

**Could the laws of physics
explain not just the universe
but also the workings of our
minds?**

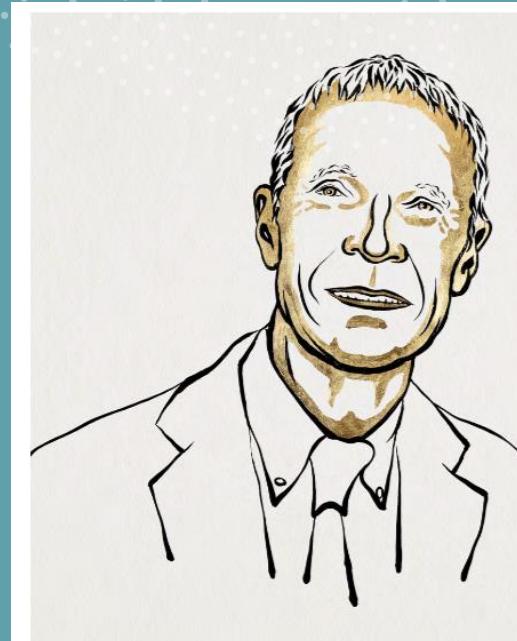
**Physical Principles in Intelligence: The 2024
Nobel Prize in Physics**

Trần Thế Truyền

Applied AI Institute (A2I2), Deakin University

truyen.tran@deakin.edu.au | truyentran.github.io

11/20/2024



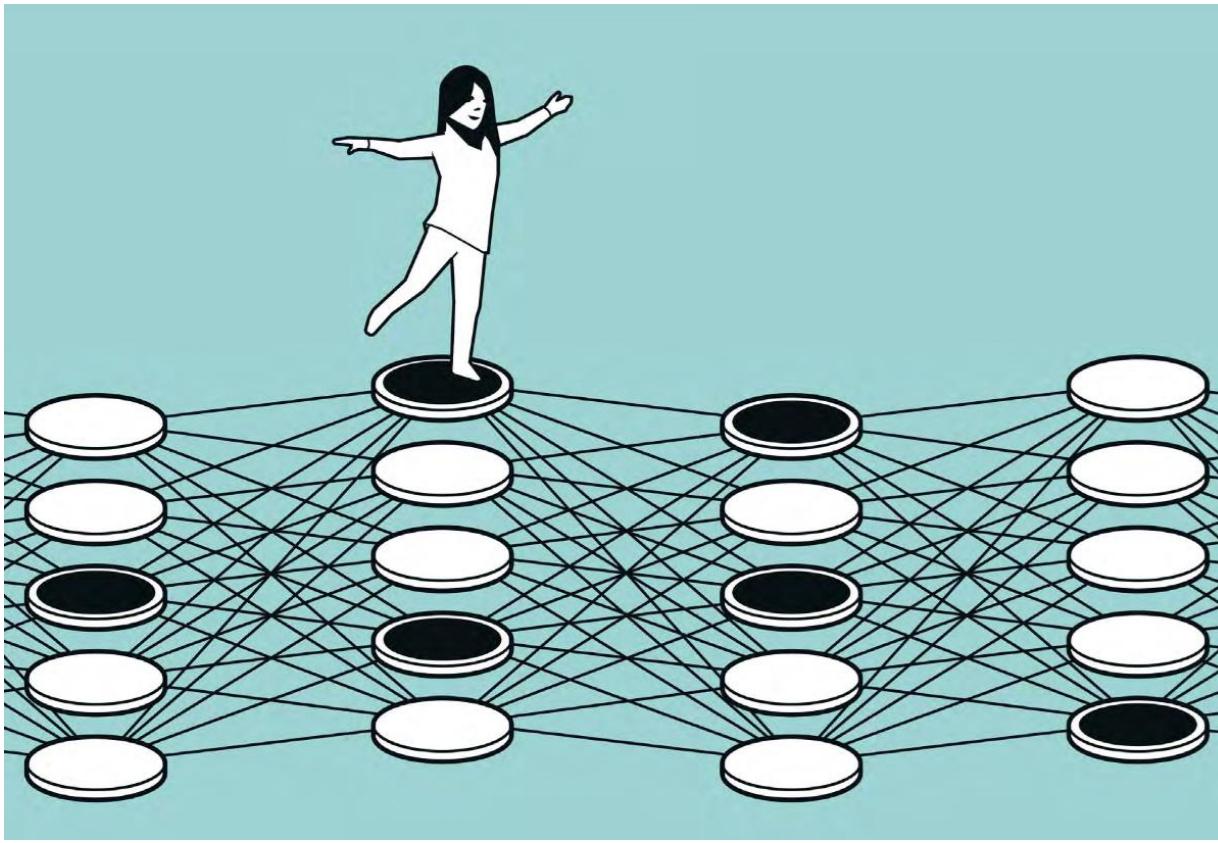
John Hopfield
Born: 1933, USA



Geoffrey E. Hinton
Born: 1947, UK

ILL. NIKLAS SELMEHED © NOBEL PRIZE OUTREACH

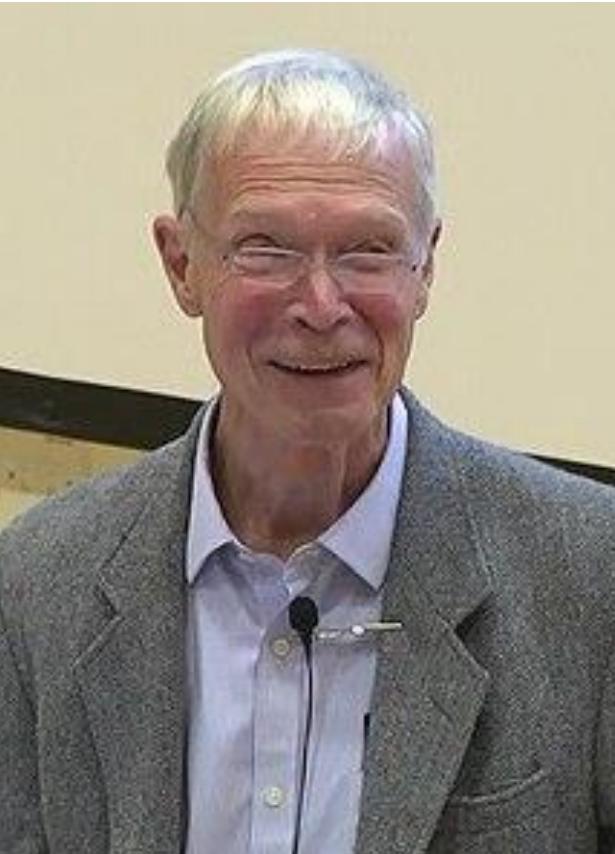




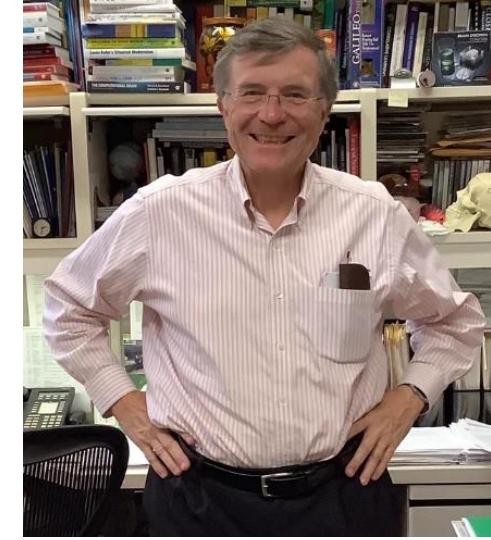
“The Nobel Prize in Physics 2024 recognizes methods that lay the foundation for the development of artificial intelligence.”

John Hopfield

- Condensed matter physics
- Statistical physics
- Biophysics
- Molecular biology
- Complex systems
- Neuroscience



John Hopfield
(born 1933)



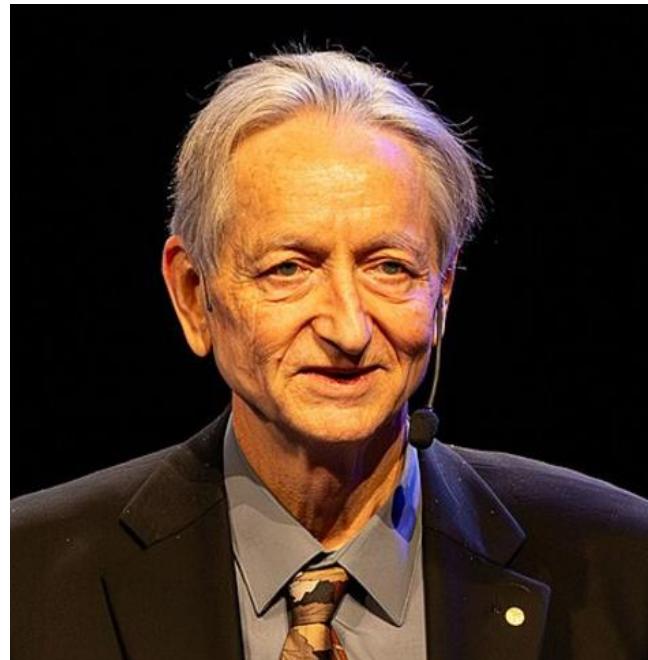
Terry Sejnowski
(born 1947)



David MacKay
(1967-2016)

Geoffrey Everest Hinton

- Psychology
- Cognitive science
- Artificial intelligence
- Ethics



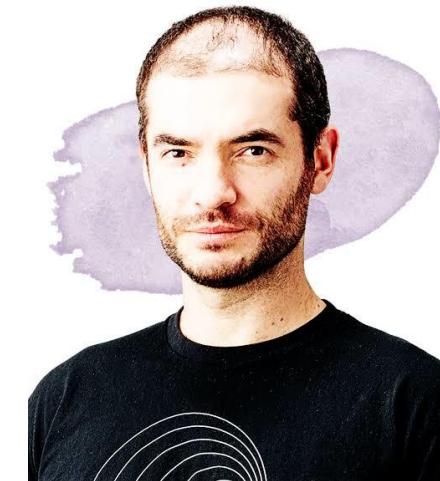
Hinton (born 1933)



Ghahramani (born 1970)



LeCun (born 1960)



Sutskever (born 1986)



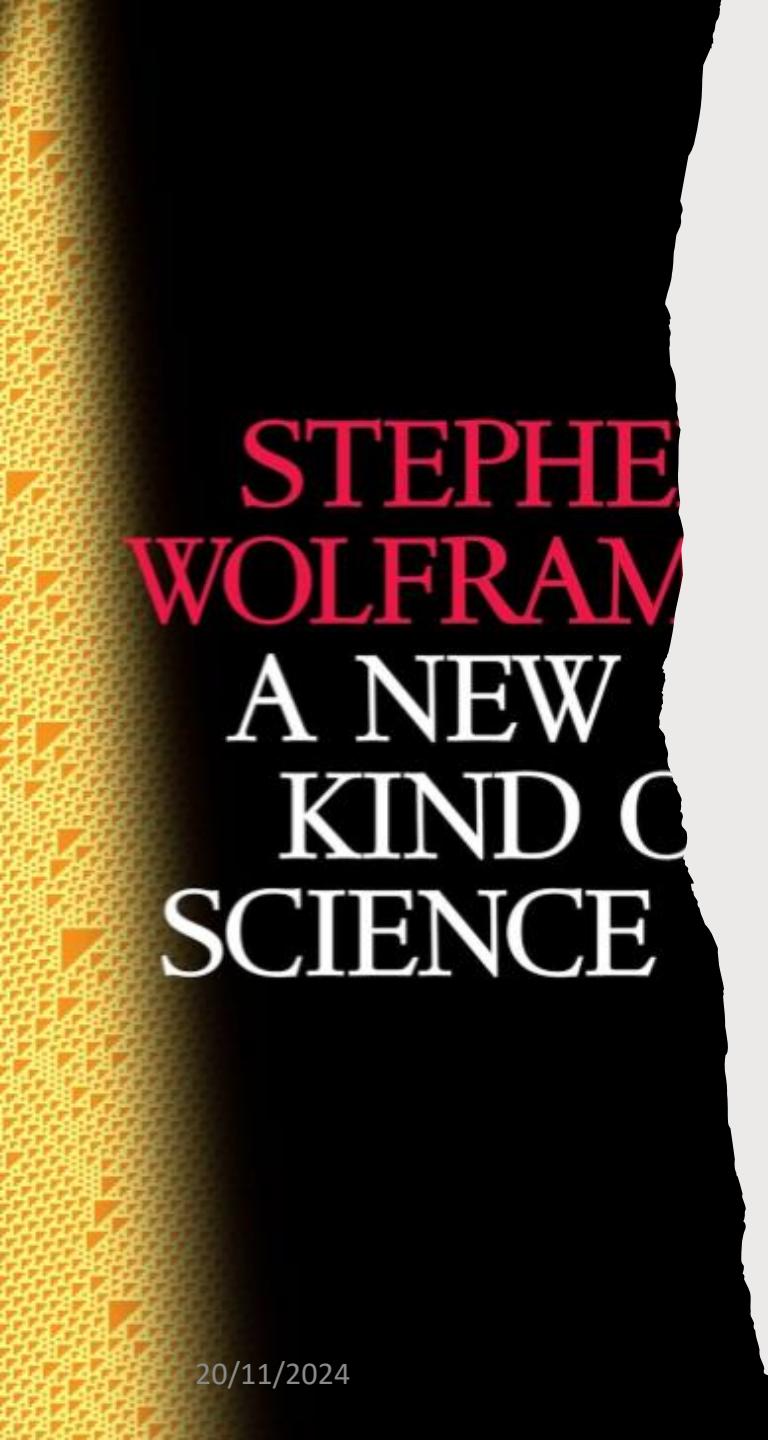
Welling (born 1968)

Agenda

- The context
- Hopfield networks
- Boltzmann machines
- Deep learning
- Discussion



ChatGPT: Generate a picture of the universe and the mind



STEPHEN WOLFRAM A NEW KIND OF SCIENCE

Physics, life, mind and computation

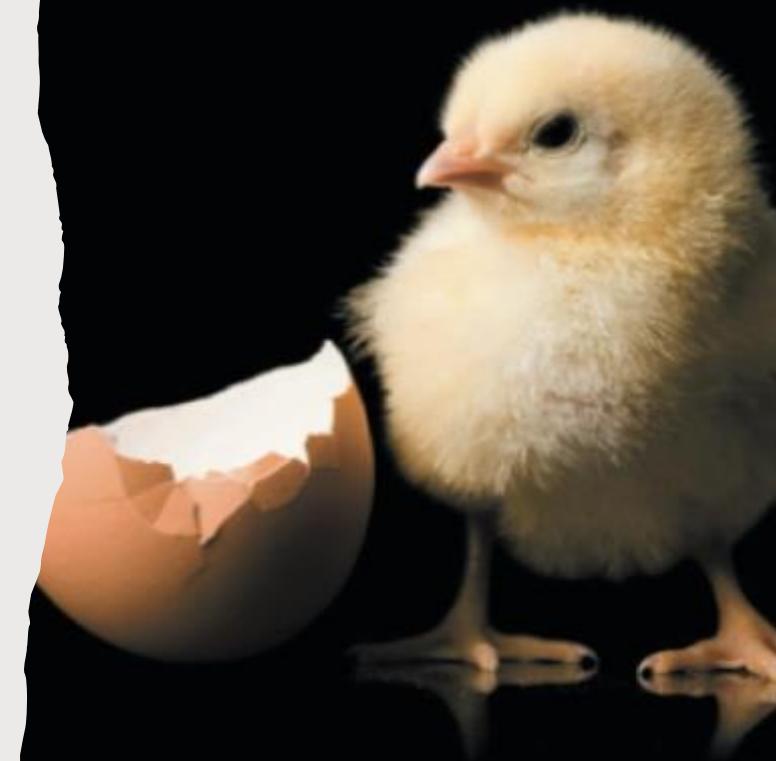
Great topics studied for thousands
of years

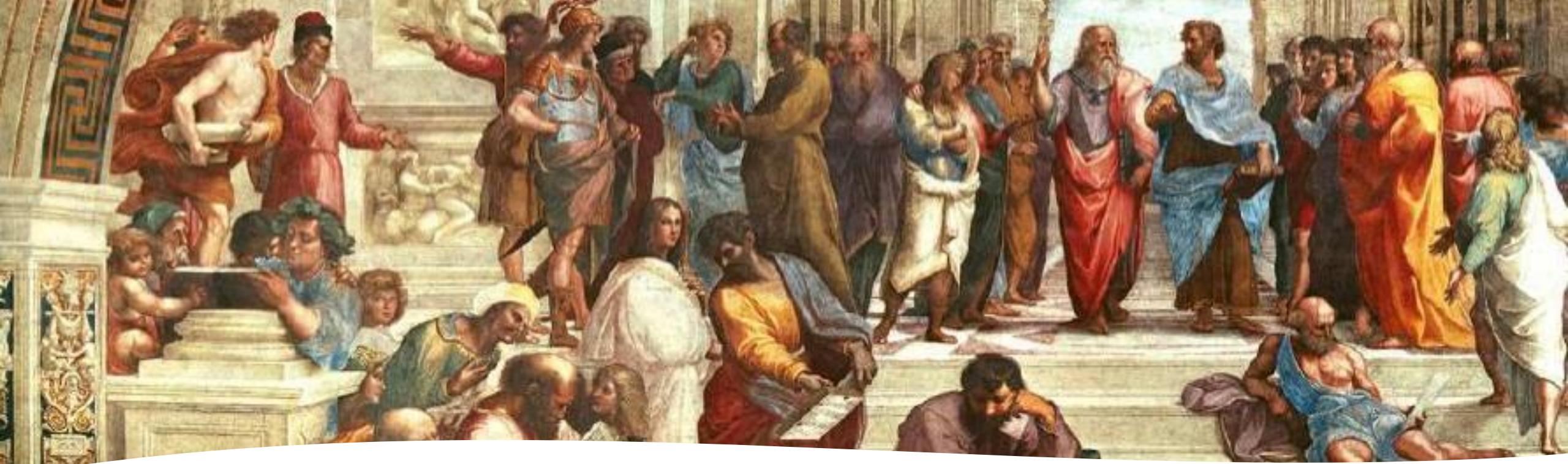
- Addressed by many traditions
(Taoism, Buddhism,
Hinduism, etc)
- Asked by philosophers,
physicists, computer
scientists, etc.

Only recently connections have been
made

- What is life?
- Is the universe a computer?
- Is mind computation?
-

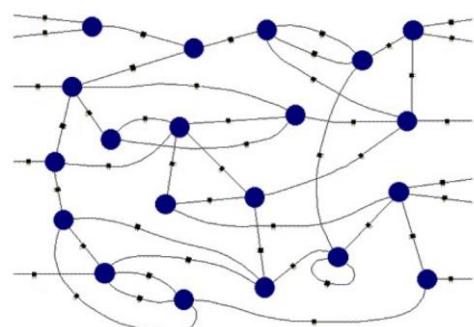
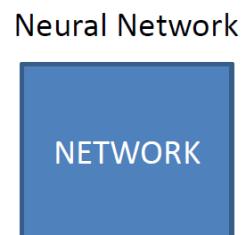
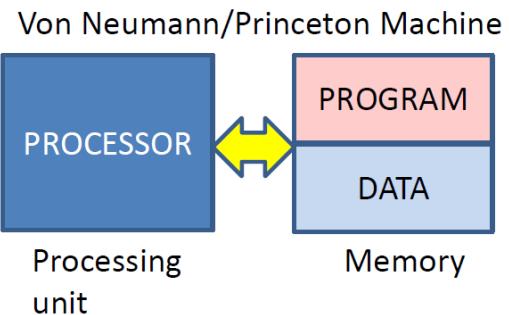
Schrödinger
What is Life?





Early models of human cognition

- **Associationism:** Humans learn through association
- 400BC-1900AD: Plato, John Locke, David Hume, David Hartley, James Mill, John Stuart Mill, Alexander Bain, Ivan Pavlov.

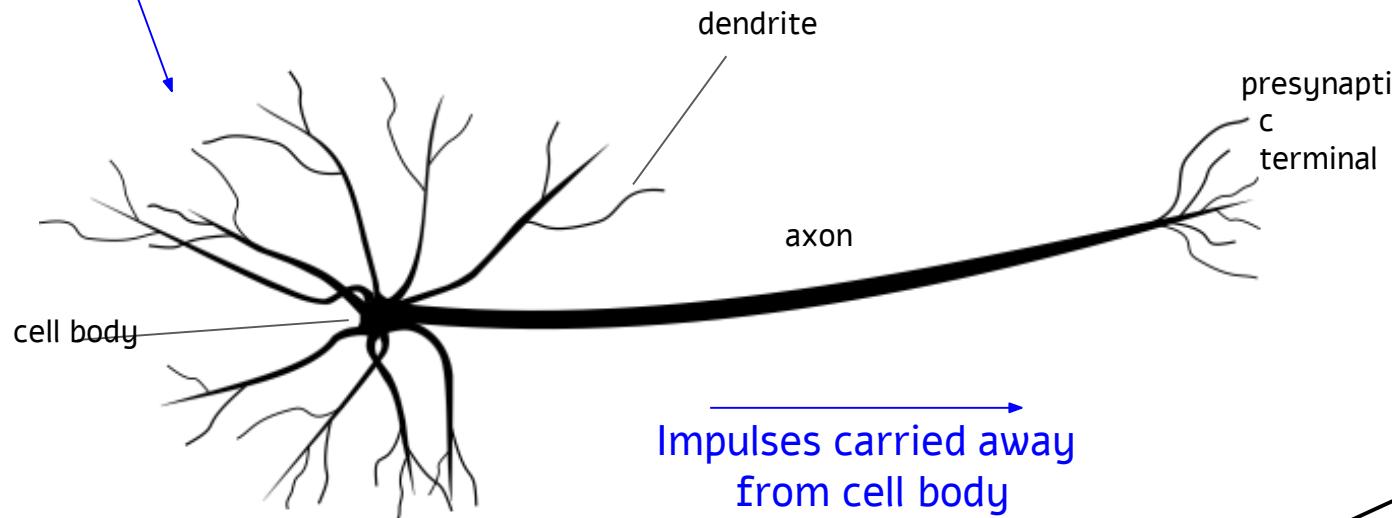


Connectionism

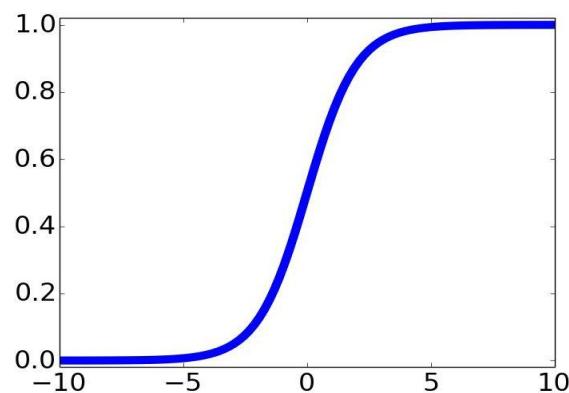
- Mid 1800s: The brain is a mass of interconnected neurons
- Alexander Bain, philosopher, psychologist, mathematician, logician, linguist, professor
 - 1873: The information is in the connections – *Mind and body*.
- Connectionist machines
 - Network of processing elements
 - All world knowledge is stored in the connections between the elements
- **Neural networks** are connectionist machines
 - As opposed to Von Neumann Machines

Neurons *in silico*

Impulses carried toward cell body



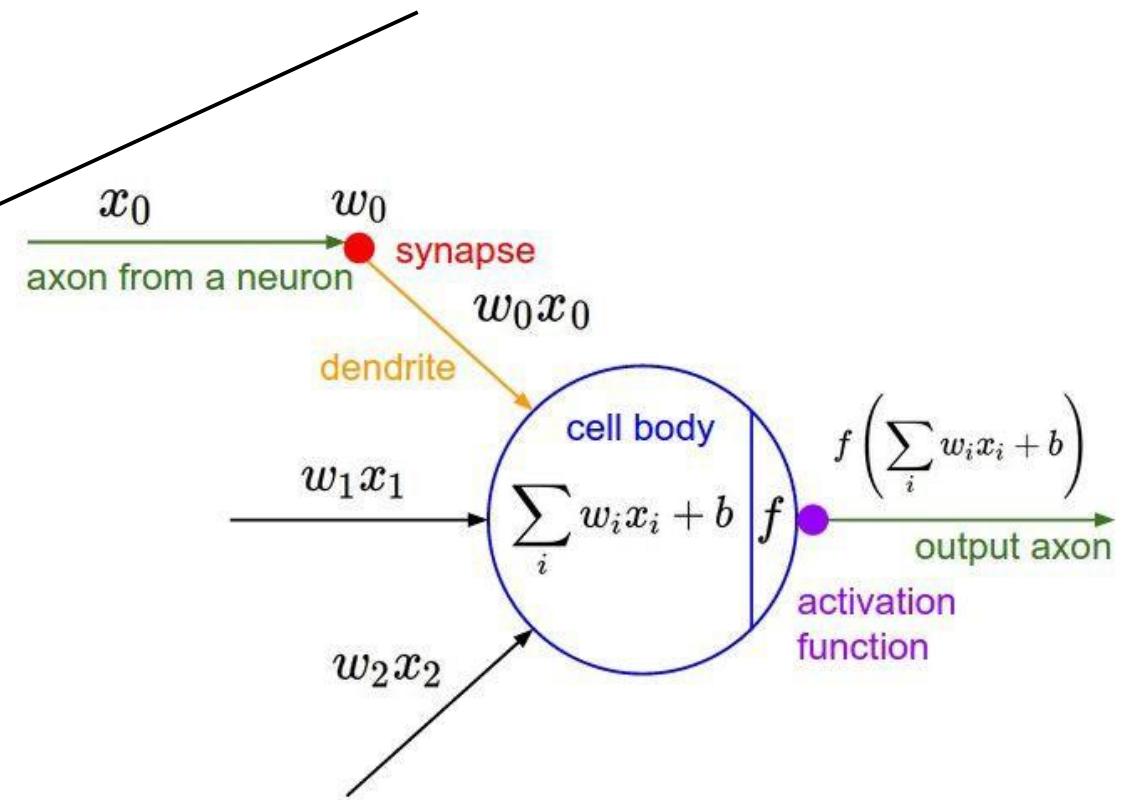
This image by Felipe Perucho
is licensed under CC-BY 3.0

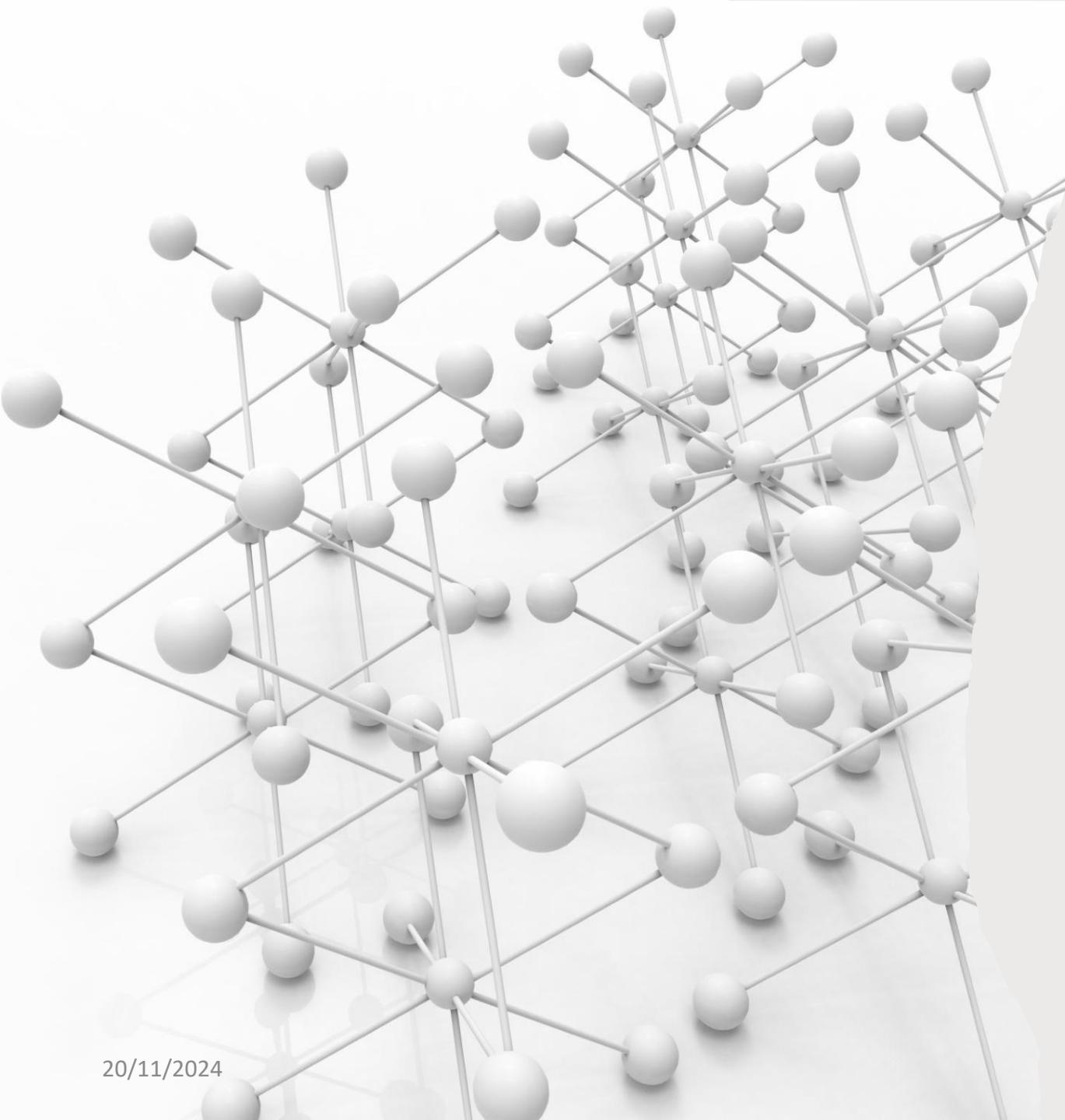


sigmoid activation
function

$$\frac{1}{1 + e^{-x}}$$

Impulses carried away
from cell body



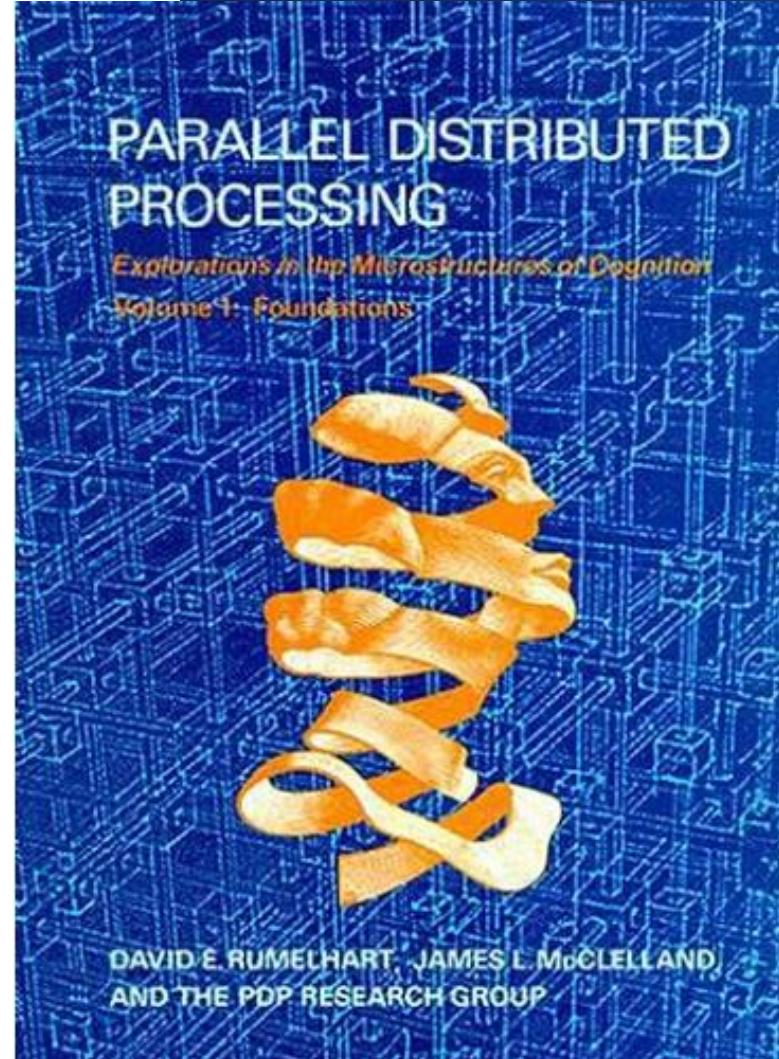


Neural networks as model of mind

- How does the mind work?
 - Perceiving
 - Remembering
 - Learning
 - Reasoning, planning
 - Doing
- Parallel Distributed Processing (1986)
- 1990s: Mind as recurrent neural network

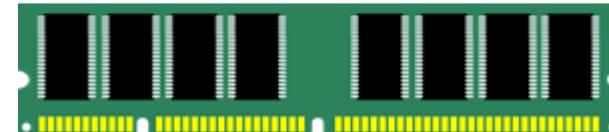
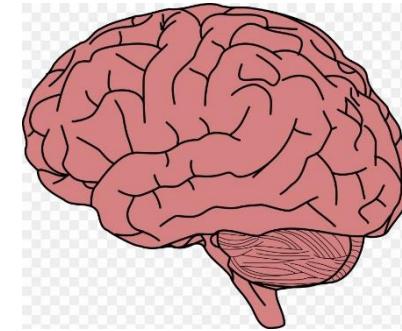
1980s: Parallel Distributed Processing

- Information is stored in many places (distributed)
 - Each concept is represented by many neurons
 - Each neuron plays role in many concepts
- Activations are sparse (enabling selectivity and invariance)
- Popular these days: Embedding of everything into the vector space



Memory is essential to intelligence

- Memory is the ability to **store**, **retain** and **recall** information
- Brain memory stores items, events and high-level structures
- Computer memory stores data and temporary variables



Associative memory is powerful

Language

"Green" means
"go," but what
does "red" mean?

Time

birthday party on
30th Jan

Object

Where is my pen?
What is the
password?

Behaviour



Semantic
memory

Episodic
memory

Working
memory

Motor
memory

~~Four~~Five things in AI

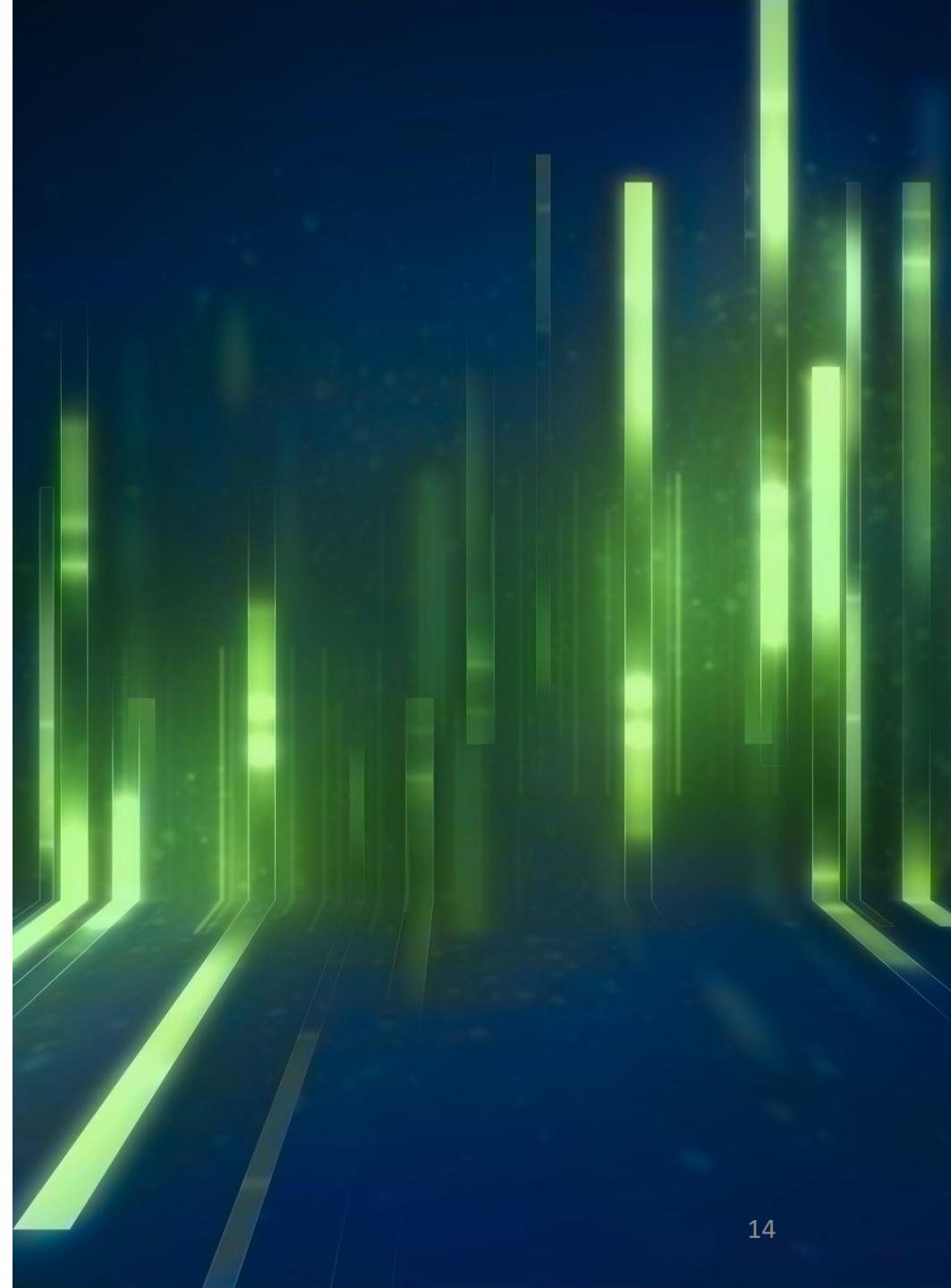
Representation

Inference

Learning

Stochasticity

Scale with compute



Agenda



The context



Hopfield networks



Boltzmann machines



Deep learning



Discussion

Questions



How do we teach the network to store *a specific* pattern or set of patterns?



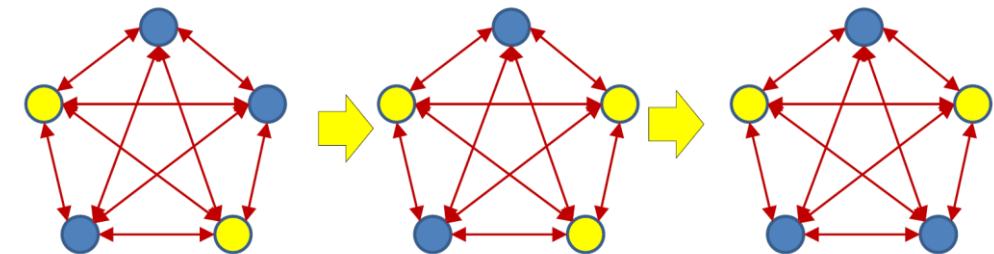
How many patterns can we store?



How to “retrieve” patterns better..

Hopfield network

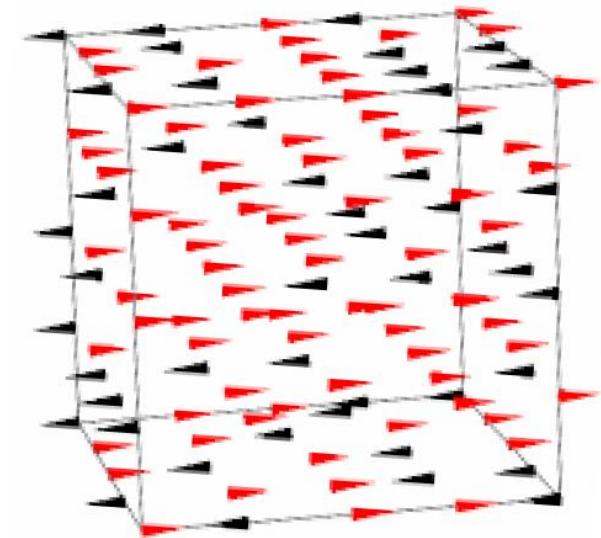
- Symmetric weights
- No self-connection
- Update either asynchronous or synchronous
- Neurons attract or repel each other
- The dynamic is deterministic



$$s_i \leftarrow \begin{cases} +1 & \text{if } \sum_j w_{ij} s_j \geq \theta_i, \\ -1 & \text{otherwise.} \end{cases}$$

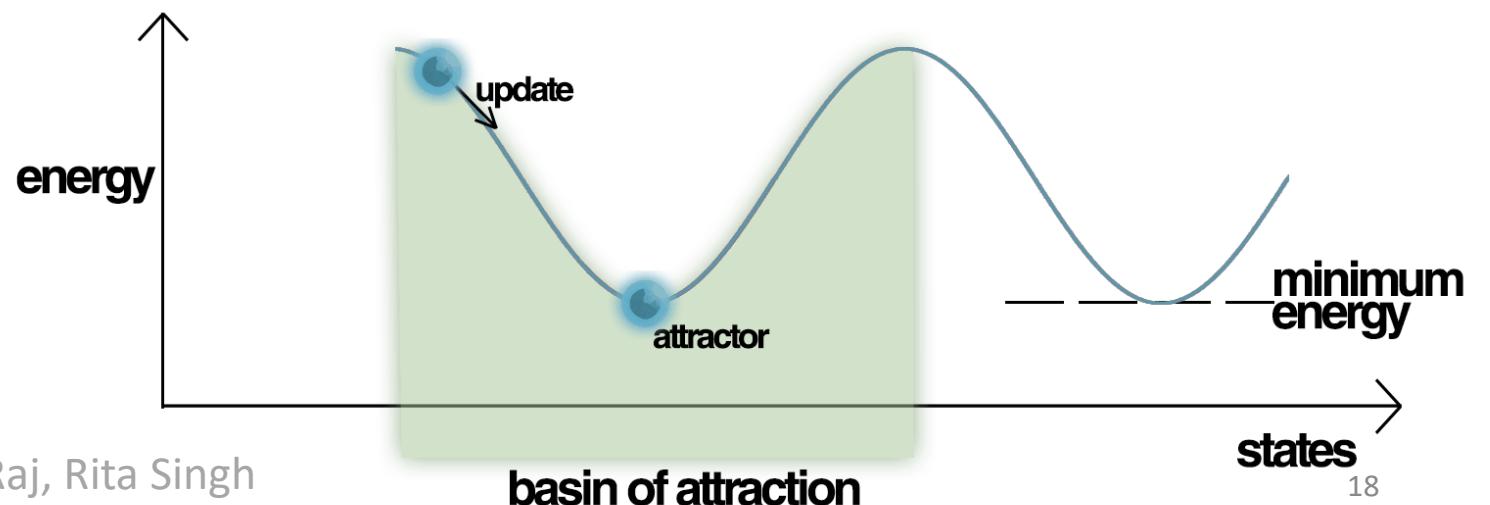
Update as energy minimization

- Start at some initial pattern (configuration)
- Let the network “runs”
- The convergence is a local attractor (stable)
- The operation is essentially Iterated Local Mode (ICM) known in (spatial) statistics, 1975.

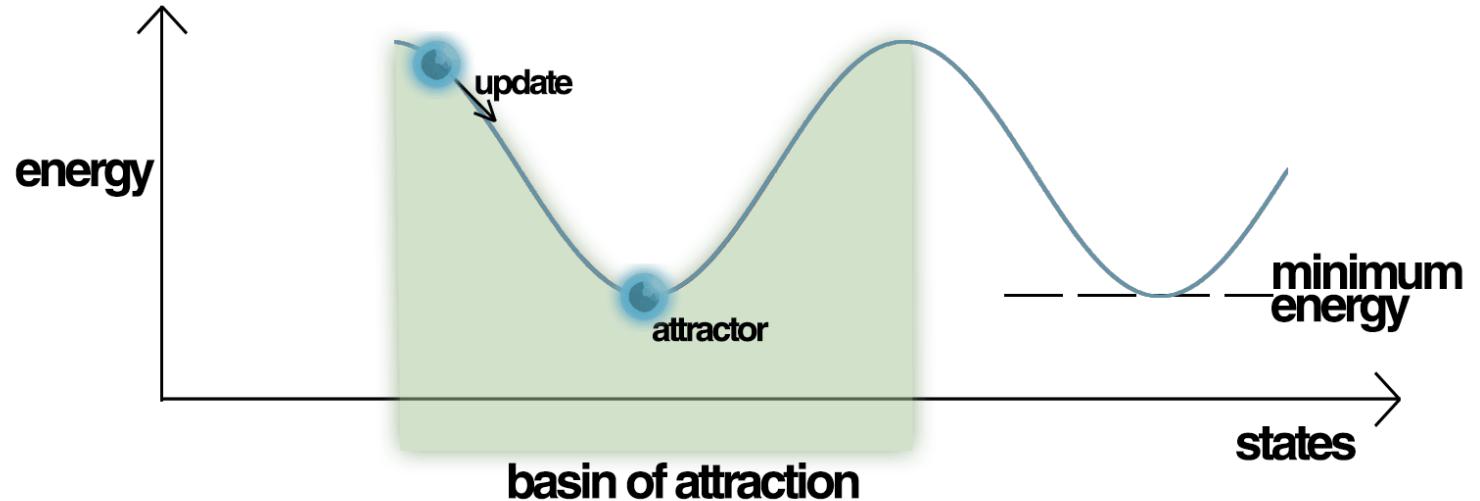


A spin glass system

$$E = -\frac{1}{2} \sum_{i,j} w_{ij} s_i s_j - \sum_i \theta_i s_i$$

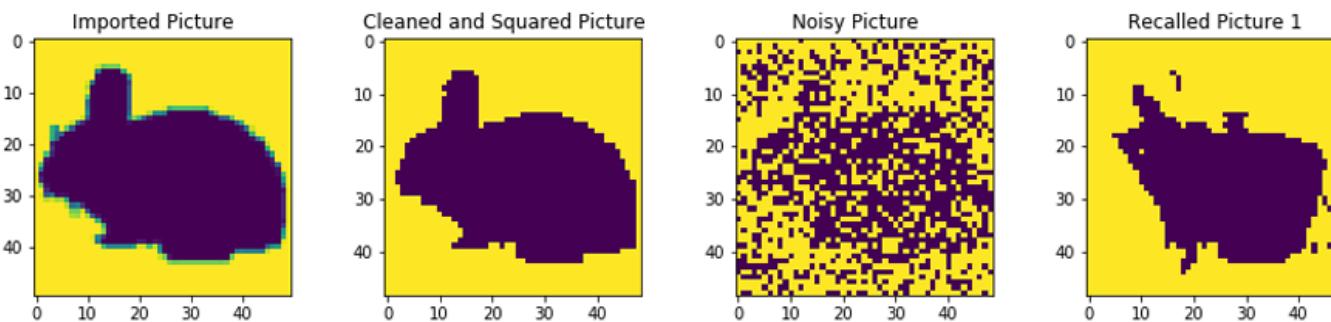
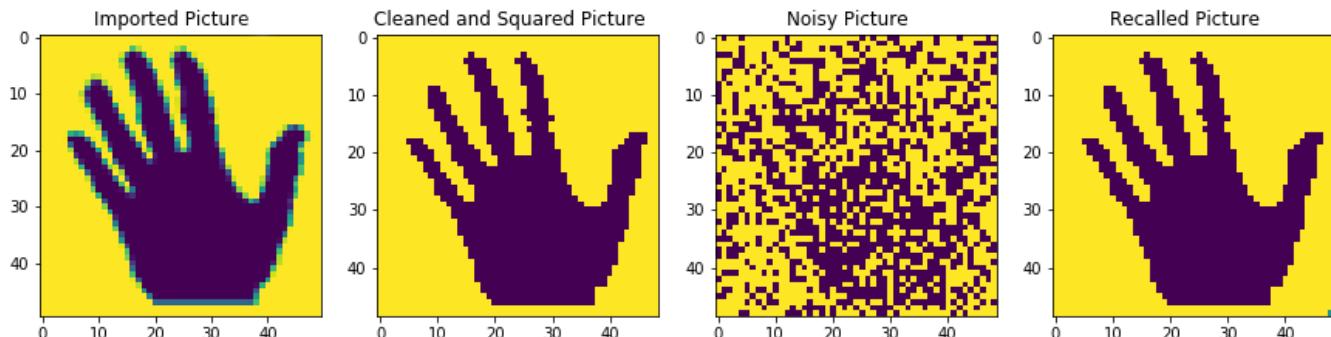
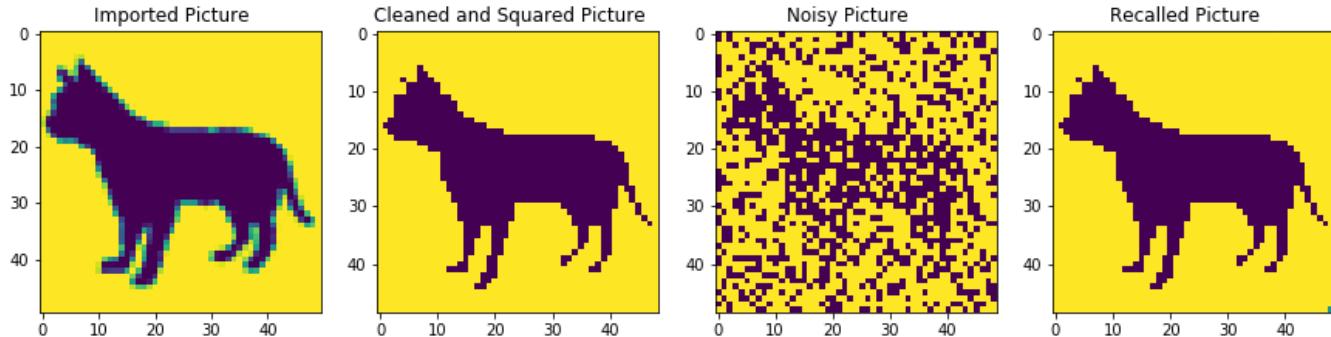


Content-addressable memory



- Each of the minima is a “stored” pattern
- => If the network is initialized close to a stored pattern, it will inevitably evolve to the pattern
- **This is a *content addressable memory***
 - Recall memory content from partial or corrupt values
 - Also called *associative memory*
- ***Indeed, current frontier AI models are!***

Examples: Image denoising



To store one pattern

Hebbian learning rule

$$w_{ji} = y_j y_i$$

“Neurons that fire together, wire together”

$$\begin{aligned} E &= - \sum_i \sum_{j < i} w_{ji} y_j y_i = - \sum_i \sum_{j < i} y_i^2 y_j^2 \\ &= - \sum_i \sum_{j < i} 1 = -0.5N(N-1) \end{aligned}$$

$$\text{sign}\left(\sum_{j \neq i} w_{ji} y_j\right) = \text{sign}\left(\sum_{j \neq i} y_j y_i y_j\right)$$

$$= \text{sign}\left(\sum_{j \neq i} y_j^2 y_i\right) = \text{sign}(y_i) = y_i$$

This is the global minimum

To store multiple patterns

$$w_{ji} = \sum_{\mathbf{y}_p \in \{\mathbf{y}_p\}} y_i^p y_j^p$$

- **Hopfield:** For a network of N neurons can store up to $\sim 0.15N$ patterns through Hebbian learning
 - Provided they are “far” enough
- **Later:** Guarantees that a network of N bits trained via Hebbian learning can store $0.14N$ random patterns with less than 0.4% probability that they will be unstable
- A better method: Energy-based methods!
 - Contrastive learning (e.g., CLIP)

Agenda



THE CONTEXT



HOPFIELD
NETWORKS



**BOLTZMANN
MACHINES**



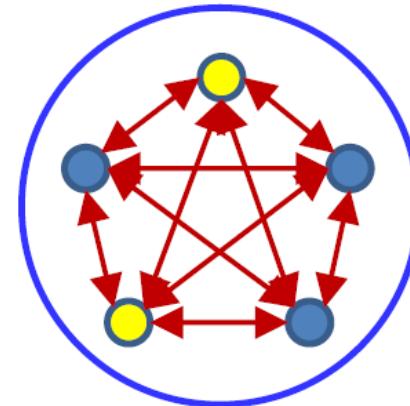
DEEP
LEARNING



DISCUSSION

A stochastic Hopfield network with hidden nodes

- Capacity of Hopfield network can be vastly increased by introducing hidden nodes
- Stochasticity gives principled ways to handle uncertainty, randomness and statistical properties
- **Observation:** The behavior of the Hopfield net is analogous to annealed dynamics of a spin glass characterized by a Boltzmann distribution
- Linked to MaxEnt principle (Max Entropy)
 - Everything else equal ...



The Helmholtz Free Energy of a System

$$F_T = \sum_s P_T(s) E_s - kT \sum_s P_T(s) \log P_T(s)$$

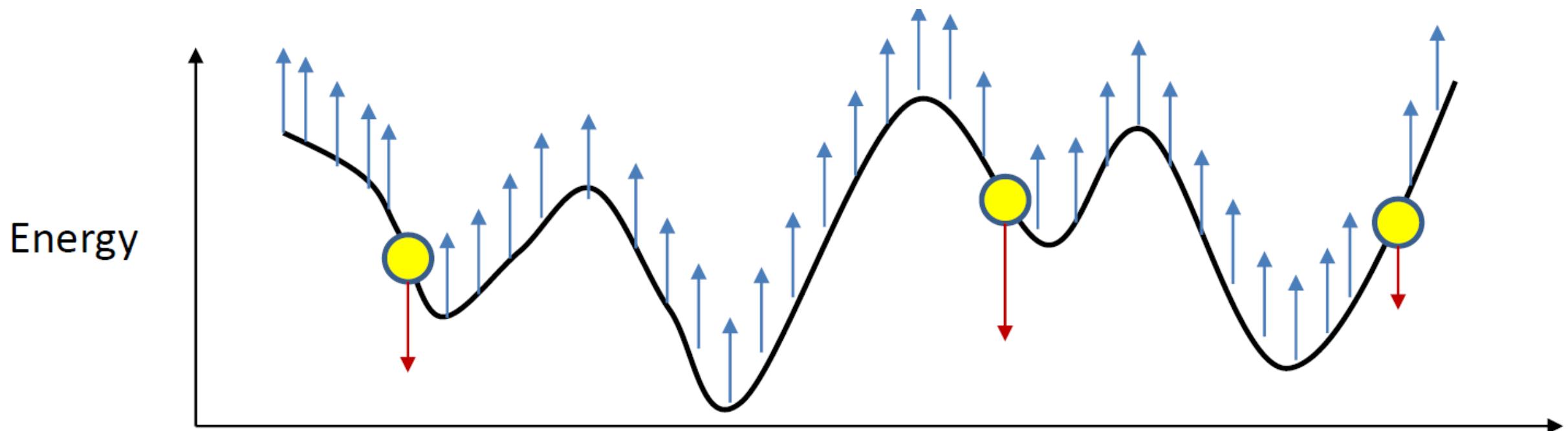


$$P_T(s) = \frac{1}{Z} \exp\left(\frac{-E_s}{kT}\right)$$

$$E(S) = - \sum_{i < j} w_{ij} s_i s_j$$

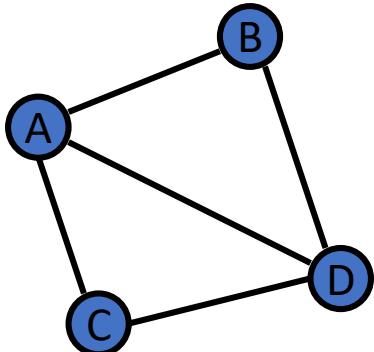
Gibbs distribution

Training a Boltzmann machine

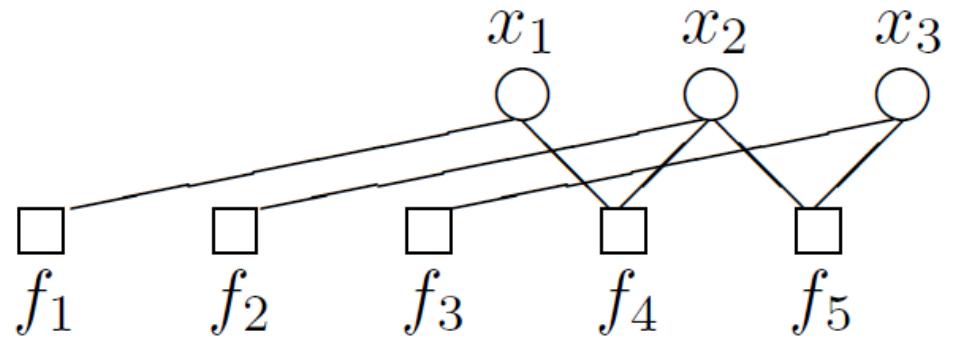


Generalising Boltzmann machines

Markov network



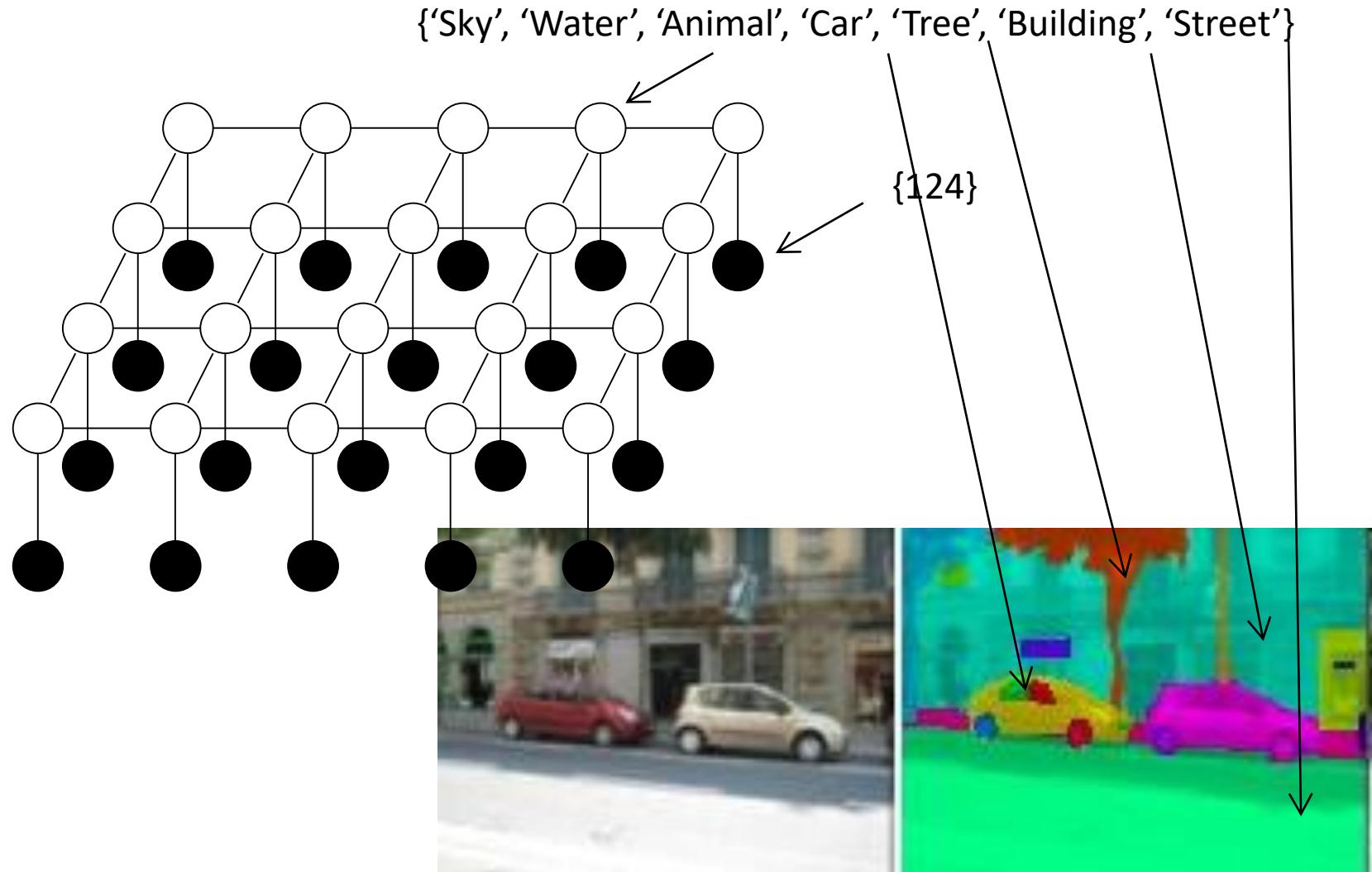
Factor graph



Decoding in
Information
Theory

$$\begin{aligned} P(x_A, x_B, x_C, x_D) &= \frac{1}{Z} \Phi(x_A, x_B, x_D) \Phi(x_A, x_C, x_D) \\ Z &= \sum_{x_A, x_B, x_C, x_D} \Phi(x_A, x_B, x_D) \Phi(x_A, x_C, x_D) \end{aligned}$$

Example: Markov random fields



Inference as Bethe free-energy minimization

- Inference problems
 - → Estimate MAP as energy minimization
 - → Compute marginal probability
 - → Compute expectation & normalisation constant
- Key solution = Belief propagation
 - = Sum-Product algorithm in factor graphs.

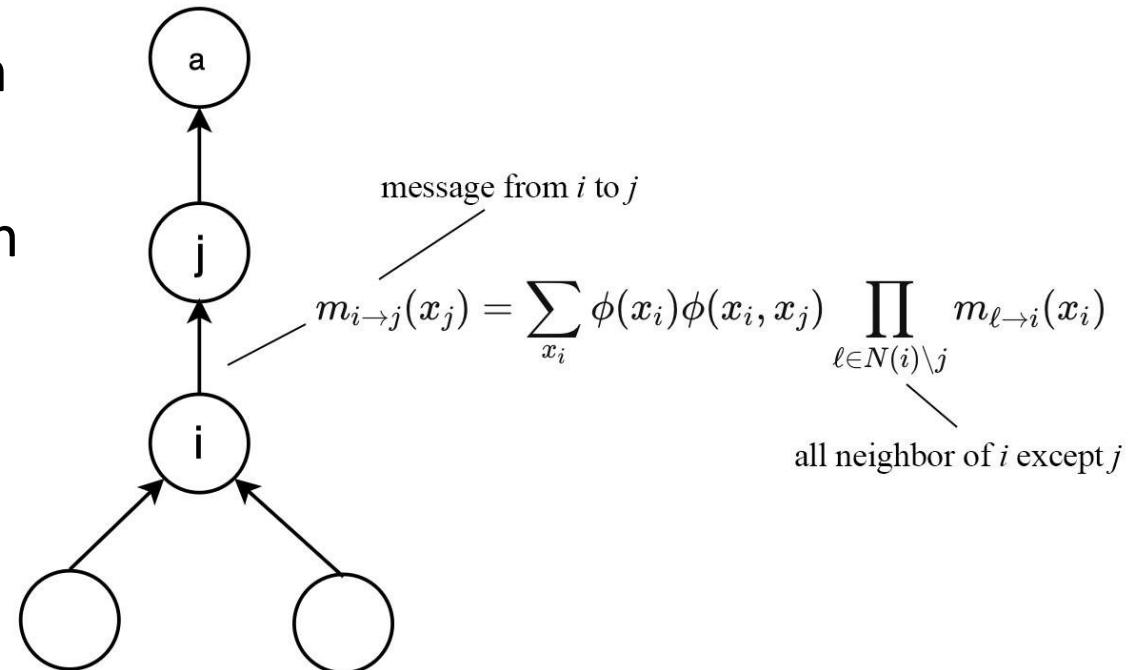
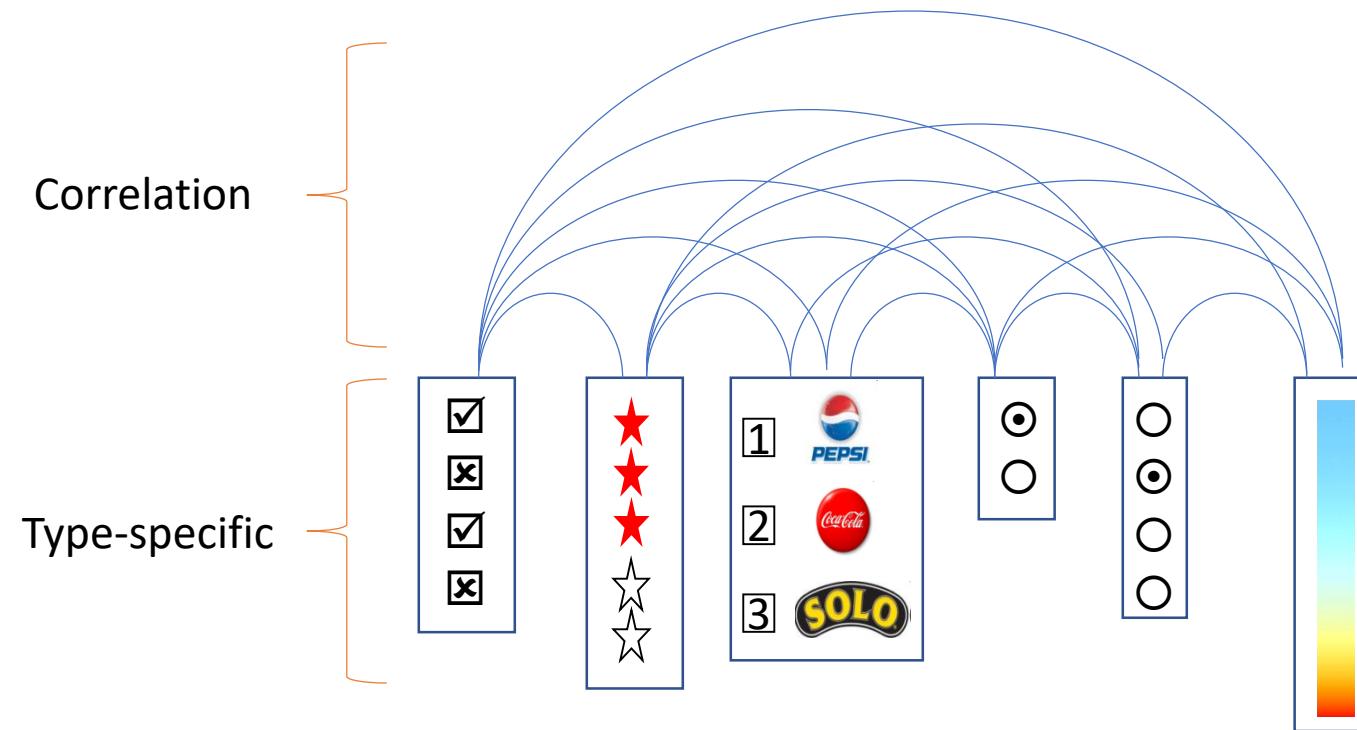


Figure credit: Jonathan Hui

Generation: Markov networks for mixed data types

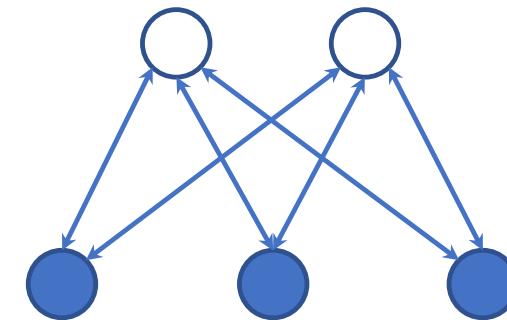


Restricted Boltzmann machines (RBMs)

- Hidden variables to denote underlying unobserved processes
- Stack of RBMs is akin to renormalization trick in physics

$$p(\mathbf{v}, \mathbf{h}; \psi) \propto \exp [-E(\mathbf{v}, \mathbf{h}; \psi)]$$

energy

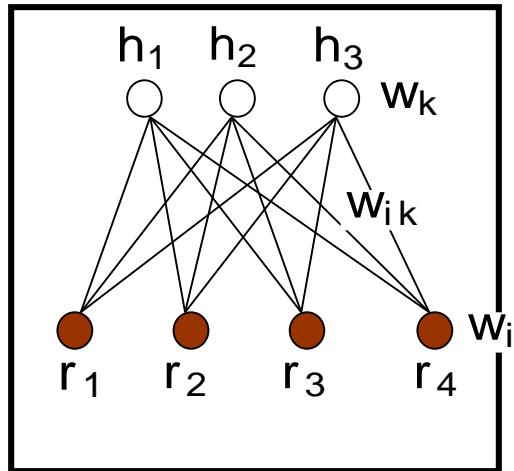


Restricted Boltzmann Machine
(~1994, 2001)

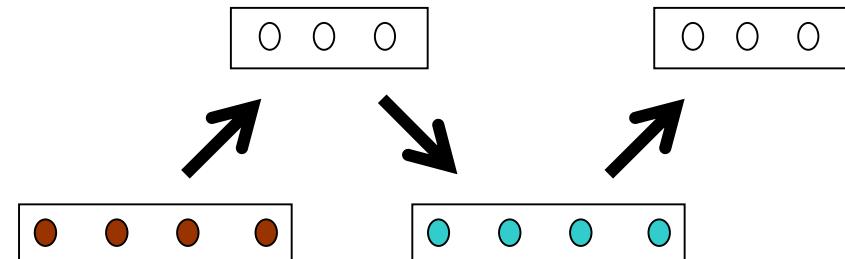
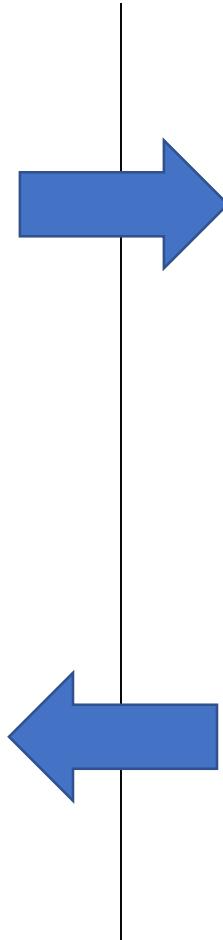
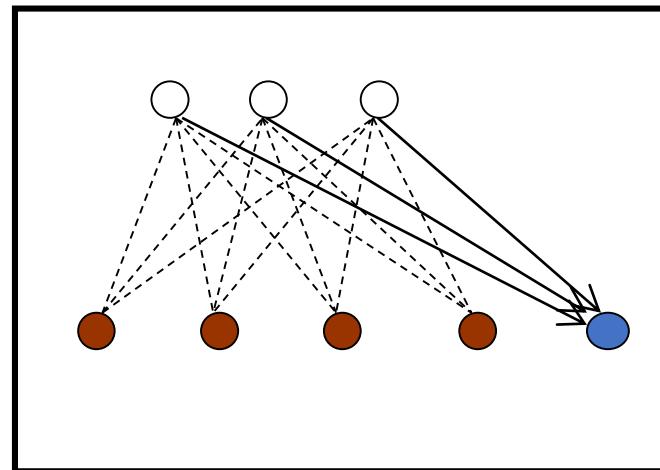
Learning and prediction

$$p(\mathbf{v}, \mathbf{h}; \psi) \propto \exp [-E(\mathbf{v}, \mathbf{h}; \psi)]$$

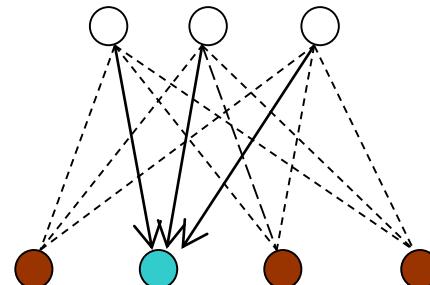
energy



prediction



CD Learning



Pseudo-likelihood Learning

Agenda



The context



Hopfield
networks



Boltzmann
machines



Deep learning

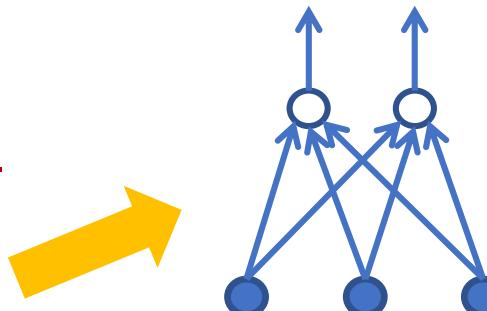
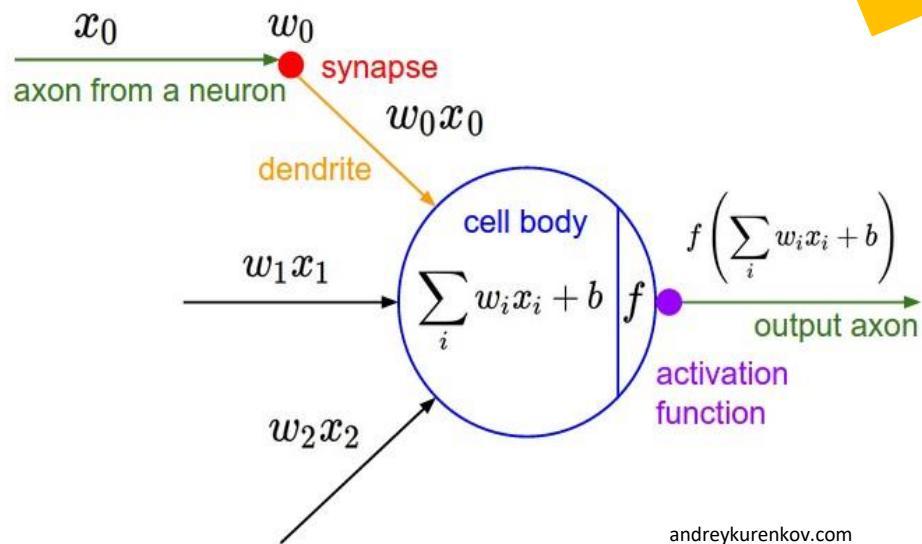


Discussion

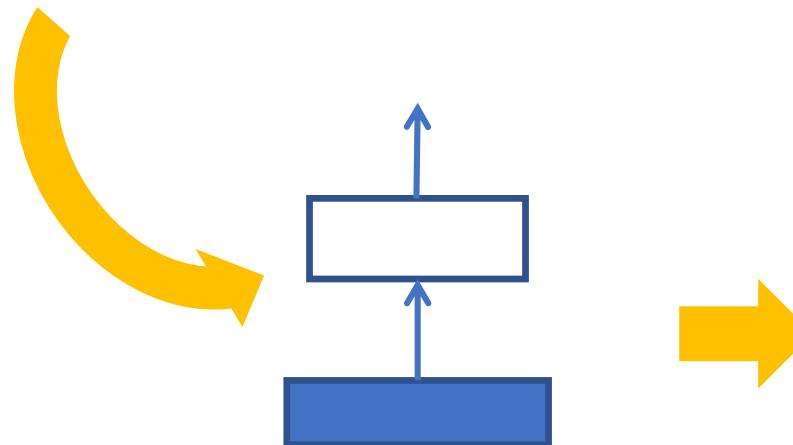
Deep models via layer stacking

Theoretically powerful, but very difficult to train!

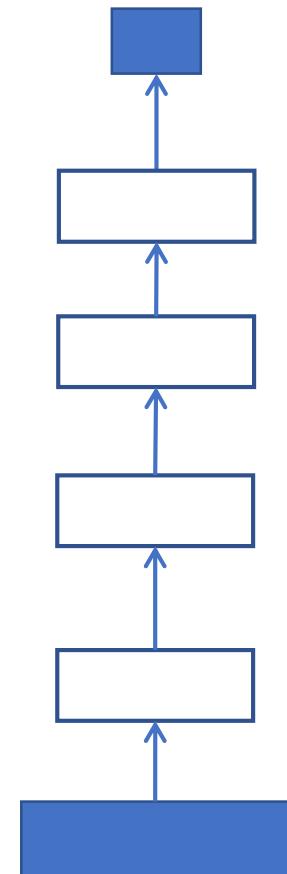
Integrate-and-fire neuron



Feature detector



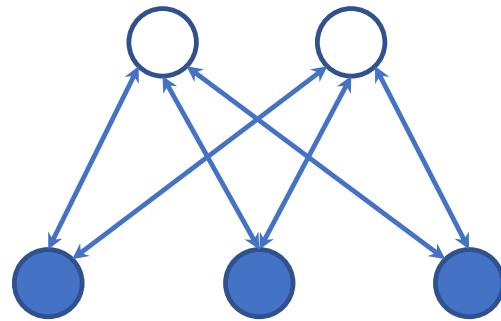
Block representation



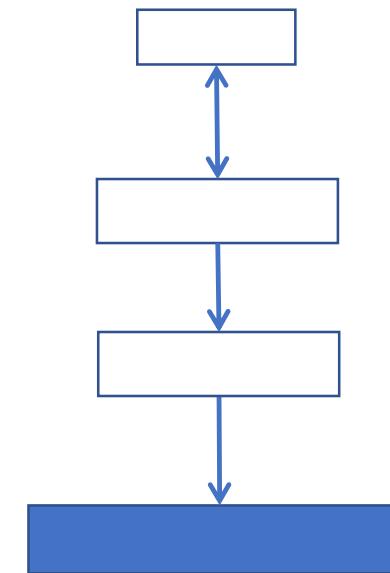
Start of deep learning: Layer-wise training

$$p(\mathbf{v}, \mathbf{h}; \psi) \propto \exp [-E(\mathbf{v}, \mathbf{h}; \psi)]$$

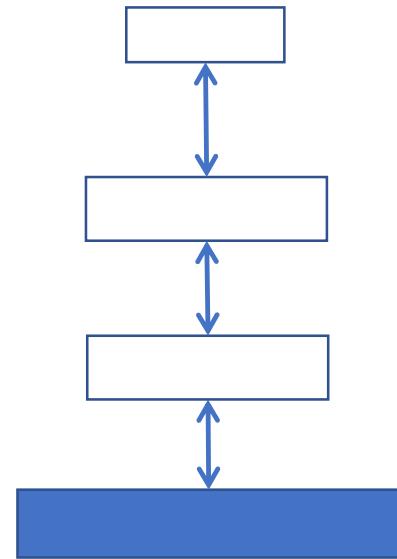
energy



Restricted Boltzmann Machine
(~1994, 2001)



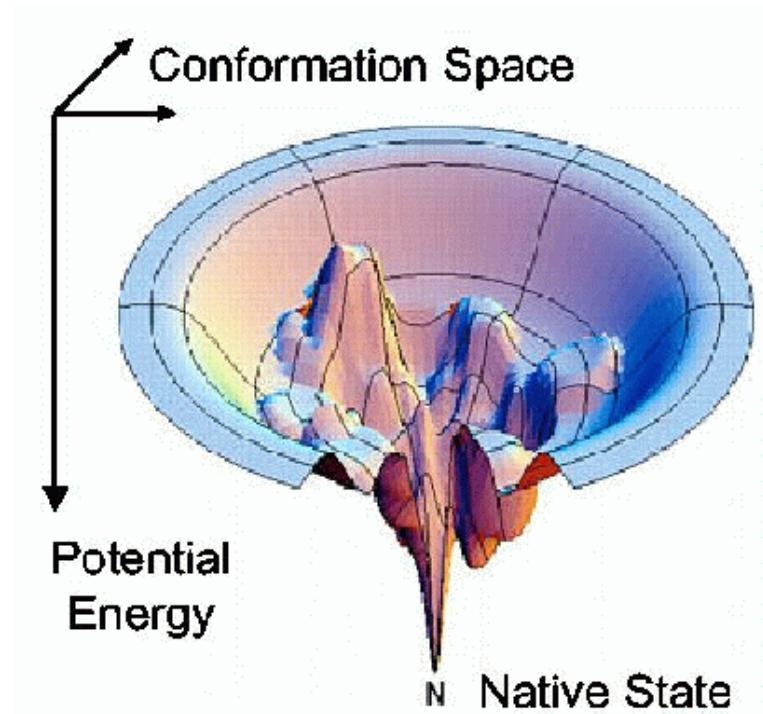
Deep Belief Net
(2006)



Deep Boltzmann Machine
(2009)

Later: Stochastic gradient descent (SGD)

- Using mini-batch to smooth out gradient
- Use large enough learning rate to get over poor local minima
- Periodically reduce the learning rate to land into a good local minima
- It sounds like **Simulated Annealing**, but without proven global minima
- Works well in practice since the **energy landscape** is a funnel



Modern SGD: Adaptive

Adagrad (Duchi et al, 2011)

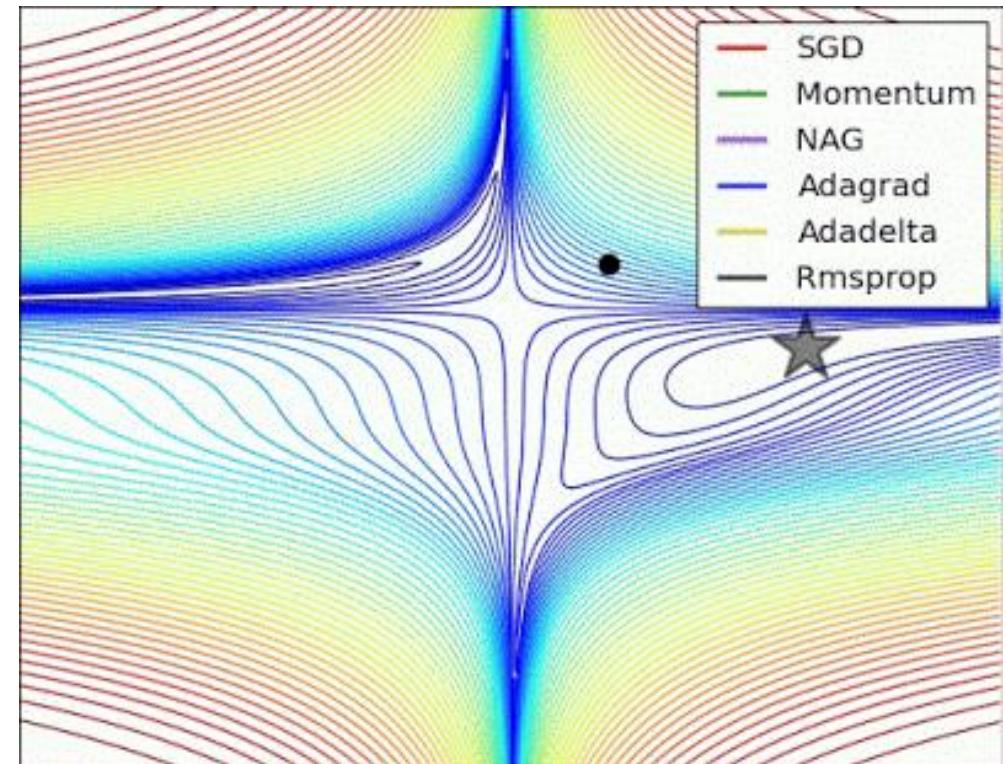
$$w_t \leftarrow w_{t-1} - \eta \frac{g_t}{\sqrt{\epsilon + \sum_{j=1}^{t-1} g_j * g_j}}$$

Init learning rate

Gradient

Smoothing factor

Previous gradients

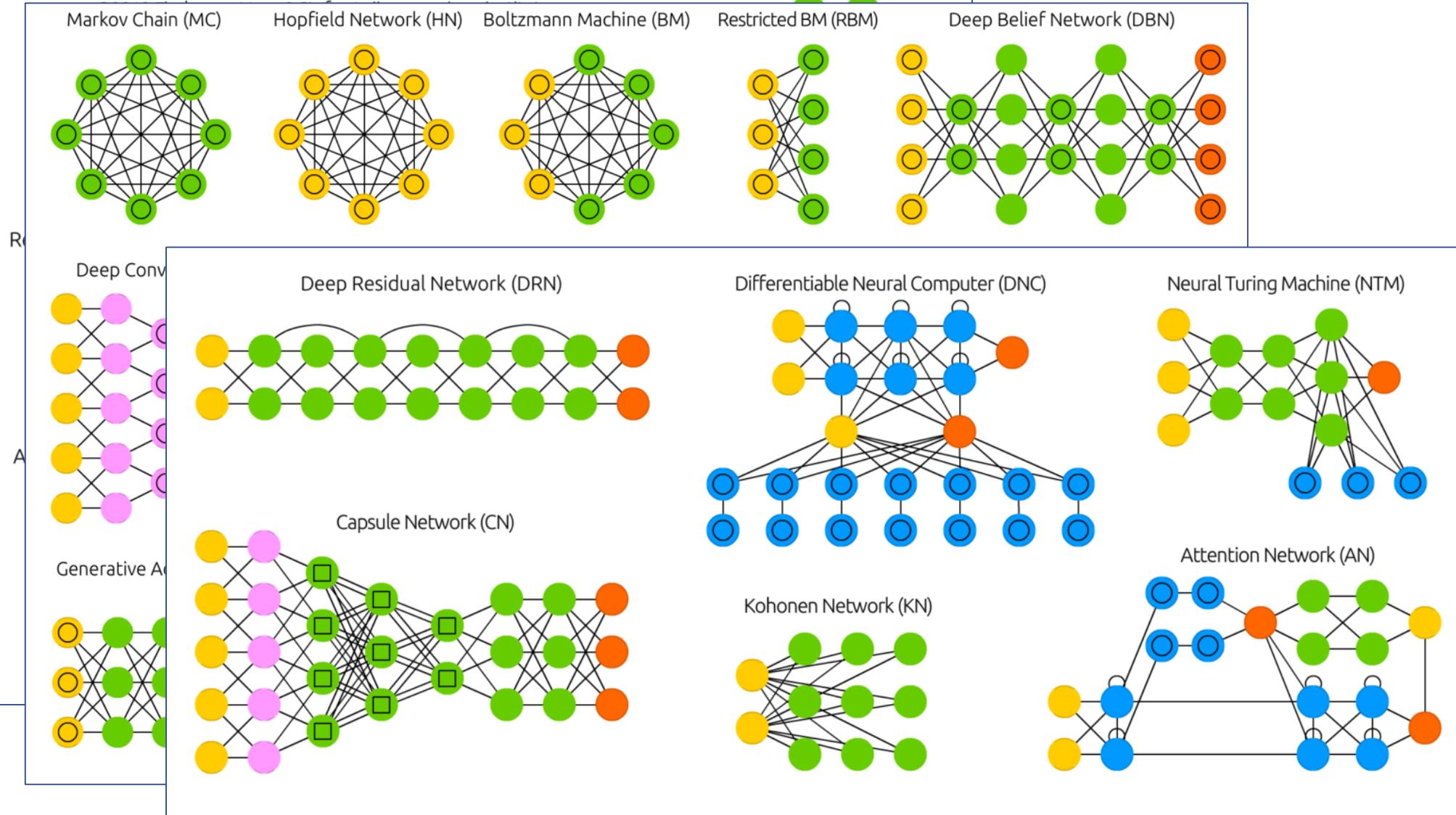


A mostly complete chart of

Neural Networks

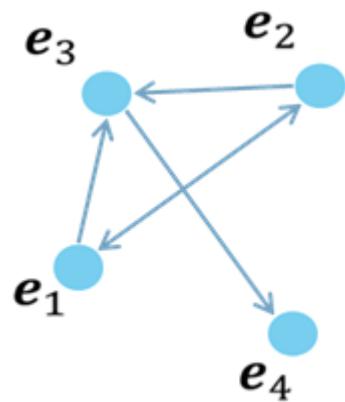
- Input Cell
- Backfed Input Cell
- △ Noisy Input Cell
- Hidden Cell
- Probabilistic Hidden Cell
- △ Spiking Hidden Cell
- Capsule Cell
- Output Cell
- Match Input Output Cell
- Recurrent Cell
- Memory Cell
- △ Gated Memory Cell
- Kernel
- Convolution or Pool

Deep Feed Forward (DFF)

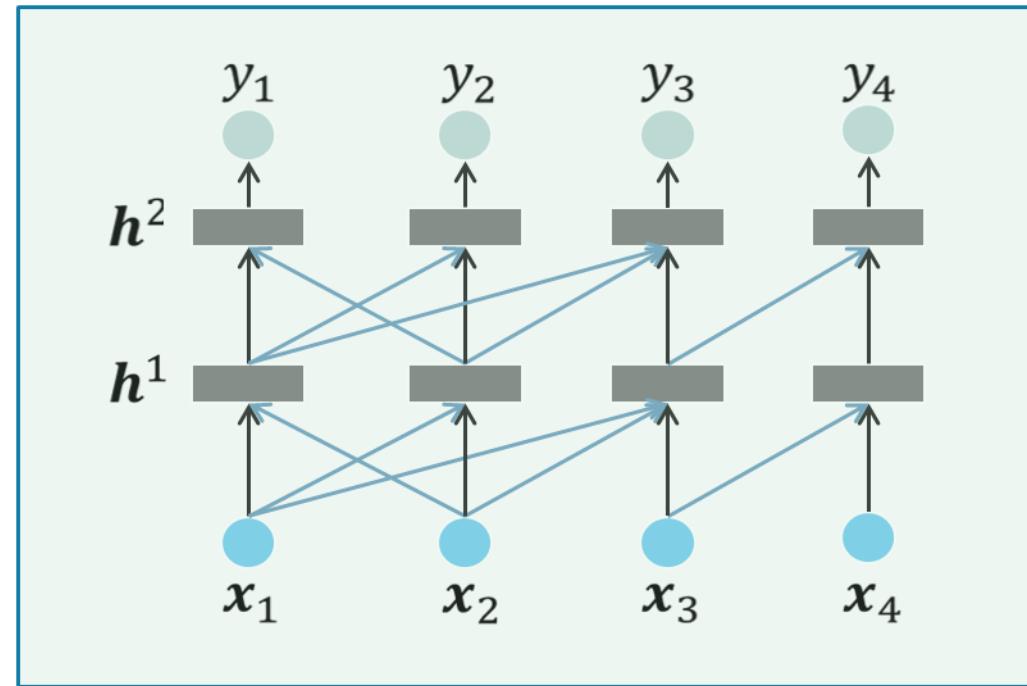


Column Networks inspired from columnar structure of brain

Relation graph

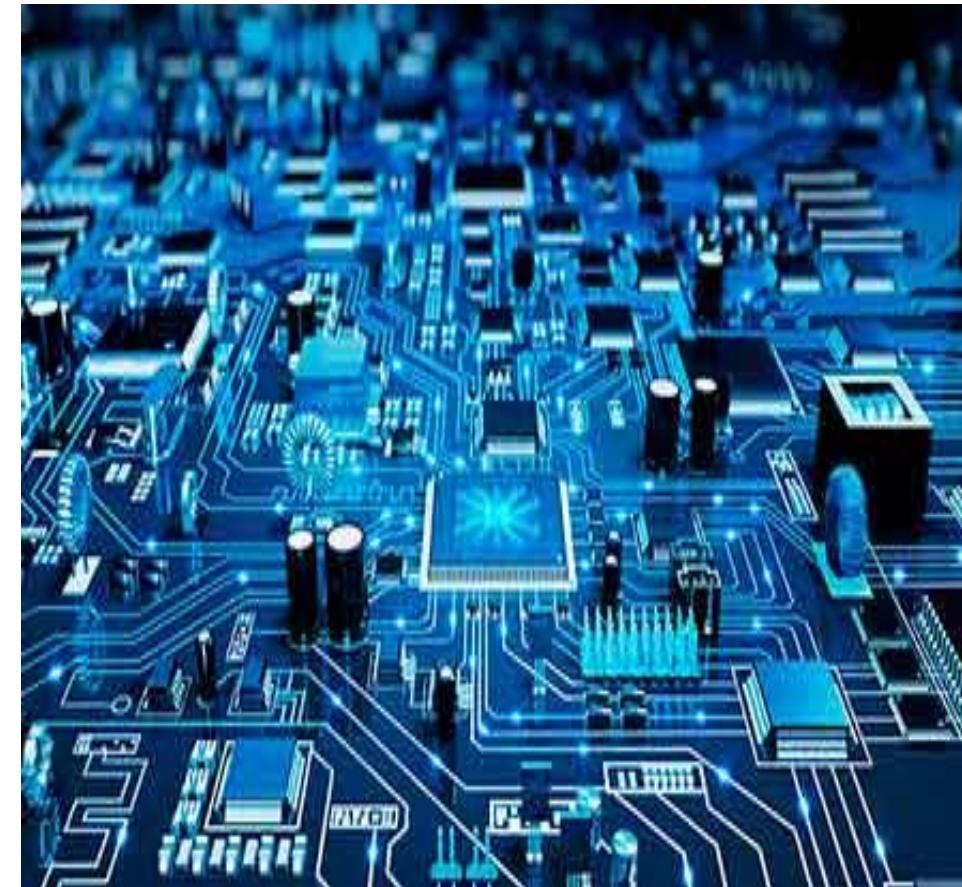


Generalized message passing



Deep learning vs electronics

- Neuron as feature detector → SENSOR, FILTER
- Multiplicative gates → AND gate, Transistor, Resistor
- Attention mechanism → SWITCH gate
- Memory + forgetting → Capacitor + leakage
- Skip-connection → Short circuit
- Computational graph → Circuit
- Compositionality → Modular design



2015 - Attention is a key for efficiency

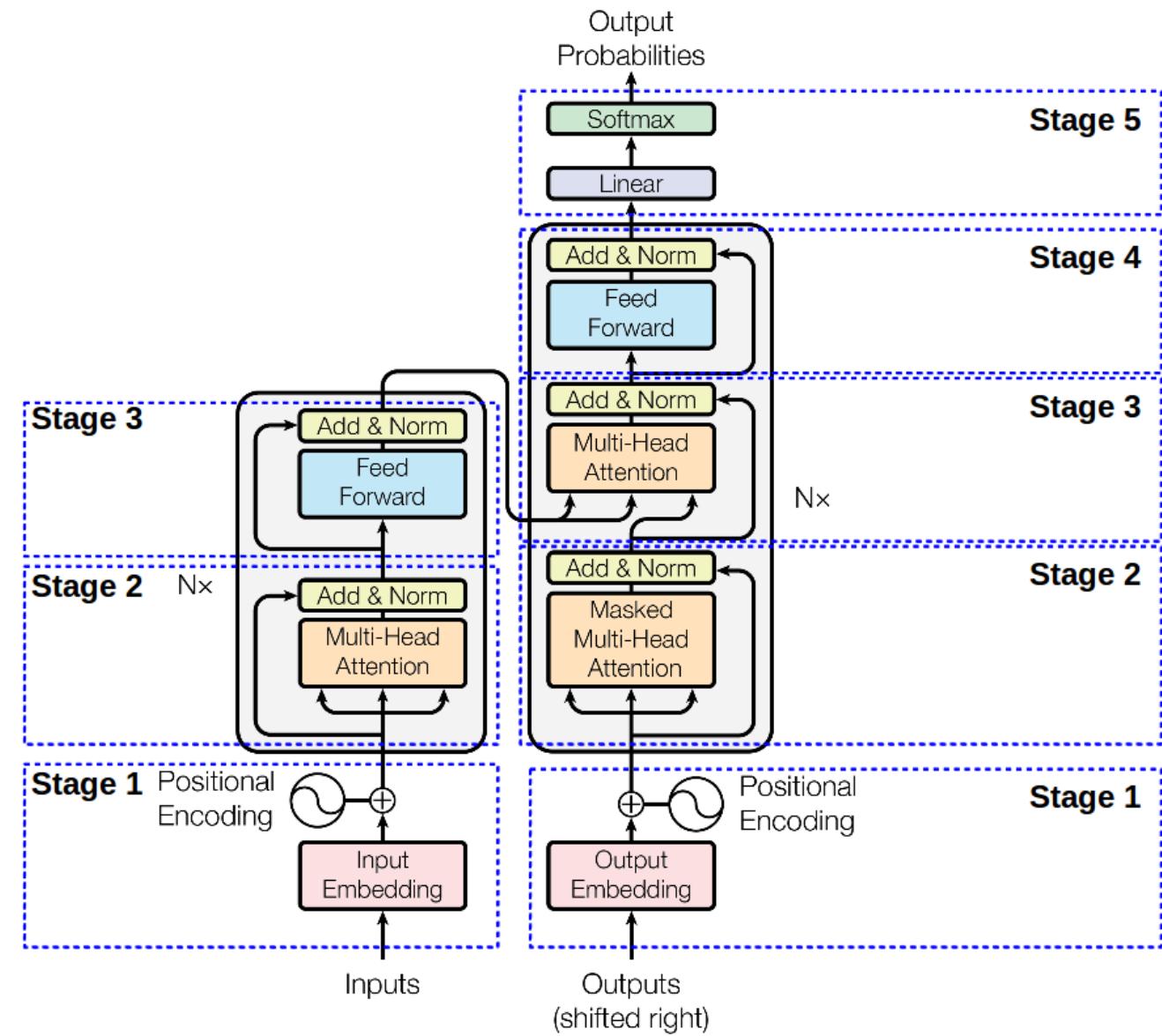
- Visual attention in human: Focus on specific parts of visual inputs to compute the adequate responses.
- Examples:
 - We focus on objects rather than the background of an image.
 - We skim text by looking at important words.
- In neural computation, we need to select the most relevance piece of information and ignore all other parts



Photo: programmersought

2017 - Transformer

- Tokenization
- Token encoding
- Position coding
- Sparsity
- Exploit spatio-temporal structure



Transformer: Key ideas

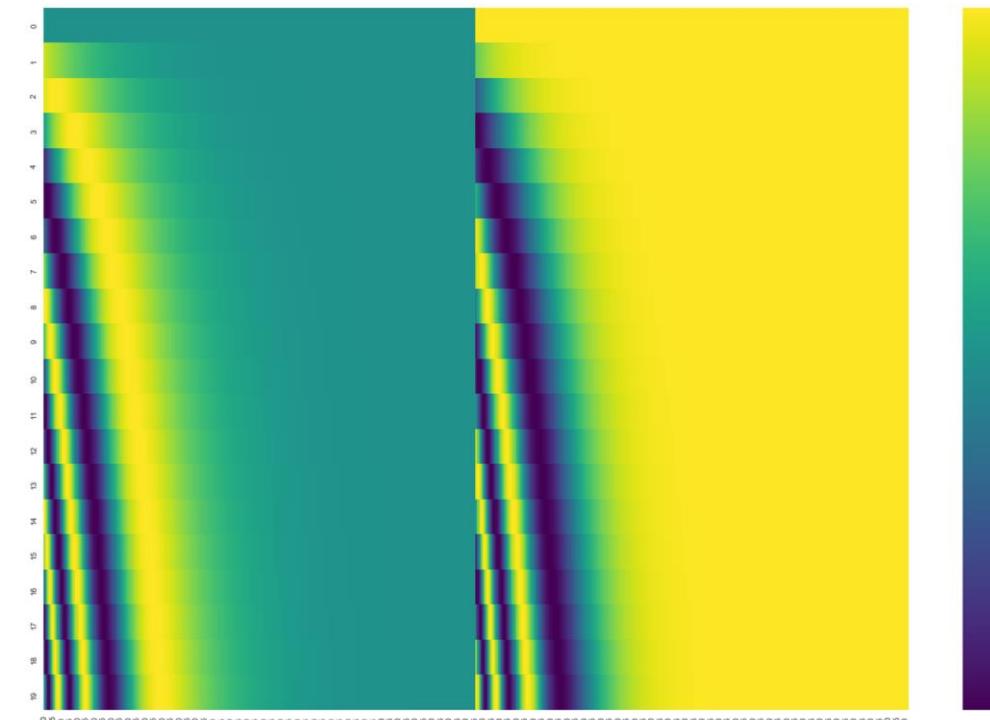
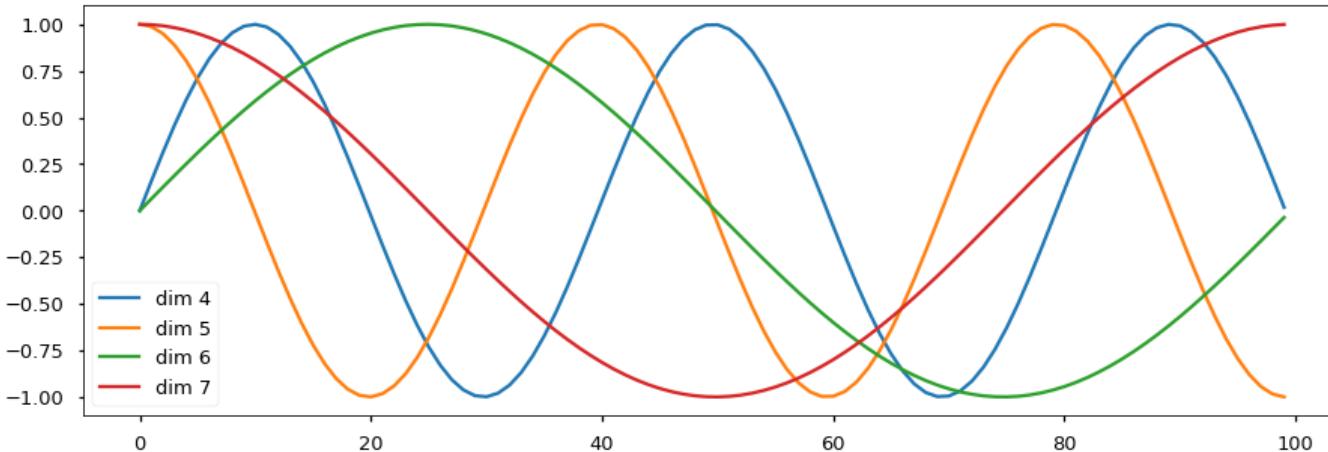
- Use self-similarity to refine token's representation (embedding).
 - “**June** is **happy**” -> June is represented as a person’s name.
 - Hidden contexts are borrowed from other sentences that share tokens/motifs/patterns, e.g., “She is happy”, “Her name is June”, etc.
 - Akin to retrieval: matching **query** to **key**.
- Context is simply other tokens co-occurring in the same text segment.
 - Related to “co-location”.
 - How big is context? → Small window, a sentence, a paragraph, the whole doc.
 - What is about relative position? → Position coding.

Positional encoding

- The Transformer relaxes the sequentiality of data
 - Positional encoding to embed sequential order in model

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d_{\text{model}}})$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d_{\text{model}}})$$



Large Language Models are Transformer trained on the entire Internet!



Yann LeCun • 3rd+

VP & Chief AI Scientist at Meta

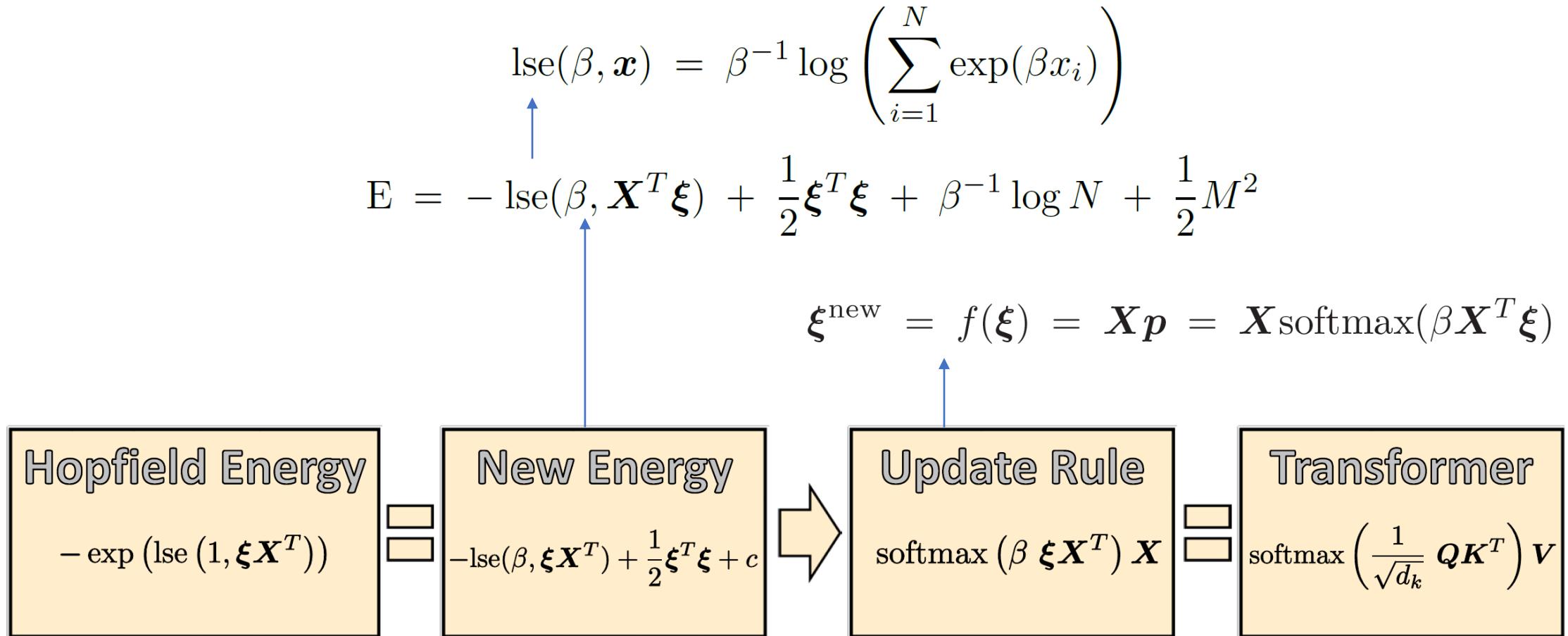
11mo •

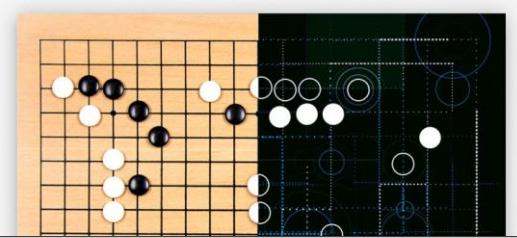
[+ Follow](#) ...

The weak reasoning abilities of LLMs are partially compensated by their large associative memory capacity.

They are a bit like students who have learned the material by rote but haven't really built deep mental models of the underlying reality.

Transformers are modern Hopfield net



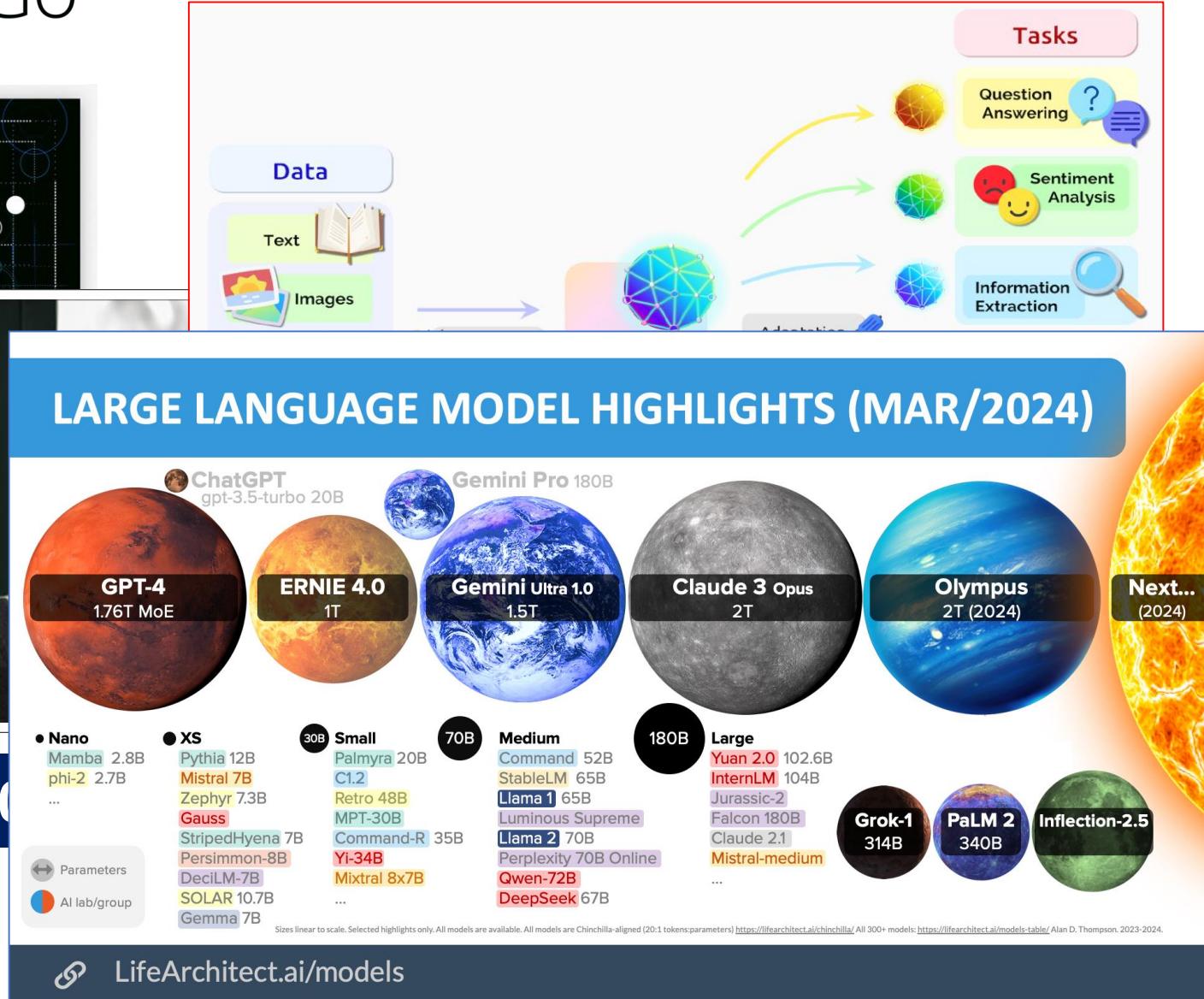


2012

Turing Awards 2012

12 years snapshot

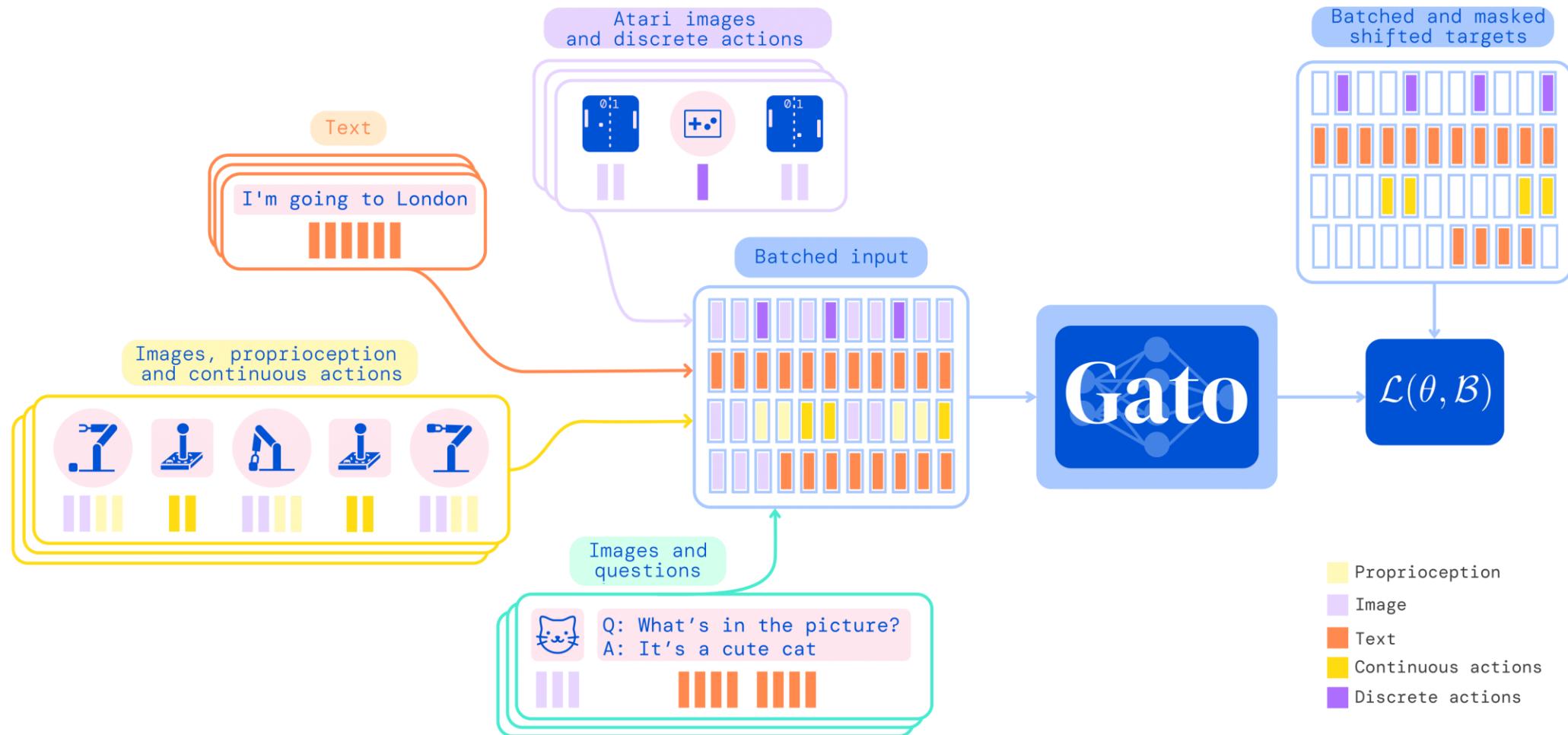
20/11/2024



2024

46

Convergence: One model for all – the case of Gato (2022)



Reed, Scott, Konrad Zolna, Emilio Parisotto, Sergio Gomez Colmenarejo, Alexander Novikov, Gabriel Barth-Maron, Mai Gimenez et al. "A generalist agent." *arXiv preprint arXiv:2205.06175* (2022).

Why one-model-for-all possible?

- The world is regular: Rules, patterns, motifs, grammars, recurrence
 - World models are learnable from data!
- Human brain gives an example
 - One brain, but capable of processing all modalities, doing plenty of tasks, and learning from different kind of training signals.
 - Thinking at high level is independent of input modalities and task-specific skills.
- Advances in modern AI:
 - Model flexibility
 - Powerful training and inference machines
 - Smart tricks

Current Generative AI



GenAIs are
compression
engine

Prompting is conditioning
for the (preference-
guided) decompression.



GenAIs are
approximate
program database

Prompting is retrieving an
approximate program that
takes input and delivers
output.



GenAIs are
World Model

We can live entirely in
simulation!

Agenda



The context



Hopfield
networks



Boltzmann
machines

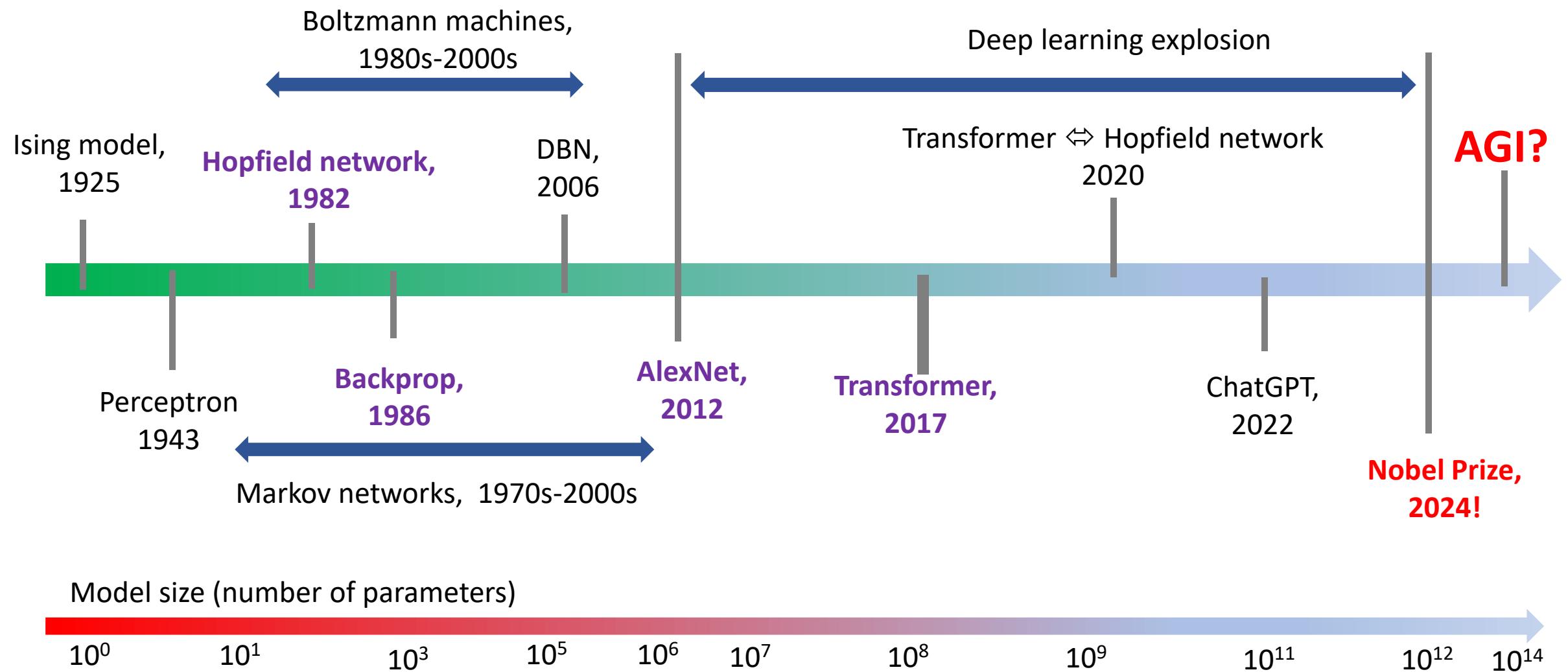


Deep learning

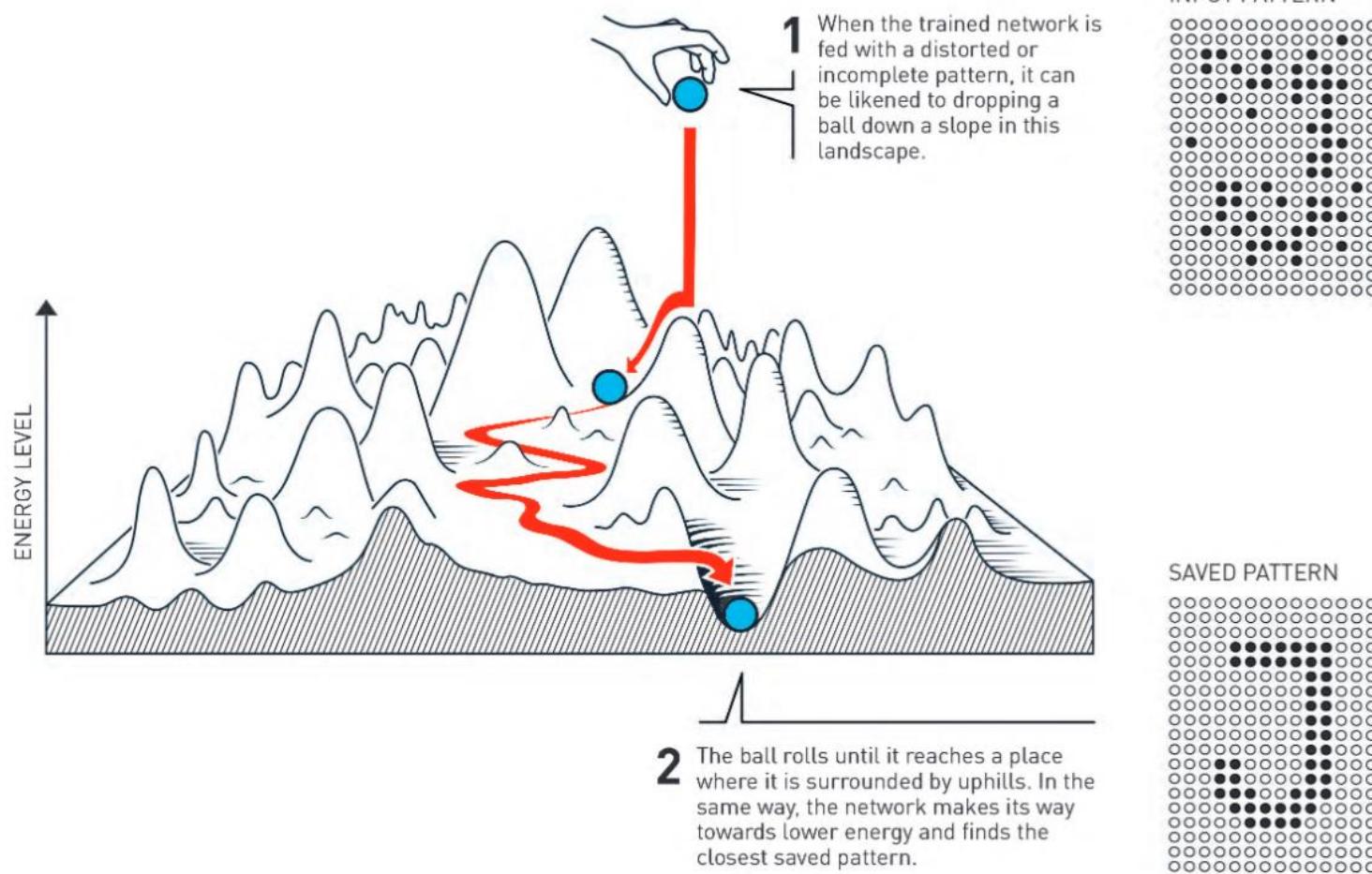


Discussion

A very brief timeline of 100 years



Memories are stored at attractors in a landscape



Recognition of the 2024 Nobel Prize in Physics

20/11/2024

The interdisciplinary nature of modern physics, crossing with:

- Computation
- Information theory
- Cognition.

Concepts foundational to AI stem from physics (energy landscapes and statistical mechanics).

- Reshaping the discipline's boundaries.
- Foundational work => practical AI systems.

Timeliness & innovation: AI has pervasive societal impact.

53

The bridges between physics, mind, and computer science

Physical principles can model cognitive processes

Memory as energy states

Learning as optimization.



Modern physics and AI treat information processing as intrinsic to the fabric of nature, akin to matter and energy.



**Information, Physics,
and Computation**

Marc Mézard
Andrea Montanari

Criticisms

AI belongs to computer science or technology, not physics.

Prioritizing applied technologies over fundamental discoveries.

Inadequate recognition of prior works

Prior to Hopfield network (1982): Lenz-Ising recurrent architecture (1925); Amari (1972), Nakano (1972).

Prior to Boltzmann machine (1985): Glauber (1963); Ivakhnenko & Lapa (1965); Edwards-Anderson (1975); Sherrington & Kirkpatrick (1975)

Prior to layer-wise training of DBN (2006): Ivakhnenko & Lapa (1965); Schmidhuber (1990, 1991)

Prior to backprop (1986): Amari & Saito (1967), Linnainmaa, 1970; Werbos (1982)

Prior to modern deep networks: Ivakhnenko's 1971

Looking into the future



Giorgio Parisi, 2021 Nobel in Physics
for complex systems
Solved Sherrington-Kirkpatrick
model (1979)

AI as a discovery tool: e.g., quantum mechanics, materials science, and complex systems.

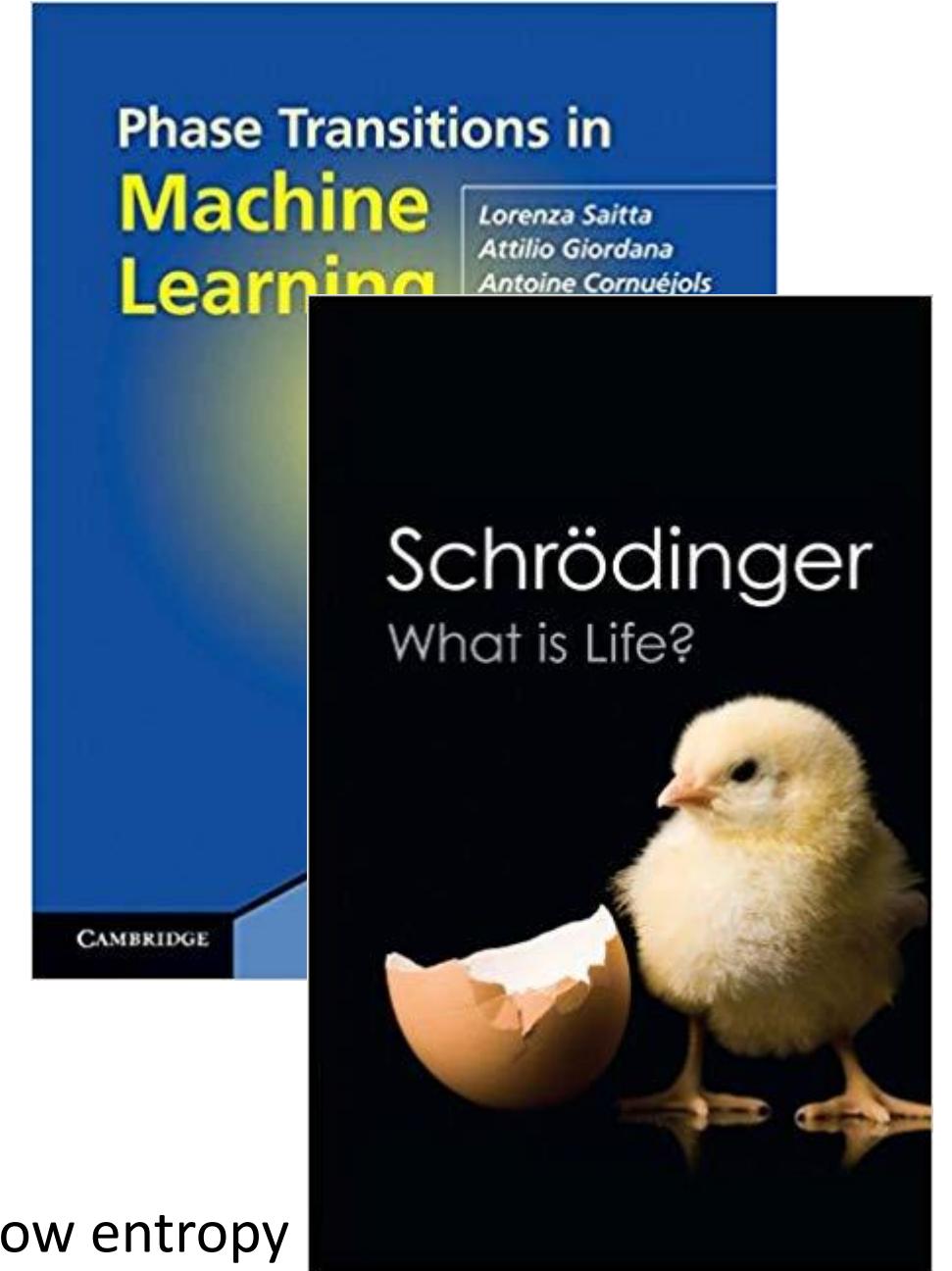
Interdisciplinary: Physics + AI + cognitive sciences for study of universe and human cognition.

Philosophical implications: Informational fabric => the nature of consciousness, intelligence, and the universe's computational structure.

Future-ready scientists: Technology + science.

AI as physics

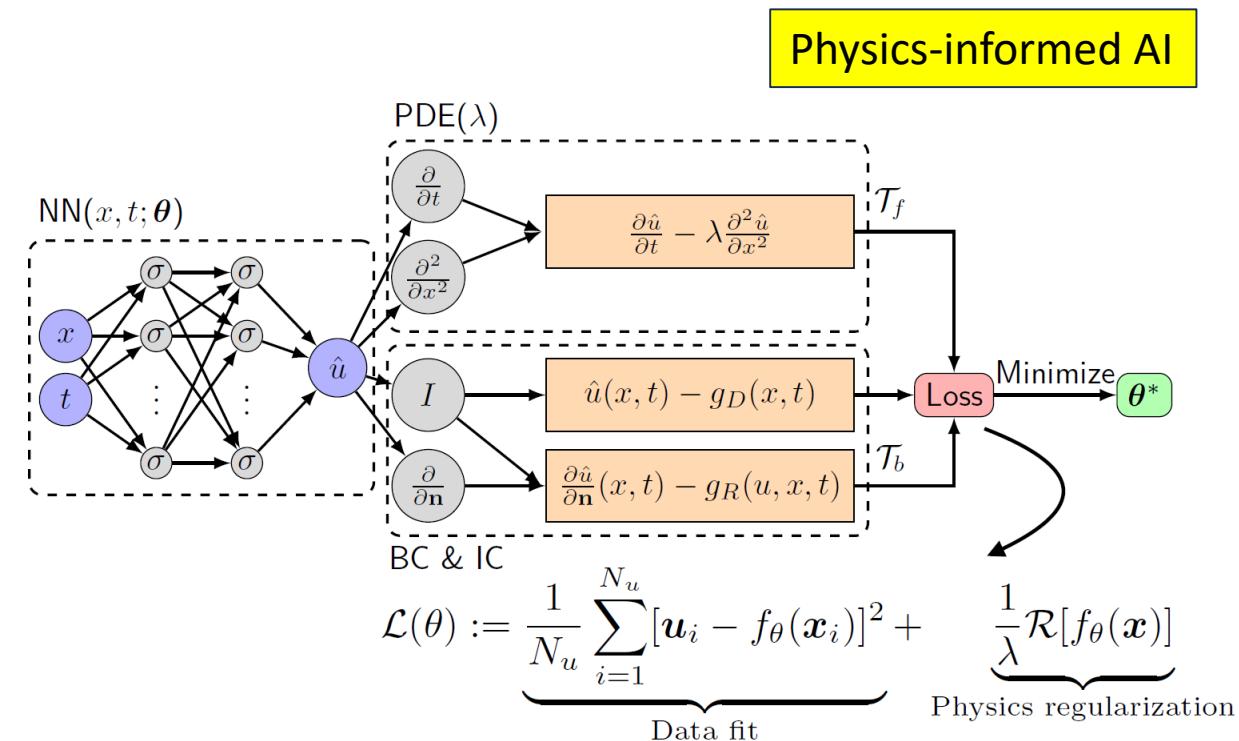
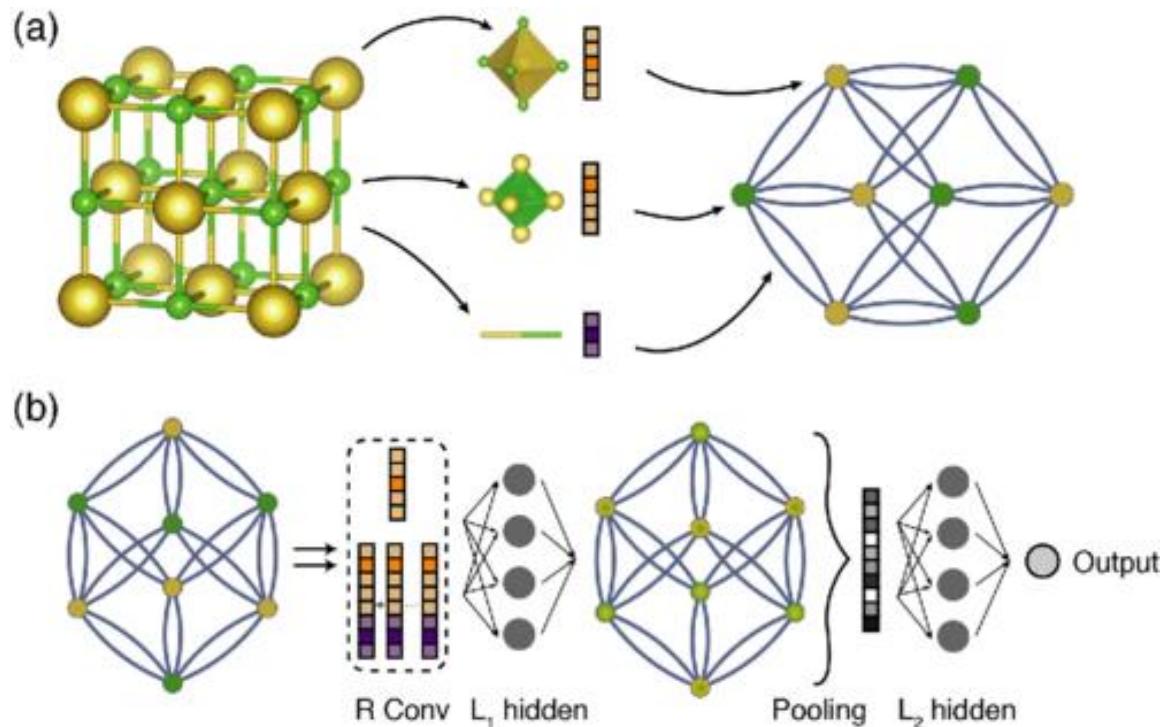
- Intelligence as self-organizing phenomena: reducing ignorance/entropy
- Neural networks as a statistical mechanical system
- Learning as variational optimization
- Inference free-energy minimization
- Phase transition may occur in AI systems
- Ultimate AI must solve the **consciousness problem**, which may require quantum physics (or a new physics)



Life has low entropy

AI for Physics

Xie, Tian, and Jeffrey C. Grossman. "Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties." *Physical review letters* 120.14 (2018): 145301.





Thank you!