins emerges not as a hypothesis but as the least-structured model that is consistent with the measured expectation values

$\Rightarrow$ The mapping to the Ising model is a mathematical equivalence, **not** an analogy, with the details of the model specified by empirical data

Many pattern (p>1)



ξ_i^μ (μ=1, …, p): patterns memorized

A natural extension: make w_{ij} a superposition of terms – one for each pattern

$\Rightarrow w_{ij} \propto \sum_{\mu=1,p} \xi_i^\mu \xi_j^\mu$

For a particular pattern ξ_i^ν to be stable, $\mathrm{sgn}(\sum_j w_{ij}\, \xi_j^\nu) = \xi_i^\nu$ for all i

the input $h_i \equiv \sum_j w_{ij}\, \xi_j^\nu = \sum_j \sum_\mu \xi_i^\mu \xi_j^\mu \xi_j^\nu = \underbrace{\xi_i^\nu}_{\text{Desired pattern}} + \underbrace{\sum_j \sum_{\mu \neq \nu} \xi_i^\mu \xi_j^\mu \xi_j^\nu}_{\text{Crosstalk term}}$

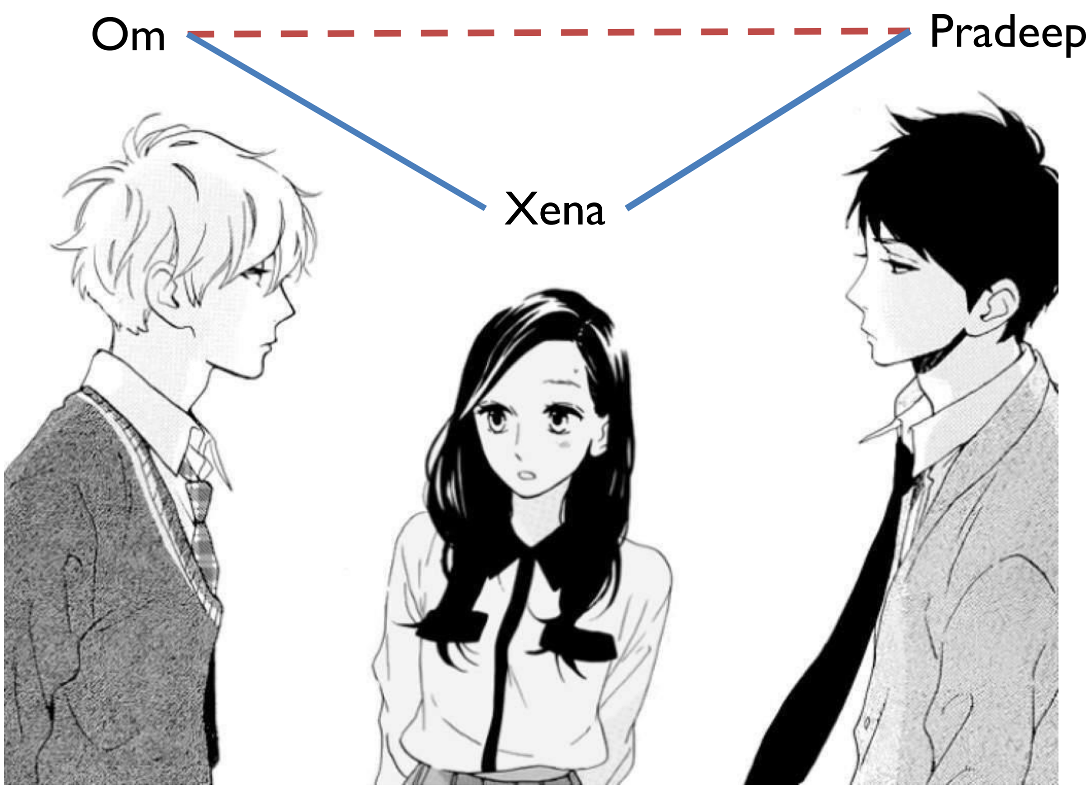
will converge to output same as the stored pattern if the magnitude of the crosstalk term < 1 (true for small p)

$\Rightarrow$ Network will correct errors in any initial pattern sufficiently close to any of the stored patterns ξ^μ (multiple attractors)

As number of patterns to be stored increases, the resulting rise in crosstalk leads to

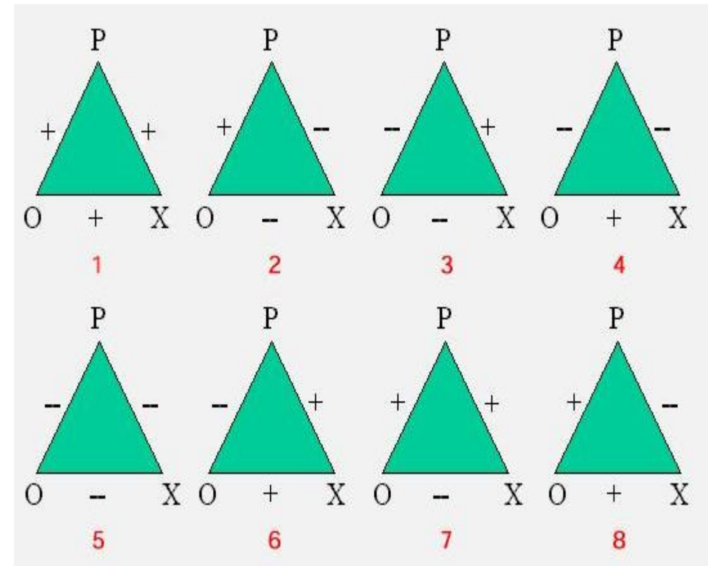
Frustration

A basic characterization of relationships between mutual acquaintances proposed by Fritz Heider



aminoapps.com

Balance



No Balance

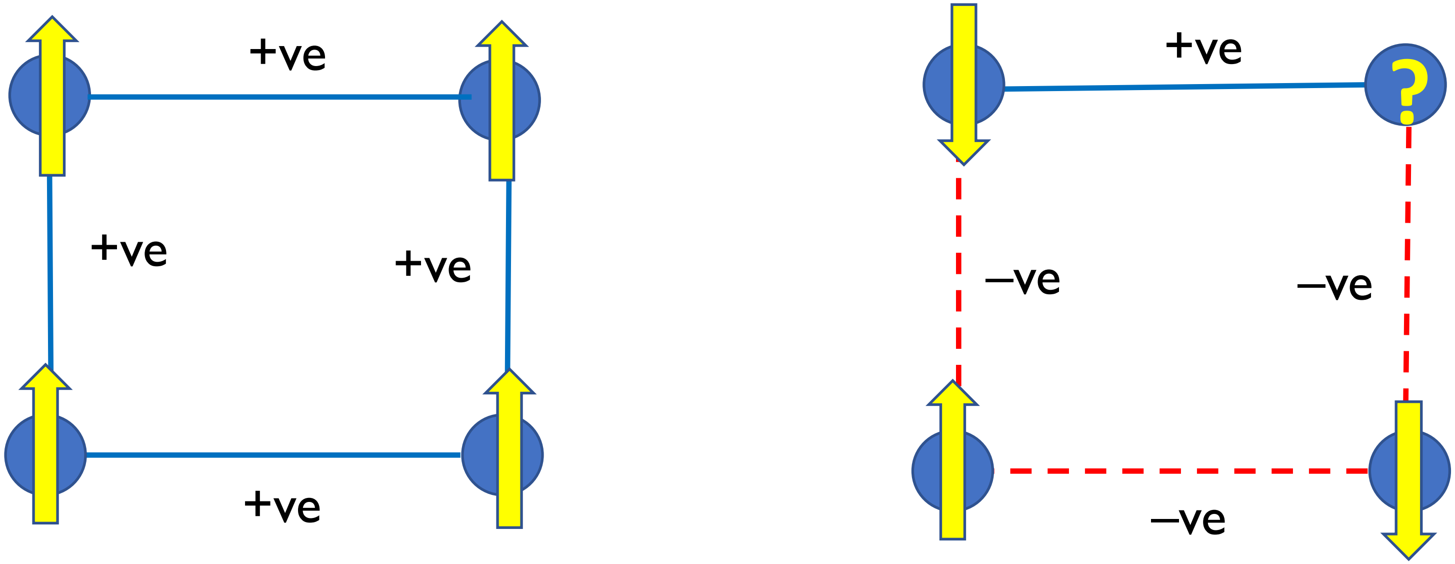
(1) Om and Xena are friends ⇒ OX: +ve interaction (link)
(2) Pradeep and Xena are friends ⇒ PX: +ve interaction (link)
(3) Om and Pradeep are enemies ⇒ OP: –ve interaction (link)

Tension

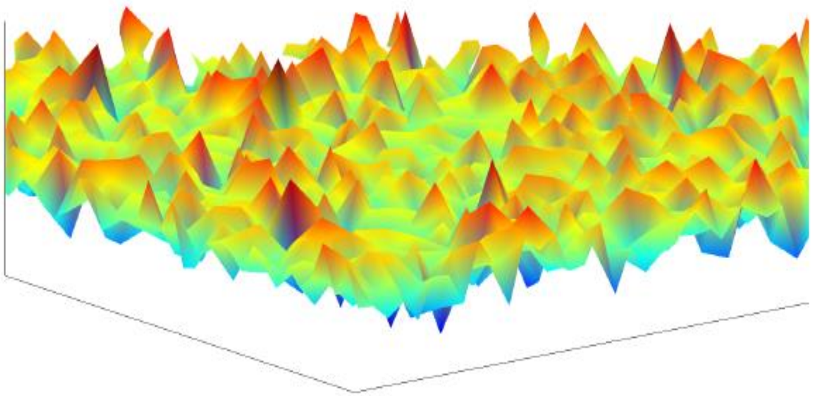
Frustration

Conflicting Constraints in Disordered Systems

Spins in binary states $(+1/-1)$ having +/ – interactions at random



Frustration results in a rugged energy landscape, with the system trapped in any one of a large number of local minima (spin glass states)

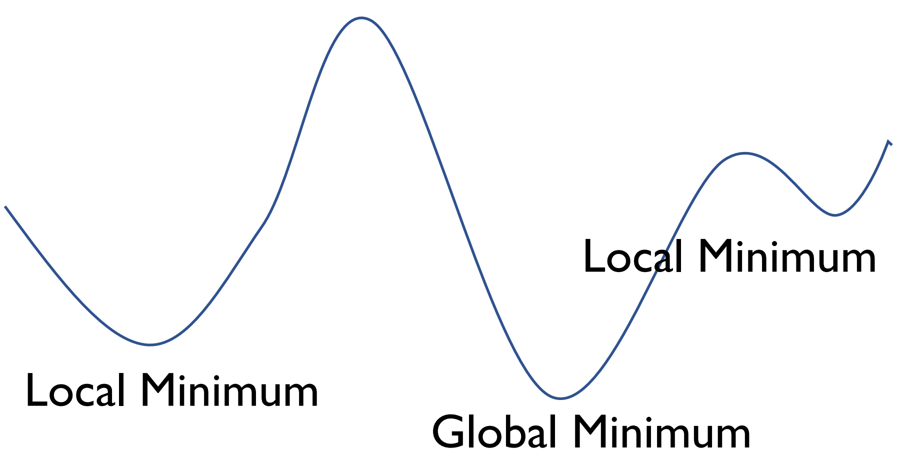


Memory Recall in Hopfield Network

- ❑ Start from arbitrary initial configuration of {x}
- ❑ What final state does the network converge ?
- ❑ Evaluate an 'energy' value associated with the network state:

$$E = -\frac{1}{2}\sum_{j}\sum_{\substack{i=1 \\ i \neq j}}^{N} w_{j,i} x_i \, x_j$$

- ❑ System converges to an attractor a local/global minimum of *E*



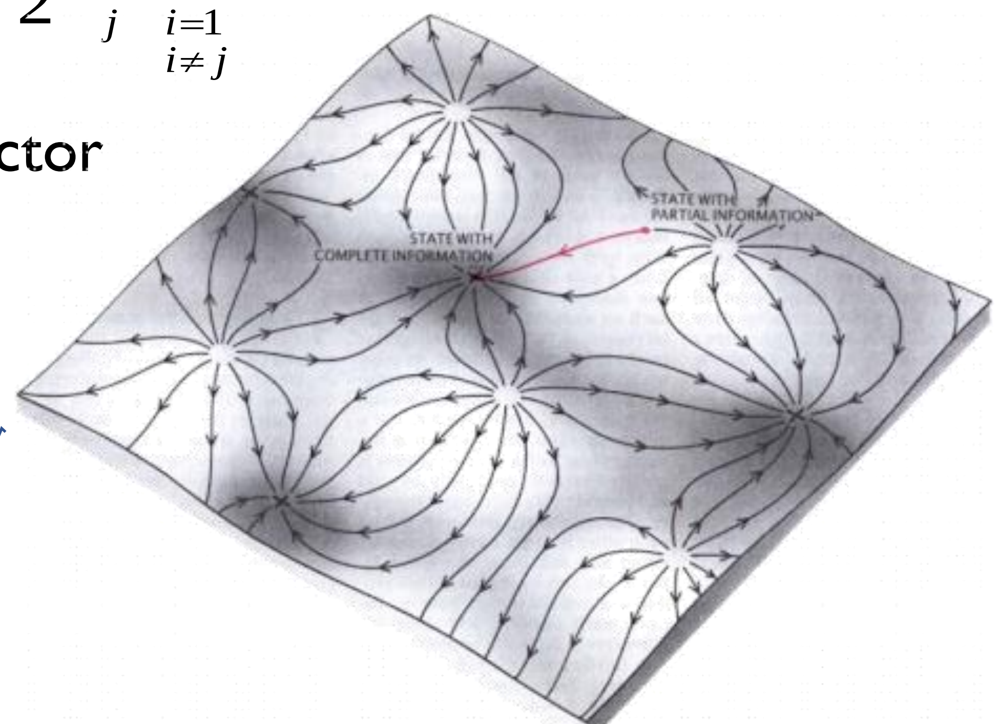


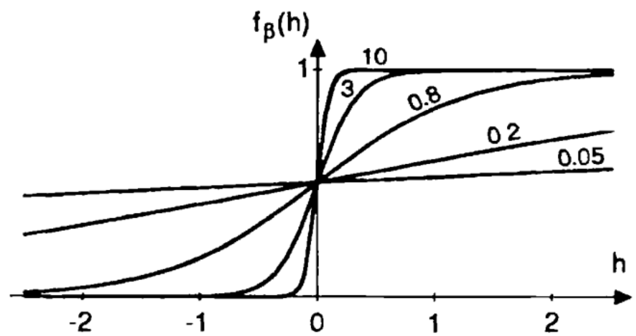
Image: Tank & Hopfield, Scientific American (1987)

T>0 or Stochastic dynamics

In neurons, fluctuations in the release of neurotransmitters in discrete vesicles $\Rightarrow$ neurons may fire even when weighted input < threshold or not fire when input > threshold

Noise $\rightarrow$ Stochasticity in neuronal firing

Amount of noise quantified by "pseudo temperature" T
T=0 $\rightarrow$ deterministic dynamics



For T>0

$$\text{Prob}\,(S_i = +1) = f_T\left(\sum_j w_{ij} S_j \right)$$

$$= 1/[1 - \exp(2 \textstyle\sum_j w_{ij} S_j / T)]$$

Storage capacity of "noisy" Hopfield Model

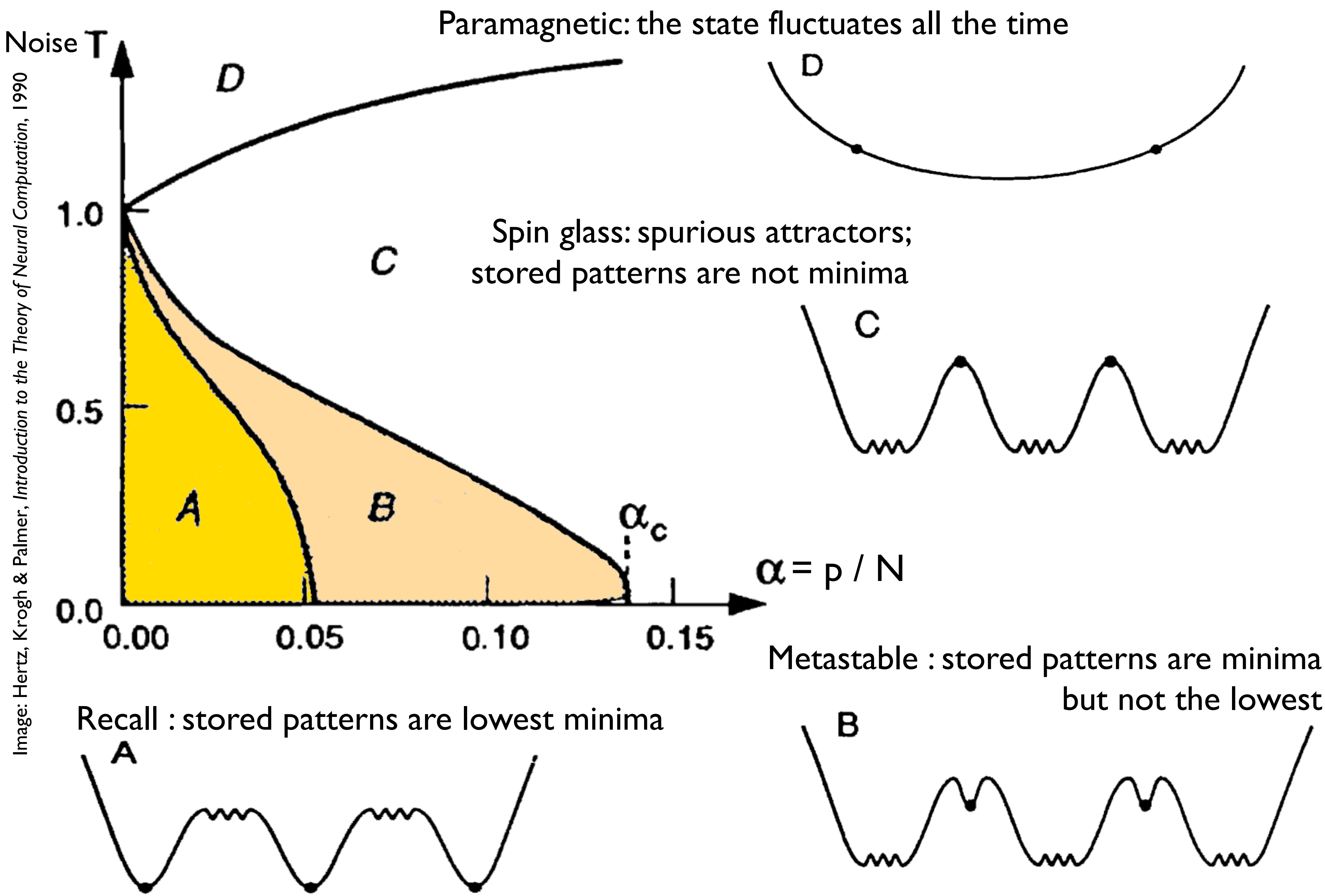


Image: Hertz, Krogh & Palmer, *Introduction to the Theory of Neural Computation*, 1990

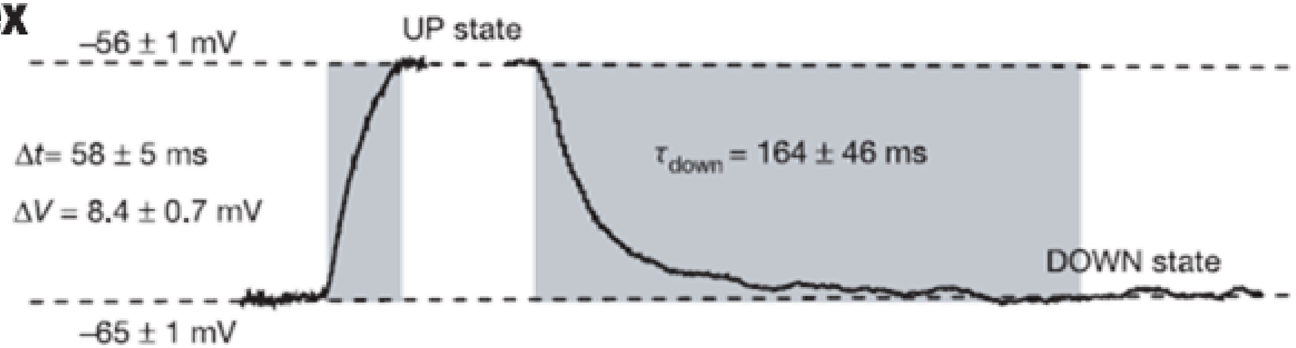
Attractor networks in the neocortex

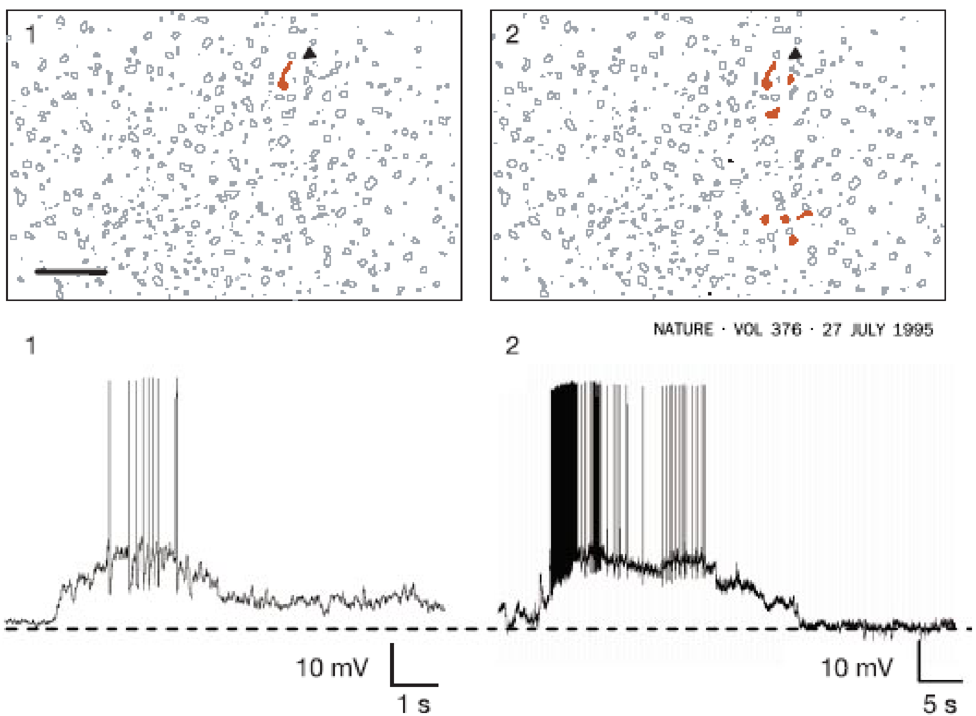
Attractor dynamics of network UP states in the neocortex

Rosa Cossart, Dmitriy Aronov & Rafael Yuste

NATURE | VOL. 423 | 15 MAY 2003 |

Using two-photon calcium imaging reconstructed dynamics of spontaneous activity of ~1400 neurons in mouse visual cortex slices





"the membrane potential of cortical neurons fluctuates spontaneously between a resting (DOWN) and a depolarized (UP) state which may also be coordinated. The elevated firing rate in the UP state follows sensory stimulation and provides a substrate for persistent activity … that might mediate working memory."

"network UP states are circuit attractors …that could implement memory states or solutions to computational problems."

Hopfield network assumes that every neuron is connected to all other neurons (Clique) – but in reality neuronal network connectivity is sparse

How would the performance of attractor networks alter for sparse connections? In particular,

Does modular structure provide advantage in dynamics of an attractor network ?

The Global Stability of Attractor Networks is **maximum** at an optimal modularity

Neeraj Pradhan

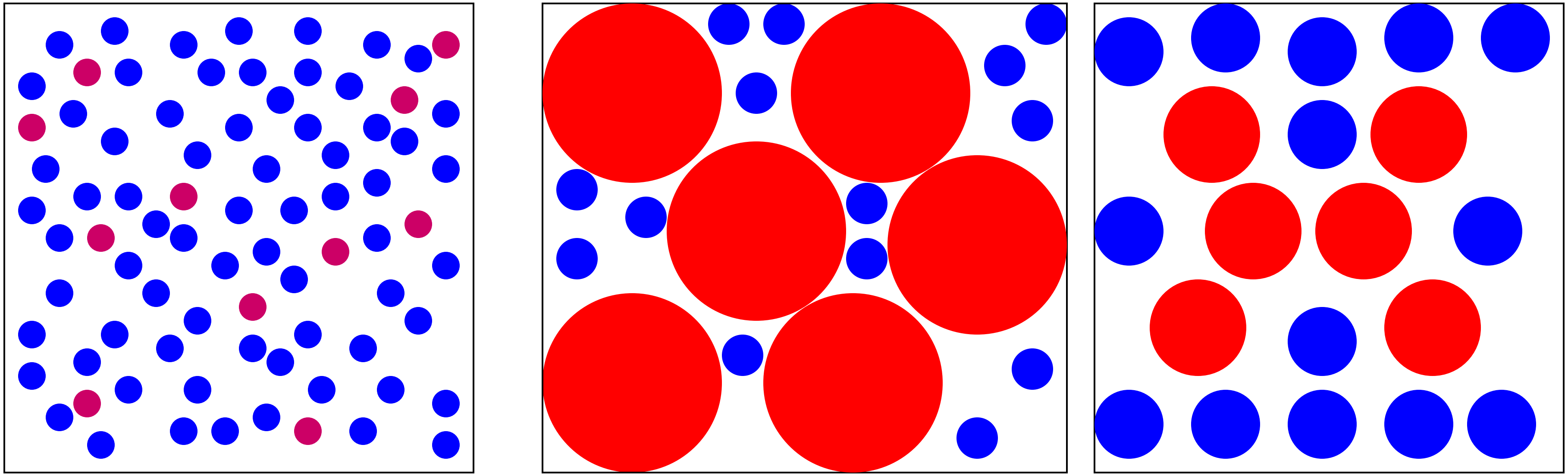
The basins of attraction of the stored patterns – a measure of global stability in attractor networks – cover largest fraction of phase space when the network has an

optimal modular structure ($r \approx r_c$)

for storing multiple patterns in a network with *N* nodes and *L* links,

The attractor landscape of the network changes with modularity

Isolated Modules	Modular	Homogeneous
$r << r_c$	$r \approx r_c$	$r = 1$
High-basin entropy	Low-basin entropy	High-basin entropy



But wait….

The Brain has a hierarchical arrangement!

Neural networks used in deep learning are inspired by the layered network organization of the brain

Input layer ↔ sensory neurons
Hidden layers ↔ interneurons
Output layer ↔ motor neurons

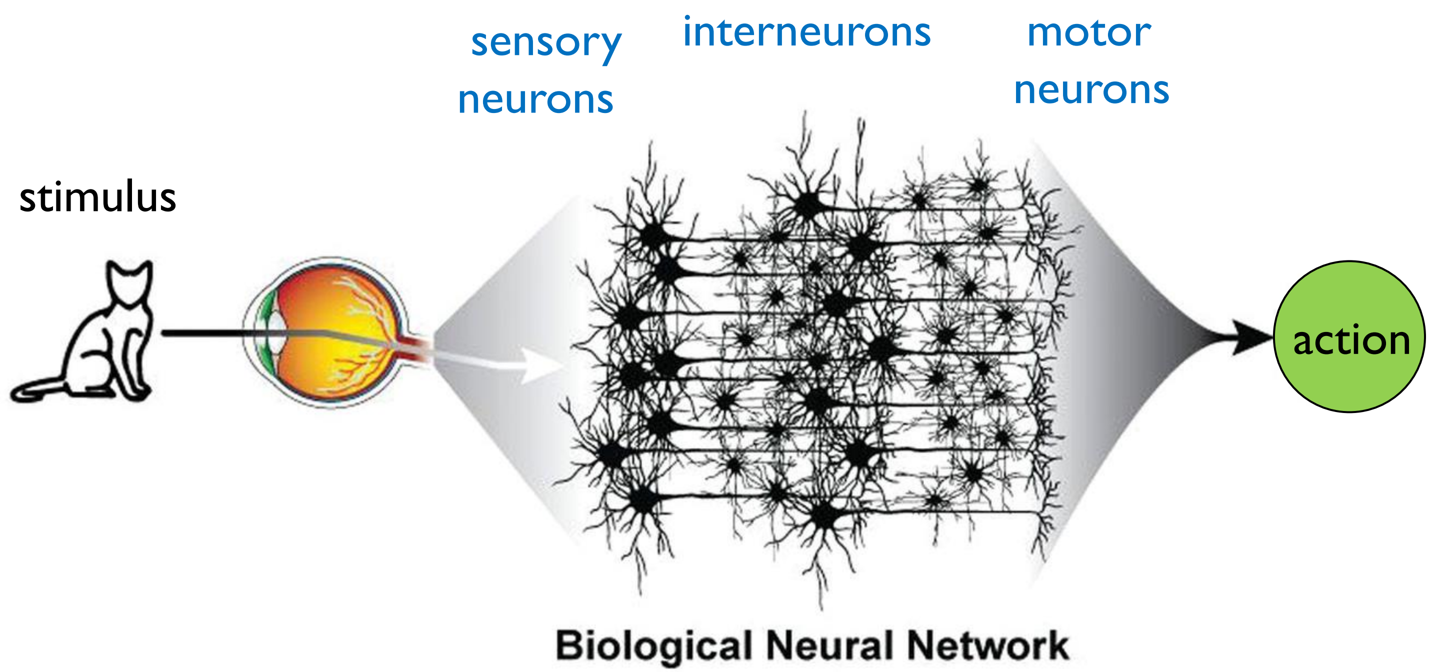
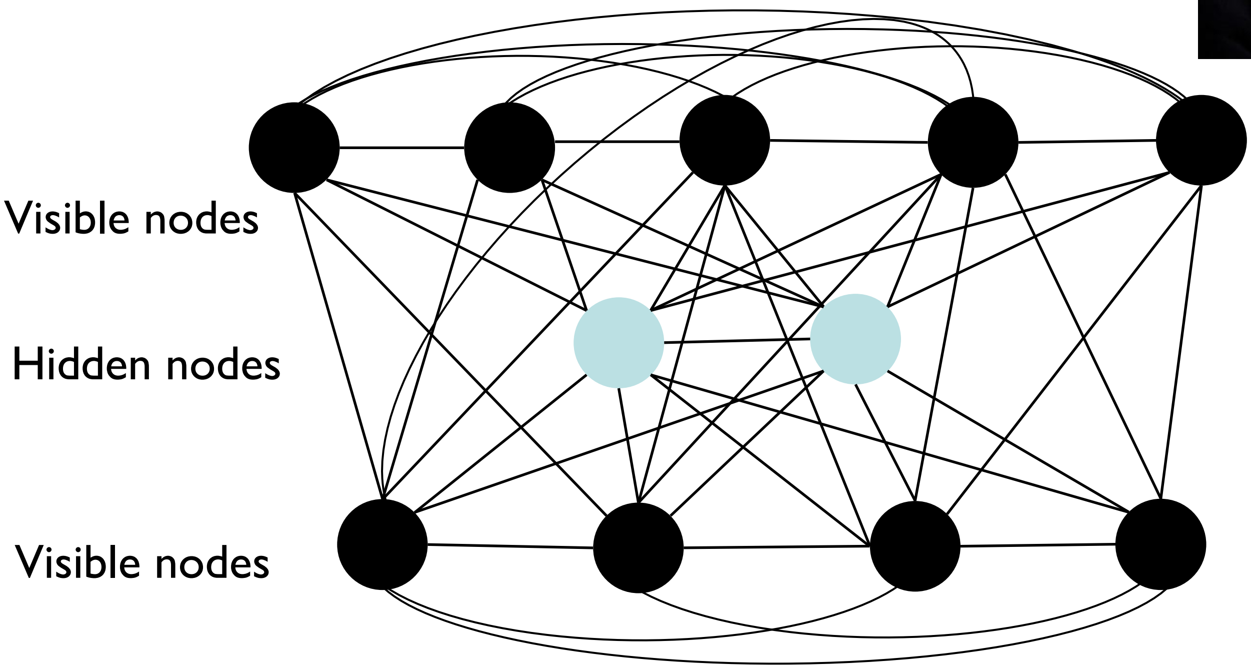


Image: G. Zaharchuk et al, Am J Neuroradiol **39** (2018) 1776

Sejnowski and Hinton came up with a solution

Why not have Hopfield model with hidden nodes that are not directly subject to observation or stimulation ?

Boltzmann Machine



As in the Hopfield model,

- every pair of nodes are connected, and
- $W_{ij} = W_{ji}$ (symmetry)

But, a subset of nodes are not accessible to stimuli

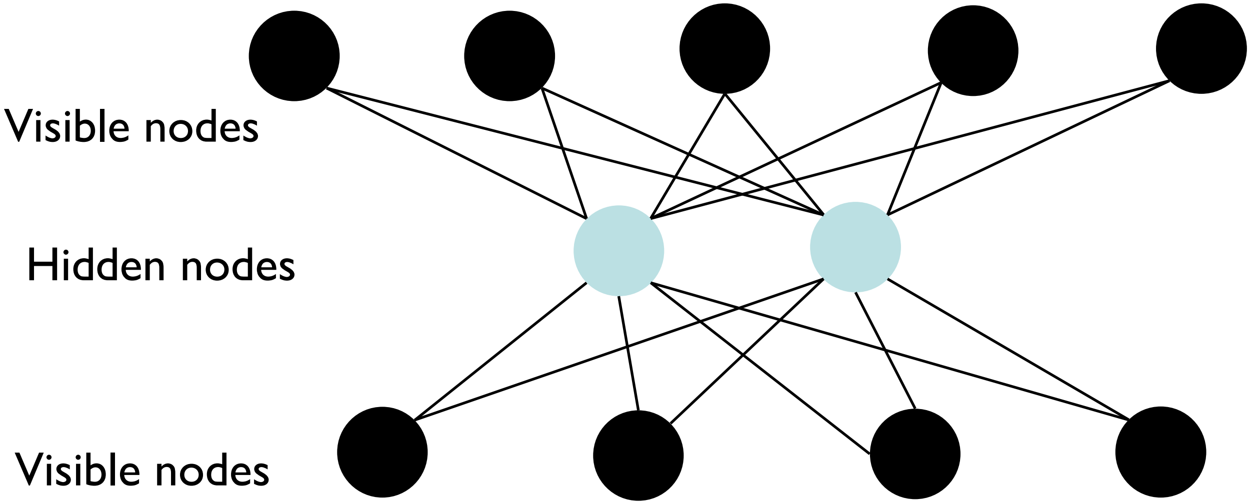
Now one can introduce higher order correlations that are not trivially linked to the mean $\langle S_i \rangle$ and pairwise correlations $\langle S_i S_j \rangle$

This is just the problem of implementing XOR in Perceptron is disguise!

But training the Boltzmann Machine is computationally expensive

So as a compromise, allow only connections between different node types (bipartite network of visible and hidden nodes)

Restricted Boltzmann Machine (RBM)



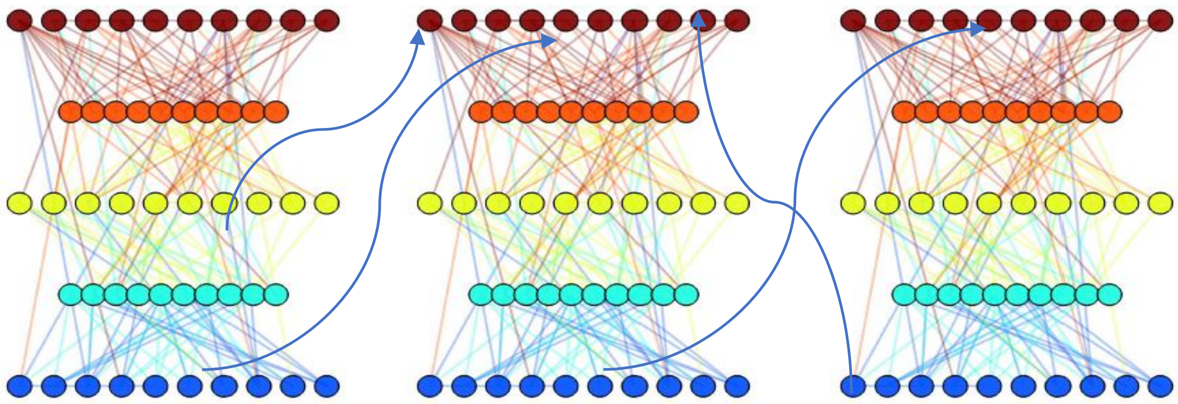
- Now every pair of nodes are **not** connected, but
- $W_{ij} = W_{ji}$ (symmetry)

In practice, RBMs are arranged in a chain – and sequentially used one after the other

The first RBM in the sequence is trained using a given stimulus and then the resulting hidden nodes is used to train the next RBM in the sequence, and so on down the chain…

Modular hierarchy

an intriguing interplay between the mesoscopic organizational principles of hierarchy & modularity



Modularity $\Leftarrow$ necessity of performing multiple independent tasks in parallel, with relatively low requirement for coordination between them$\rightarrow$ sub-networks, each characterized by high intra-connection density facilitating recurrent communication

Hierarchy $\Leftarrow$ if the function typically requires performing several steps in sequence (such that each step needs to be finished before initiating the next), possibly coordinating across multiple input streams $\rightarrow$ efficient serial processing, often in conjunction with feed-back and feed-forward connection across the levels

Functionally, **modular hierarchies** provide a basis for systematic integration of information $\Rightarrow$ allow for distributive processing in a network that otherwise has a markedly modular organization and hence would have appeared to support a segregated (specialized) mode of processing

Thanks to

Department of Atomic Energy,
Government of India

IMSc Center of Excellence in
Complex Systems & Data Science