2026 ACC Tutorial

Neurocomputation

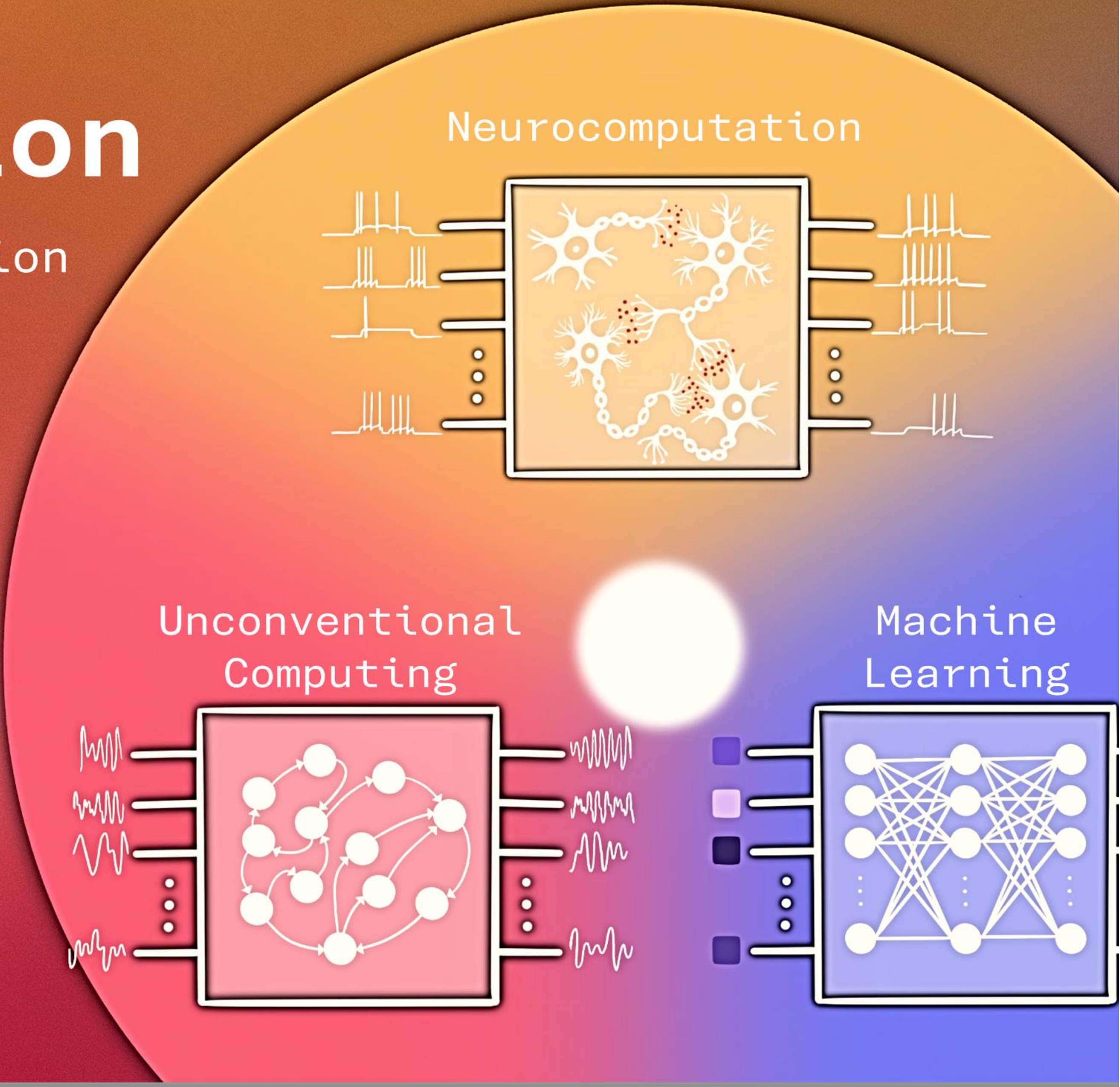
Learning, Dynamics, and Optimization

Francesco Bullo

Dmitry Krotov

Arthur Montanari

Adilson E. Motter



New Orleans, Louisiana

AI / DL → Control systems

WeB03 – Bridging Learning-Enabled Control with Modern Neural Network Verifier: A Unified Framework with α,β-CROWN

1:00 - 3:00 pm | Grand Salon 3

Deep learning–based controllers have gained substantial popularity due to their strong empirical performance. However, their deployment in safety-critical scenarios remains a major concern. In this tutorial, we bridge learning-based control with the modern neural network verifier α,β-CROWN, a state of-the-art method that scales via GPU-parallel symbolic linear bound propagation and branch-and-bound refinement.

AI / DL ← Control theory

WeC03 - Neurocomputation: Learning, Dynamics, and Optimization

3:30 - 5:00 pm | Grand Salon 3

Recent advances at the intersection of control theory and neuroscience have revealed novel mechanisms by which dynamical systems perform computation. These advances encompass a wide range of conceptual and computational ideas, used for model learning and training, memory retrieval, data-driven control, and optimization. This tutorial will highlight neuro-inspired approaches to computation that aim to improve scalability, robustness, and energy efficiency across such tasks, bridging the gap between artificial and natural systems.

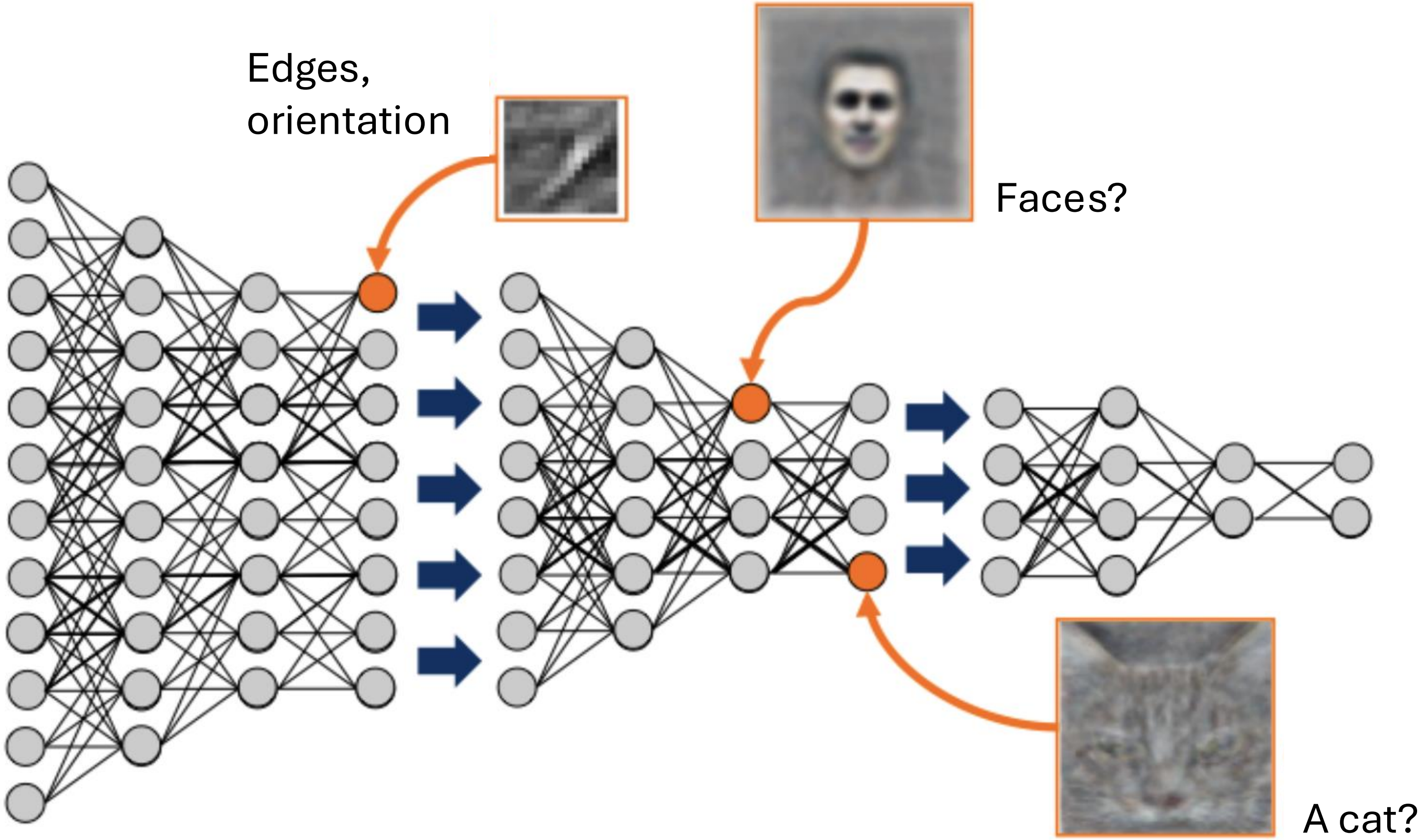
arXiv:2604.05042v1 [cs.LG] 6 Apr 2026

Energy-Based Dynamical Models for Neurocomputation, Learning, and Optimization

Arthur N. Montanari, Francesco Bullo, Dmitry Krotov, and Adilson E. Motter

Big challenges in AI models

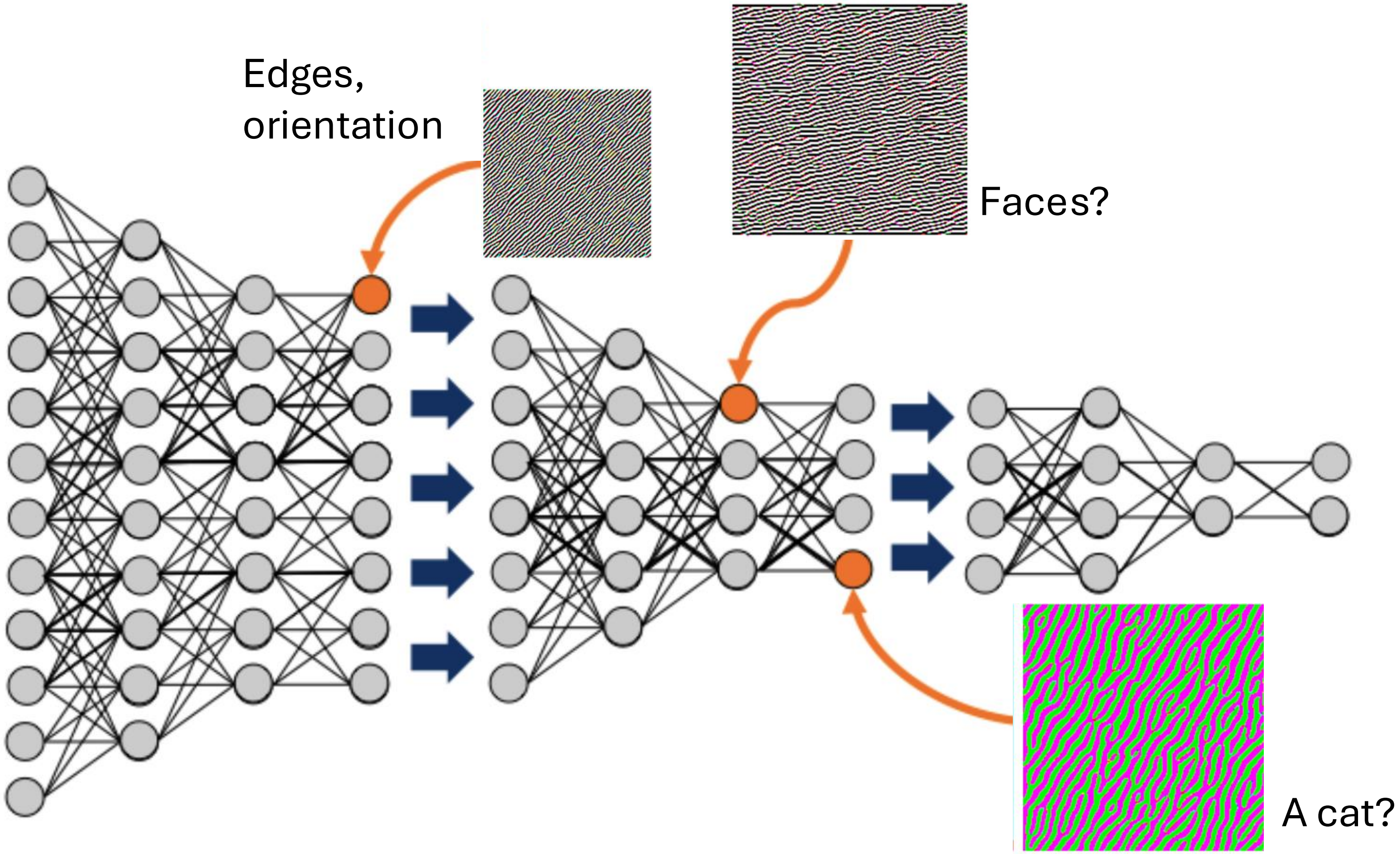
Interpretability?



C Olah, A Mordvintsev, L Schubert. *Distill* (2017).
https://distill.pub/2017/feature-visualization/

Big challenges in AI models

Interpretability?

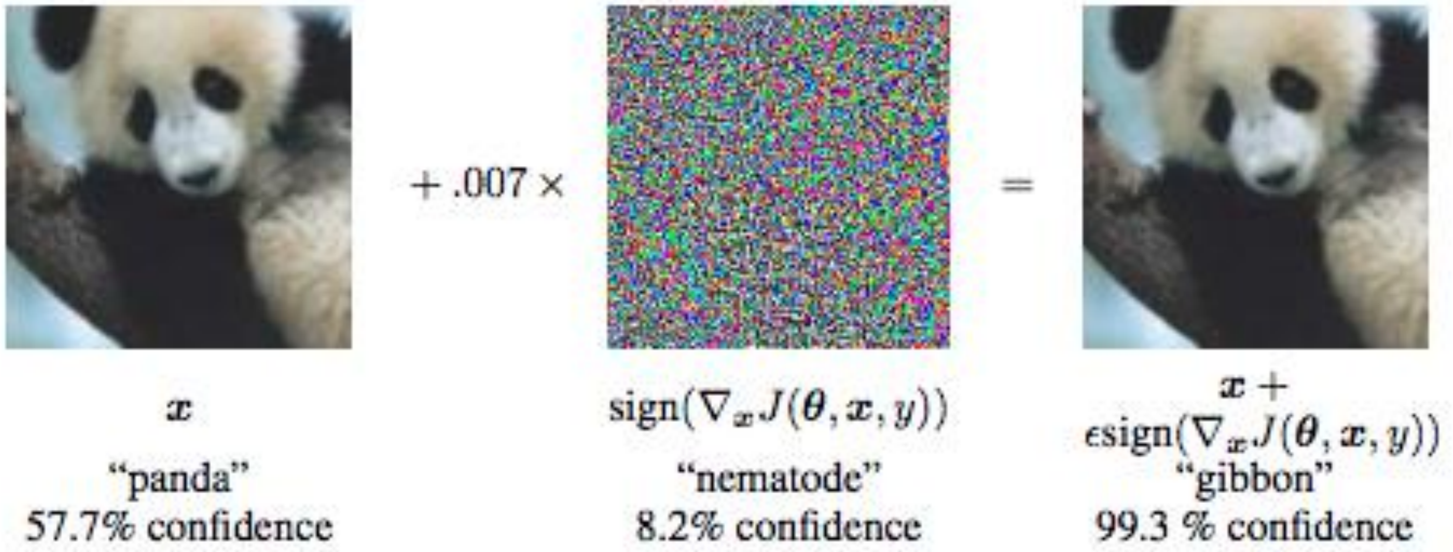


C Olah, A Mordvintsev, L Schubert. *Distill* (2017).
https://distill.pub/2017/feature-visualization/

Big challenges in AI models

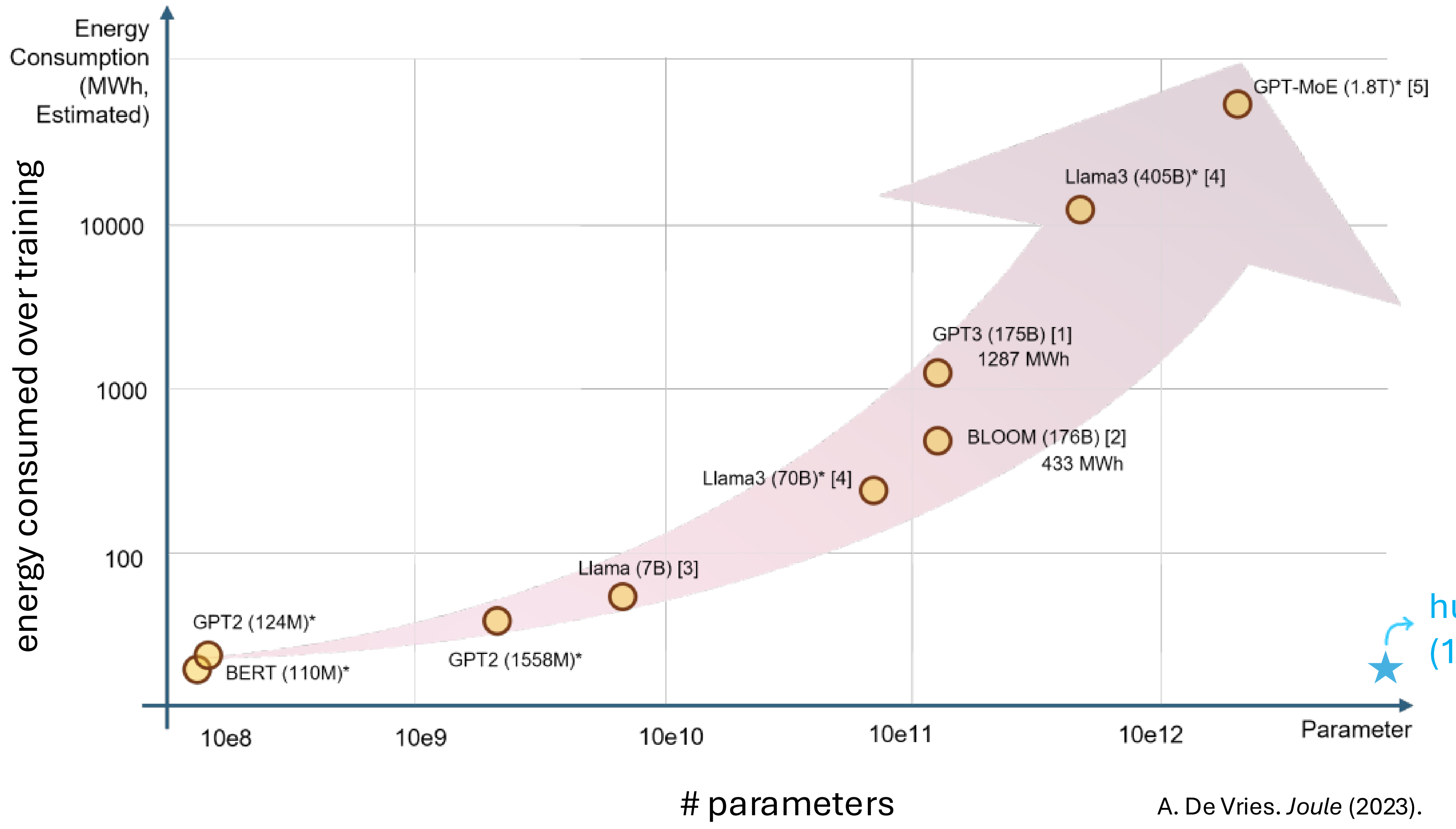
Interpretability?
Reliability?

J Su, DV Vargas, K Sakurai.
IEEE Trans. Evolutionary Computation (2019).



IJ Goodfellow, J Shlens, C Szegedy.
ICLR (2015).

Big challenges in AI models



Interpretability?

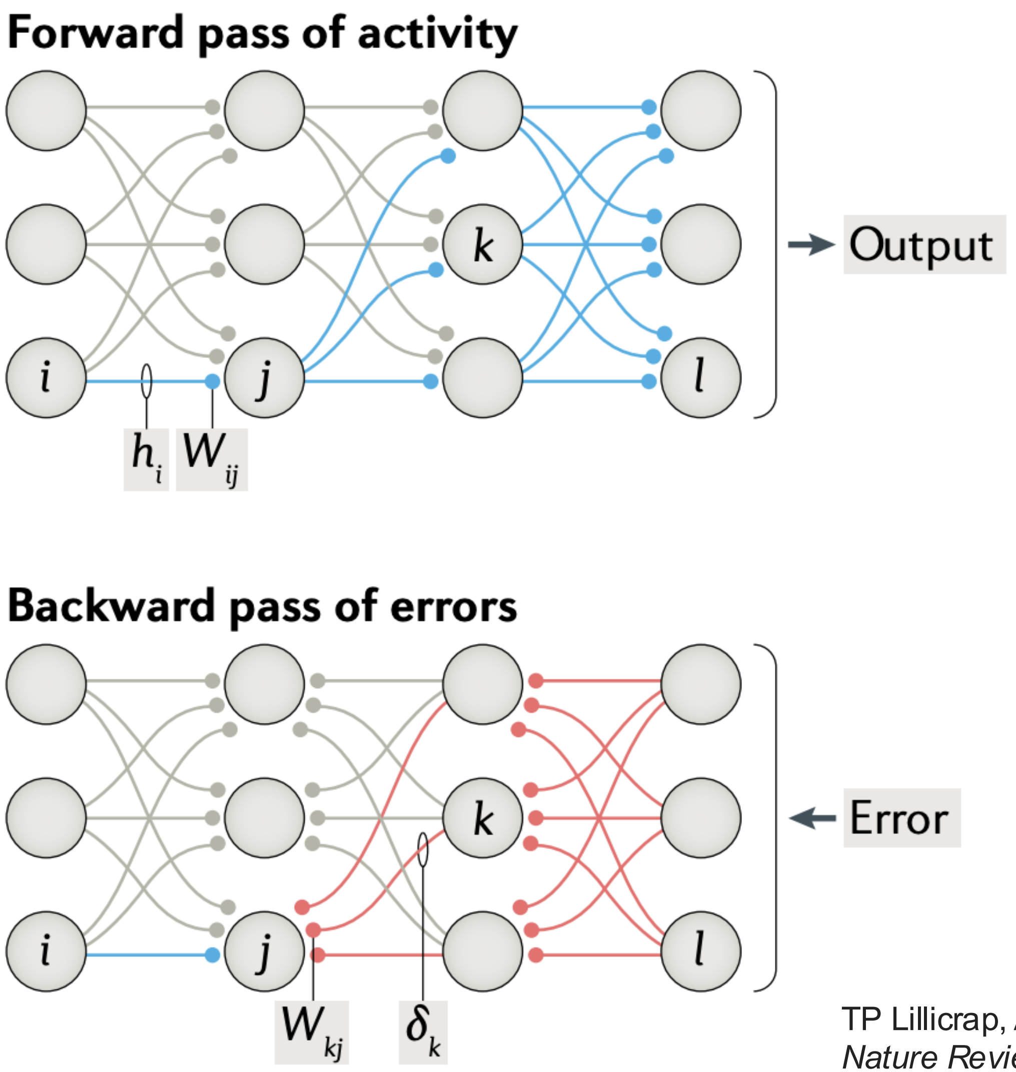
Reliability?

Scalability?

A. De Vries. *Joule* (2023).

Y Li, et al. *arXiv:2409.11416* (2024).

Big challenges in AI models



Interpretability?
Reliability?
Scalability?
Biological plausibility?

Global error propagation?
Separate forward and backward phases?
Symmetric connections?

TP Lillicrap, A Santoro, L Marris, CJ Akerman, G Hinton. *Nature Reviews Neuroscience* (2020).

ACC 2026 New Orleans

Neurocomputation:
Learning, Dynamics, and Optimization

3:30 - 4:00 pm. **Arthur N. Montanari**, Northwestern University
Recurrent neural networks and oscillator models for learning and optimization.

4:00 - 4:30 pm. **Dmitry Krotov,** Dynamical Mind
Dense associative memory for novel AI architectures.

4:30 - 5:00 pm. **Francesco Bullo,** UC Santa Barbara
Positive competitive neural networks for sparse reconstruction.

Building a recurrent neural network for...

Interpretability?

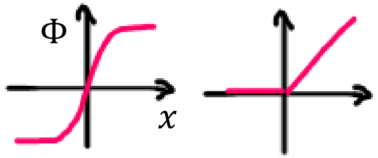
Reliability?

Scalability?

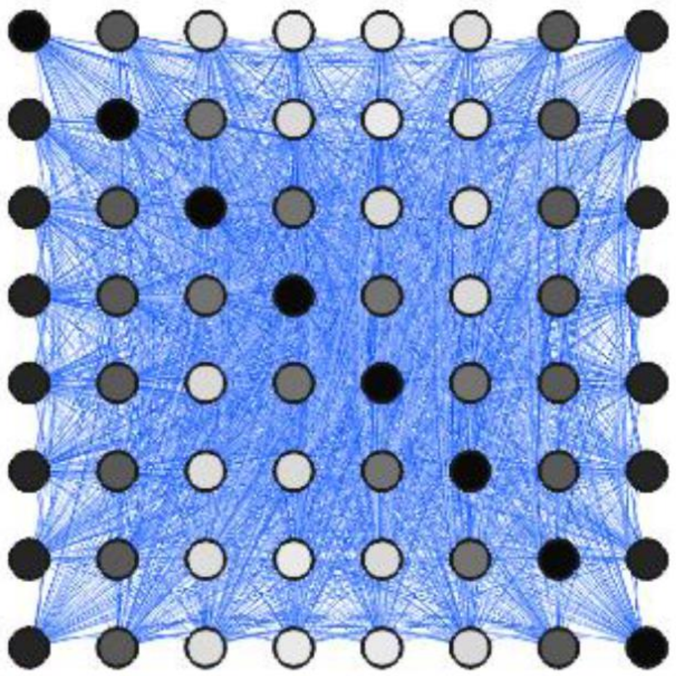
Biological plausibility?

Recurrent neural network

$$\tau \dot{x}_i = -d_i x_i + \sum_{j=1}^{N} W_{ij}\Phi(x_j) + B_{ij}u_j$$

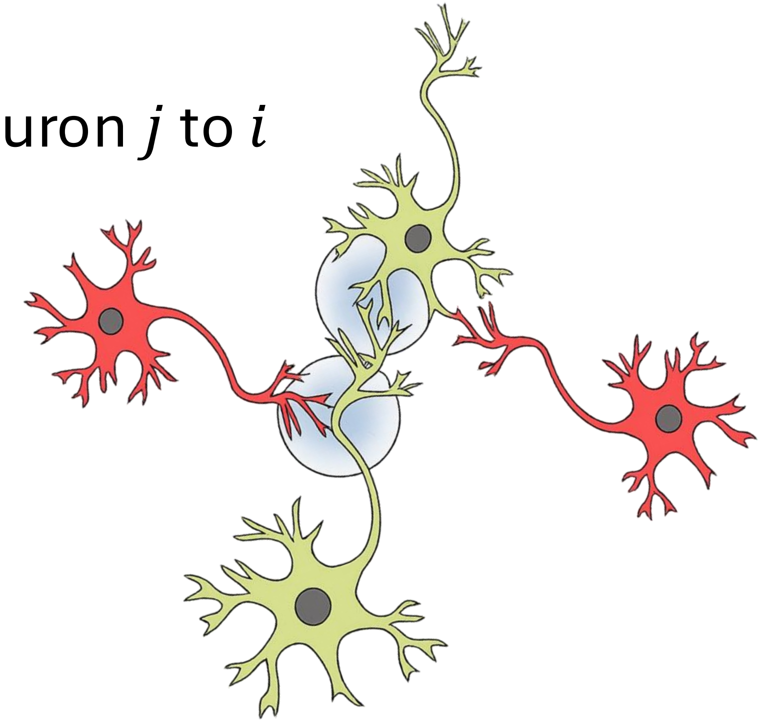


Interpretability?
Reliability?
Scalability?
Biological plausibility?



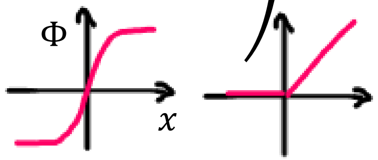
W_{ij} → synaptic weight from neuron j to i

x_i → neuron activity

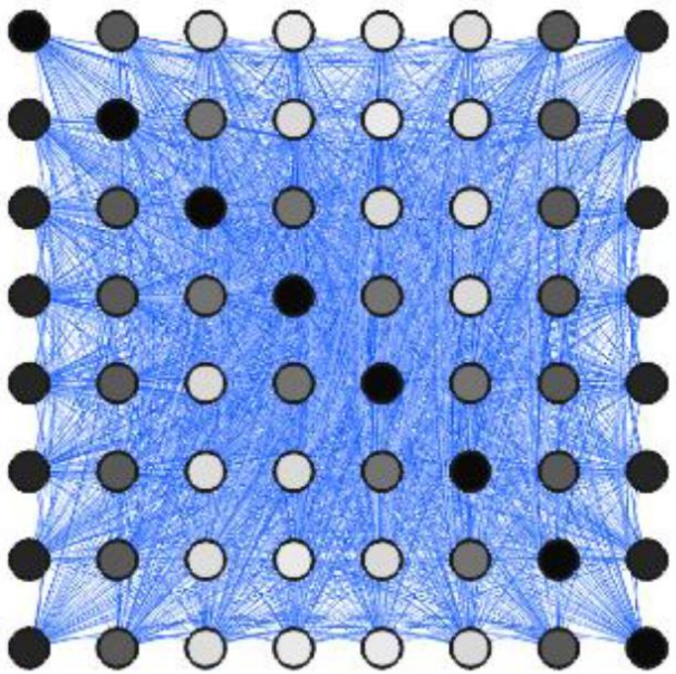


Recurrent neural network

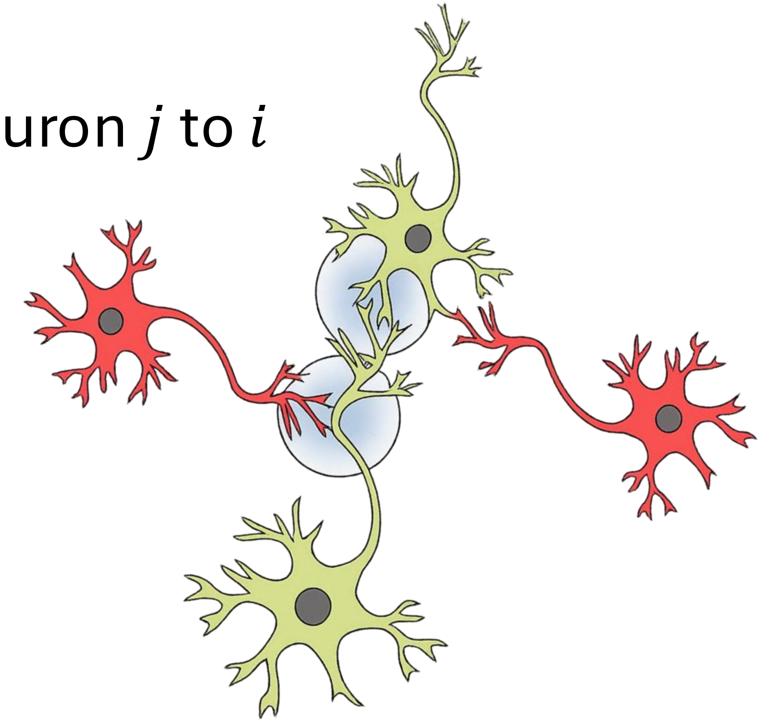
$$\tau \dot{x}_i = -d_i x_i + \Phi\left(\sum_{j=1}^{N} W_{ij} x_j + B_{ij} u_j\right)$$



Interpretability?
Reliability?
Scalability?
Biological plausibility?

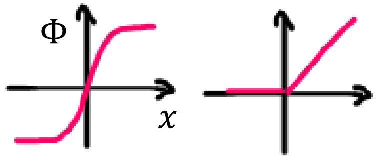


W_{ij} → synaptic weight from neuron j to i
x_i → neuron activity

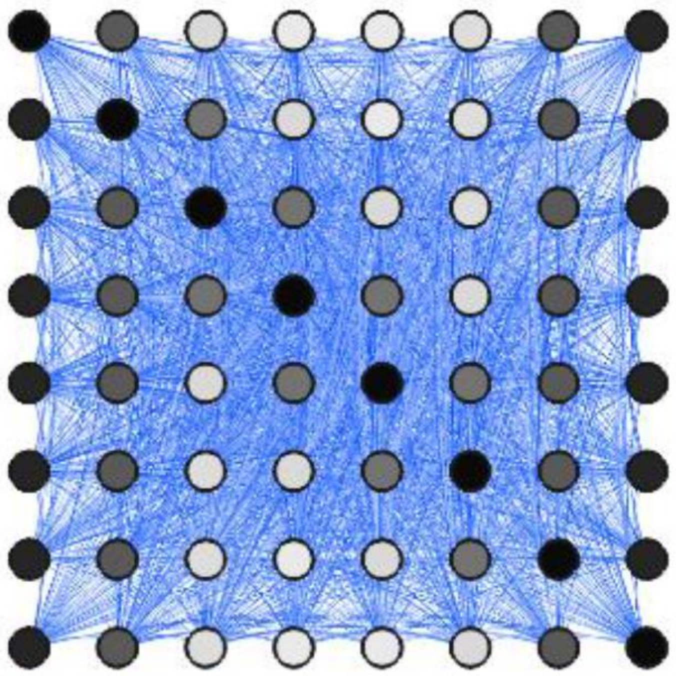


Recurrent neural network

$$\tau \dot{x}_i = -d_i x_i + \sum_{j=1}^{N} W_{ij}\Phi(x_j) + B_{ij}u_j$$



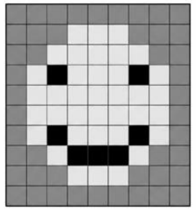
Interpretability?
Reliability?
Scalability?
Biological plausibility?



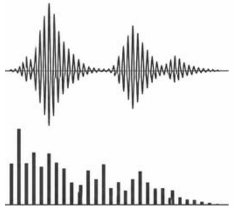
W_{ij} → synaptic weight from neuron j to i

x_i → neuron activity

$\boldsymbol{x} \in \mathbb{R}^N$ → image, sound, text, any data

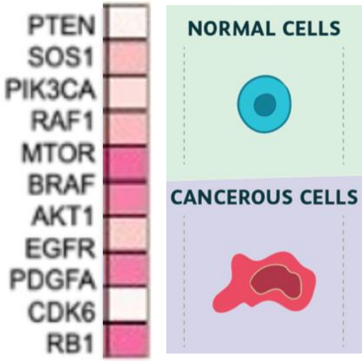


each neuron is a pixel



each neuron is a frequency coefficient

words are embedded as vectors, and each neuron represents a component



Recurrent neural network

$$\tau \dot{x}_i = -d_i x_i + \sum_{j=1}^{N} W_{ij} \Phi(x_j) + B_{ij} u_j$$

Nonlinear control system

$$\begin{cases} \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) \\ \boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}) \end{cases}$$

Interpretability?

Reliability?

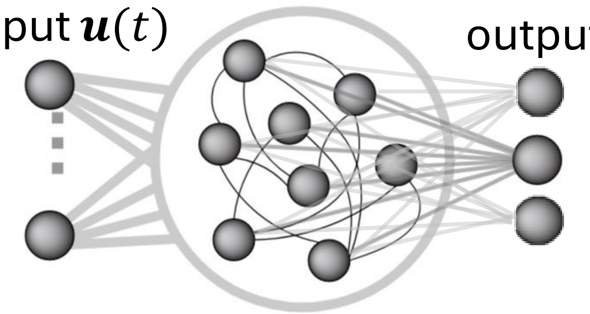
Scalability?

Biological plausibility?

hidden states $\boldsymbol{x}(t)$

input $\boldsymbol{u}(t)$

output $\boldsymbol{y}(t) = \Phi_y(W_y \boldsymbol{x}(t) + \boldsymbol{b}_y)$



Recurrent neural network

$$\tau \dot{x}_i = -d_i x_i + \sum_{j=1}^{N} W_{ij} \Phi(x_j) + B_{ij} u_j$$

Nonlinear control system

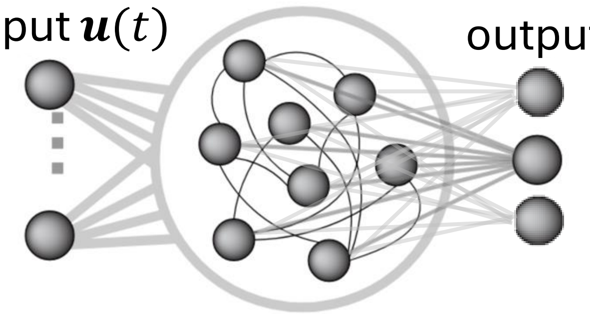
$$\begin{cases} \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) \\ \boldsymbol{y} = \boldsymbol{h}(\boldsymbol{x}) \end{cases}$$

Interpretability?
Reliability?
Scalability?
Biological plausibility?

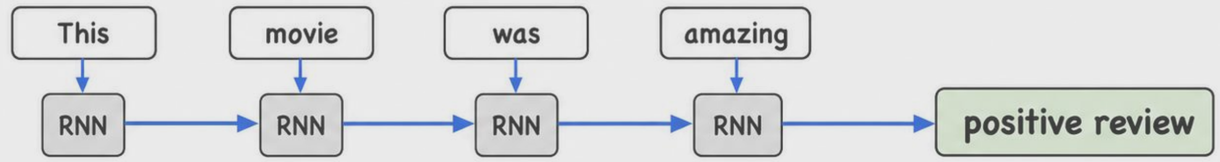
hidden states $\boldsymbol{x}(t)$

input $\boldsymbol{u}(t)$

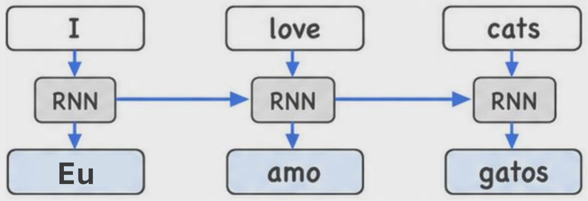
output $\boldsymbol{y}(t) = \Phi_y(W_y \boldsymbol{x}(t) + \boldsymbol{b}_y)$



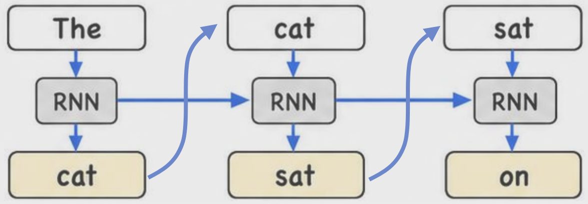
One-to-many classification
(e.g., sentiment analysis)



Many-to-many classification
(e.g., text translation)



Prediction
(e.g., text generation)



Where did they come from?

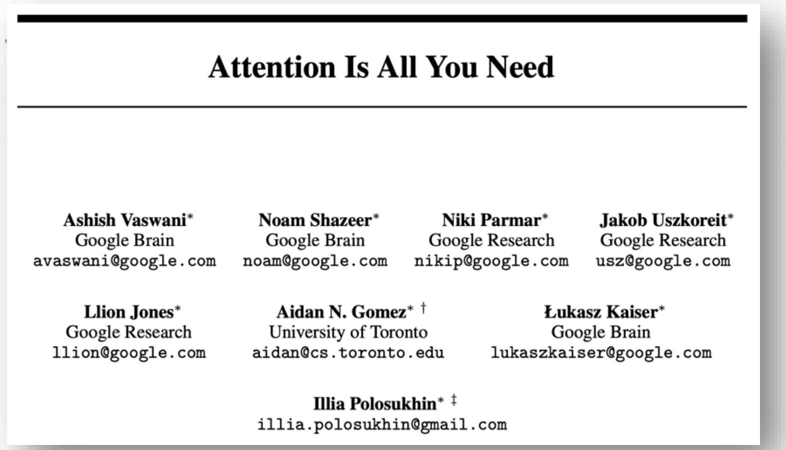
90s and before: neural networks, RBMs, Hopfield model

2000s: CNNs and RNNs

2010s: DNNs

2020s: transformers

???



Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

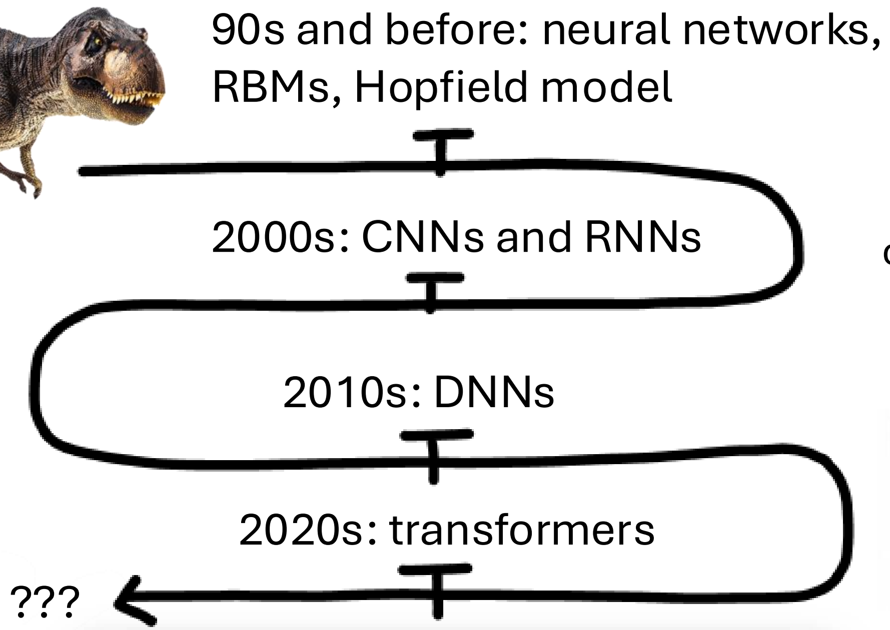
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

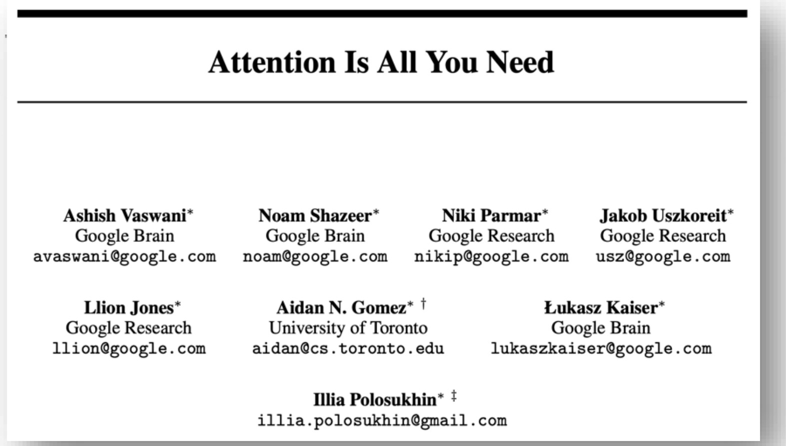
Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Interpretability?
Reliability?
Scalability?
Biological plausibility?

And where do we go?





Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

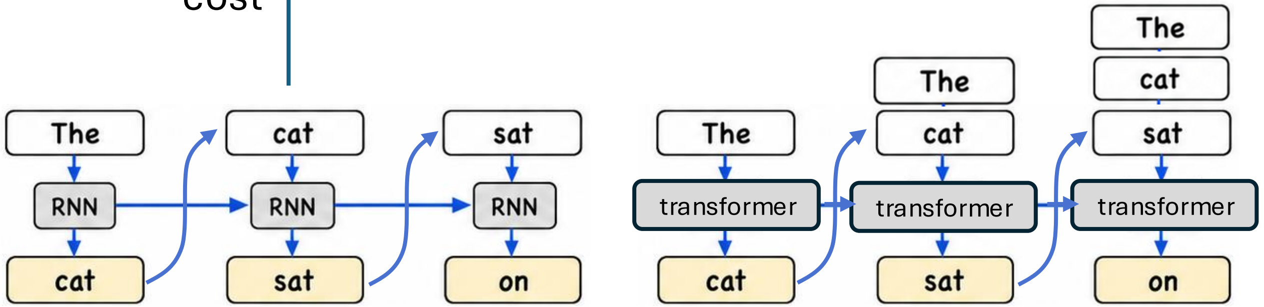
	RNNs	Transformers
memory cost	$\mathcal{O}(1)$	$\mathcal{O}(T)$
computational cost	$\mathcal{O}(T)$	$\mathcal{O}(T^2)$

Interpretability?

Reliability?

Scalability?

Biological plausibility?



And where do we go?

90s and before: neural networks, RBMs, Hopfield model

2000s: CNNs and RNNs

2010s: DNNs

2020s: transformers

?

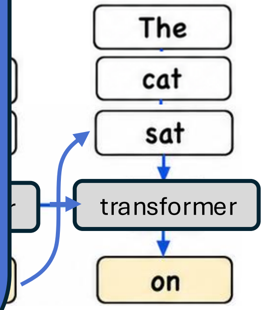
	RNNs	Transformers
memory cost	$\mathcal{O}(1)$	$\mathcal{O}(T)$

Interpretability?

Reliability?

Scalability?

Plausibility?



We can derive minimal [RNN models] that (1) use fewer parameters than their traditional counterparts, (2) are fully parallelizable during training, and (3) achieve surprisingly competitive performance on a range of tasks, rivalling recent models including Transformers.

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com
Noam Shazeer* Google Brain noam@google.com
Niki Parmar* Google Research nikip@google.com
Jakob Uszkoreit* Google Research usz@google.com
Llion Jones* Google Research llion@google.com
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu
Łukasz Kaiser* Google Brain lukaszkaiser@google.com
Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Hopfield Networks is All You Need

Hubert Ramsauer* Bernhard Schäfl* Johannes Lehner* Philipp Seidl* ...s Holzleitner* ...vid Kreil† ...r* Sepp Hoc...

Dense Associative Memory for Pattern Recognition

Dmitry Krotov
Simons Center for Systems Biology
Institute for Advanced Study
Princeton, USA
krotov@ias.edu

John J. Hopfield
Princeton Neuroscience Institute
Princeton University
Princeton, USA
hopfield@princeton.edu

Were RNNs All We Needed?

Leo Feng Mila – Université de Montréal & Borealis AI leo.feng@mila.quebec
Frederick Tung Borealis AI frederick.tung@borealisai.com
Mohamed Osama Ahmed Borealis AI mohamed.o.ahmed@borealisai.com
Yoshua Bengio Mila – Université de Montréal yoshua.bengio@mila.quebec
Hossein Hajimirsadeghi Borealis AI hossein.hajimirsadeghi@borealisai.com

Finding an energy (Lyapunov) function

Interpretability?
Reliability?
Scalability
Biological plausibility?

Continuous-time RNN

$$\tau \dot{x}_i = -d_i x_i + \sum_{j=1}^{N} W_{ij} \Phi(x_j) + B_{ij} u_j$$

Finding an energy (Lyapunov) function

Interpretability?

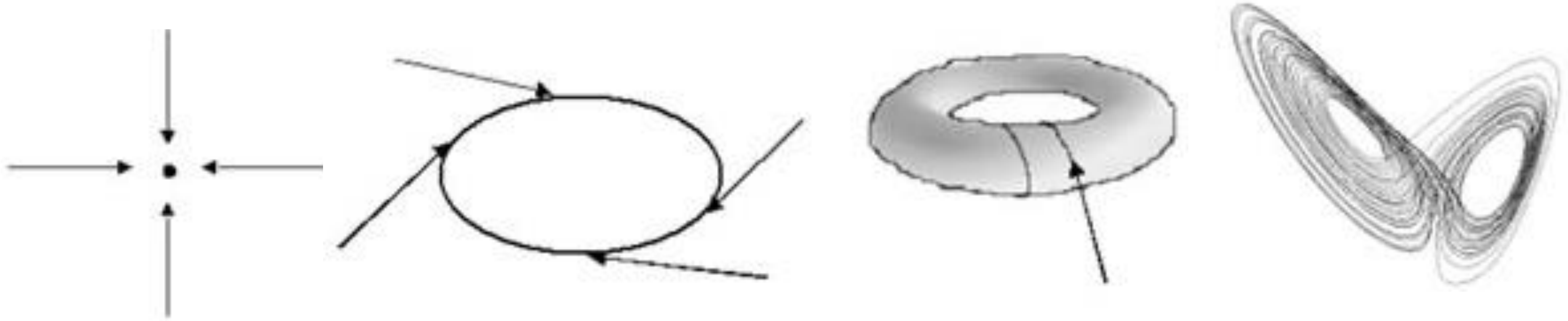
Reliability?

Scalability

Biological plausibility?

Continuous-time RNN

$$\tau \dot{\boldsymbol{x}} = -D\boldsymbol{x} + W\Phi(\boldsymbol{x}) + B\boldsymbol{u}$$



Finding an energy (Lyapunov) function

Interpretability?

Reliability?

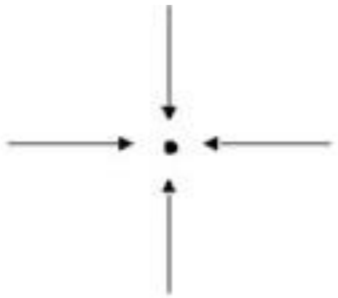
Scalability

Biological plausibility?

Continuous-time

Hopfield model

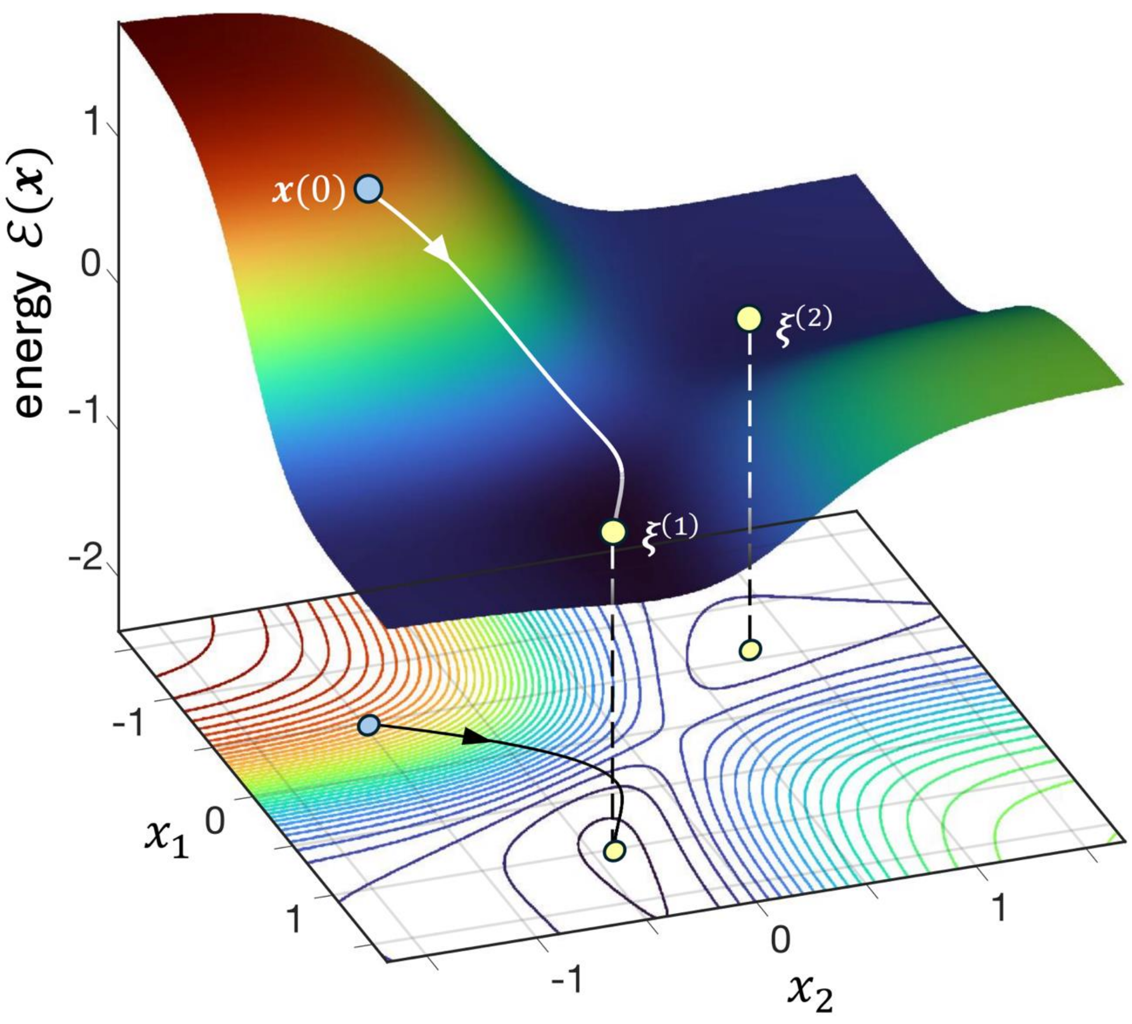
$$\tau\dot{\boldsymbol{x}} = -D\boldsymbol{x} + W\Phi(\boldsymbol{x}) + B\boldsymbol{u}$$



Theorem. If $W = W^T$ and Φ is continuously differentiable, then the system is described by the following energy function:

$$\mathcal{E}(\boldsymbol{x}) = -\frac{1}{2}\boldsymbol{\Phi}(\boldsymbol{x})^{\mathsf{T}} W \boldsymbol{\Phi}(\boldsymbol{x}) + (D\boldsymbol{x} - B\boldsymbol{u})^{\mathsf{T}}\boldsymbol{\Phi}(\boldsymbol{x}) - \sum_{i=1}^{N} d_i \int_0^{x_i} \Phi(w)\mathrm{d}w.$$

Finding an energy (Lyapunov) function



Interpretability?

Reliability?

Scalability

Biological plausibility?

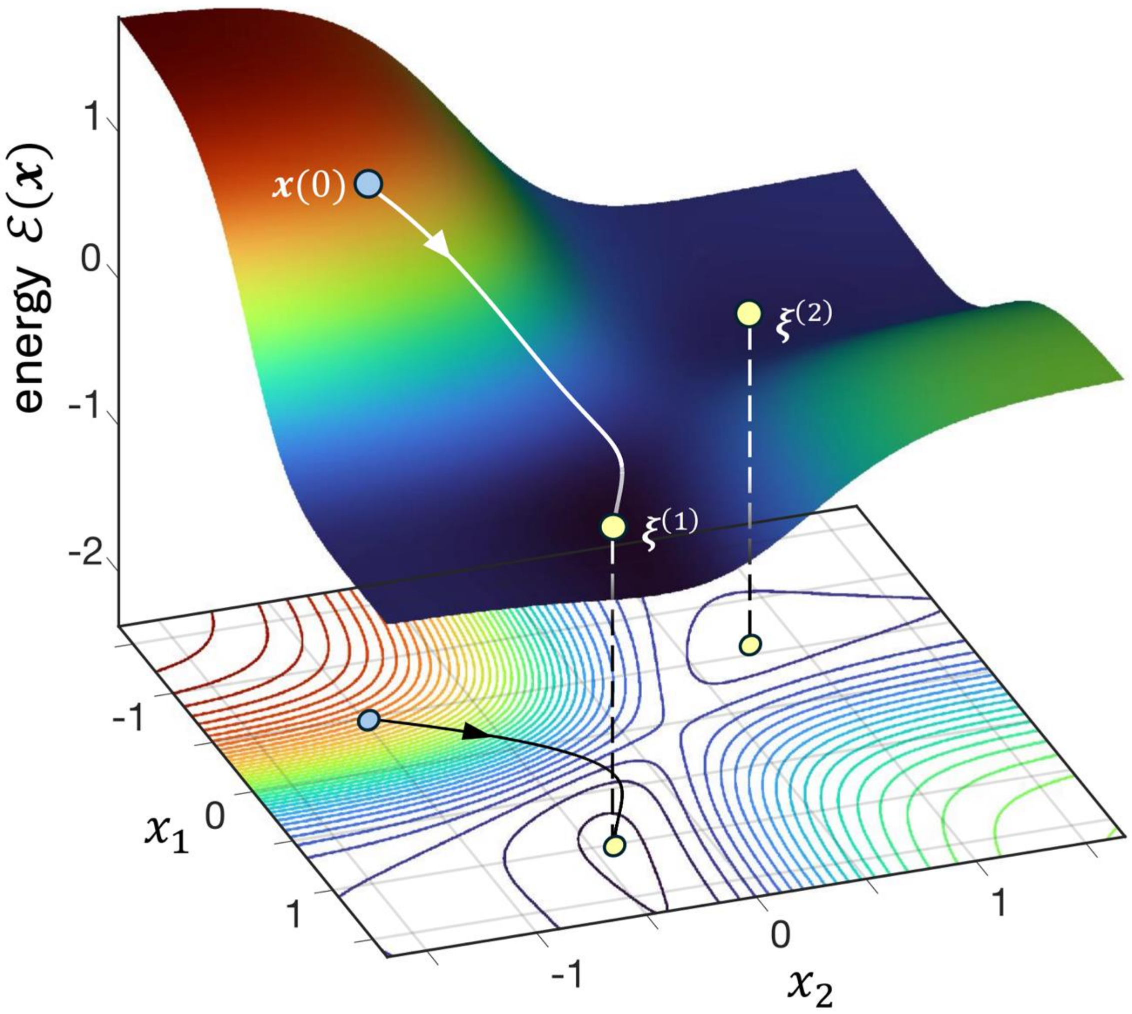
Continuous-time
Hopfield model

$$\tau\dot{\boldsymbol{x}} = -D\boldsymbol{x} + W\Phi(\boldsymbol{x}) + B\boldsymbol{u}$$
$$= -M(\boldsymbol{x})\,\nabla_{\boldsymbol{x}}\mathcal{E}(\boldsymbol{x})$$

Theorem. If $W = W^T$ and Φ is continuously differentiable, then the system is described by the following energy function:

$$\mathcal{E}(\boldsymbol{x}) = -\frac{1}{2}\boldsymbol{\Phi}(\boldsymbol{x})^{\mathsf{T}}W\boldsymbol{\Phi}(\boldsymbol{x}) + (D\boldsymbol{x} - B\boldsymbol{u})^{\mathsf{T}}\boldsymbol{\Phi}(\boldsymbol{x}) - \sum_{i=1}^{N} d_i \int_0^{x_i} \Phi(w)\mathrm{d}w.$$

Finding an energy (Lyapunov) function



Interpretability?

Reliability?

Scalability

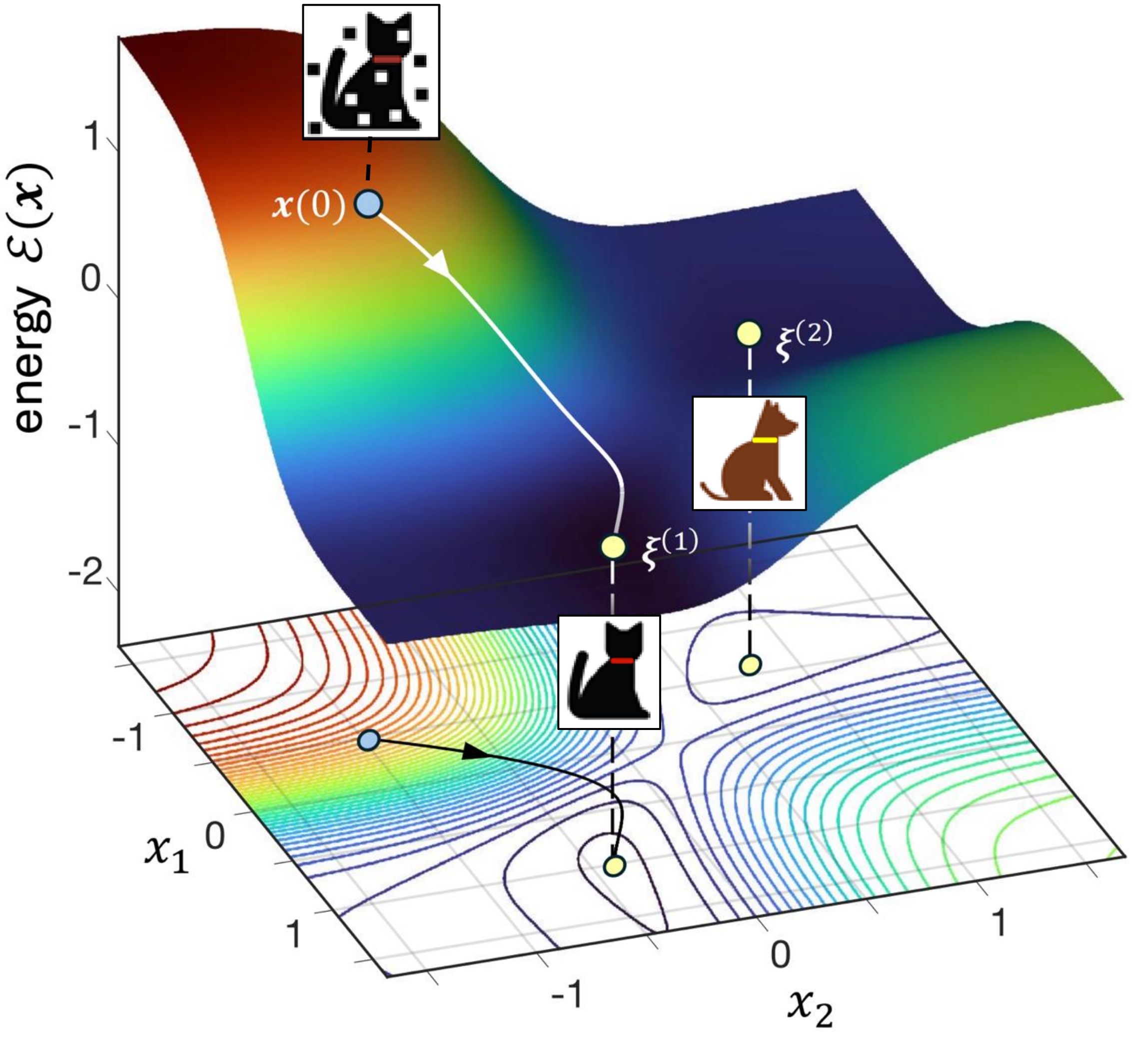
Biological plausibility?

Continuous-time
Hopfield model

$$\tau\dot{\boldsymbol{x}} = -D\boldsymbol{x} + W\Phi(\boldsymbol{x}) + B\boldsymbol{u}$$
$$= -M(\boldsymbol{x})\nabla\mathcal{E}(\boldsymbol{x})$$

$\mathcal{E}(\boldsymbol{x})$ represents a computational problem where its minima $\boldsymbol{x}^*$ encode the desired solutions

Associative memory model



Interpretability?

Reliability?

Scalability

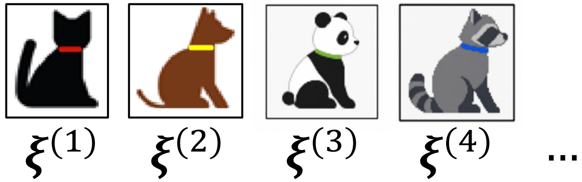
Biological plausibility?

Continuous-time *Hopfield model*

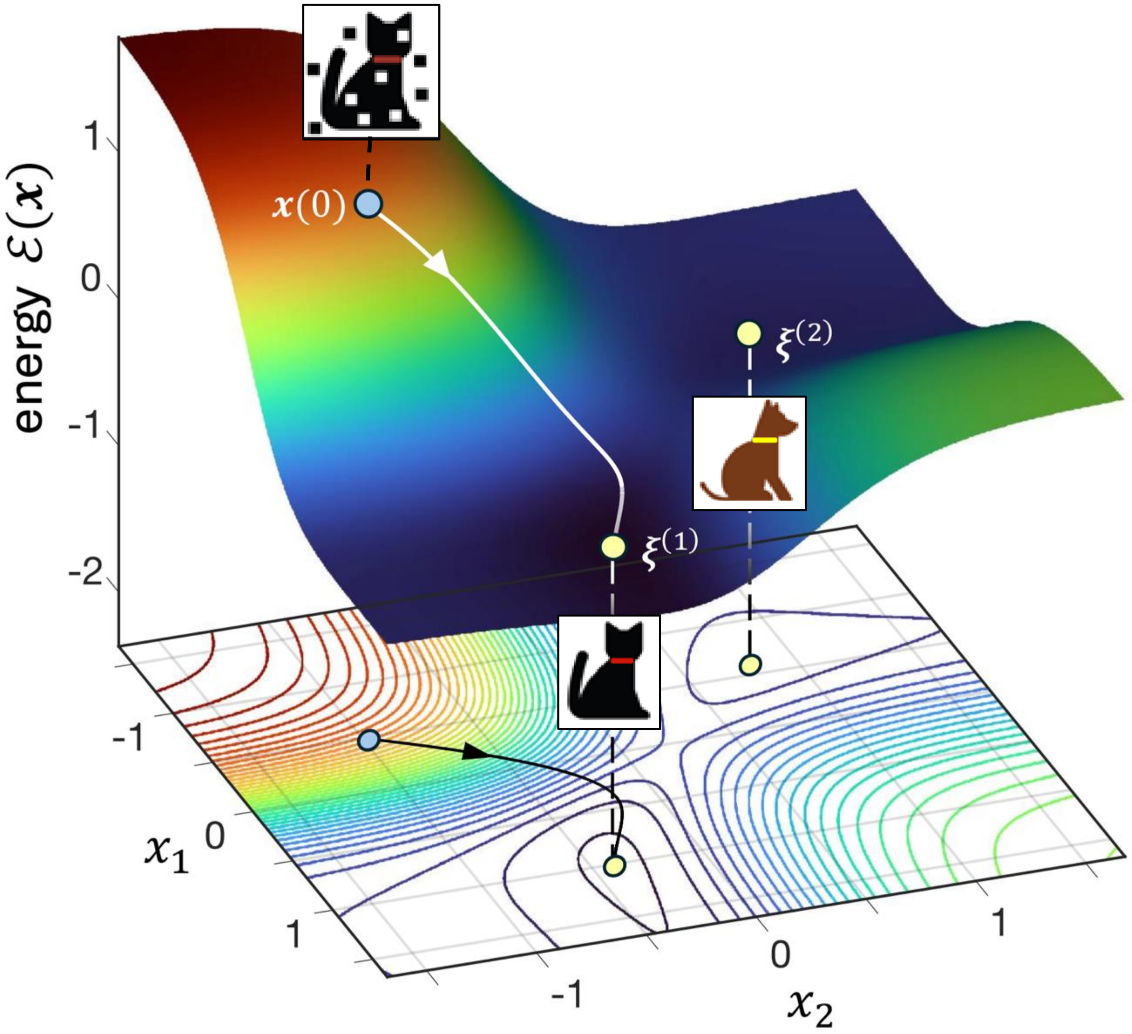
$$\tau \dot{\boldsymbol{x}} = -D\boldsymbol{x} + W\Phi(\boldsymbol{x}) + B\boldsymbol{u}$$
$$= -M(\boldsymbol{x})\nabla\mathcal{E}(\boldsymbol{x})$$

$\mathcal{E}(\boldsymbol{x})$ represents a computational problem where its minima $\boldsymbol{x}^*$ encode the desired solutions.

For instance, let each minima $\boldsymbol{x}^*$ correspond to a memory pattern $\boldsymbol{\xi}^{(\mu)} \in \mathbb{R}^N$ sought to be stored.



Associative memory model



Interpretability?

Reliability?

Scalability

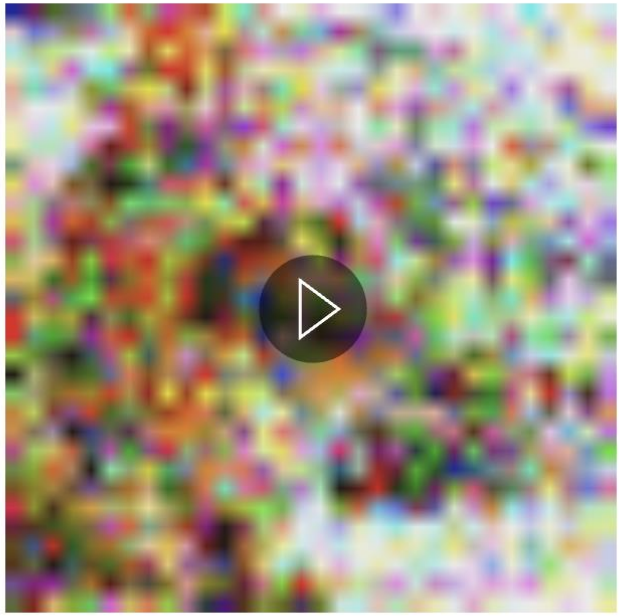
Biological plausibility?

Continuous-time

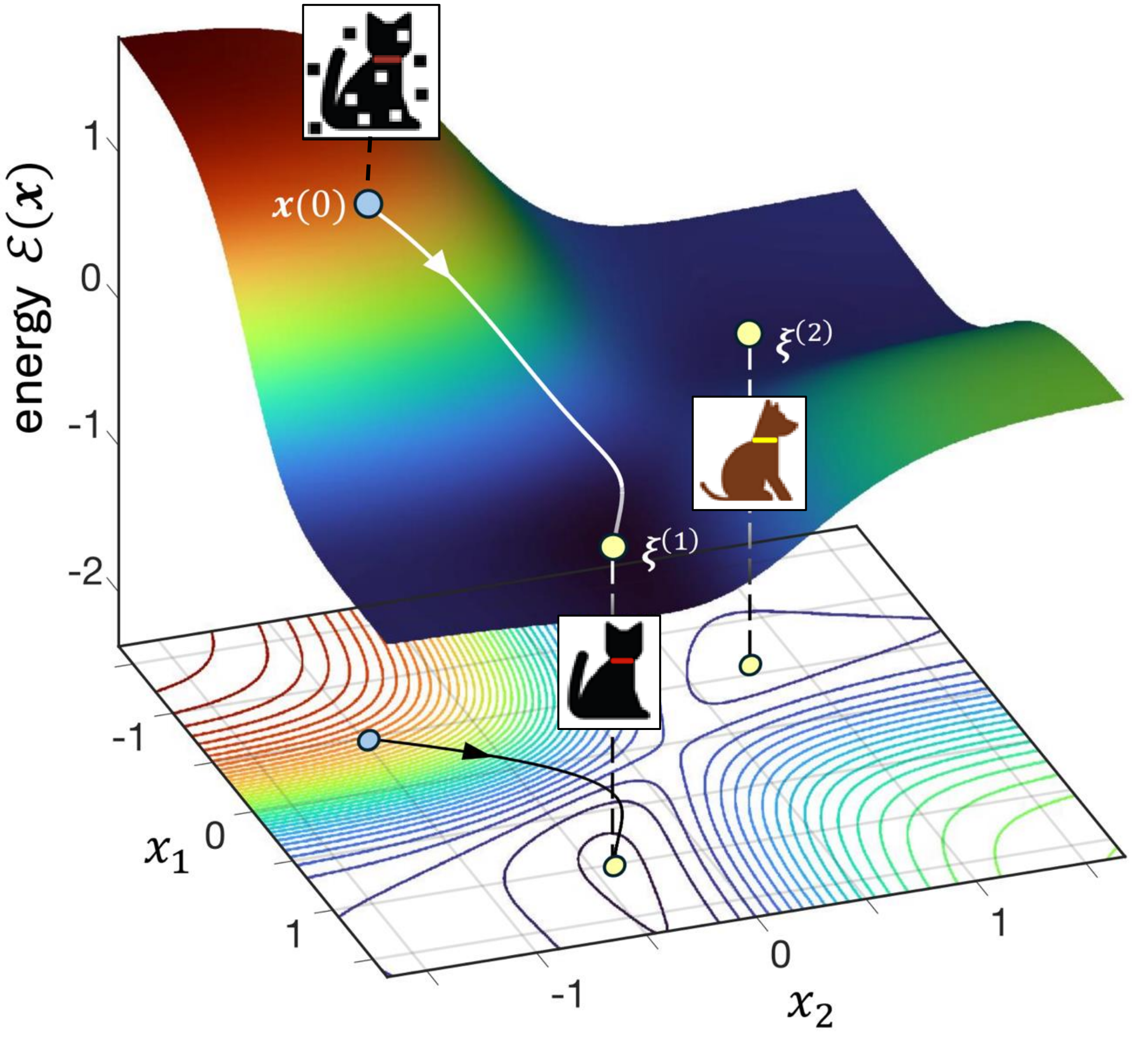
Hopfield model

$$\tau\dot{\boldsymbol{x}} = -D\boldsymbol{x} + W\Phi(\boldsymbol{x}) + B\boldsymbol{u}$$

$$= -M(\boldsymbol{x})\nabla\mathcal{E}(\boldsymbol{x})$$



Associative memory model



Interpretability?

Reliability?

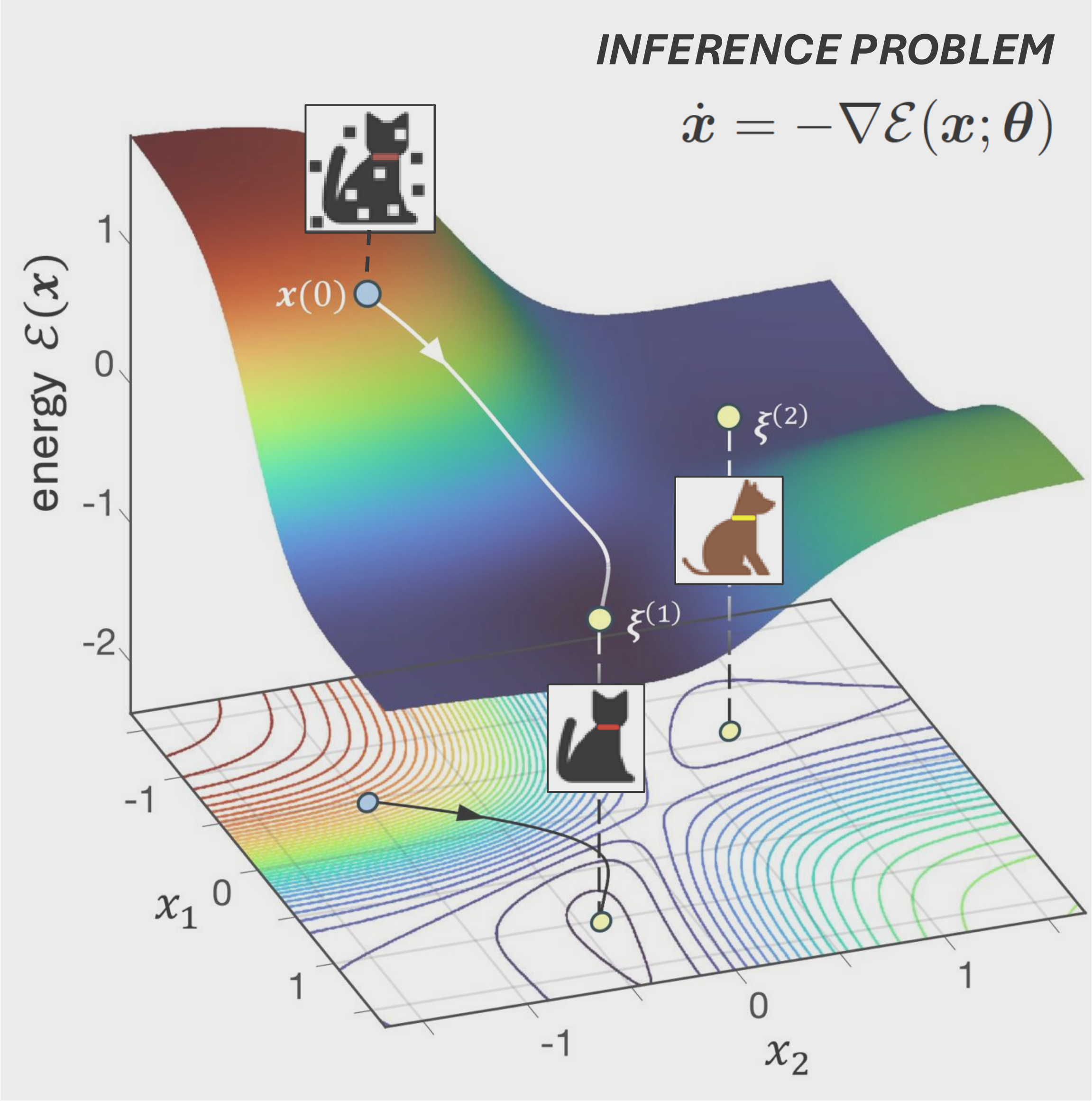
Scalability

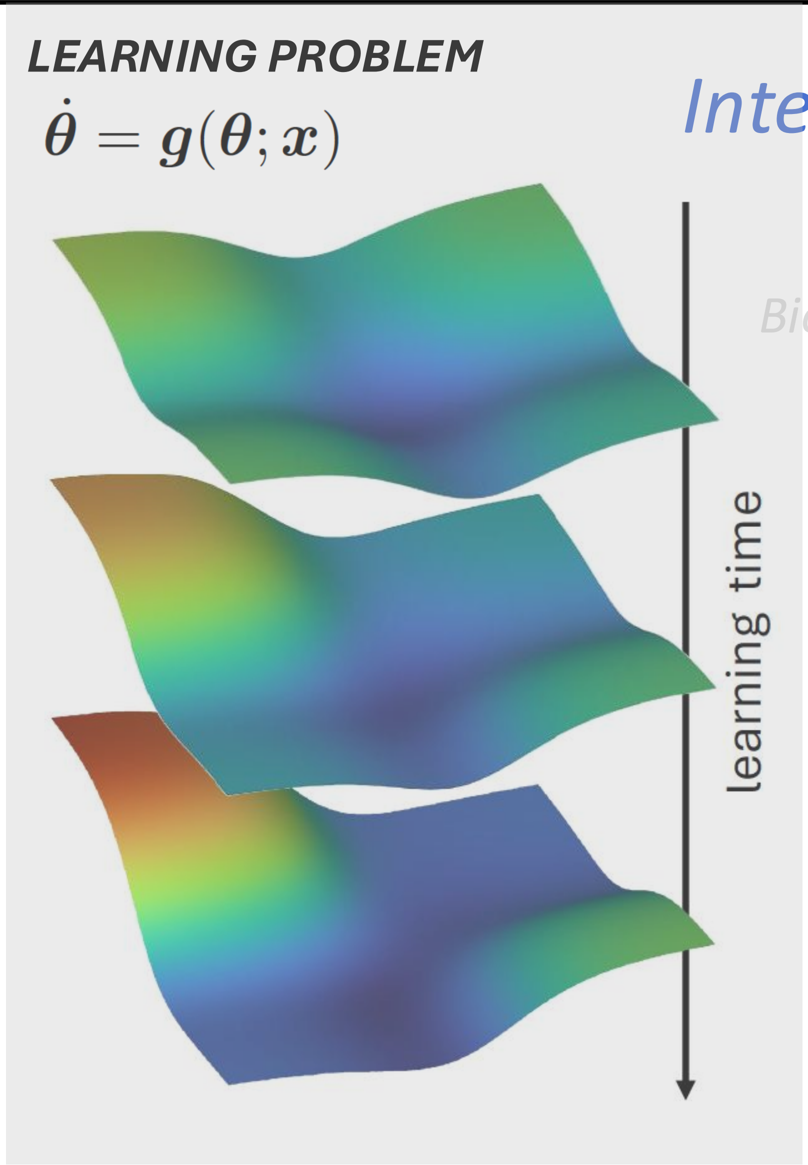
Biological plausibility?

Continuous-time
Hopfield model

$$\tau\dot{\boldsymbol{x}} = -D\boldsymbol{x} + W\Phi(\boldsymbol{x}) + B\boldsymbol{u}$$
$$= -M(\boldsymbol{x})\nabla\mathcal{E}(\boldsymbol{x})$$

How this neural network learns?





Interpretability?

Reliability?

Scalability

Biological plausibility?

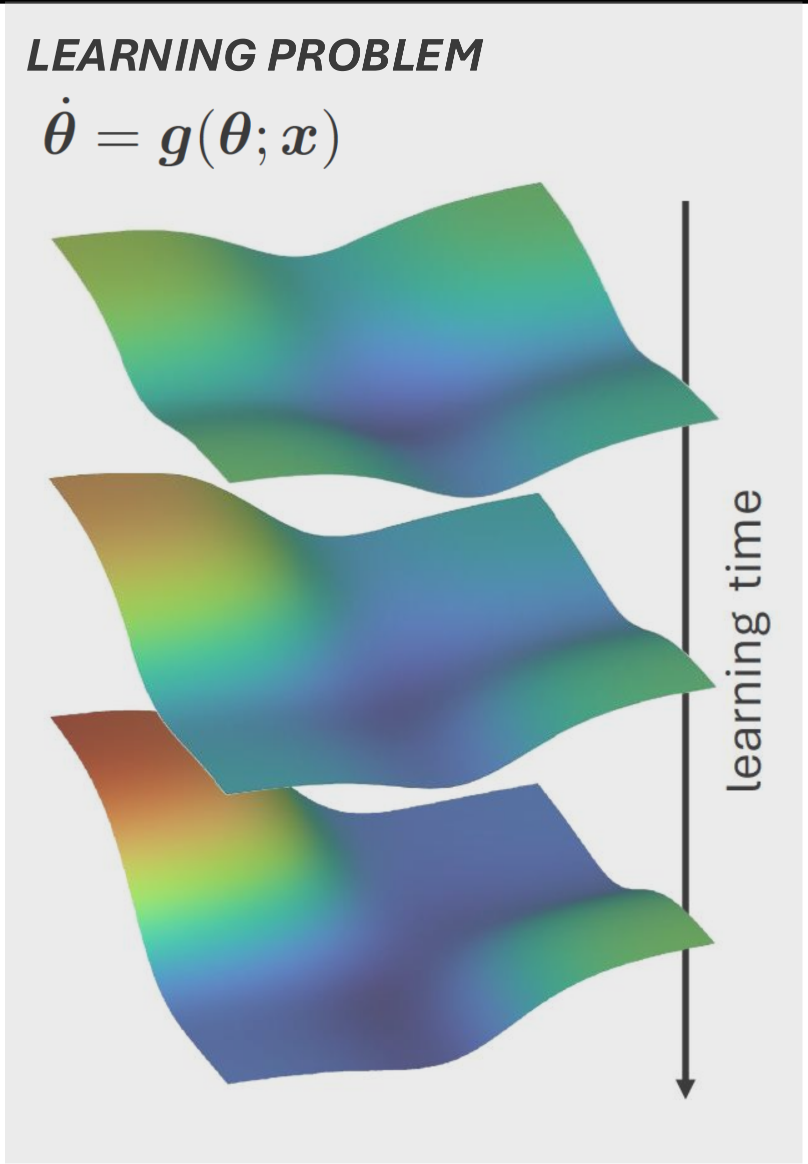
How this neural network learns?

Let $\boldsymbol{\xi}^{(\mu)}$, for $\mu = 1, \ldots, K$, be the set of patterns to be stored.

$$E(\boldsymbol{x}; W) = -\sum_{i,j} W_{ij} x_i x_j = \boldsymbol{x}^T W \boldsymbol{x}$$

Recursive learning procedure for each pattern to find the best weight that minimizes the energy associated with a pattern $\boldsymbol{\xi}^{(\mu)}$

$$\frac{\partial E^{(\mu)}}{\partial W_{ij}} = -x_i x_j =: \ -\xi_i^{(\mu)} \xi_j^{(\mu)}$$



Interpretability

Reliability?

Scalability

Biological plausibility?

Hebbian learning

Let $\boldsymbol{\xi}^{(\mu)}$, for $\mu = 1, \dots, K$, be the set of patterns to be stored.

$$E(\boldsymbol{x}; W) = -\sum_{i,j} W_{ij} x_i x_j = \boldsymbol{x}^T W \boldsymbol{x}$$

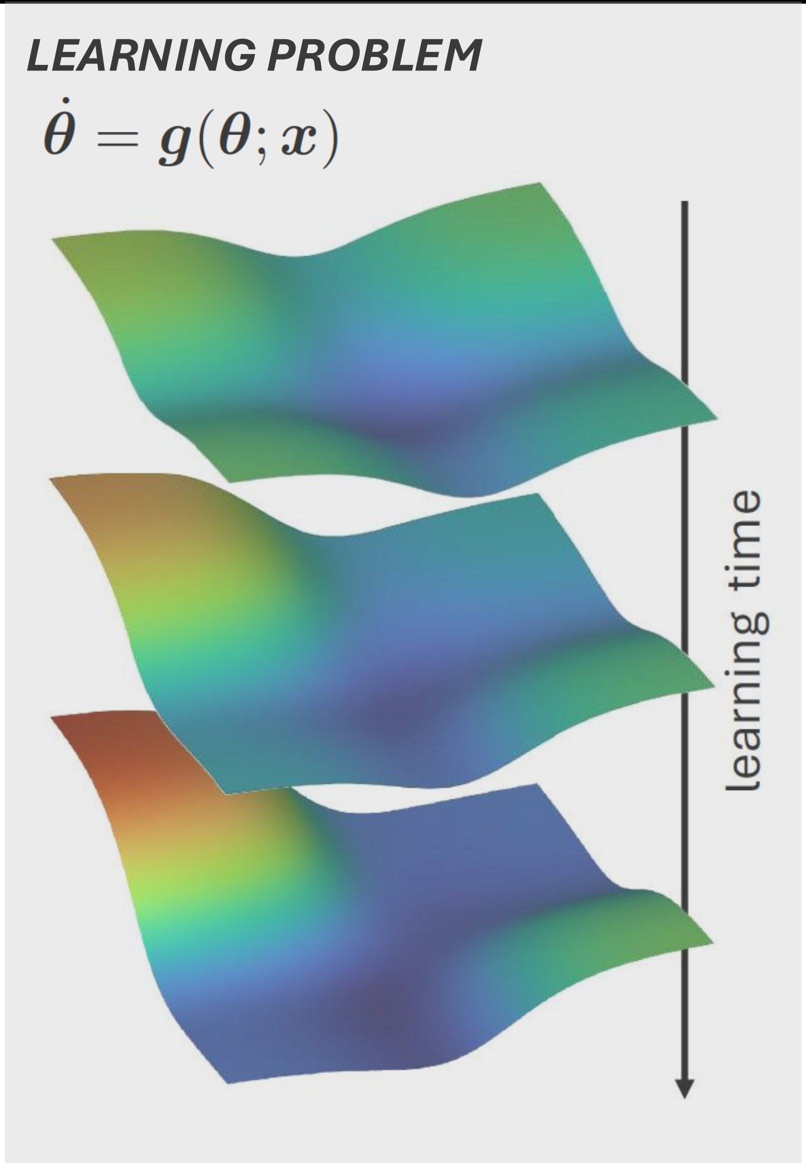
Recursive learning procedure for each pattern to find the best weight that minimizes the energy associated with a pattern $\boldsymbol{\xi}^{(\mu)}$

$$\frac{\partial E^{(\mu)}}{\partial W_{ij}} = -x_i x_j =: \; -\xi_i^{(\mu)} \xi_j^{(\mu)}$$

if both neurons are active, then increasing W_{ij} lowers the energy of that pattern.

Neurons that fire together, wire together

Donald Hebb



Interpretability

Reliability?

Scalability

Biological plausibility?

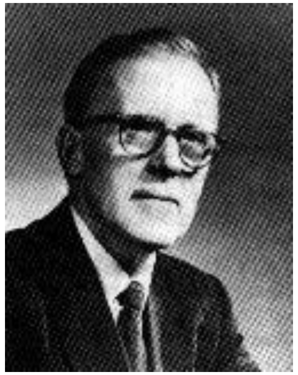
Hebbian learning

Let $\boldsymbol{\xi}^{(\mu)}$, for $\mu = 1, \dots, K$, be the set of patterns to be stored.

$$E(\boldsymbol{x}; W) = -\sum_{i,j} W_{ij} x_i x_j = \boldsymbol{x}^T W \boldsymbol{x}$$

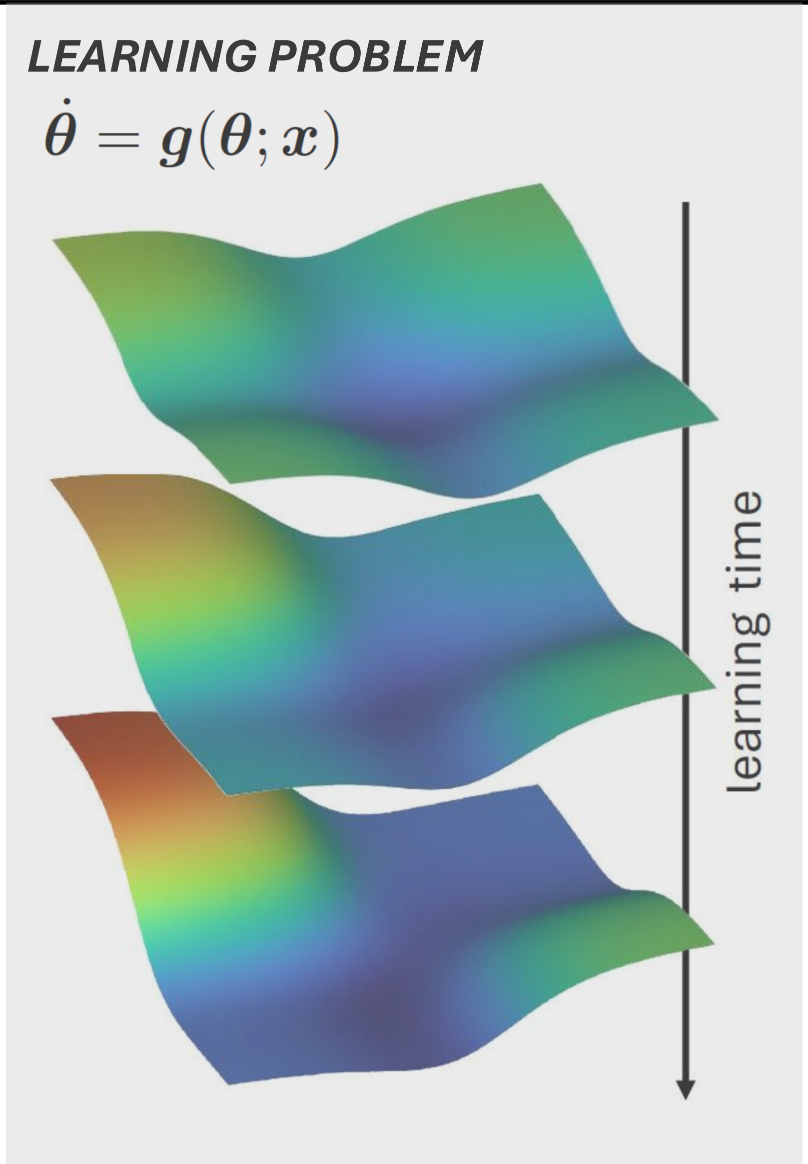
Recursive learning procedure for each pattern to find the best weight that minimizes the energy associated with a pattern $\boldsymbol{\xi}^{(\mu)}$

$$\frac{\partial E^{(\mu)}}{\partial W_{ij}} = -x_i x_j =: \; -\xi_i^{(\mu)} \xi_j^{(\mu)}$$



Donald Hebb

$$W_{ij} = \frac{1}{N}\left(\xi_i^{(1)}\xi_j^{(1)} + \xi_i^{(2)}\xi_j^{(2)} + \dots\right)$$
$$= \frac{1}{K}\sum_{\mu=1}^{K} \xi_i^{(\mu)}\xi_j^{(\mu)}$$
$$= \frac{1}{K}\boldsymbol{\Xi}\boldsymbol{\Xi}^T$$



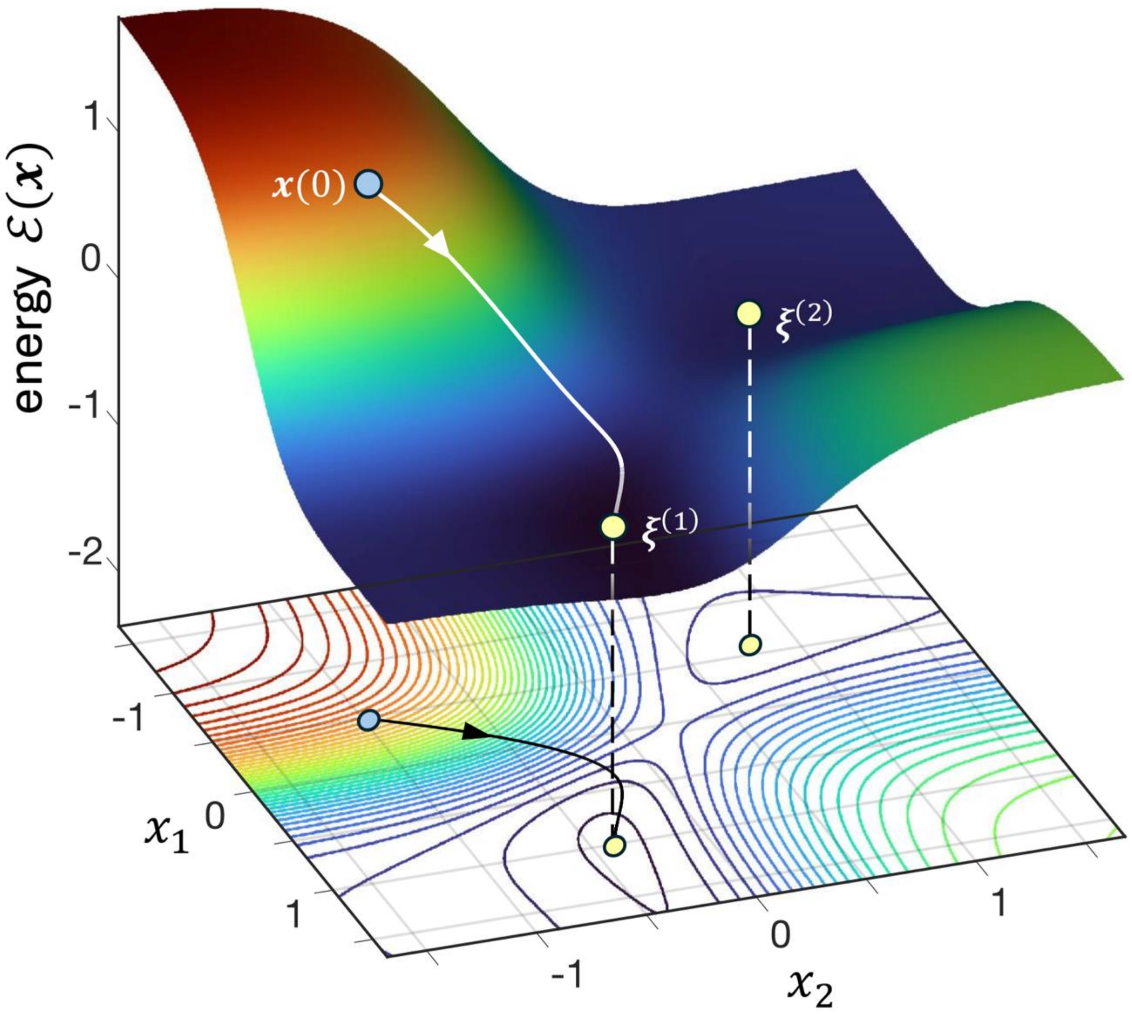
Interpretability

Reliability?

Scalability

Biological plausibility?

Energy-based dynamical model (EDM)



$\dot{\boldsymbol{x}} = f(\boldsymbol{x}; \boldsymbol{\theta})$ such that
$\dot{\mathcal{E}}(\boldsymbol{x}) = \nabla\mathcal{E}(\boldsymbol{x})^T \dot{\boldsymbol{x}} \leq 0,$

Interpretability
Reliability?
Scalability
Biological plausibility

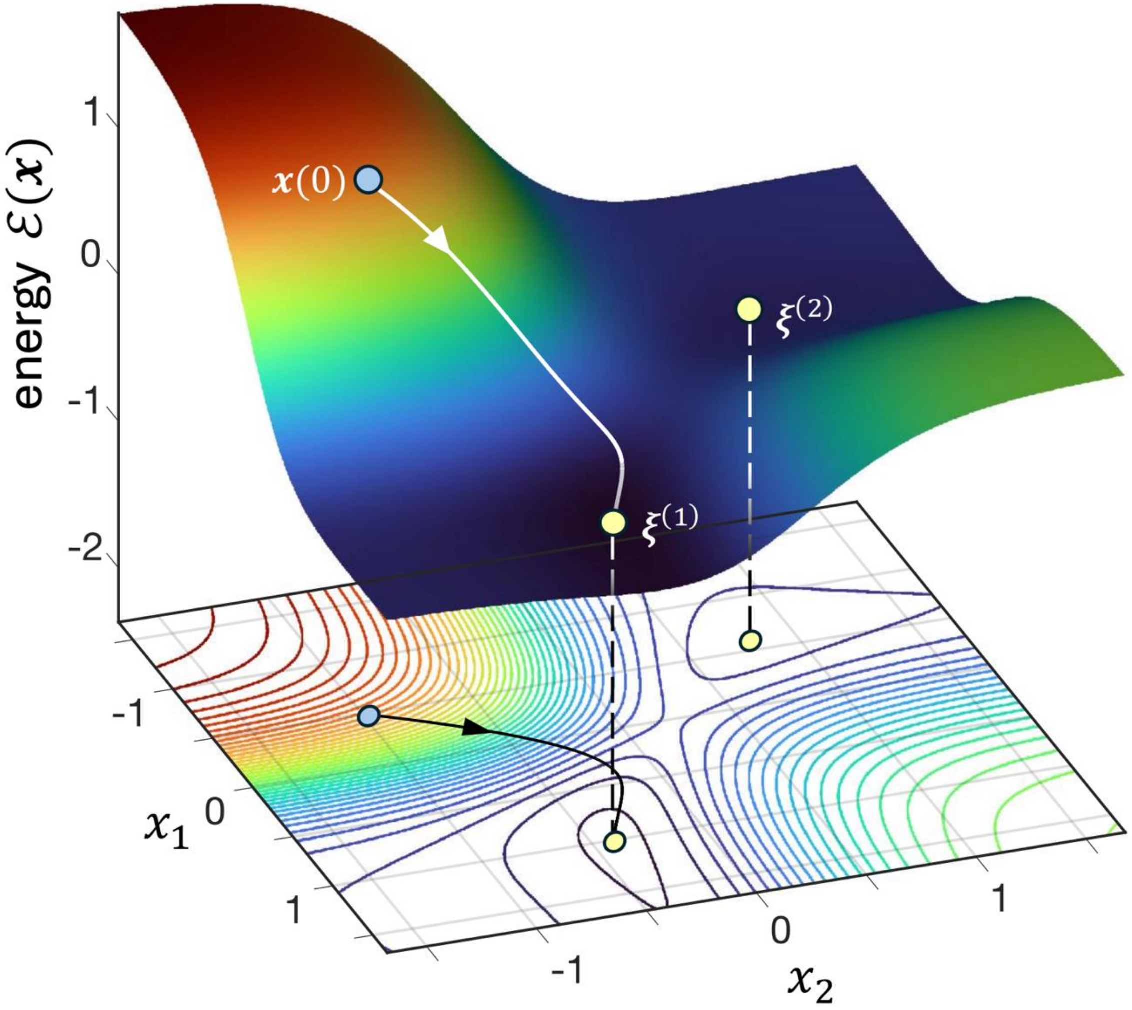
INFERENCE

- gradient-flow: $\dot{x} = -\nabla\mathcal{E}(x)$
- pre-conditioned flow: $\dot{x} = -M(x)\nabla\mathcal{E}(x)$
- stochastic flow: $dx_t = -\nabla\mathcal{E}(x_t)dt + \sqrt{2T}dw_t$
- projected gradient flow: $\dot{x} = \Pi_{\mathcal{C}}\left(-\nabla\mathcal{E}(x)\right)$

LEARNING

- Hebbian learning
- Contrastive Hebbian learning
- Equilibrium propagation
- Score matching

Energy-based dynamical model (EDM)



$\dot{\boldsymbol{x}} = f(\boldsymbol{x}; \boldsymbol{\theta})$ such that
$\dot{\mathcal{E}}(\boldsymbol{x}) = \nabla\mathcal{E}(\boldsymbol{x})^T \dot{\boldsymbol{x}} \leq 0,$

Interpretability
Reliability?
Scalability
Biological plausibility

INFERENCE

- gradient-flow: $\dot{x} = -\nabla\mathcal{E}(x)$
- pre-conditioned flow: $\dot{x} = -M(x)\nabla\mathcal{E}(x)$
- stochastic flow: $dx_t = -\nabla\mathcal{E}(x_t)dt + \sqrt{2T}dw_t$
- projected gradient flow: $\dot{x} = \Pi_{\mathcal{C}}\left(-\nabla\mathcal{E}(x)\right)$

$\mathcal{E}(\boldsymbol{x})$ represents a computational problem where its minima $\boldsymbol{x}^*$ encode the desired solutions } **OPTIMALITY**

co-design

$\mathcal{E}(\boldsymbol{x})$ represents a Lyapunov function, guaranteeing stability and convergence } **STABILITY**

Designing EDMs

$\mathcal{E}(\boldsymbol{x})$ represents a computational problem where its minima $\boldsymbol{x}^*$ encode the desired solutions → OPTIMALITY

co-design

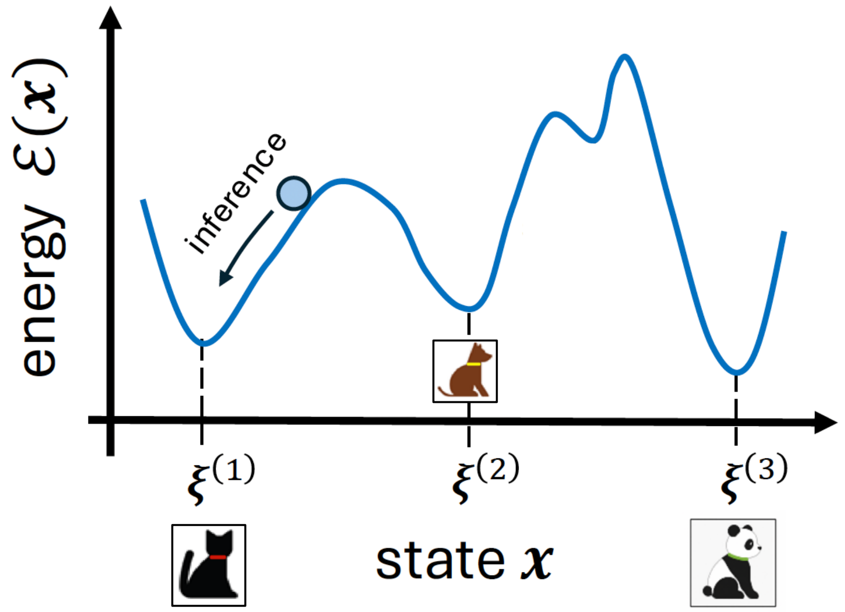
$\mathcal{E}(\boldsymbol{x})$ represents a Lyapunov function, guaranteeing stability and convergence → STABILITY

Interpretability

Reliability?

Scalability

Biological plausibility



Design objectives

1) Store a large number of memory patterns — *see Krotov's talk*
2) Basins of attraction must be sufficiently large
3) Inference must respect certain constraints — *see Bullo's talk*
4) Generalize to unseen data — *see tutorial paper (Boltzmann machines)*

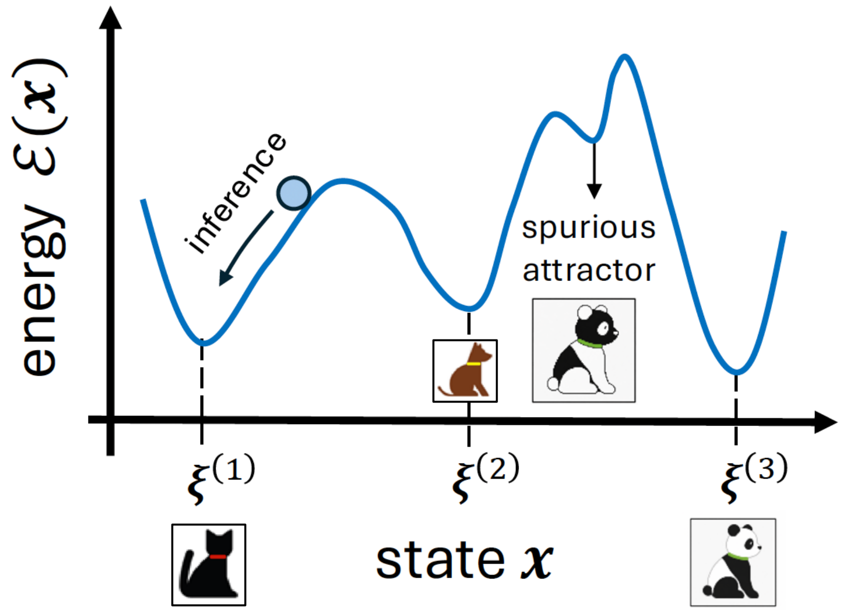
Designing EDMs

$\mathcal{E}(\boldsymbol{x})$ represents a computational problem where its minima $\boldsymbol{x}^*$ encode the desired solutions } OPTIMALITY

co-design

$\mathcal{E}(\boldsymbol{x})$ represents a Lyapunov function, guaranteeing stability and convergence } STABILITY

Interpretability
Reliability?
Scalability
Biological plausibility



Design objectives

1) Store a large number of memory patterns — *see Krotov's talk*
2) Basins of attraction must be sufficiently large
3) Inference must respect certain constraints — *see Bullo's talk*
4) Generalize to unseen data — *see tutorial paper (Boltzmann machines)*
5) Mitigate spurious attractors

Oscillator model

Restricting to binary patterns $\boldsymbol{x} \in \{-1, +1\}^N$,
we have the energy function

$$\mathcal{E}(\boldsymbol{x}) = -\frac{1}{2}\sum_{i,j=1}^{N} W_{ij} x_i x_j = \boldsymbol{x}^T W \boldsymbol{x}\,.$$

Now, consider the regularized energy function for $\boldsymbol{\phi} \in \mathbb{S}^N$:

$$\mathcal{E}(\boldsymbol{\phi}) = -\frac{1}{2}\sum_{i,j=1}^{N} W_{ij}\cos(\phi_j - \phi_i) - \frac{\kappa}{4N}\sum_{i,j=1}^{N}\cos\left(2(\phi_j - \phi_i)\right)$$

Interpretability

Reliability?

Scalability

Biological plausibility

Oscillator model

Restricting to binary patterns $\boldsymbol{x} \in \{-1, +1\}^N$, we have the energy function

$$\mathcal{E}(\boldsymbol{x}) = -\frac{1}{2}\sum_{i,j=1}^{N} W_{ij} x_i x_j = \boldsymbol{x}^T W \boldsymbol{x}.$$

$$|\phi_j^* - \phi_i^*| = \begin{cases} 0, & \text{if } \xi_i^{(\mu)} = \xi_j^{(\mu)} \\ \pi, & \text{if } \xi_i^{(\mu)} \neq \xi_j^{(\mu)} \end{cases}$$

Interpretability

Reliability?

Scalability

Biological plausibility

Now, consider the regularized energy function for $\boldsymbol{\phi} \in \mathbb{S}^N$:

$$\mathcal{E}(\boldsymbol{\phi}) = -\frac{1}{2}\sum_{i,j=1}^{N} W_{ij} \cos(\phi_j - \phi_i) - \frac{\kappa}{4N}\sum_{i,j=1}^{N} \cos\left(2(\phi_j - \phi_i)\right)$$

Oscillator model

Restricting to binary patterns $\boldsymbol{x} \in \{-1, +1\}^N$, we have the energy function

$$\mathcal{E}(\boldsymbol{x}) = -\frac{1}{2}\sum_{i,j=1}^{N} W_{ij} x_i x_j = \boldsymbol{x}^T W \boldsymbol{x}.$$

$$|\phi_j^* - \phi_i^*| = \begin{cases} 0, & \text{if } \xi_i^{(\mu)} = \xi_j^{(\mu)} \\ \pi, & \text{if } \xi_i^{(\mu)} \neq \xi_j^{(\mu)} \end{cases}$$

Interpretability

Reliability?

Scalability

Biological plausibility

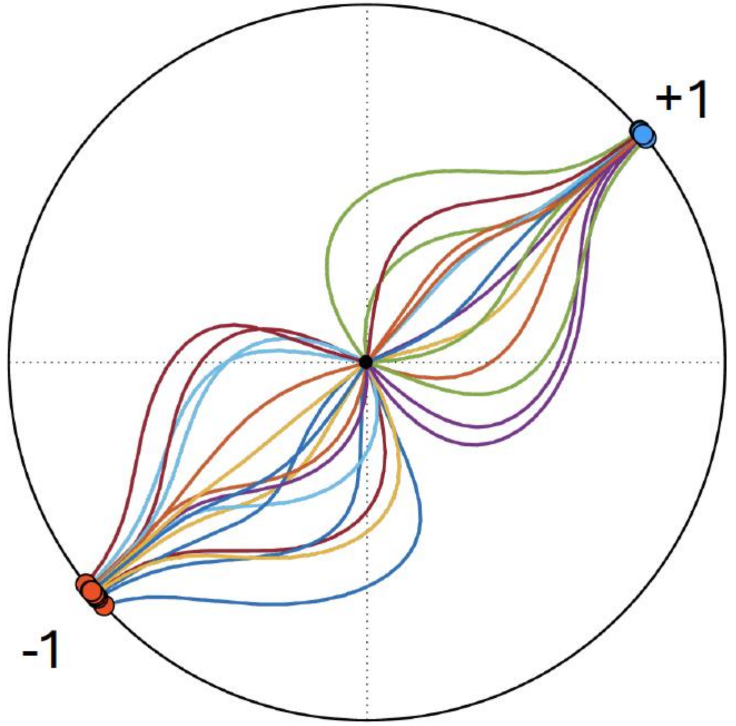
Now, consider the regularized energy function for $\boldsymbol{\phi} \in \mathbb{S}^N$:

$$\mathcal{E}(\boldsymbol{\phi}) = -\frac{1}{2}\sum_{i,j=1}^{N} W_{ij}\cos(\phi_j - \phi_i) - \frac{\kappa}{4N}\sum_{i,j=1}^{N}\cos\left(2(\phi_j - \phi_i)\right)$$

Oscillatory EDM:

$$\dot{\phi} = -\nabla\mathcal{E}(\boldsymbol{\phi})$$

$$\dot{\phi}_i = \omega + \sum_{j=1}^{N} W_{ij}\sin(\phi_j - \phi_i) + \frac{\kappa}{N}\sum_{j=1}^{N}\sin\left(2(\phi_j - \phi_i)\right).$$



Oscillator model

Restricting to binary patterns $\boldsymbol{x} \in \{-1, +1\}^N$, we have the energy function

$$\mathcal{E}(\boldsymbol{x}) = -\frac{1}{2}\sum_{i,j=1}^{N} W_{ij} x_i x_j = \boldsymbol{x}^T W \boldsymbol{x}.$$

$$|\phi_j^* - \phi_i^*| = \begin{cases} 0, & \text{if } \xi_i^{(\mu)} = \xi_j^{(\mu)} \\ \pi, & \text{if } \xi_i^{(\mu)} \neq \xi_j^{(\mu)} \end{cases}$$

Interpretability

Reliability?

Scalability

Biological plausibility

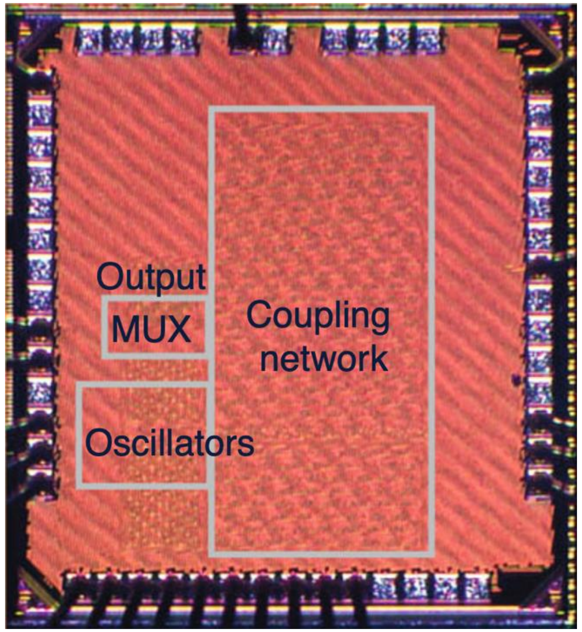
Now, consider the regularized energy function for $\boldsymbol{\phi} \in \mathbb{S}^N$:

$$\mathcal{E}(\boldsymbol{\phi}) = -\frac{1}{2}\sum_{i,j=1}^{N} W_{ij} \cos(\phi_j - \phi_i) - \frac{\kappa}{4N}\sum_{i,j=1}^{N} \cos\big(2(\phi_j - \phi_i)\big)$$

Oscillatory EDM:

$$\dot{\phi} = -\nabla\mathcal{E}(\boldsymbol{\phi})$$

$$\dot{\phi}_i = \omega + \sum_{j=1}^{N} W_{ij} \sin\big(\phi_j - \phi_i\big) + \frac{\kappa}{N}\sum_{j=1}^{N} \sin\big(2(\phi_j - \phi_i)\big).$$



A Mallick, MK Bashar, DS Truesdell, BH Calhoun, S Joshi, N Shukla. *Nature Communications* (2020).

Oscillator model

Interpretability
Reliability?
Scalability
Biological plausibility

Oscillatory EDM:

$$\dot{\phi} = -\nabla \mathcal{E}(\boldsymbol{\phi})$$

$$\dot{\phi}_i = \omega + \sum_{j=1}^{N} W_{ij} \sin(\phi_j - \phi_i) + \frac{\kappa}{N} \sum_{j=1}^{N} \sin\left(2(\phi_j - \phi_i)\right).$$

Theorem. Let the weights be assigned by Hebbian learning, $W = \frac{1}{K} \Xi \Xi^T$. The equilibrium point $\phi^* \equiv \xi^{(\mu)}$ is stable if $\lambda_{\max}(J) < 2\kappa$.

T Nishikawa, YC Lai, FC Hoppensteadt. *PRL* (2004).

Oscillator model

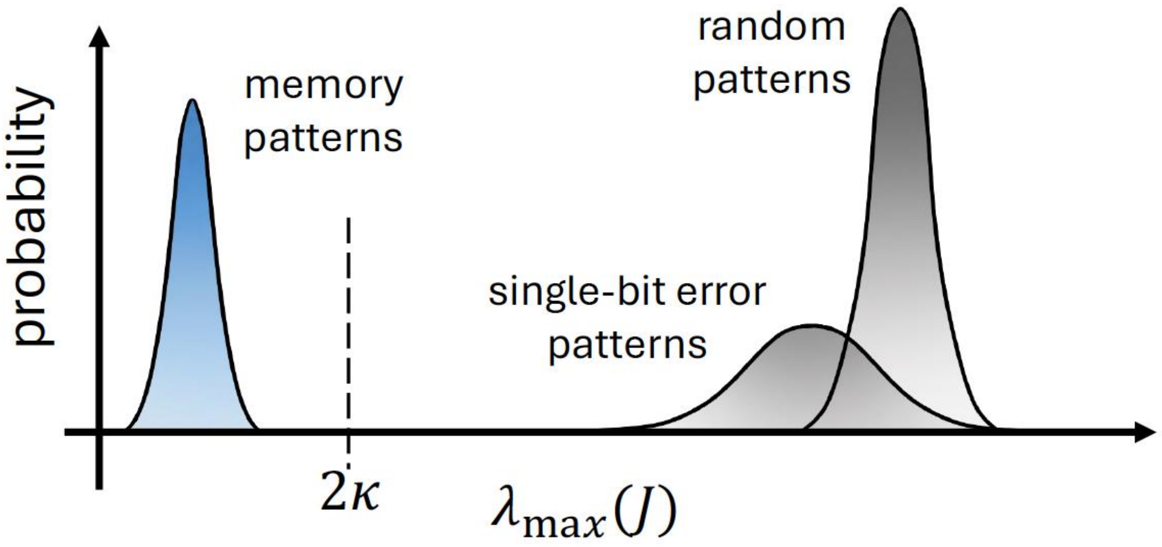
Oscillatory EDM:

Interpretability
Reliability?
Scalability
Biological plausibility

$$\dot{\phi} = -\nabla \mathcal{E}(\boldsymbol{\phi})$$

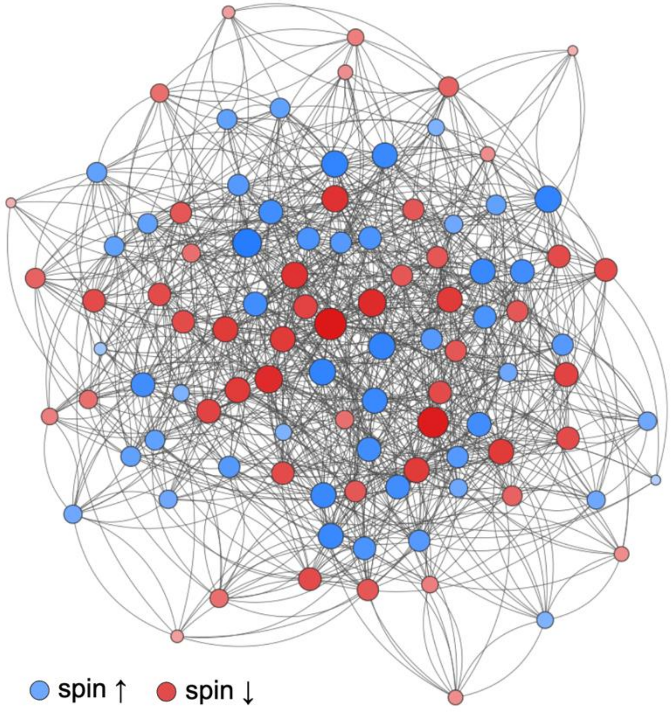
$$\dot{\phi}_i = \omega + \sum_{j=1}^{N} W_{ij} \sin(\phi_j - \phi_i) + \frac{\kappa}{N} \sum_{j=1}^{N} \sin\big(2(\phi_j - \phi_i)\big).$$

Theorem. Let the weights be assigned by Hebbian learning, $W = \frac{1}{K} \Xi \Xi^T$. The equilibrium point $\phi^* \equiv \xi^{(\mu)}$ is stable if $\lambda_{\max}(J) < 2\kappa$.



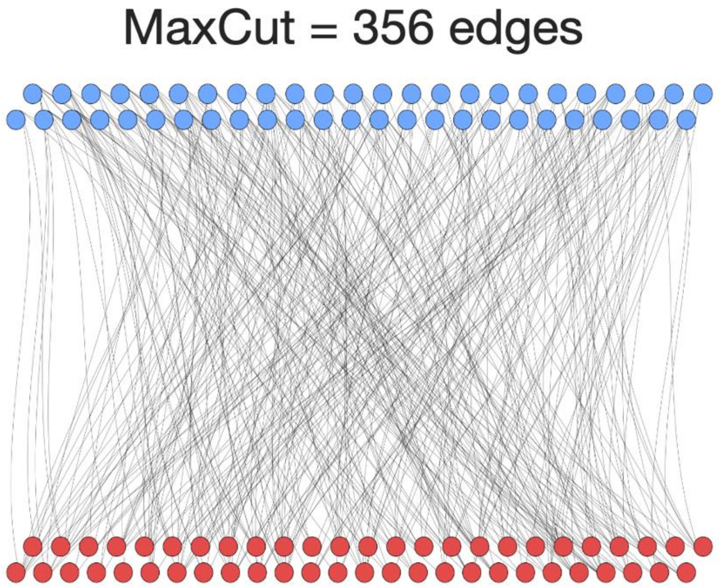
κ is a design parameter to mitigate spurious attractors by forcing only low-energy states (desired memories) to be stable

Solving combinatorial optimization problems

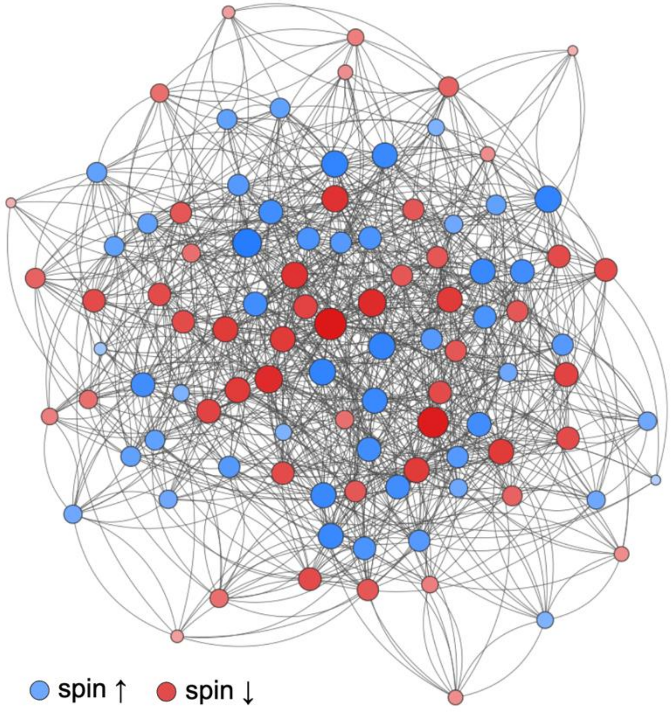


???

2^N possible solutions



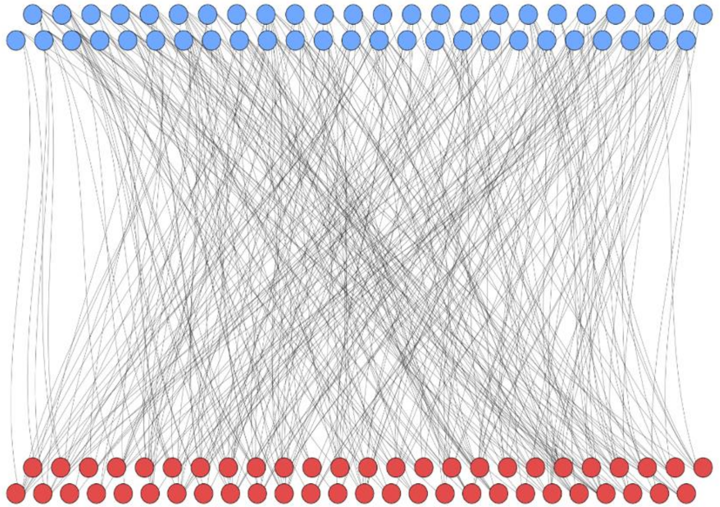
Solving combinatorial optimization problems



???

2^N possible solutions

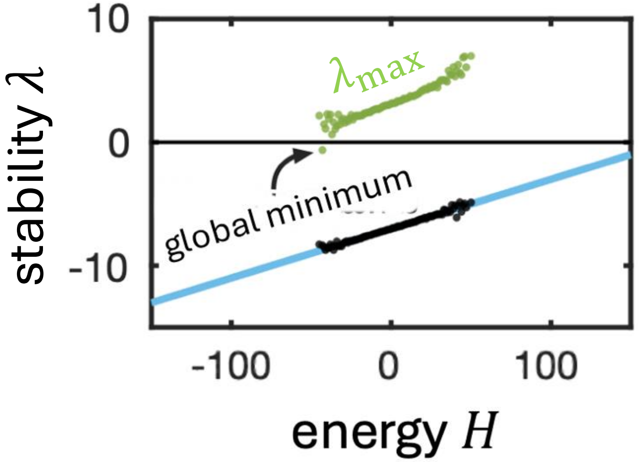
MaxCut = 356 edges



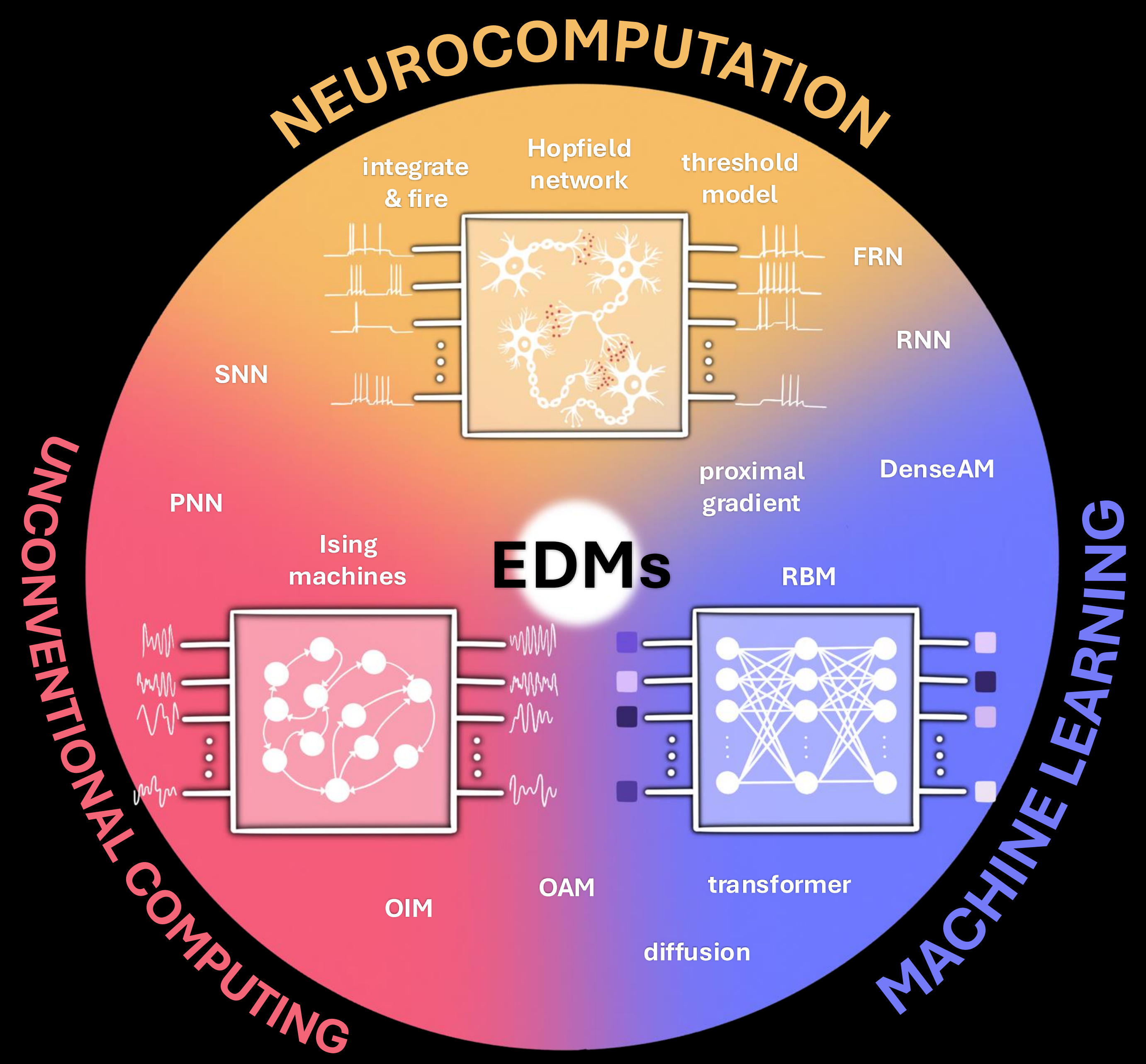
Oscillatory EDM to solve this problem:

$$\dot{\phi}_i = \text{gradient descent on Ising Hamiltonian}$$

$$= \omega + \sum_{j=1}^{N} W_{ij} \sin(\phi_j - \phi_i) + \kappa \sin(2\phi_i)$$



A Allibhoy, **AN Montanari**, F Pasqualetti, AE Motter. *IEEE CDC* (2025).


NEUROCOMPUTATION
UNCONVENTIONAL COMPUTING
MACHINE LEARNING
EDMs
integrate & fire
Hopfield network
threshold model
FRN
RNN
SNN
PNN
DenseAM
proximal gradient
Ising machines
RBM
OAM
OIM
transformer
diffusion

Acknowledgments

montanariarthur.com

slides available at my website

arXiv:2604.05042v1 [cs.LG] 6 Apr 2026

Energy-Based Dynamical Models for Neurocomputation, Learning, and Optimization

Arthur N. Montanari, Francesco Bullo, Dmitry Krotov, and Adilson E. Motter

3:30 - 4:00 pm. **Arthur N. Montanari**, Northwestern University
Recurrent neural networks and oscillator models for learning and optimization.

4:00 - 4:30 pm. **Dmitry Krotov**, Dynamical Mind
Dense associative memory for novel AI architectures.

4:30 - 5:00 pm. **Francesco Bullo**, UC Santa Barbara
Positive competitive neural networks for sparse reconstruction.