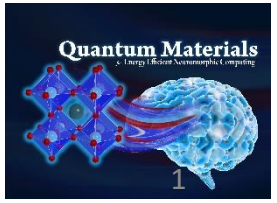
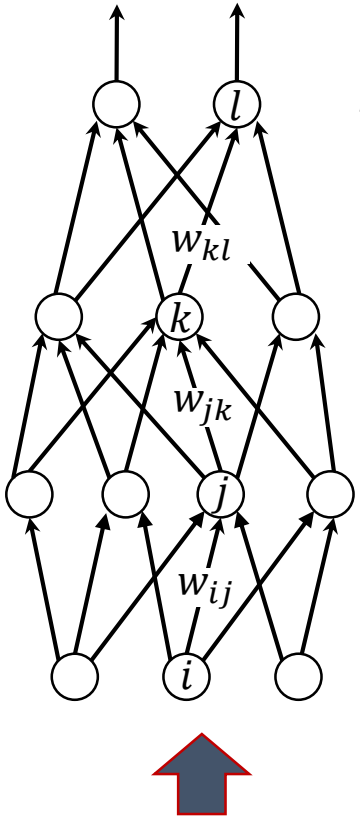
A New Spin on Neural Networks

Julie Grollier

CNRS/Thales, France



There is a long standing link between neural networks and spin systems



$$y_l = \rho(z_l)$$

$$z_l = \sum w_{kl} y_k$$

$$y_k = \rho(z_k)$$

$$z_k = \sum w_{jk} y_j$$

$$y_j = \rho(z_i)$$

$$z_j = \sum w_{ij} y_i$$

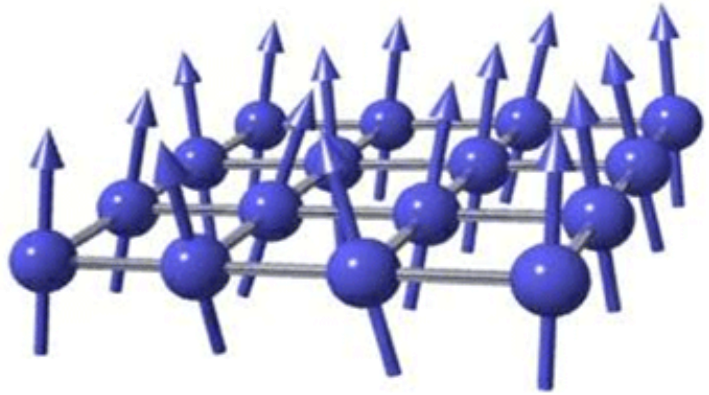
Neurons ⟷ Spins

Non-linear activation function

Synapses ⟷ Coupling

w_{ij} J_{ij}

Tunable parameter

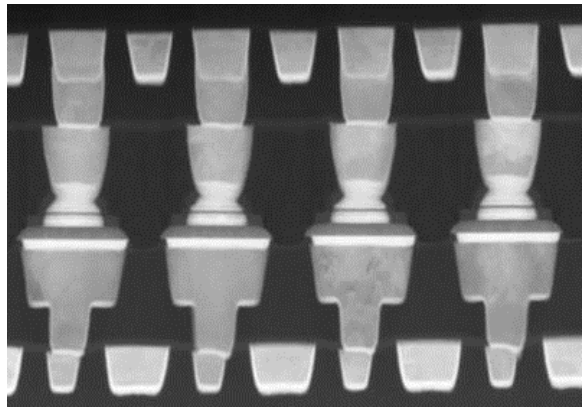


$$E = -\sum_{i<j} J_{ij} s_i s_j - \sum_i b_i s_i$$

J. J. Hopfield, PNAS 79, 2554 (1982)

We are at a convergence point for spin-based neural networks

Progress in nanotechnologies



Intel: MRAM integrated into 22nm FinFET CMOS

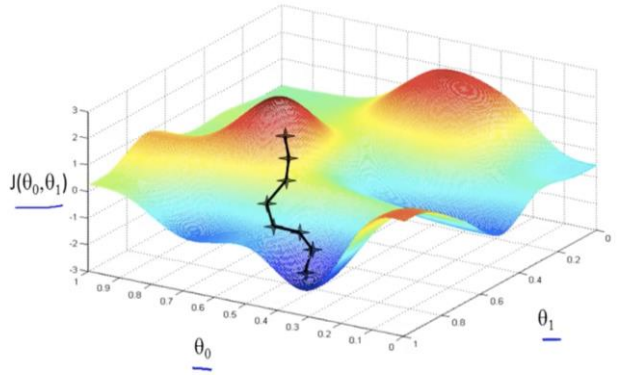
Physical spin systems for AI

Demands of applications: ultra-low power AI

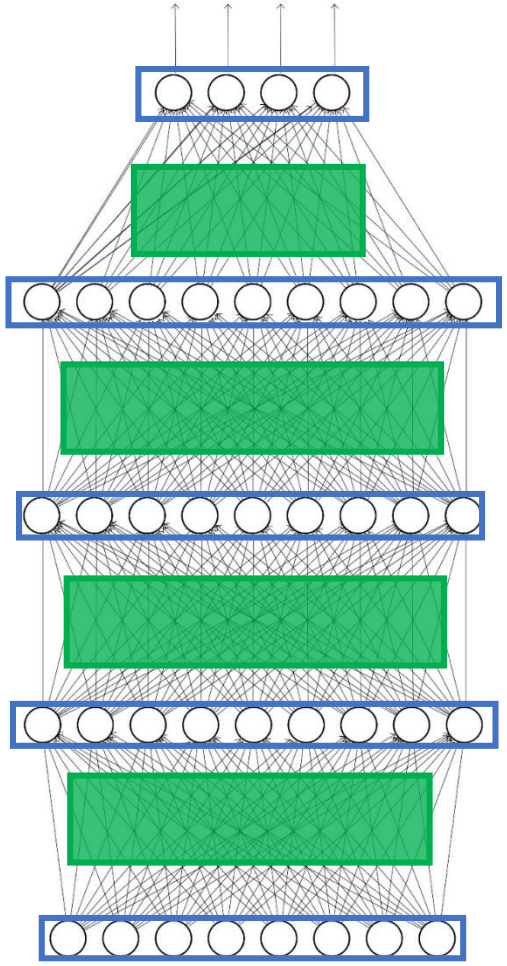
- Human brain : 20 W
- Training the GPT-3 language model: 190,000 kWh

→ 1,000 years of brain operation!

Progress in energy-based AI algorithms

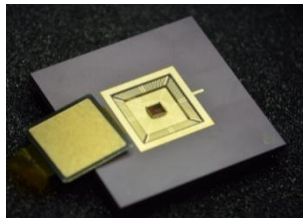


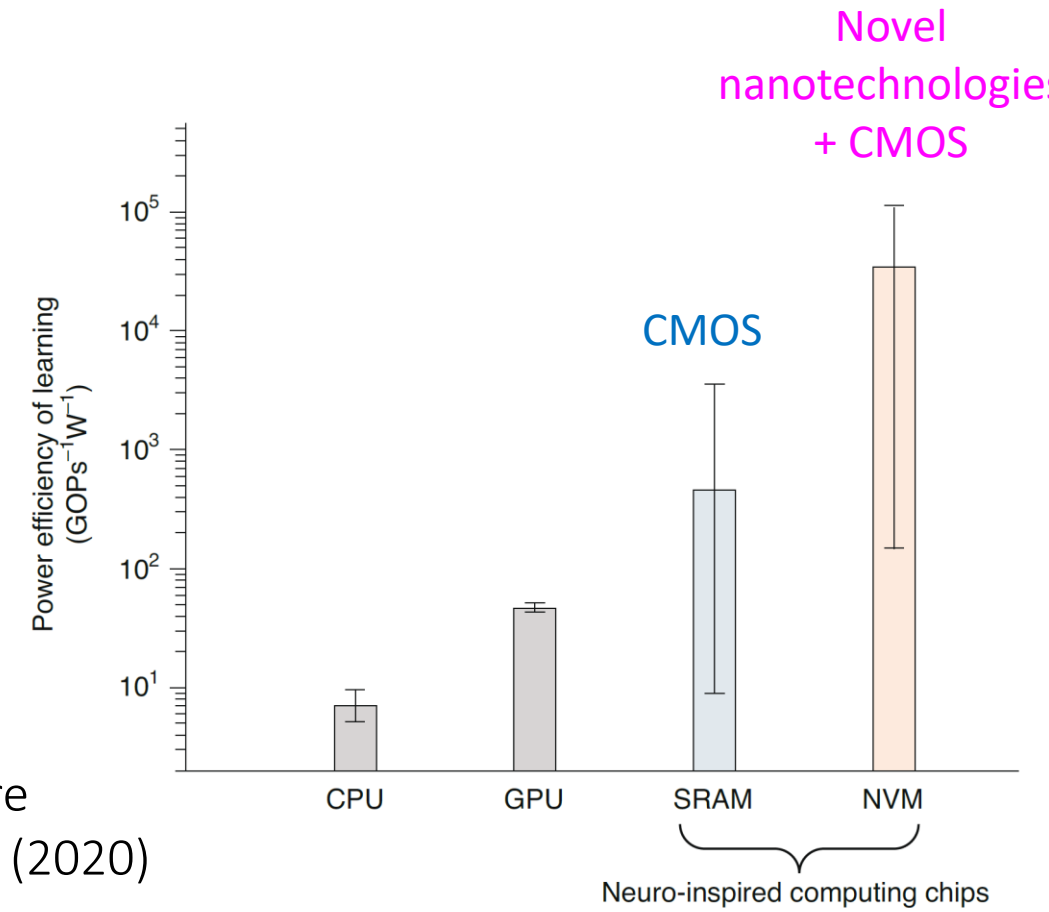
Neuromorphic chips can reduce the energy consumption of AI by several orders of magnitude



Electronic neurons

Electronic synapses



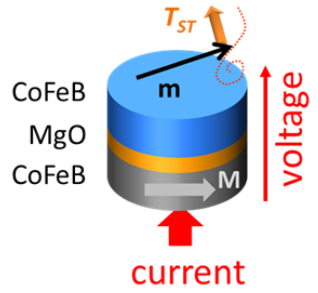


Zhang et al, Nature Electronics 3, 371 (2020)

They can solve environmental problem, speed up AI, unlock embedded AI

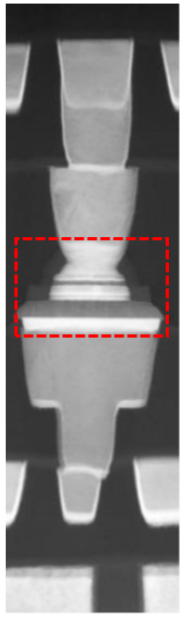
The dynamical properties of spintronic devices are an asset for neuromorphic computing

Magnetic tunnel junction



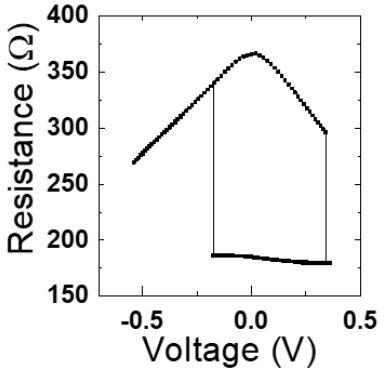
Non-volatility
Endurance
Reproducibility

Samsung IEDM 2018, **28nm** Logic



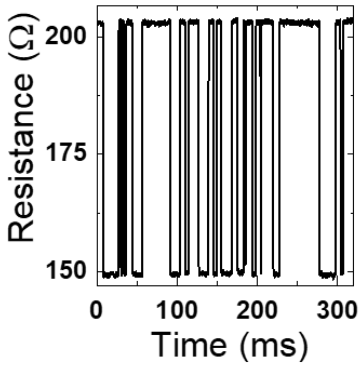
Non-volatile switching

Binary memory



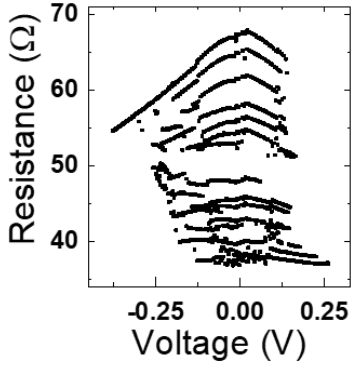
Thermal effects

Tunable stochasticity



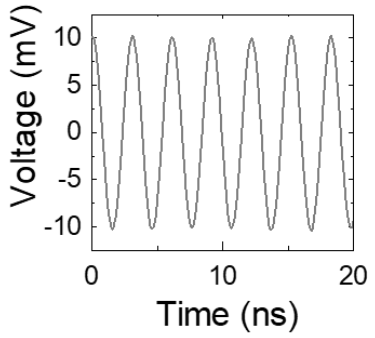
Magnetic textures

Memristor



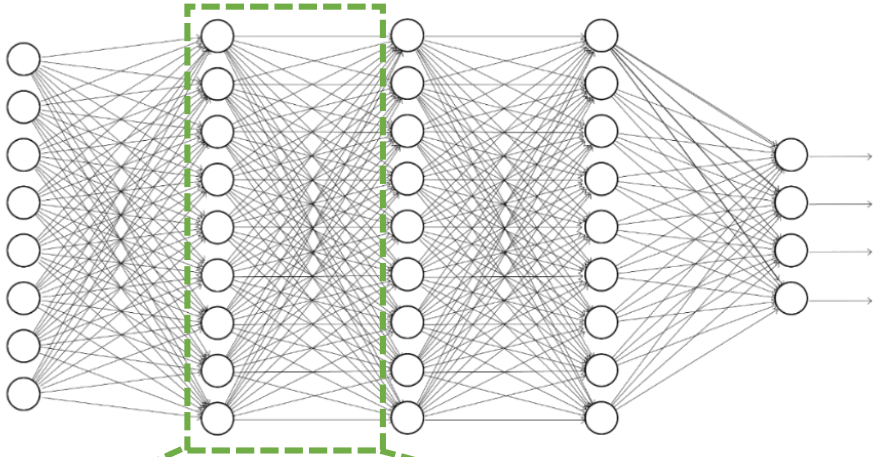
Magnetization dynamics

Non-linear oscillations



J. Grollier, D. Querlioz, K. Y. Camsari, K. Everschor-Sitte, S. Fukami, M. D. Stiles, Nature Electronics 3, 360 (2020)

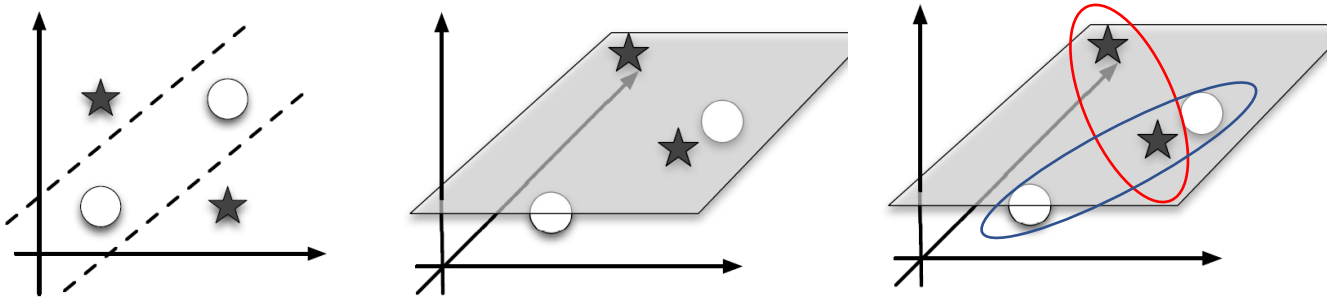
Pattern recognition requires separating then grouping data



Input data

Make data more different

Group data in classes



dog

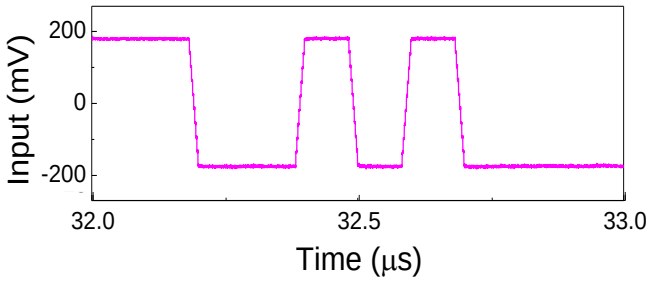
mop

We can use the complex dynamics of a spintronic nano-oscillator to separate input spoken digits, leading to a recognition rate > 99.6%

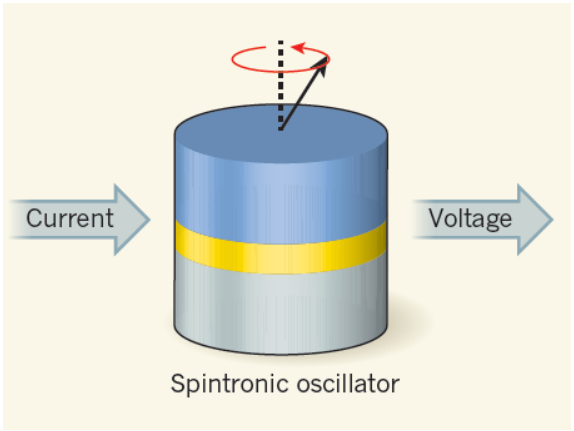
Neuron (time-multiplexed)

Synapses (computer)

Audio file

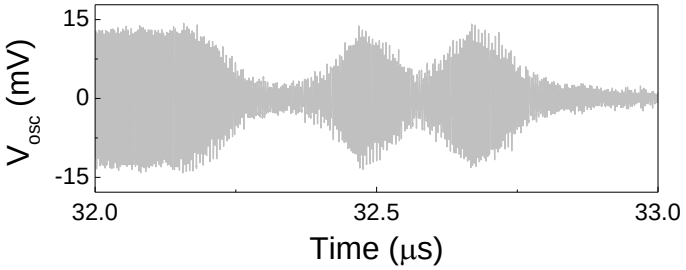


TI-46 database, 5 female speakers, cochlear pre-processing



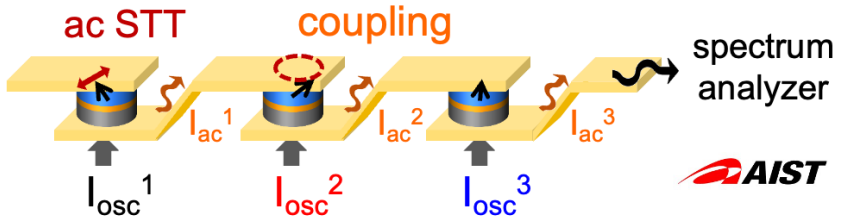
Reservoir computing

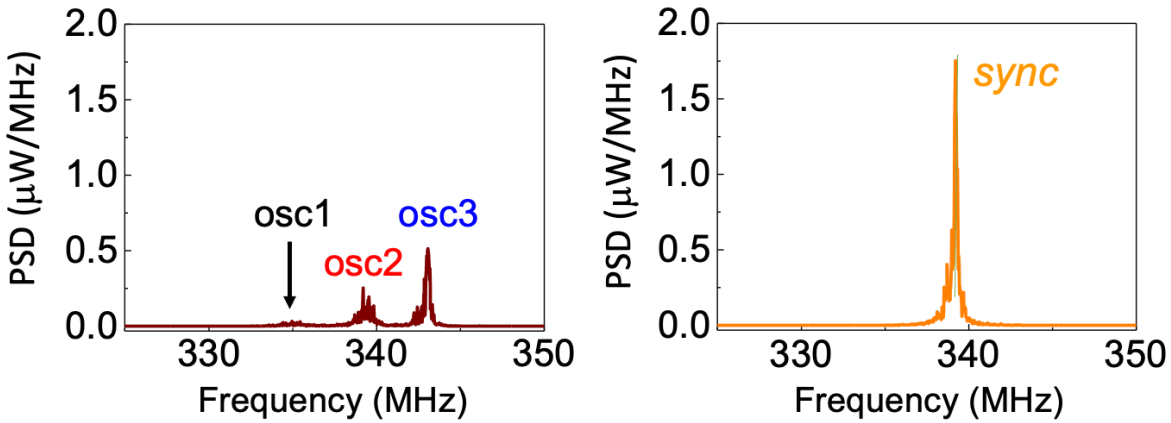
More complex than input → separation

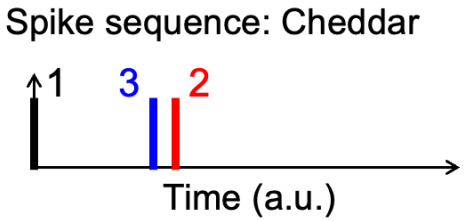


J. Torrejon, M. Riou, F. Abreu Araujo et al, Nature 547, 428 (2017)

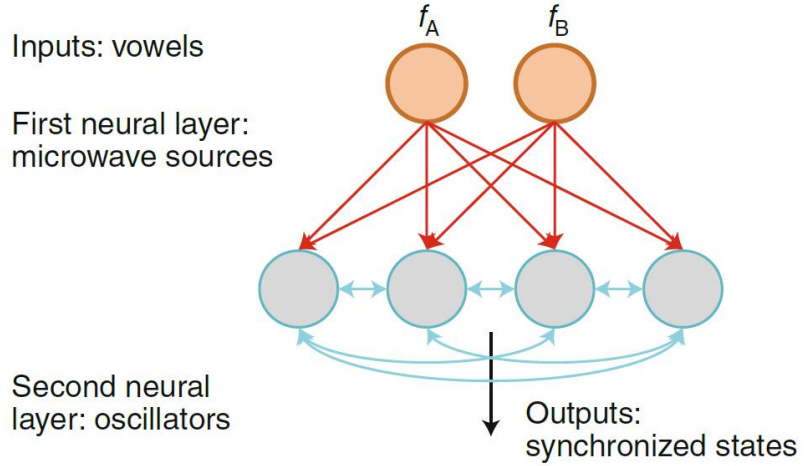
Coupled oscillators classify inputs patterns through synchronization

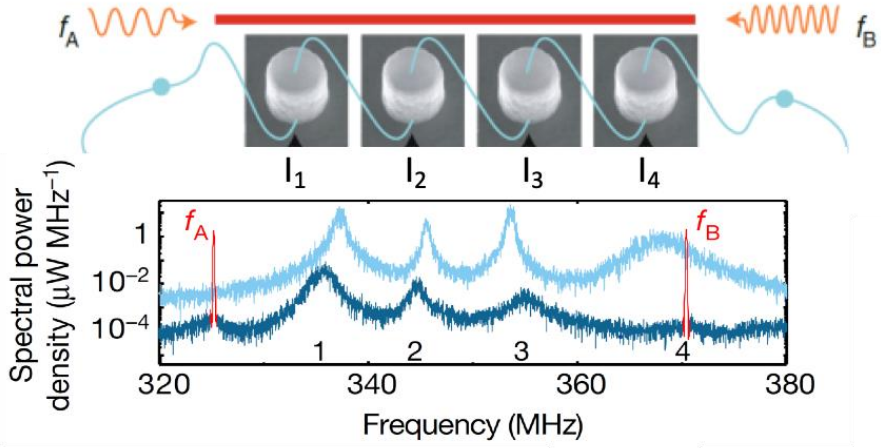






M. Romera, P. Talatchian et al, Nature Com 13, 883 (2022)



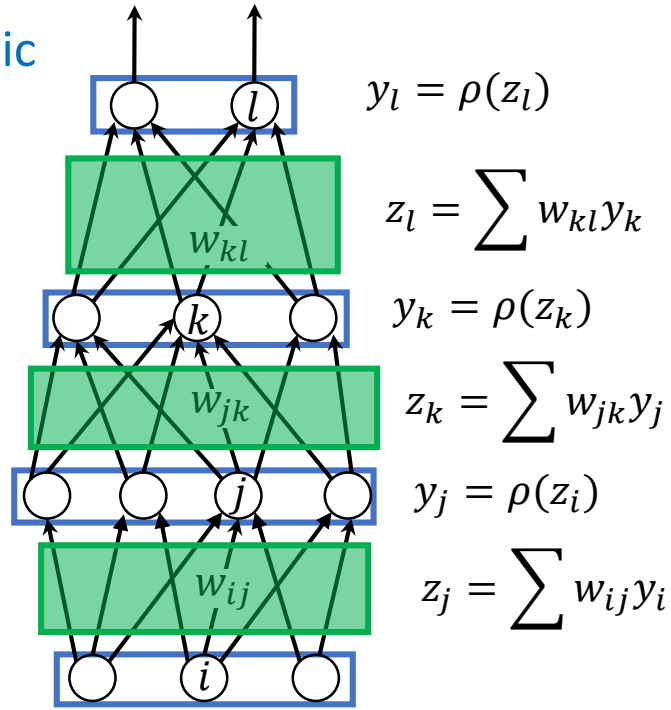


M. Romera, P. Talatchian et al, Nature 563, 230 (2018)

Challenges for building hardware neural networks are the number of components and their connectivity

Electronic neurons

Electronic synapses



$$y_l = \rho(z_l)$$

$$z_l = \sum w_{kl} y_k$$

$$y_k = \rho(z_k)$$

$$z_k = \sum w_{jk} y_j$$

$$y_j = \rho(z_i)$$

$$z_j = \sum w_{ij} y_i$$

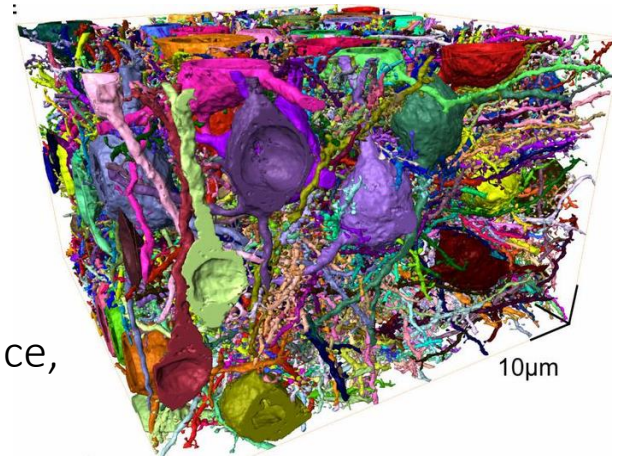
Image recognition: 10^8 neurons and synapses

Brain: 10^{11} neurons, 10^{15} synapses

→ Huge number of nanodevices on chip

Brain : 10^4 synapses/neurons

→ Huge connectivity

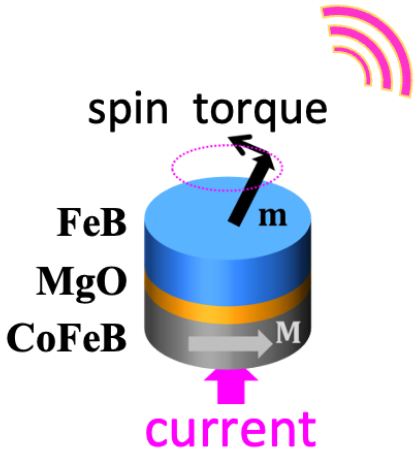


Motta et al, Science, 366, 6469 (2019)

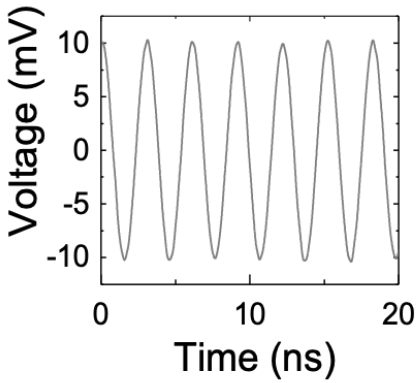
1 – Radio-frequency spintronic neural networks

2 – Spin-based neural network learning algorithms

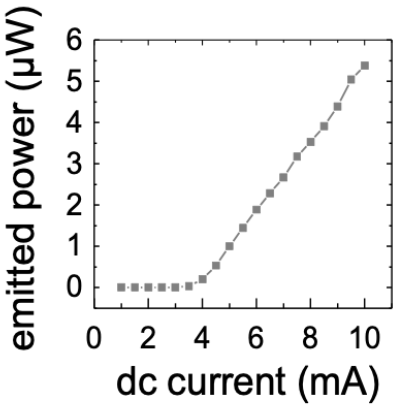
Our idea is to build deep neural networks where components communicate through radio-frequency signals

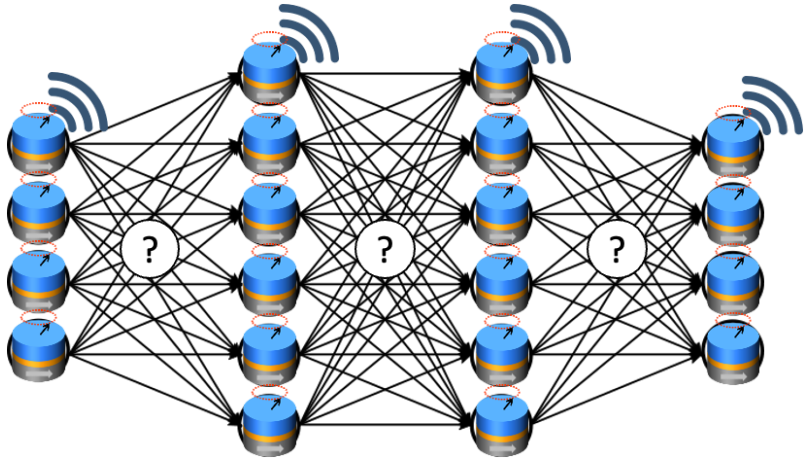


10 MHz → 20 GHz



Like a ReLu





The spintronic oscillator operates as a neuron that converts DC to RF
→ we need a synapse that converts RF to DC

N. Leroux, A. Mizrahi et al, Radio-Frequency Multiply-and-Accumulate Operations with Spintronic Synapses, Phys. Rev. Applied 15, 034067 (2021)

Research performed by:

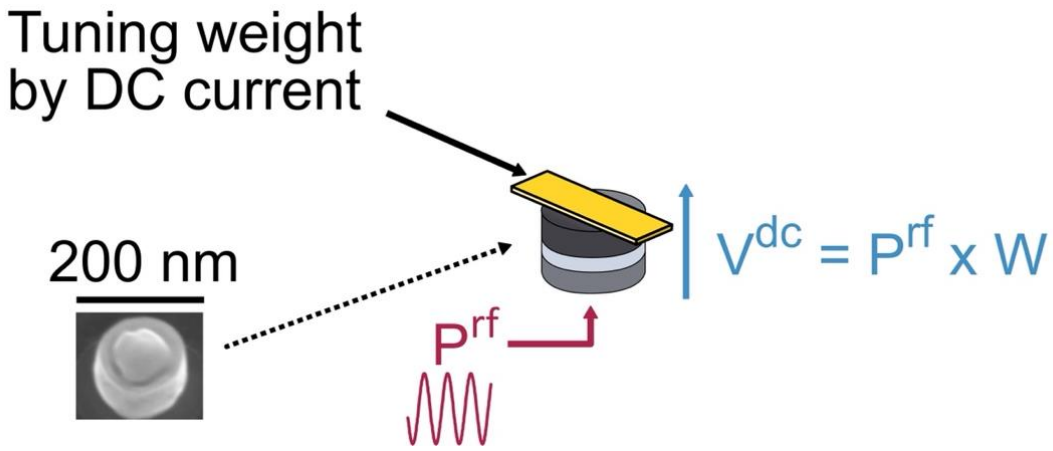
CNRS/Thales, France

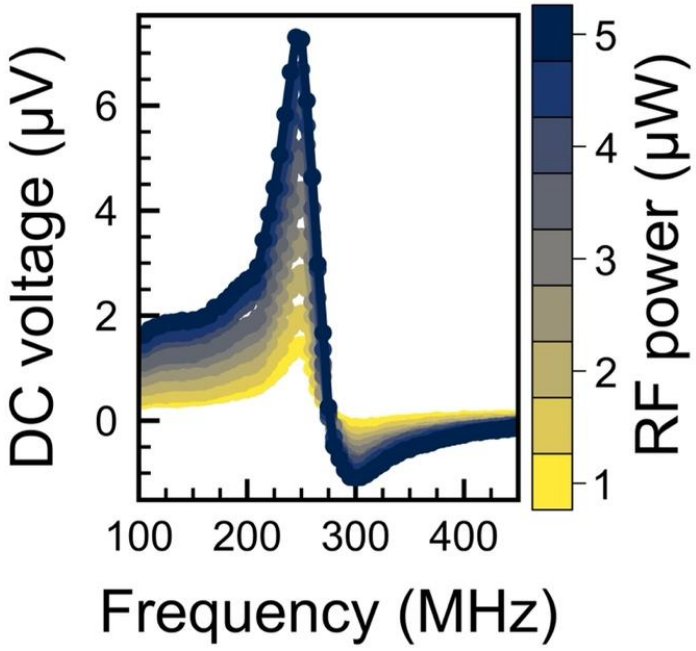
Nathan Leroux, Andrew Ross, Danijela Marković, Dedalo Sanz Hernandez, Erwan Plouet, Erwann Martin, Teodora Petrisor, Paolo Bortolotti, Juan Trastoy, Alice Mizrahi

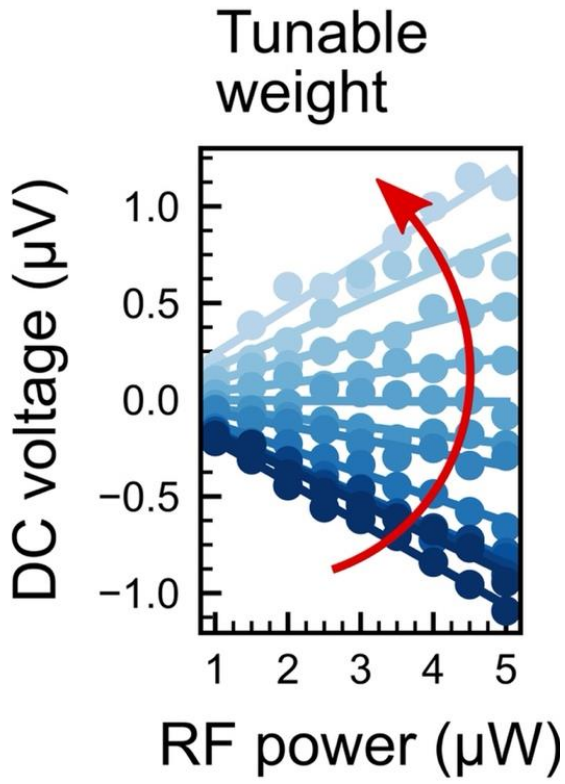
INL, Portugal

Leandro Martins, Alex Jenkins, Ricardo Ferreira

Magnetic tunnel junctions multiply input RF signals by a synaptic weight



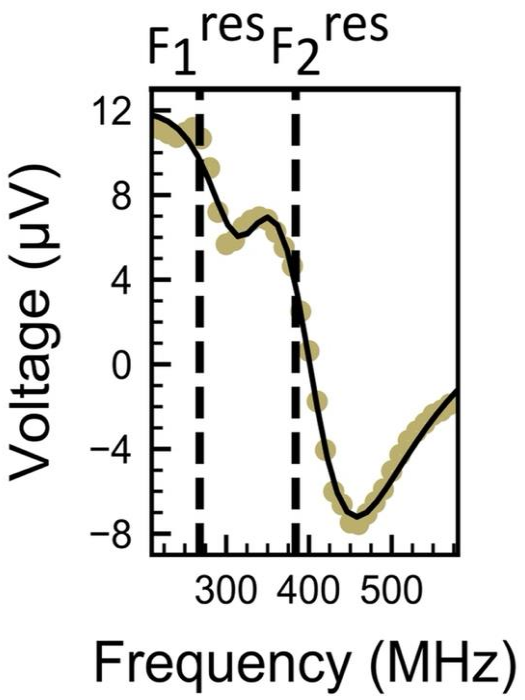


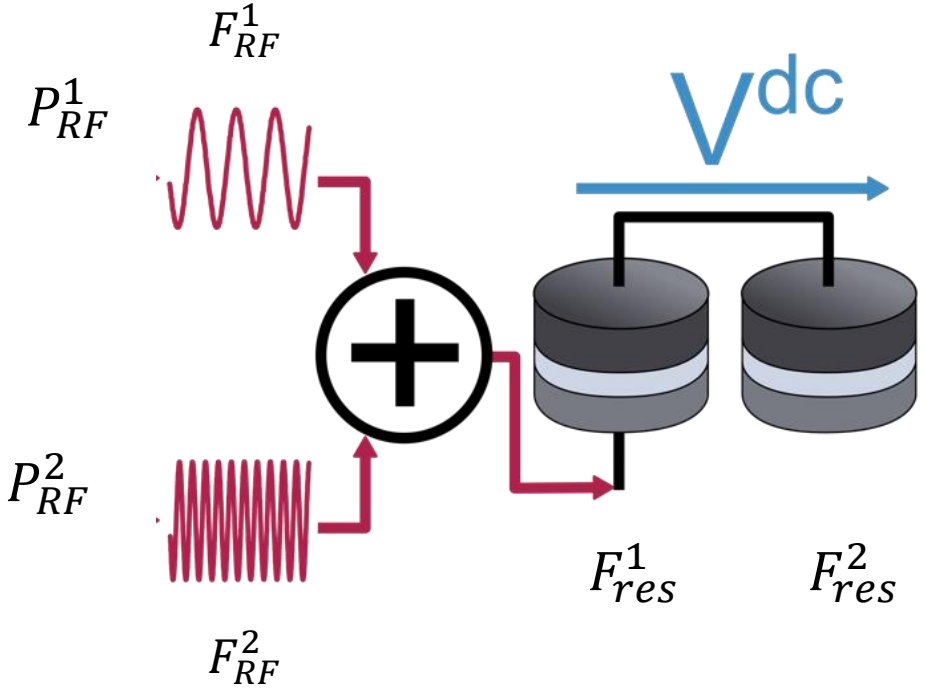


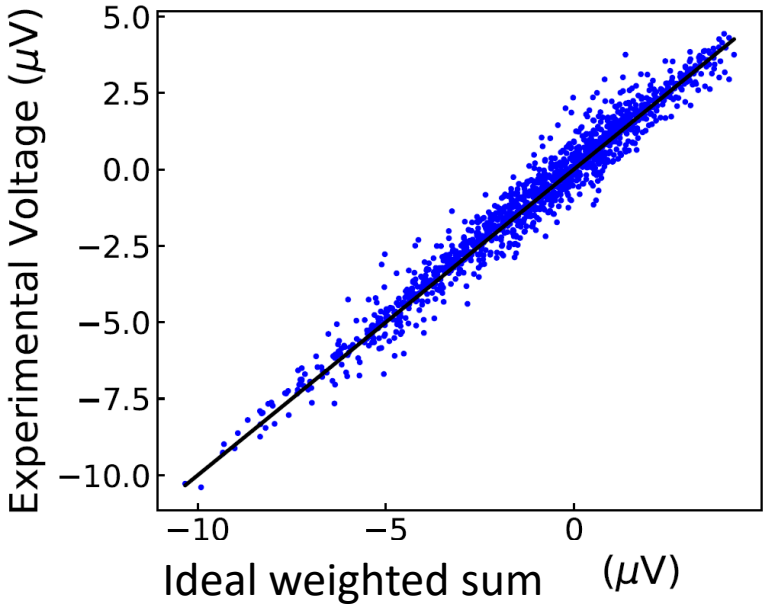
Nathan Leroux, Alice Mizrahi et al, *Physical Review Applied* 15, 034067 (2021)

Weighted sum operation on RF signals is achieved through frequency multiplexing

$$V_{dc} = \sum_i P_{RF}^i W(F_{res}^{1i})$$

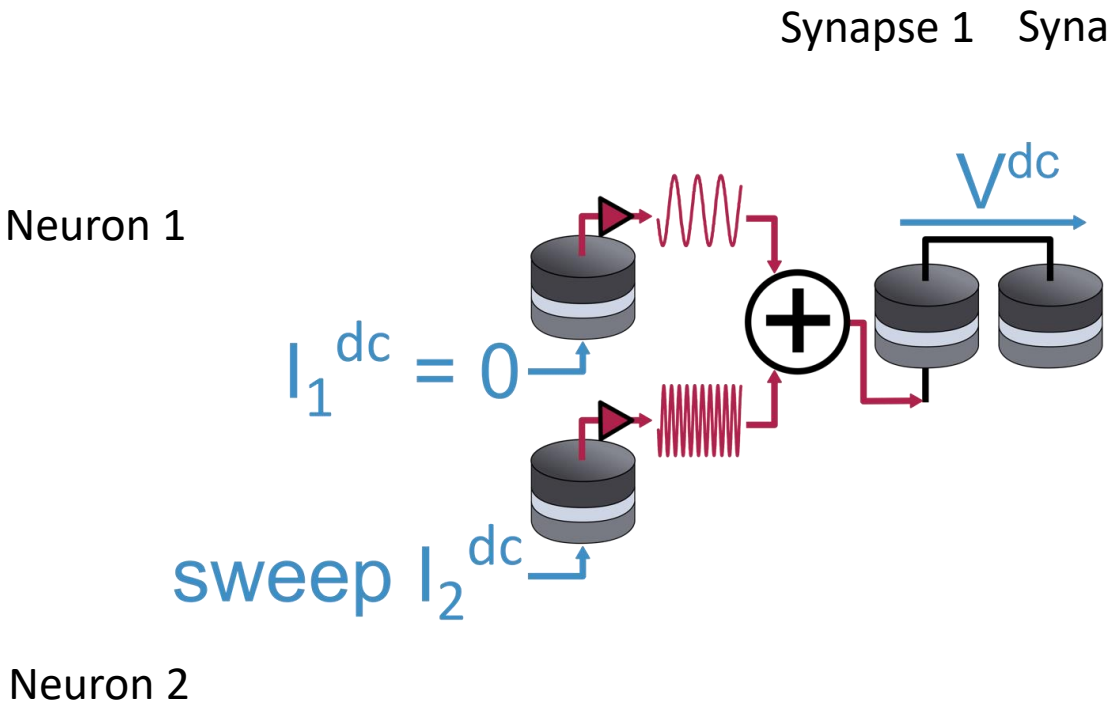


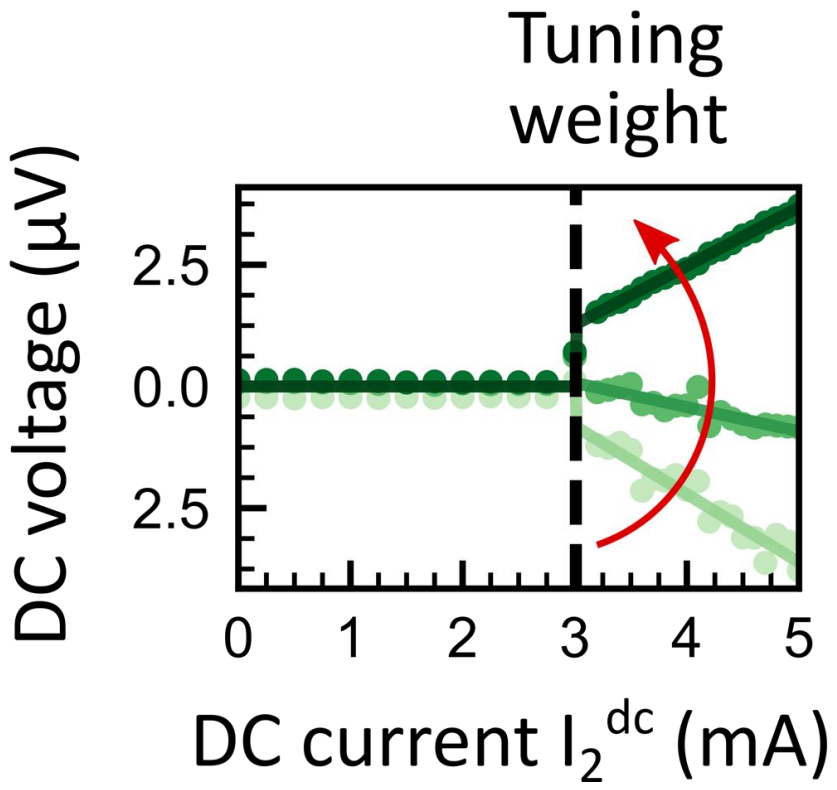




Nathan Leroux, Alice Mizrahi *et al Neuromorph. Comput. Eng.* **1** 011001 (2021)

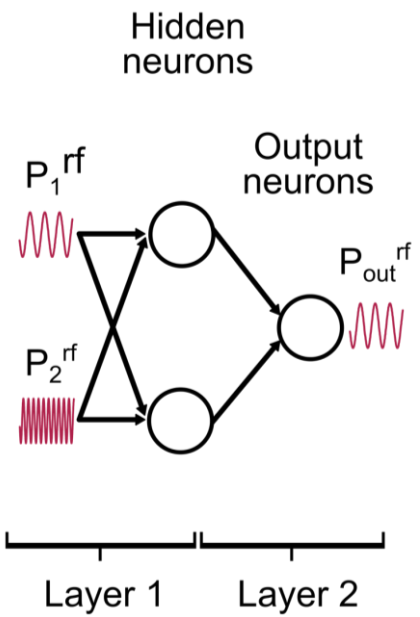
Neurons-to-synapses connection

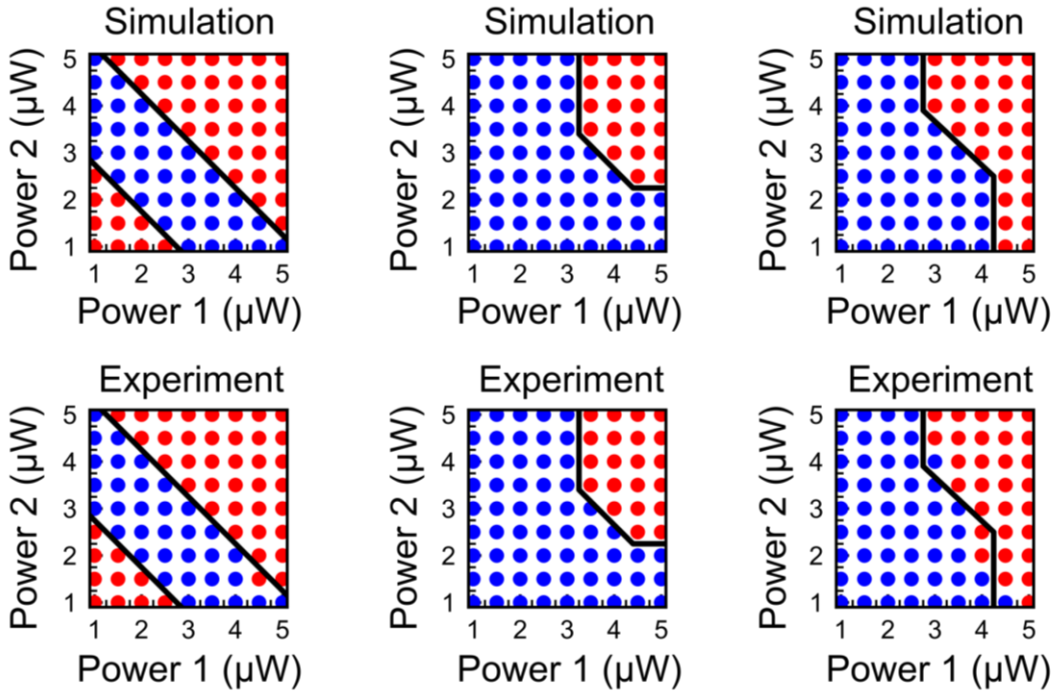


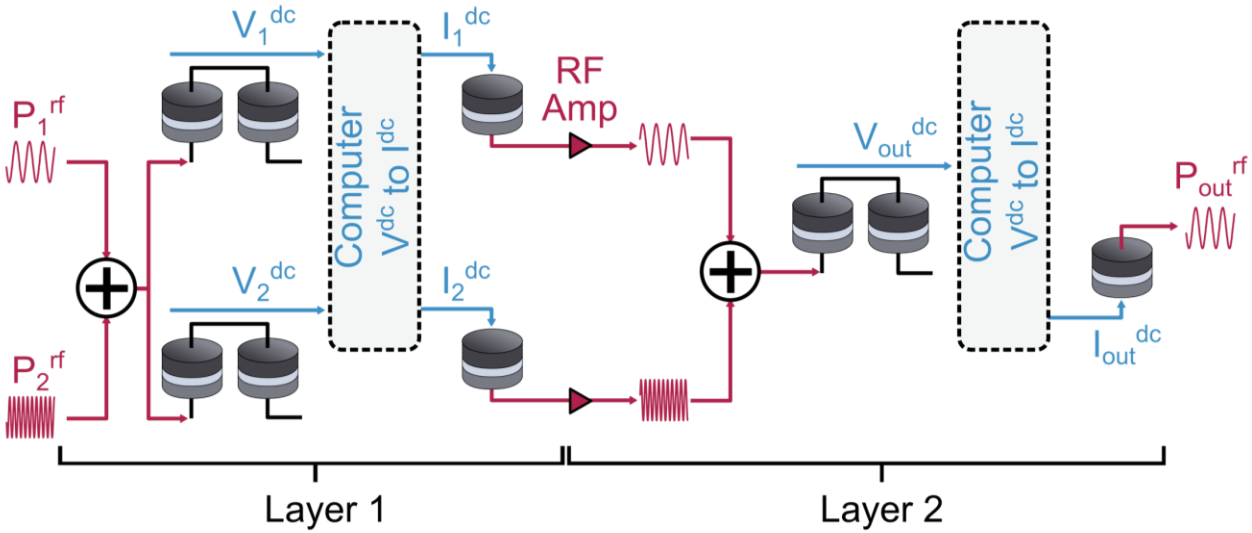


The output of neuron 2 is tuned by synaptic weight 2

Fully spintronic hardware neural network







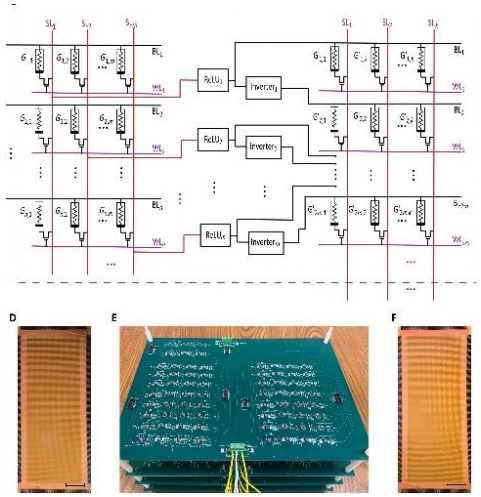
97.7% accuracy

Comparison with other nanotechnologies

We have built the first multilayer fully-nano neural network

Memristor nano-synapses connected by off-the-shelf neuron circuits

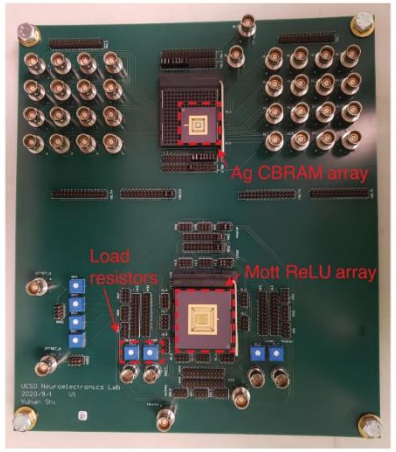
Nano-Synapses → *Not nano* → ***Nano-Synapses***

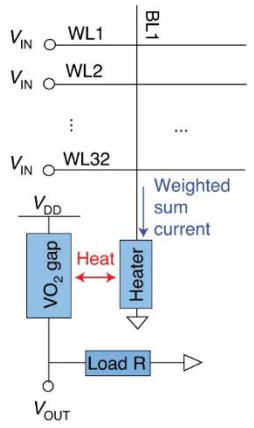


Kiani et *al., Sci. Adv.*, 7, 48 (2021)

Memristor Nano-synapses connected **to** memristor nano-neurons

Nano-Synapses → ***Nano-Neurons***

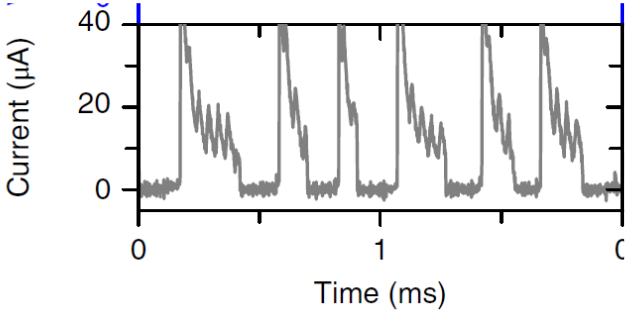




Oh *et al., Nat. Nano.* 16 (2021)

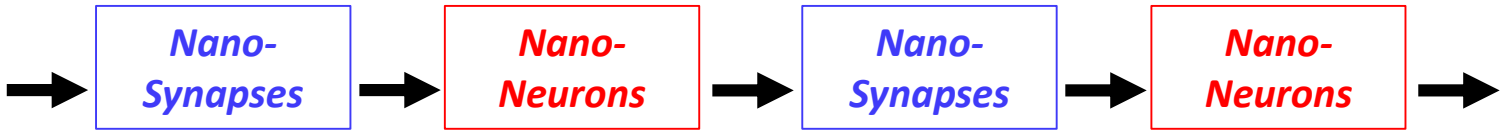
Connecting nano-neurons to nano-synapses is a challenge

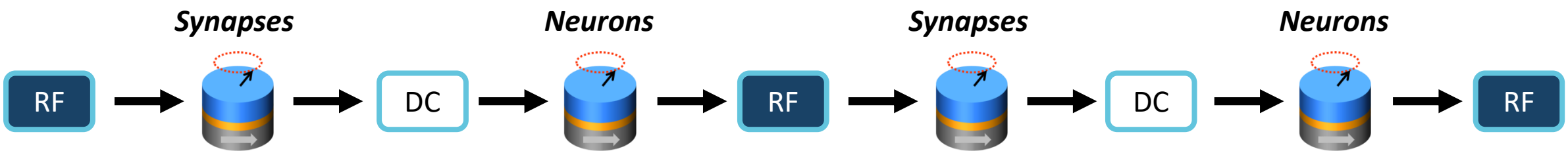
Nano-Neurons → ***Nano-Synapses***



Z. Wang et al, Nature Electronics 1, 137 (2018)

Spintronic RF-to-DC and DC-to-RF conversions enable a nice propagation of signals along the multilayer stack





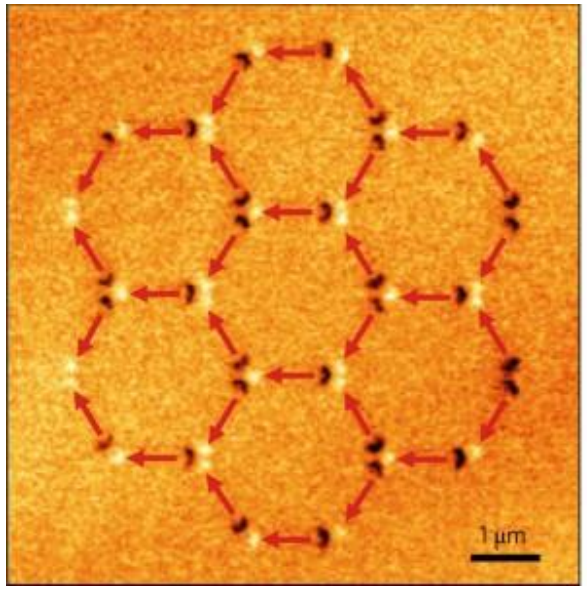
⇒ Same device for neurons and synapses

⇒ Stack layers alternating DC and RF for depth

Connectivity

Most physical systems are naturally locally connected

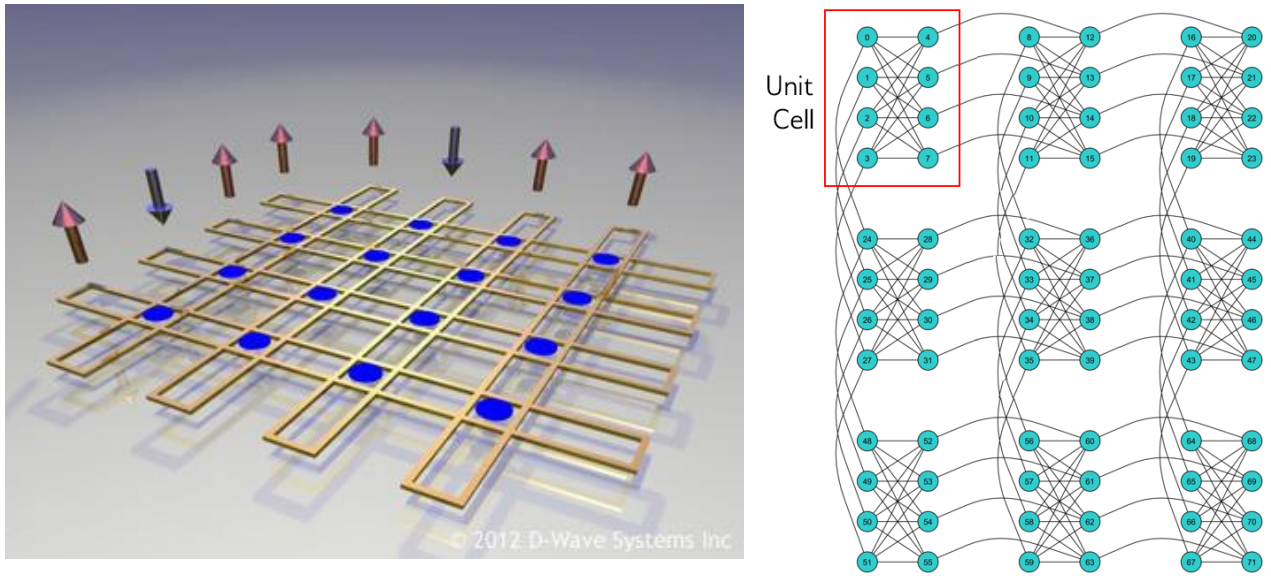
Spin ice lattice



J. C. Gartside et al, Nature Nano 13, 53 (2018)

Each magnet is connected to 4 others

D Wave squid Qubits

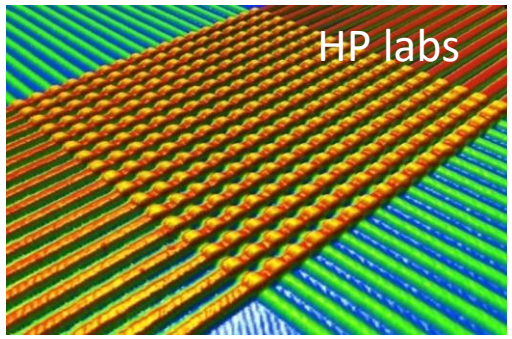


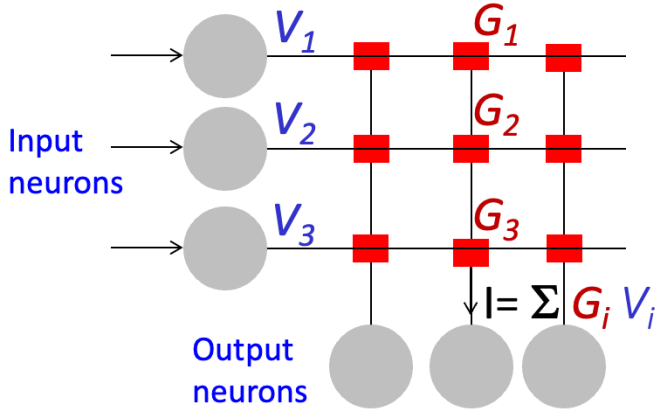
8 qubits connected all to all through JJs

2000 spins in total, but only ~50 spins can be coupled all to all through embedding

Our concept is a sneak path free alternative to crossbar arrays

Crossbar arrays

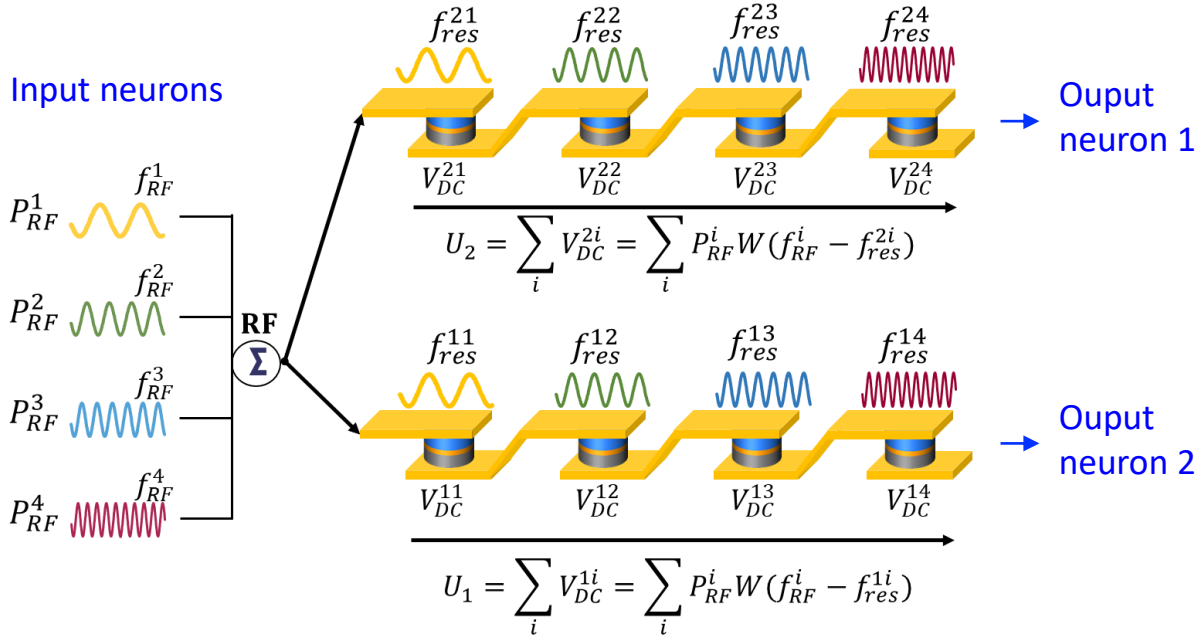




- Requirement: high OFF/ON ratio devices
- sneak paths

~ 100 synapses per neuron

RF frequency multiplexing



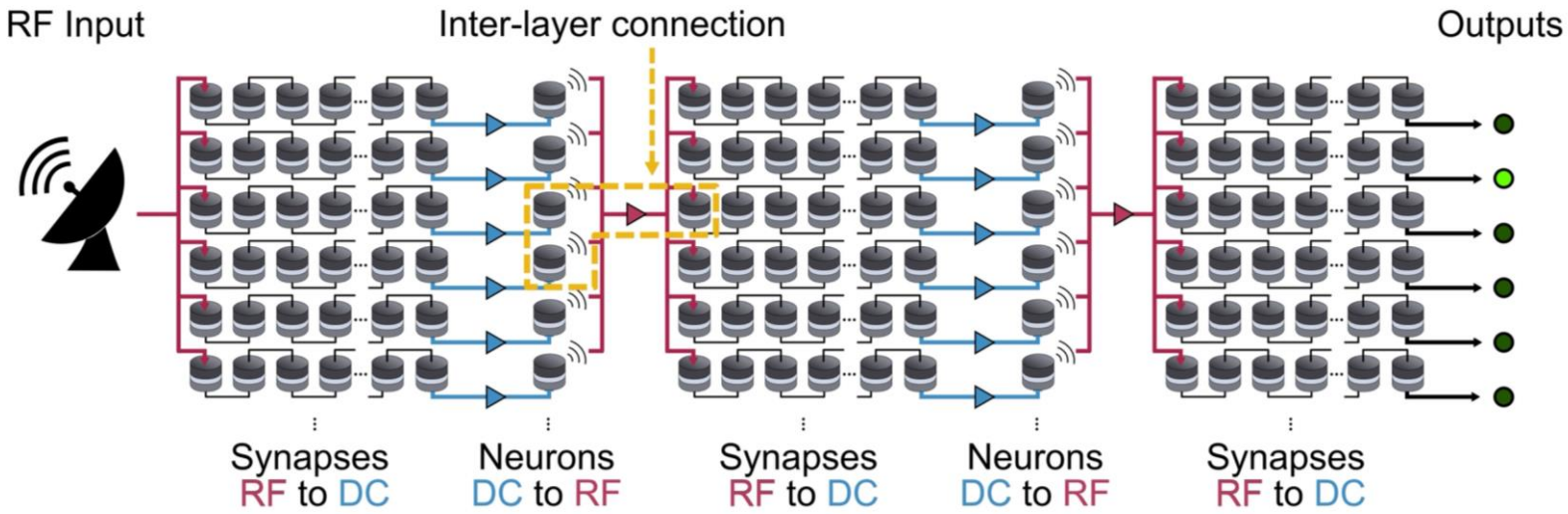
- Requirement: good RF signal propagation
- **No** sneak paths

~ 1000 synapses per neuron

Towards deep RF spintronic neural networks

- **First experimental** demonstration of **communication** between **multiple layers** using nano-devices
- **Non-linear classification** of two inputs with **high accuracy**

10 fJ/synapse and 100 fJ/neuron for MTJs with 20 nm diameter: 100 fold energy gain compared to GPUs

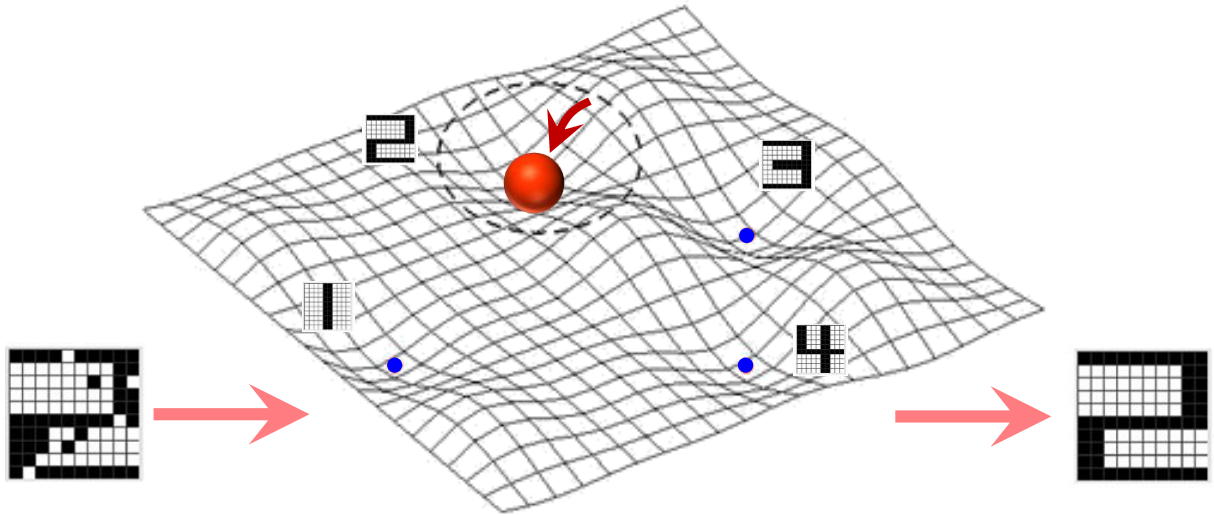


Theory weighted sum: N. Leroux, A. Mizrahi *et al. Phys. Rev. Appl.* **15**, 034067 (2021)
Experiment weighted sum: N. Leroux, A. Mizrahi *et al*., Neuromorph. Comput. Eng. **1**, 011001 (2021)
Full network experiment: A. Ross, N. Leroux *et al*. *Manuscript in Preparation*
Simulations: N. Leroux *et al Neuromorph. Comput. Eng.* **2,** 034002 (2022)

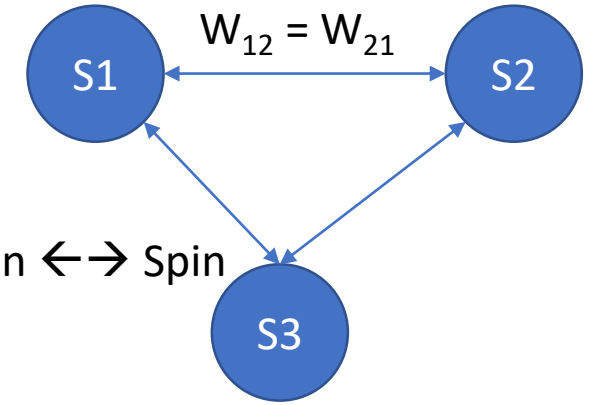
1 – Radio-frequency spintronic neural networks

2 – Spin-based neural network learning algorithms

Hopfield nets compute by minimizing an Ising energy



$$E = -\sum_{i<j} w_{ij} s_i s_j - \sum_i b_i s_i$$



Learning rule: $\Delta w_{ij} \mathrel{+}= s_i^p s_j^p$

for each pattern p,
one after the other

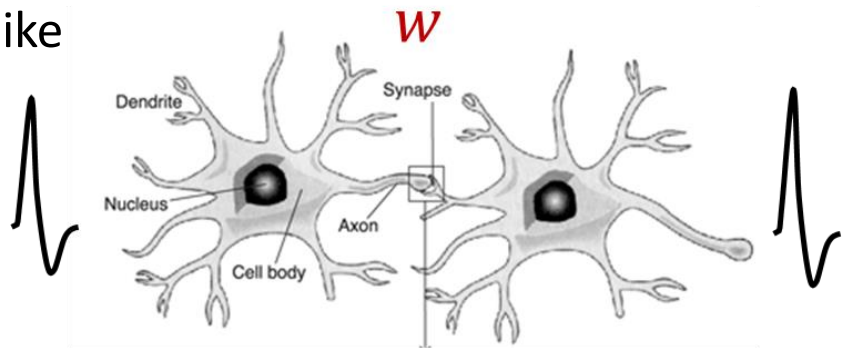
$$E(pattern\ p) = -\sum_{i<j} \left(s_i^p s_j^p\right)^2 < 0$$

J. J. Hopfield, PNAS 79, 2554 (1982)

Modern algorithms minimize the error at the network output

Local learning rules

Brain-like



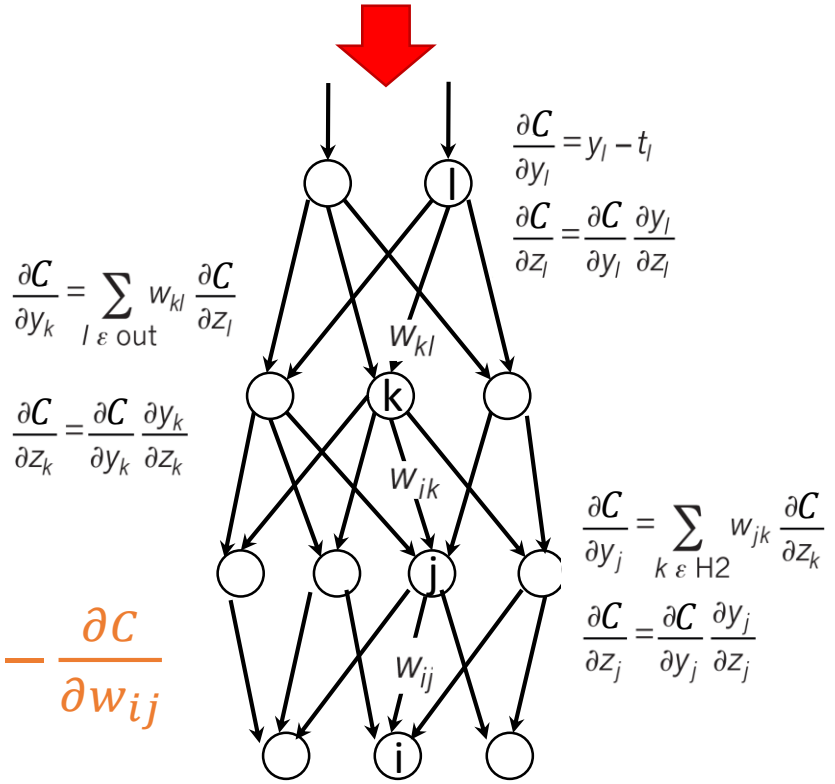
Hebbian learning rules (1949): who fires together wires together

$$\Delta w_{ij} \mathrel{+}= s_i^p s_j^p$$

Easy to implement in hardware
Low performance on complex tasks

Backpropagation of errors

Cost function: $C = \frac{1}{2}(y_l - t_l)^2$

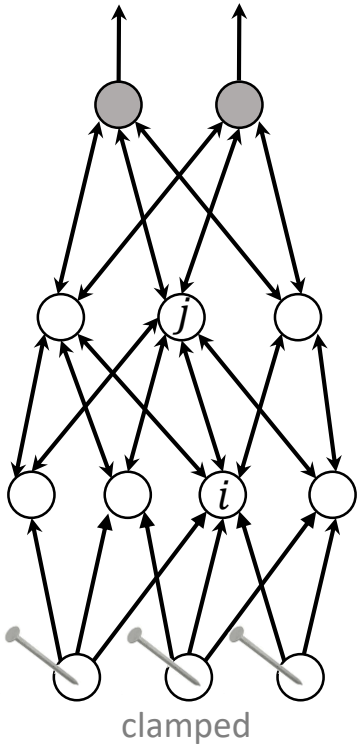


$$\Delta w_{ij} \propto -\frac{\partial C}{\partial w_{ij}}$$

Hard to implement in hardware
High performance on complex tasks

Equilibrium propagation minimizes both the energy and the output error of the system

Inference : free phase



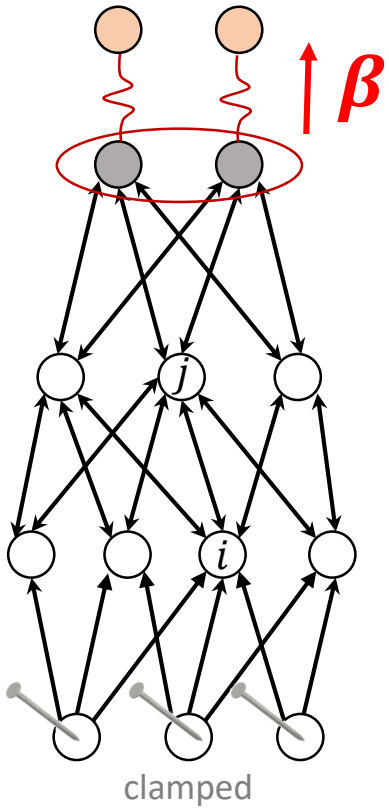
Equilibrium: $s_i{}^0$

Nudging toward the desired solution

$$\boldsymbol{F = E + \beta C}$$

Cost function: C

$$= \frac{1}{2}(y_l - t_l)^2$$



→ Spring like force on output neurons

$$\sim -(y_l - t_l)$$

Equilibrium: $s_i{}^\beta$

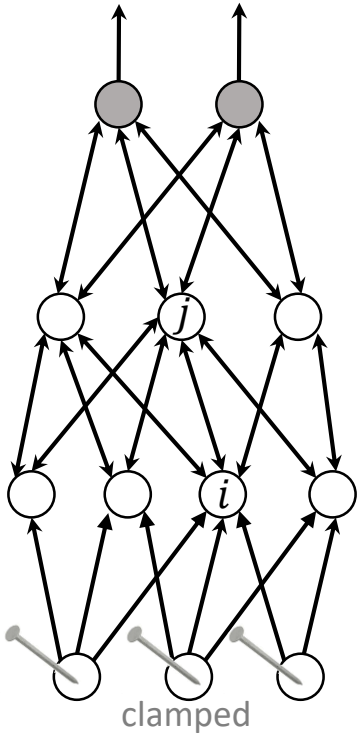
The perturbation at the output propagates through the network

Scellier & Bengio, fnins 2017

$$\Delta w_{ij} = -s_i{}^0 s_j{}^0 + s_i{}^\beta s_j{}^\beta$$

Equilibrium propagation directly backpropagates gradients through neuronal dynamics

Inference : free phase

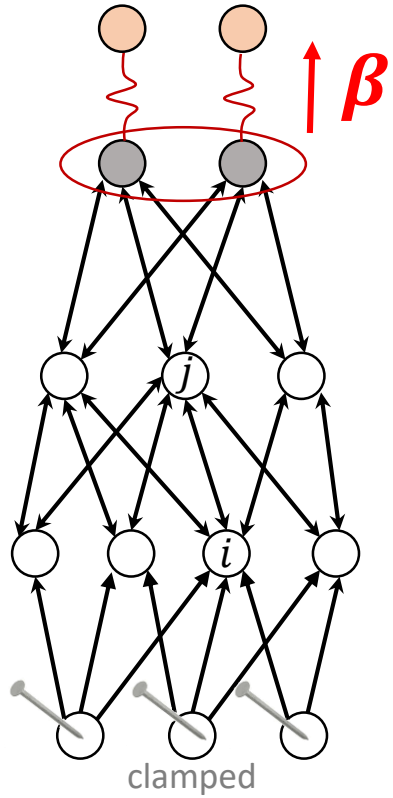


$$\frac{\partial E}{\partial s_i}(s_i^0) = 0$$

Wait for equilibrium, measure ${s_i}^0$

The perturbation at the output propagates through the network

Propagation of gradients : nudging phase $\boldsymbol{F = E + \beta C}$



$$\frac{\partial E}{\partial s_i}\left(s_i^\beta\right) + \beta\frac{\partial C}{\partial s_i} = 0$$

Wait for equilibrium, measure ${s_i}^\beta$

Difference between the two equilibrium states → $\Delta s_i \propto -\frac{\partial C}{\partial s_i}$ → $\Delta w_{ij} \propto -\frac{\partial C}{\partial w_{ij}}$

Our recent work: bridging Equilibrium Propagation and hardware

CNRS/Thales, France

Jérémie Laydevant, Erwann Martin, Dongshu Liu, Shuai Li, Danijela Marković, Julie Grollier

C2N, France

Maxence Ernoult, Axel Laborieux, Damien Querlioz

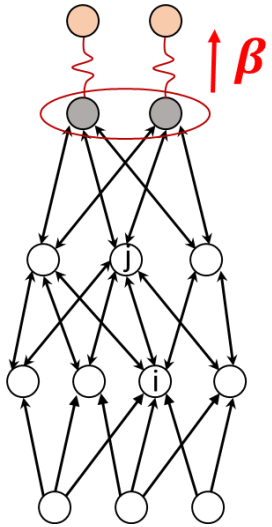
Thales, France

Erwann Martin, Teodora Petrisor

Mila, Canada

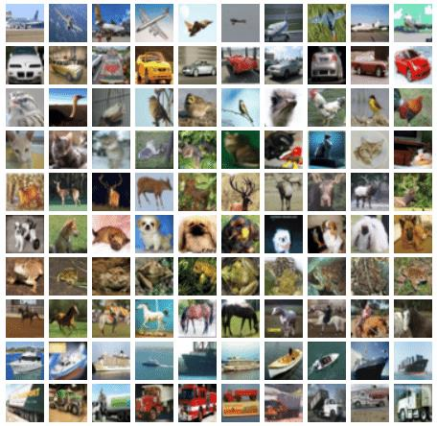
Benjamin Scellier, Yoshua Bengio

The local learning rule of Equilibrium Propagation (EP) leads to high recognition rates on image benchmark tasks



Equivalence of EP local learning rule with the gradients of Backpropagation Through Time (maths + simulation results)

$$\Delta W_{ij} = -\rho(s_i)\,\rho(s_j) + \rho^{\beta}(s_i)\,\rho^{\beta}(s_j)$$



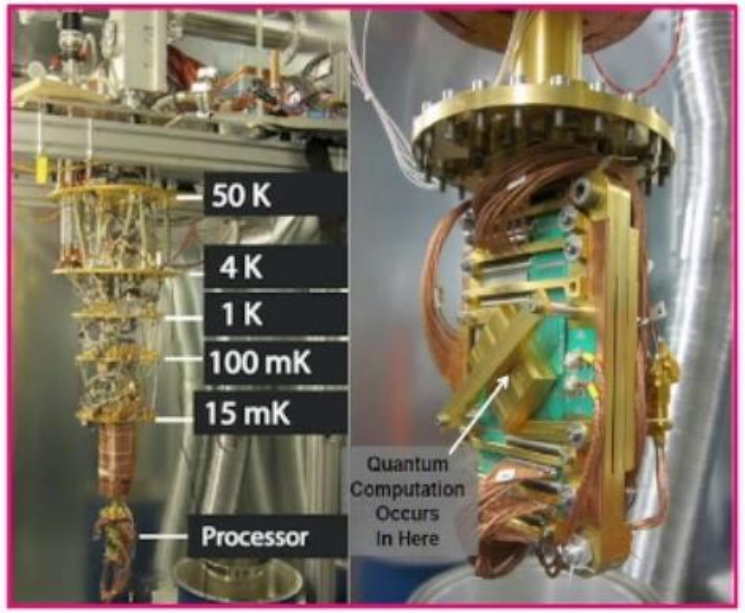
EP implements convolutions and scales to CIFAR-10

Modified 3-phase learning rule:
EP : 11.68% test error
BPTT : 11.1 %

M. Ernoult, J. Grollier, D. Querlioz, Y. Bengio, B. Scellier, NeurIPS (2019)

A. Laborieux, M. Ernoult, B. Scellier, Y. Bengio, J. Grollier, D. Querlioz, Frontiers Neuro. 15, 129 (2021)

We have implemented a full neural network and trained it to solve MNIST through Equilibrium propagation on the D'Wave quantum computer



Ising machine

Work of Jérémie Laydevant and Danijela Marković

1st challenge: design a binary version of Equilibrium propagation compatible with Ising machines

- D-wave is an Ising machine based on 2-states qubits s_i

$$E = -\frac{1}{2}\sum_{i\neq j} w_{ij} s_i s_j \; - \sum_i b_i s_i$$

- Equilibrium Propagation uses analog neurons $\rho(s_i)$

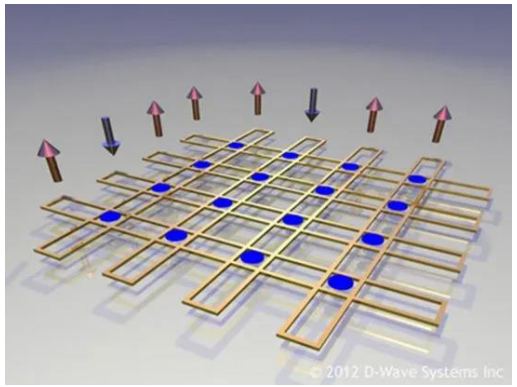
$$E = -\frac{1}{2}\sum_{i\neq j} w_{ij}\rho(s_i)\rho(s_j) - \sum_i b_i \rho(s_i) + \frac{1}{2}\sum_i {s_i}^2$$

→ We have demonstrated in software that EP trains Binary Neural Networks by using advanced Machine Learning methods

- Binary synapses (CIFAR 10)
- Binary synapses and neurons (MNIST)
- Ternary gradients

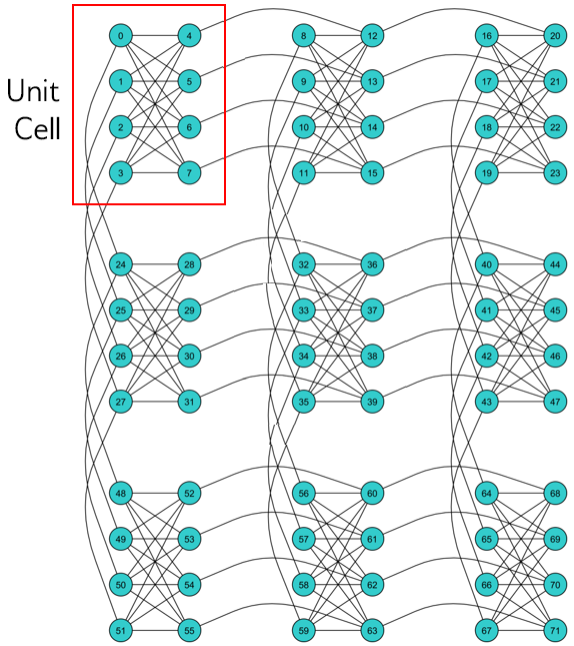
J. Laydevant, M. Ernoult, D. Querlioz, J. Grollier, arxiv 2103.08953, published in the Conference on Computer Vision and Pattern CVPR (2021)

2nd challenge: mapping a neural network that can classify handwritten digits from the MNIST database to the chip layout



Clusters of 4x4 coupled superconductive qubits

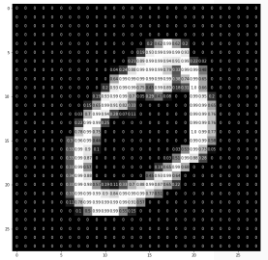
5000 qubits

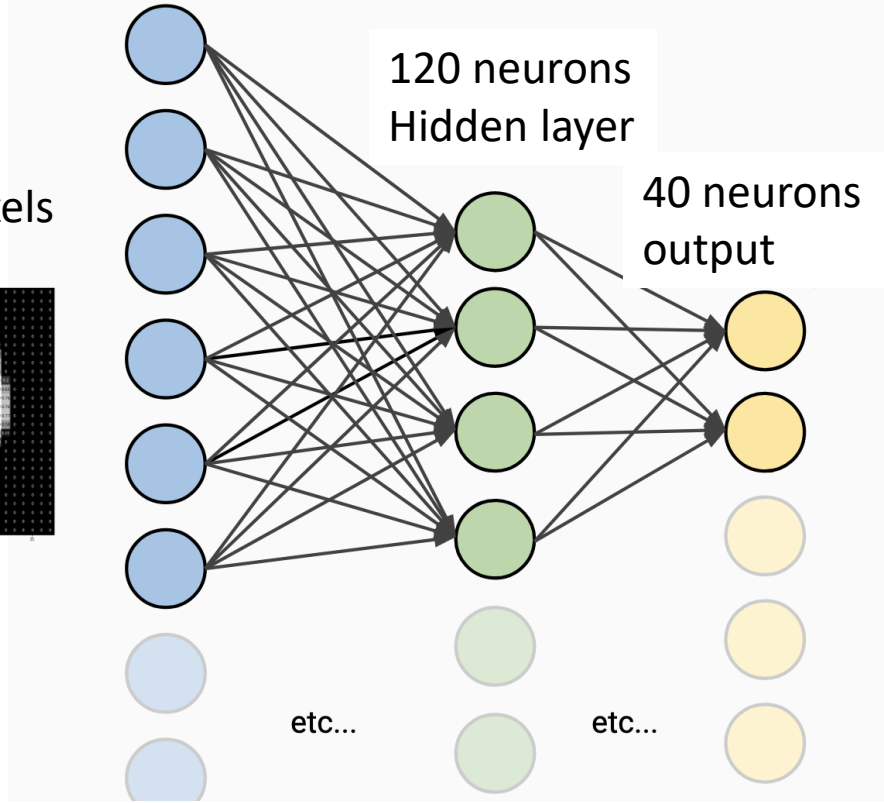


embedding

784 (28x28) inputs

28x28 pixels

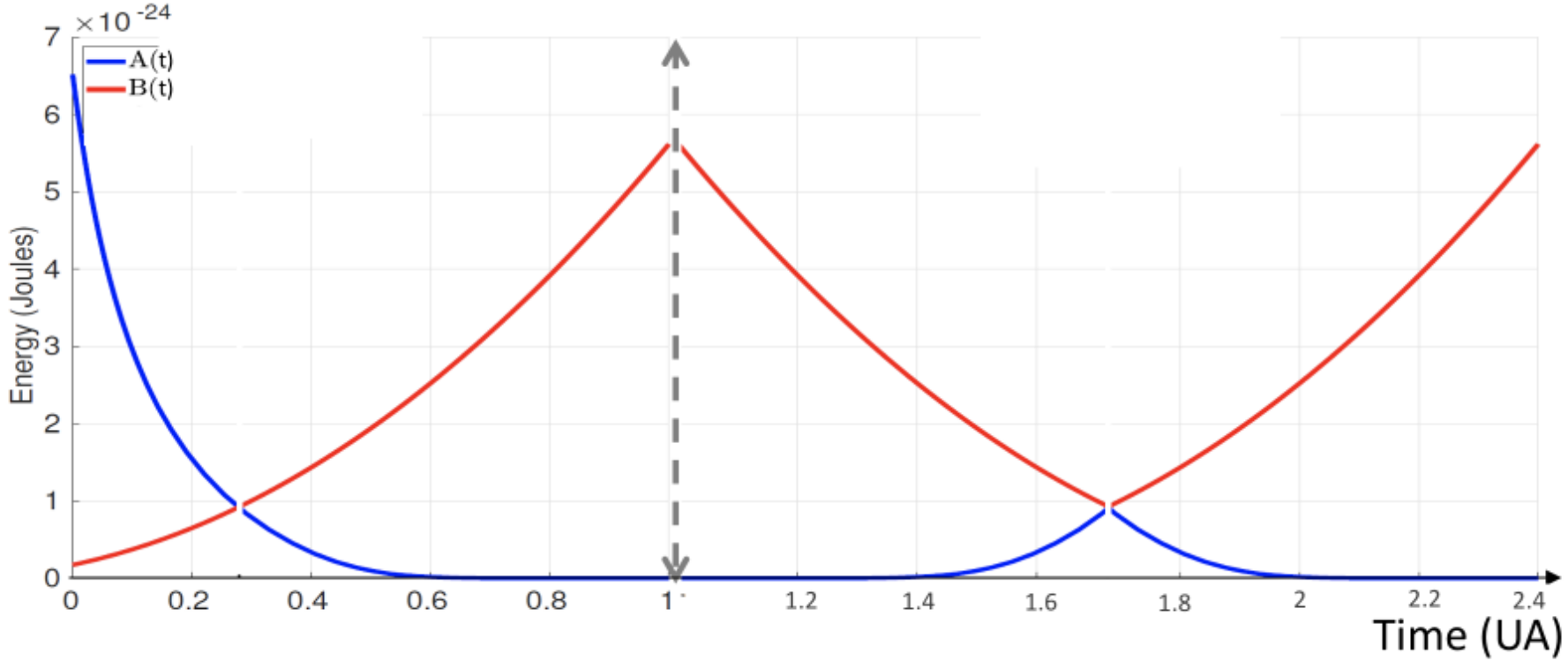




On chip

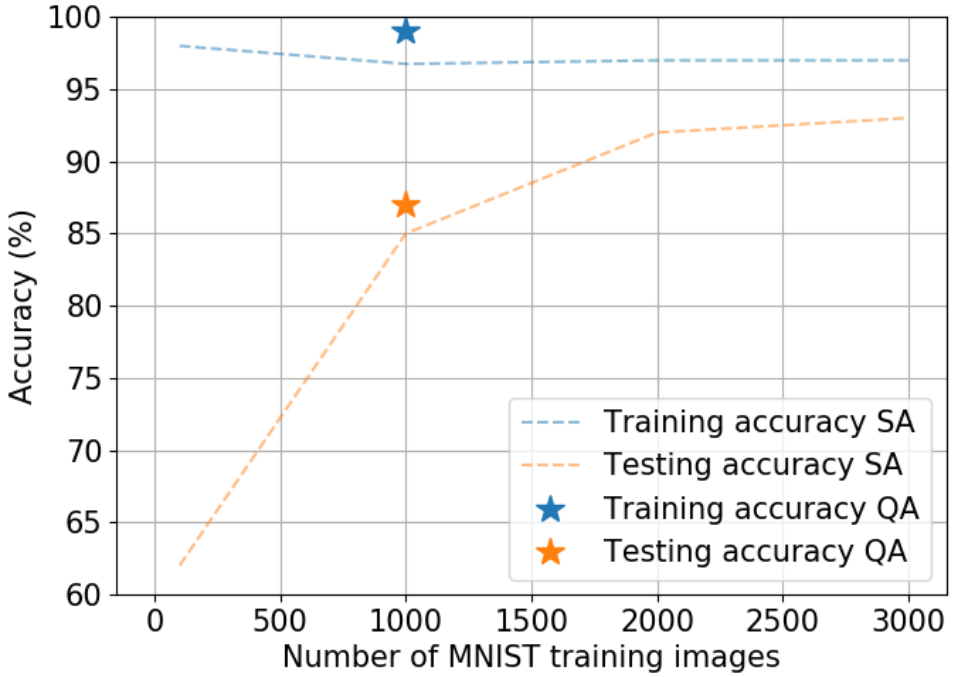
1 qubit ≠ 1 neuron due to the chip layout

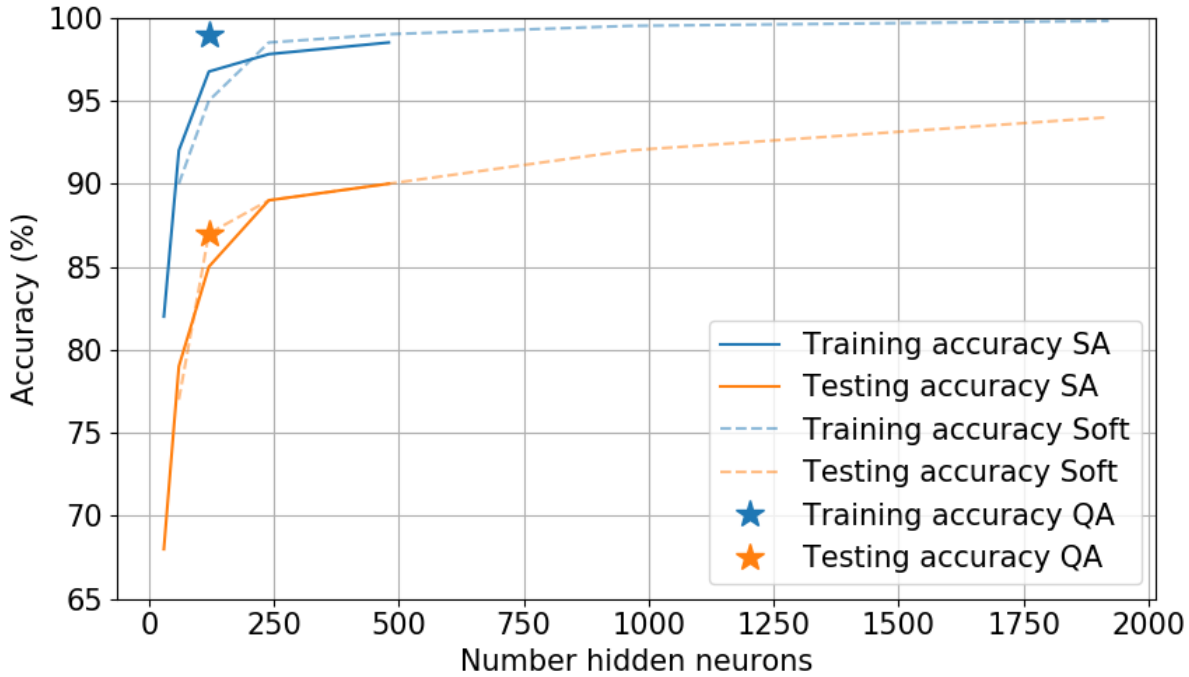
3rd challenge: hacking the quantum annealing procedure to reach the two different equilibrium points needed for EP



$$\mathcal{H}(t) = A(t) \cdot \mathcal{H}_0(\sigma_x) + B(t) \cdot \mathcal{H}_{Ising}^{free}(\sigma_z) \qquad \mathcal{H}(t) = A(t) \cdot \mathcal{H}_0(\sigma_x) + B(t) \cdot \mathcal{H}_{Ising}^{nudge}(\sigma_z)$$

First end-to-end supervised training of a neural network in an Ising machine (D-Wave) with software-equivalent accuracy (85%)

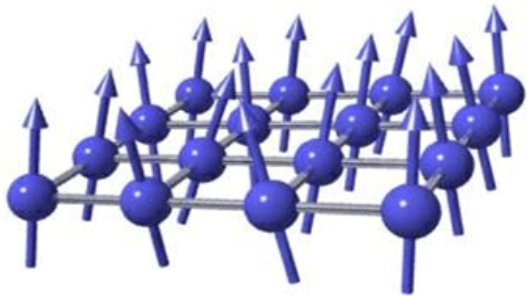




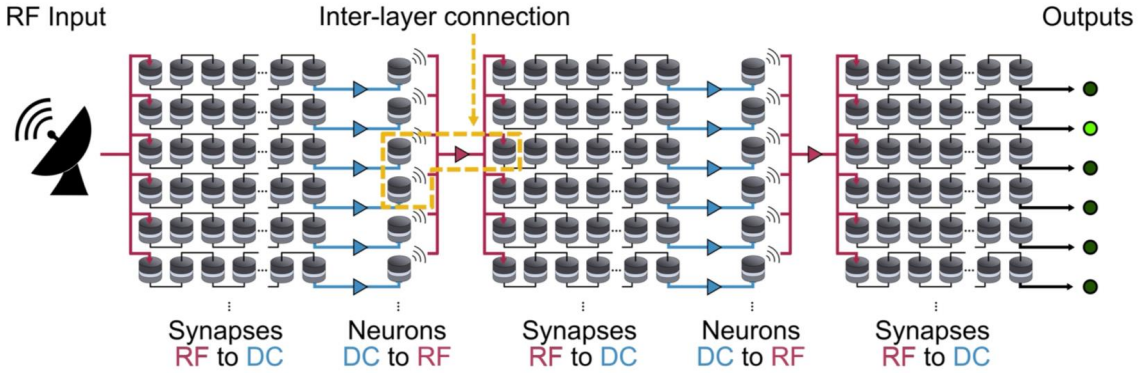
In the future, the accuracy can be improved by increasing:

- the number of presented images in the database
- the number of hardware spins

Spin-based learning algorithms can solve standard AI tasks and achieve SOTA accuracy

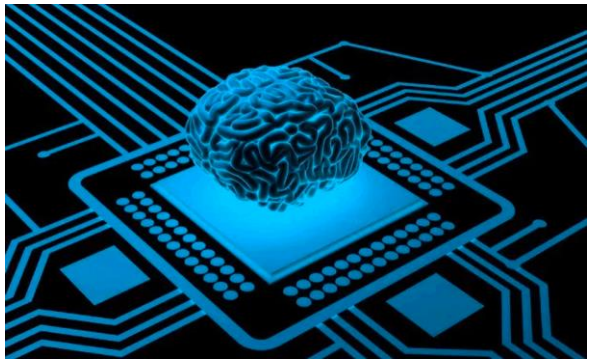


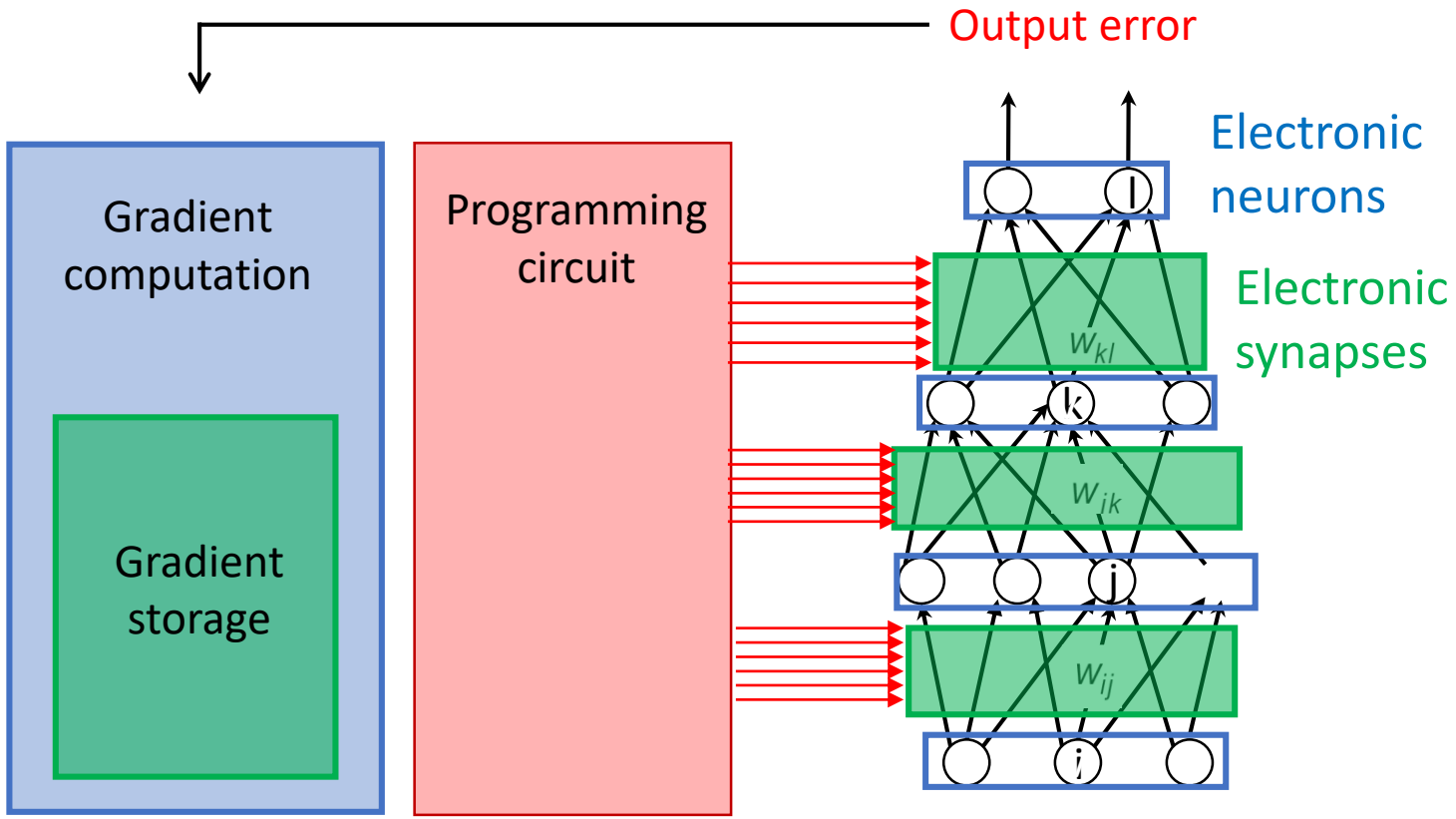
Spintronic nanodevices communicating through RF signals can implement large scale and low power hardware neural networks



Learning with high accuracy and low power is the main challenge of neuromorphic chips

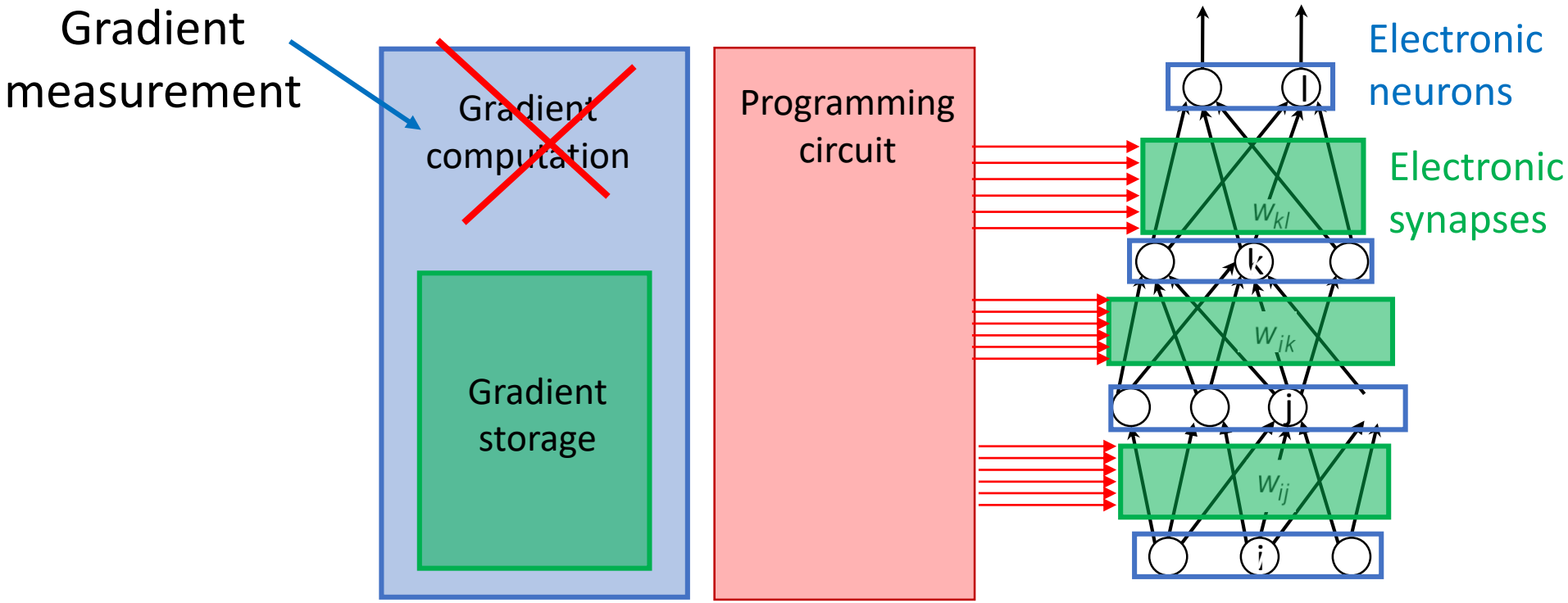
Neuromorphic chips today:
Limited to Handwritten digit classification





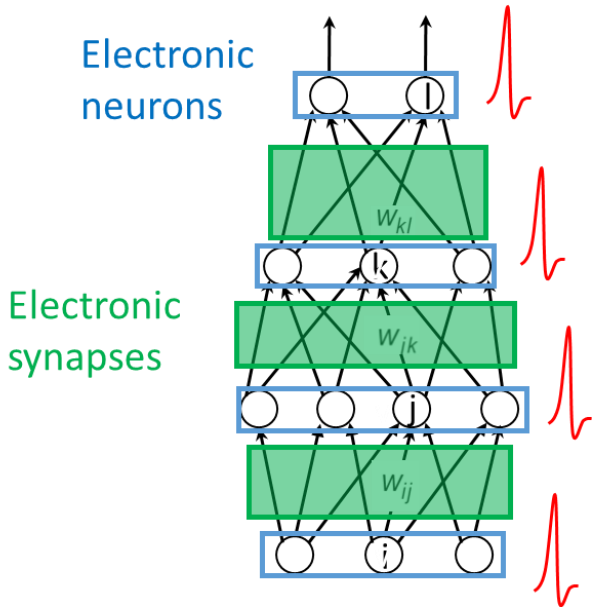
On-chip learning circuits are greedy in energy and area

Equilibrium propagation replaces gradient computation by gradient measurement



Can we do better ?

Can we train a hardware neural network through it physics, without additional « artificial circuits », with low power, and with high accuracy despite component variability, like in the brain ?



Intrinsic learning

Eqspike: E Martin, M Ernoult, J Laydevant, S Li, D Querlioz, T Petrisor, J Grollier, iScience 24, 102222 (2021)