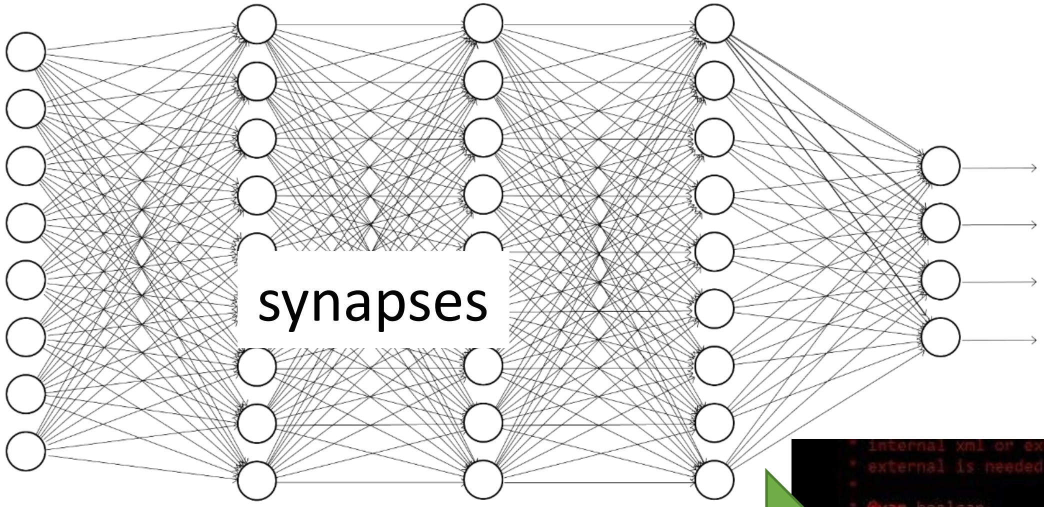
Neuromorphic computing: overview and challenges

Julie Grollier

Unité Mixte de Physique CNRS/Thales

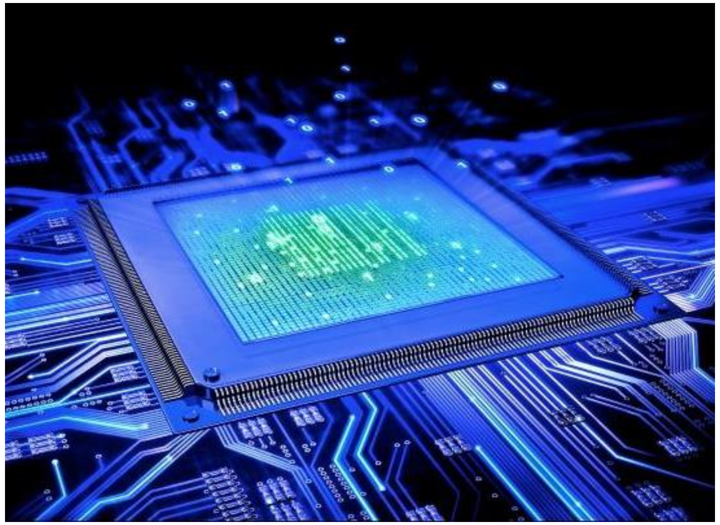
Neural networks run on unoptimized hardware



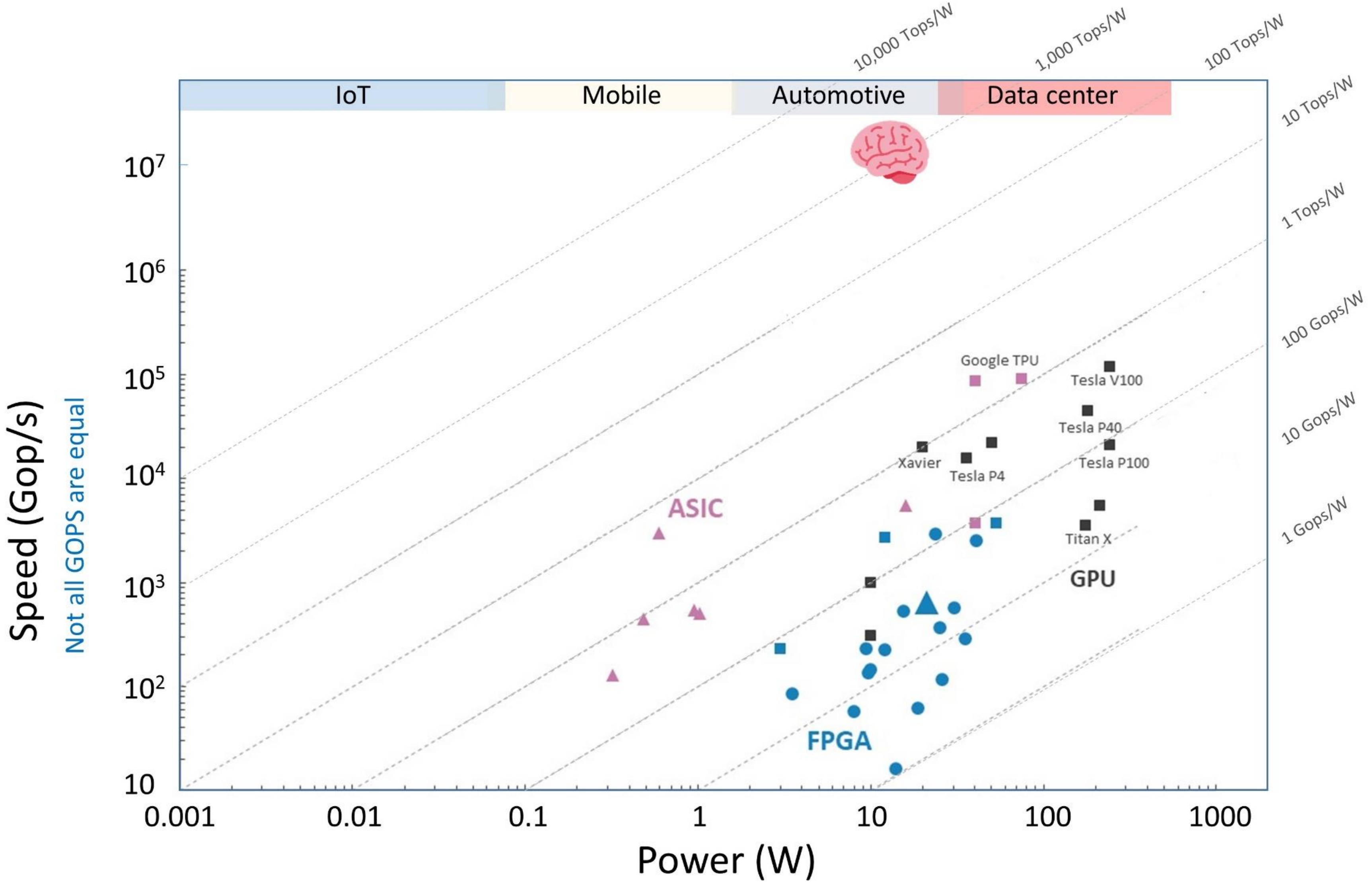
neurons neurons

0011011

GPUs, TPUs, FPGAs



Current CMOS processors cannot run future AI

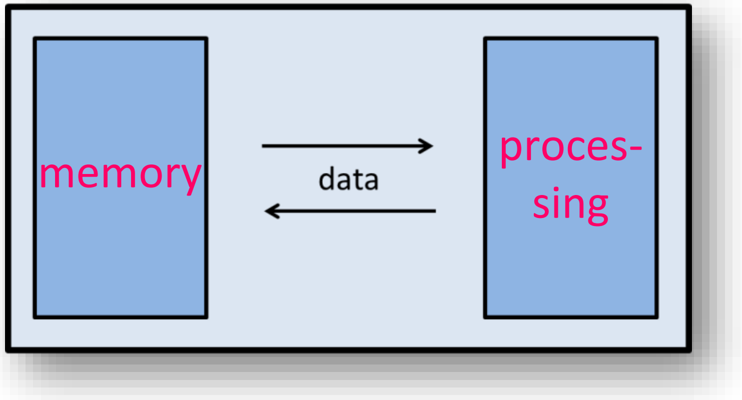


[Based on https://nicsefc.ee.tsinghua.edu.cn/projects/neural-network-accelerator/]

Entangling memory and processing allows for fast and energy efficient computing

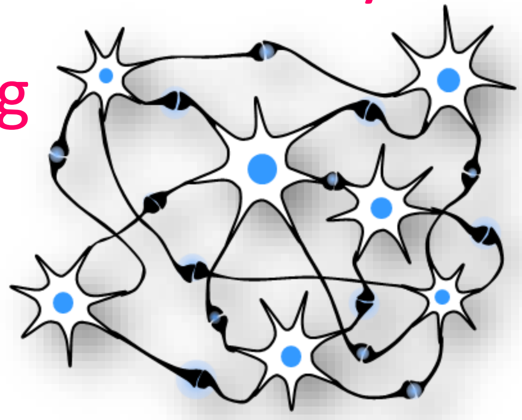
Digital computer

CPUs, GPUs, TPUs, FPGAs



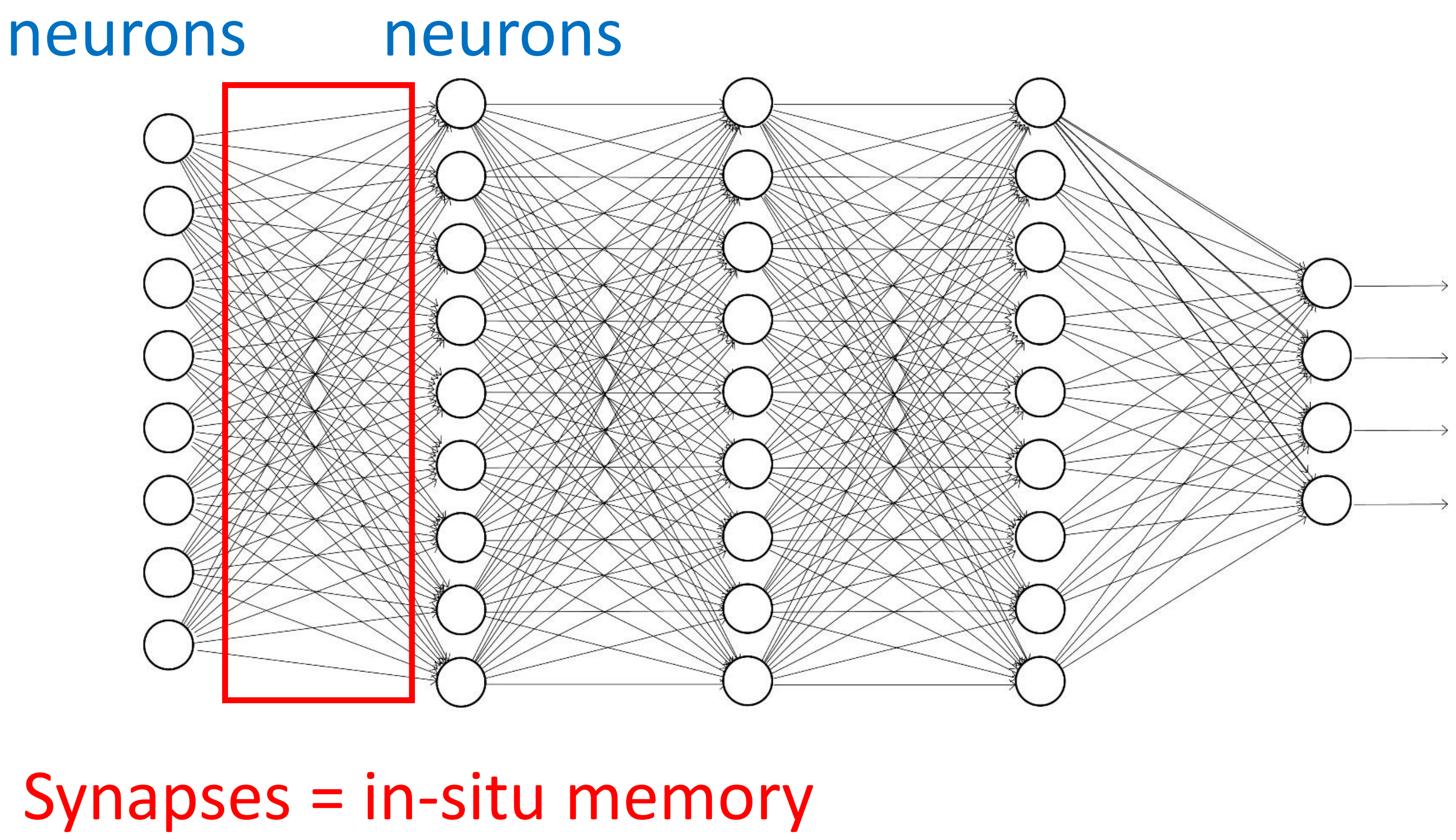
100 W/cm^2

Brain

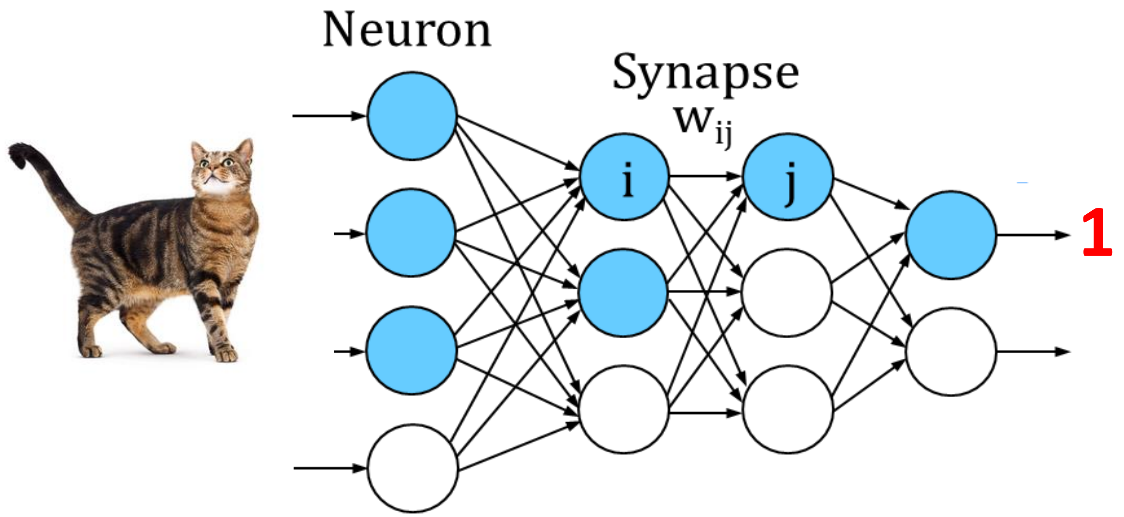


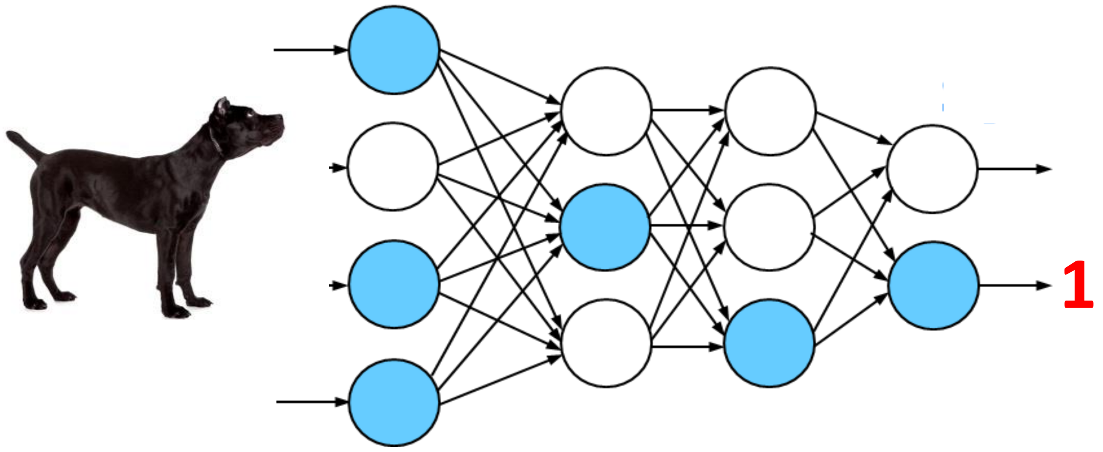
20 W in total !

Chips implementing the layered structure of the network can enable fast, low energy computing with real-time learning

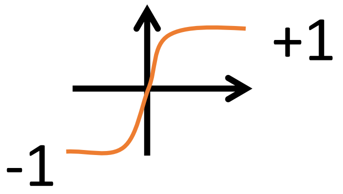


Neural networks compute through non-linearity, memory and plasticity

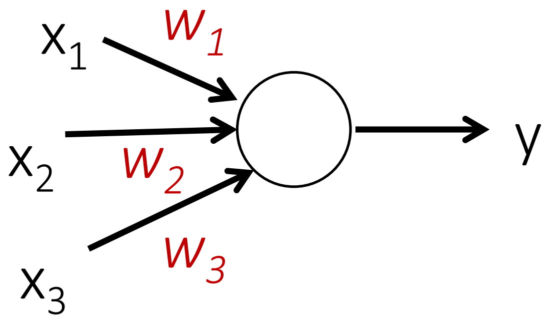




- **Neurons**: non-linear



- **Synapses:** analog valves (weights w)



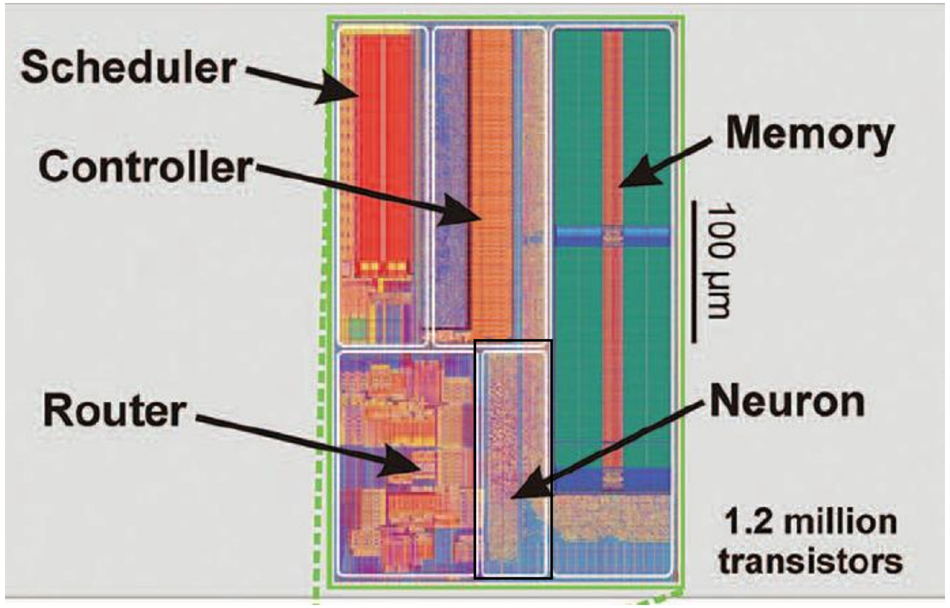
$$y = f\,(\,w_1\,x_1 + w_2\,x_2 + w_3\,x_3)$$

CMOS neurons and synapses are complex circuits

- A transistor is nanoscale but it is just a switch
- CMOS does not provide memory (volatile)

CMOS neuron **10-100 µm**

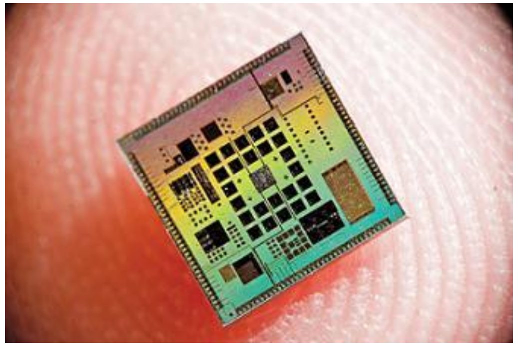
CMOS synapse **10 µm**

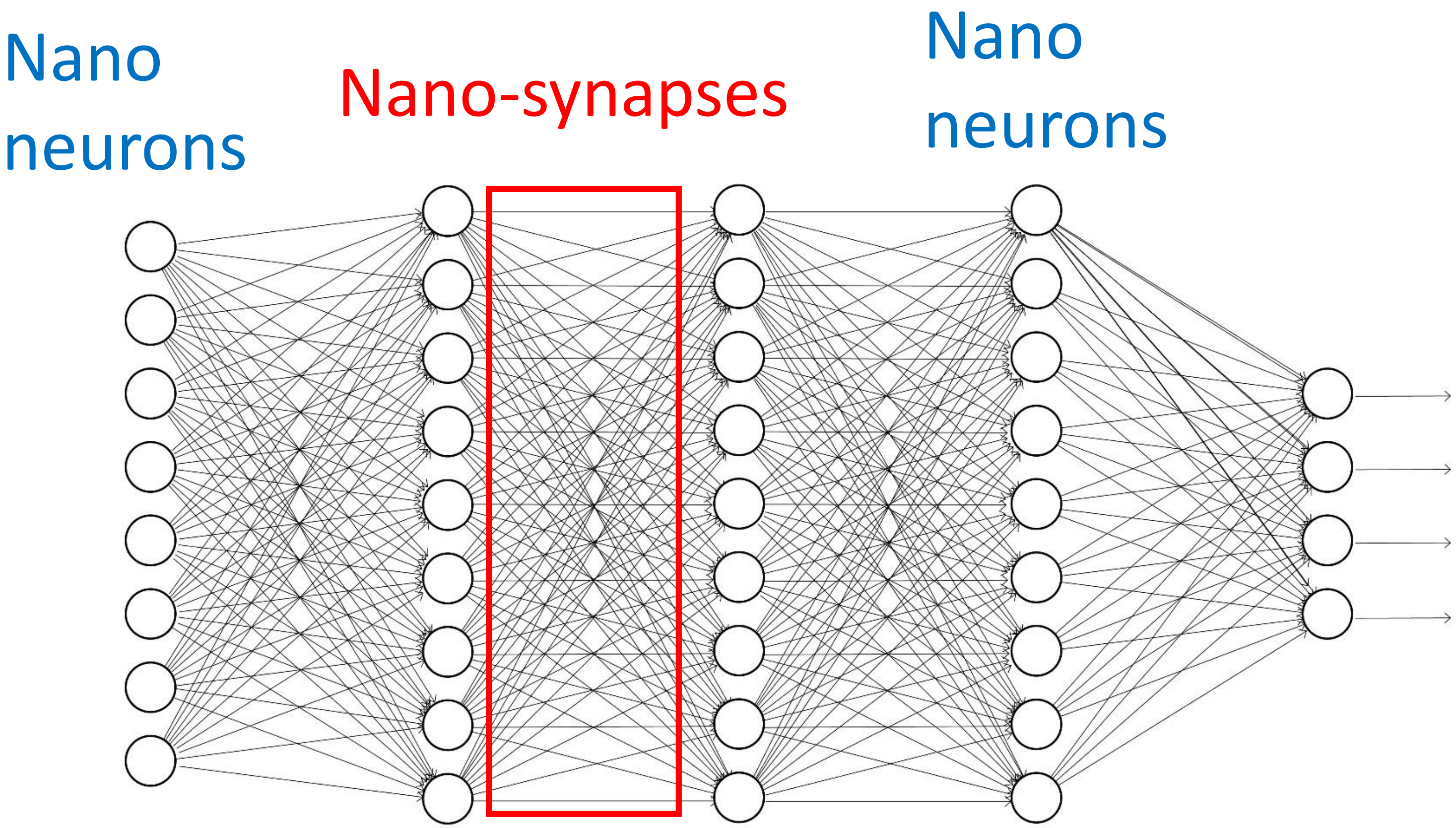


Merolla et al, *Science* **345**, 668 (2014)

Brainscales 20 wafer machine. 4M neurons, 1B synapses

Can we build small neuromorphic chips that run deep neural networks ?

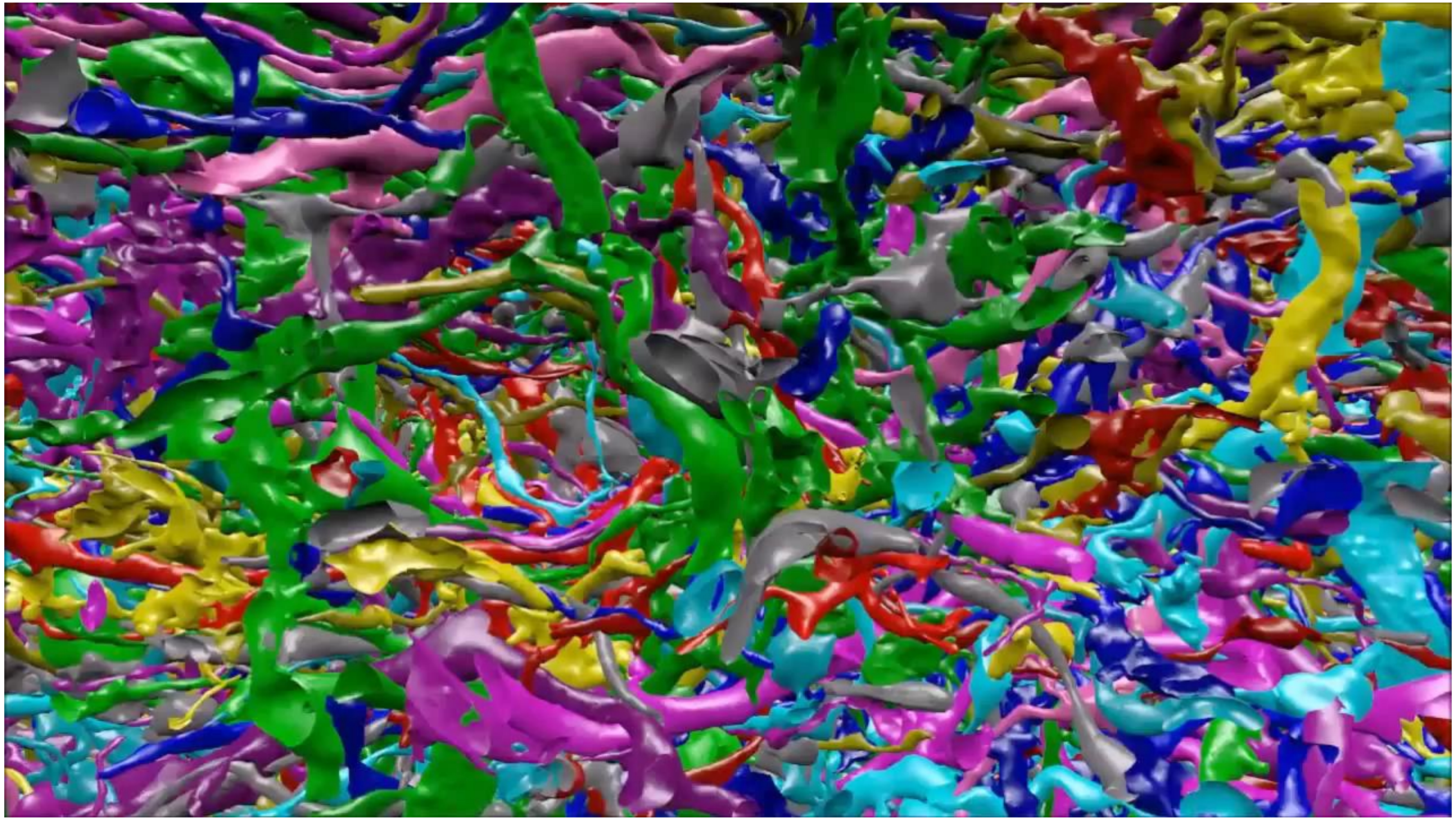




Hundred millions of neurons and synapses in a 1 cm^2 chip
→ Each device smaller than 1 μm^2

Future AI will be massively interconnected

Brain: 10^4 synapses/neurons



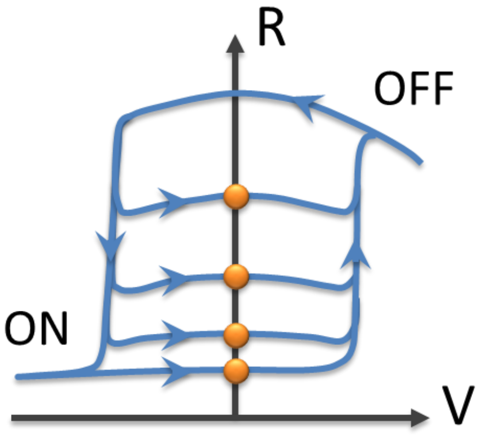
Moritz Helmstaedter lab, retina flight 2013

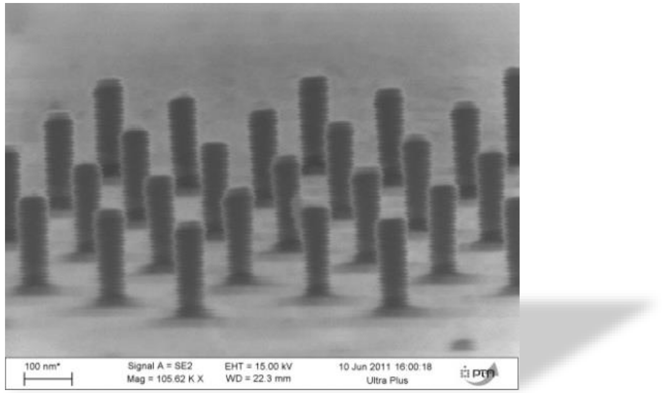
Memristive synapses with CMOS neurons

→ connecting through wiring

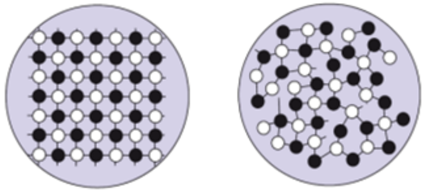
Synapses can be imitated with memristors

Chua, IEEE Trans. Circuit Theory (1971)



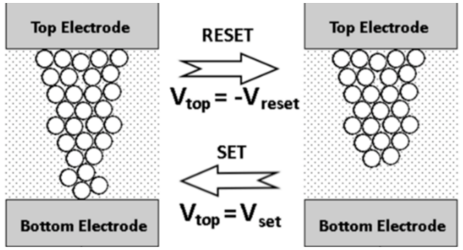


Phase change



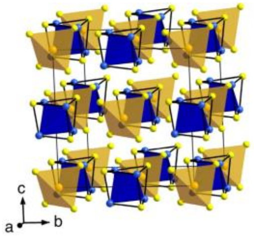
Kuzum et al, Nanotechnology (2013)

Filamentary switching



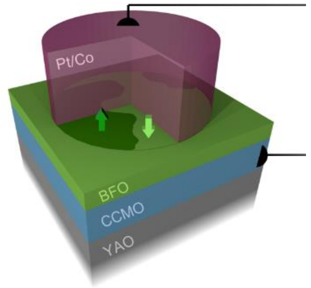
Yang et al., Nature Nano. (2013)

Mott insulators



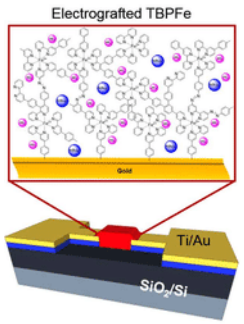
Rozenberg et al., PRL (2018)

Ferroelectric



Chanthbouala et al, Nature Mat. (2012)

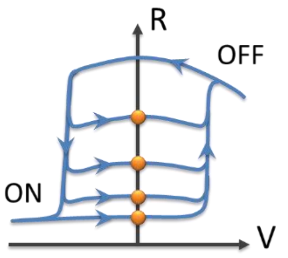
Organic

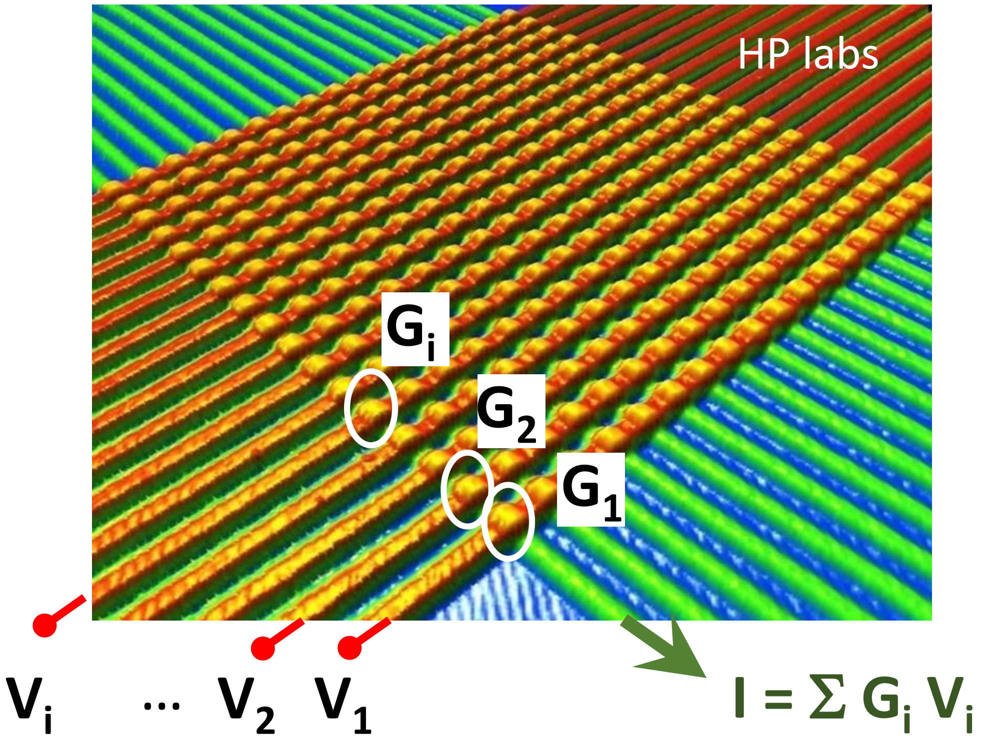


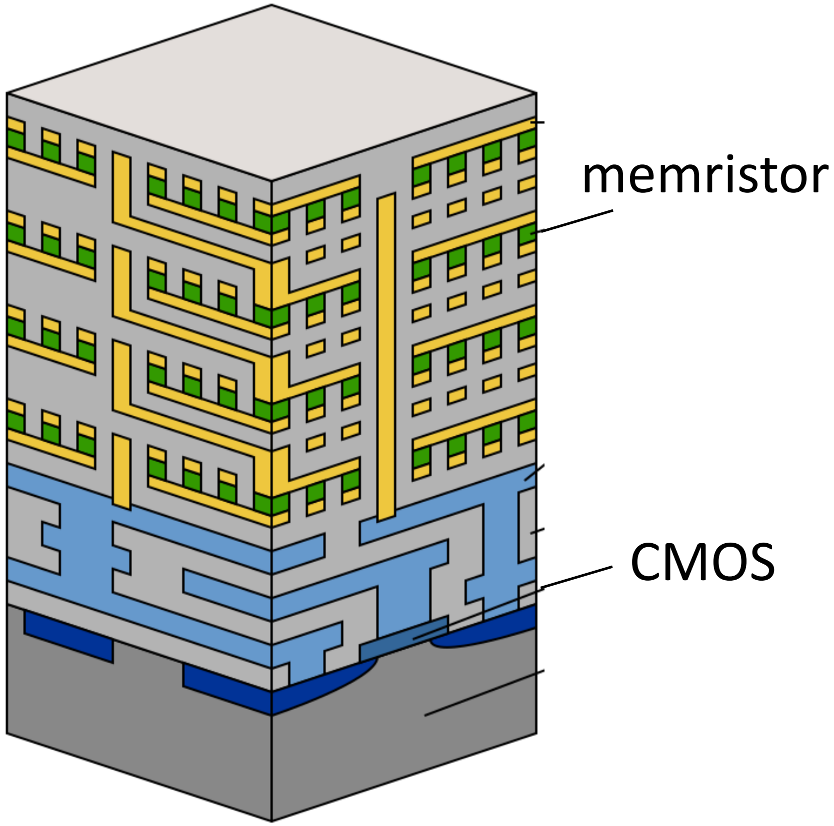
Cabaret et al, IEEE ICN (2014)

CMOS neurons + Memristive synapses

10 000 synapses per neuron ?



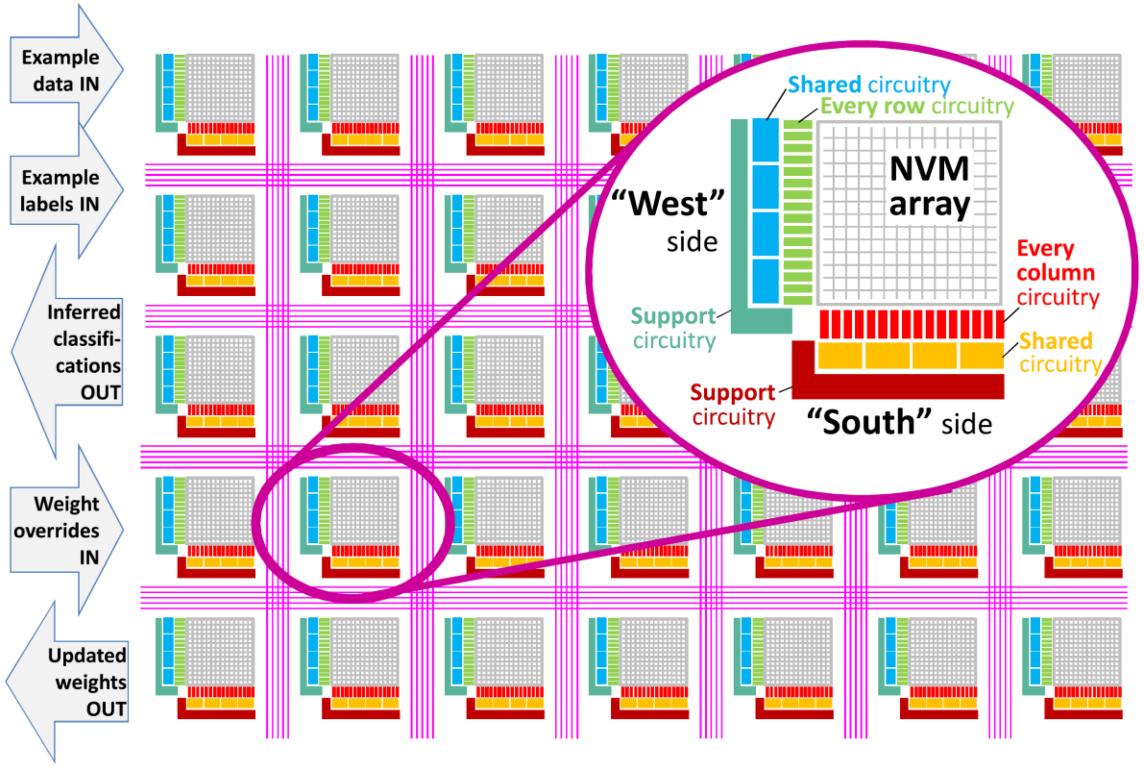




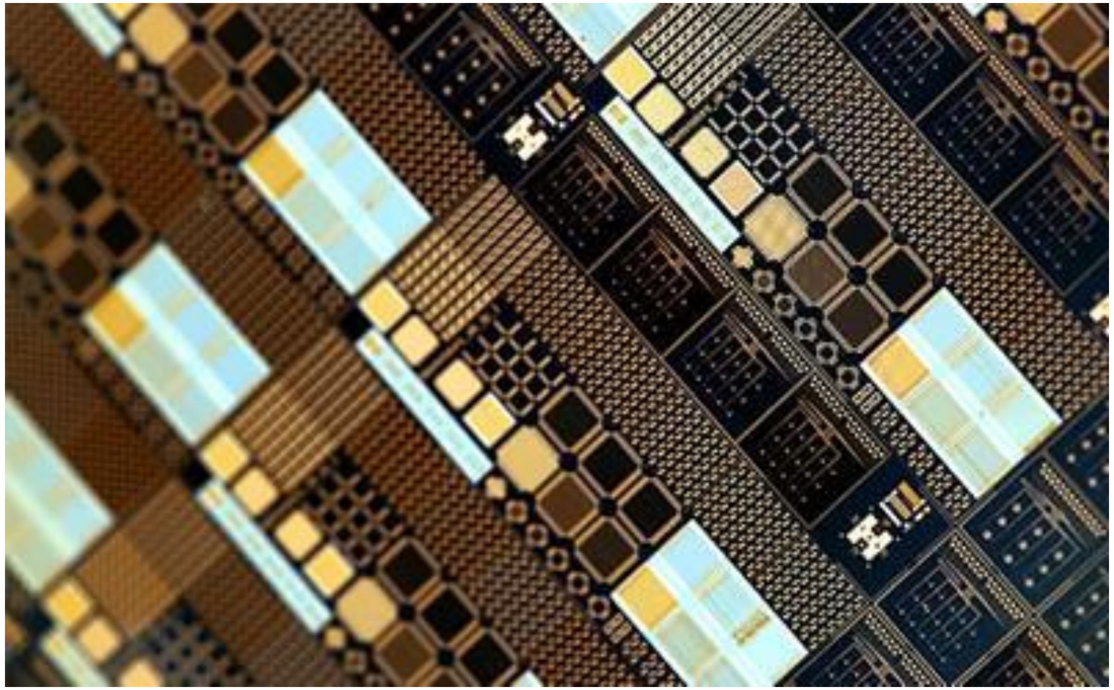
Strukov and Williams, PNAS 106, 20155 (2009)

Memristors open a path to energy-efficient real-time learning, upcoming chips will implement inference first

Claims : 100 x faster than GPUs and 100 x less energy consumption



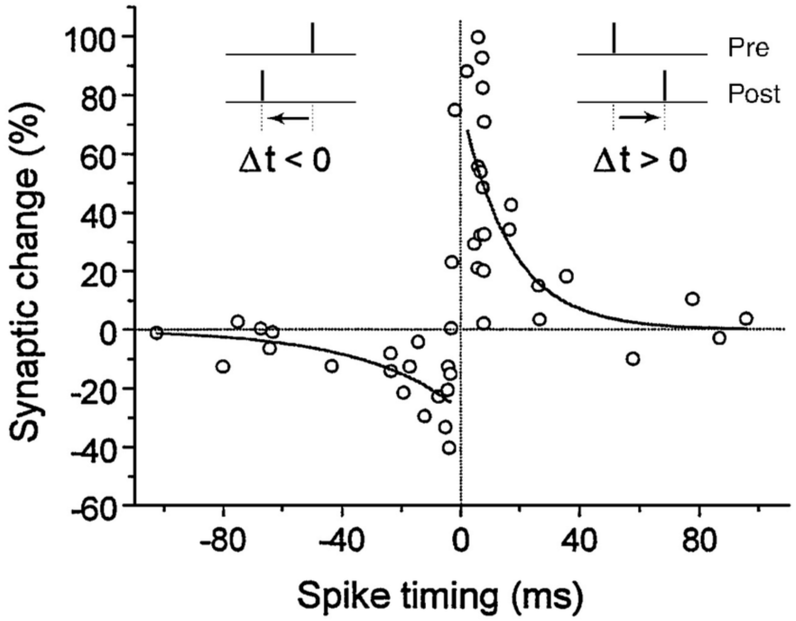
S. Ambrogio et al, Nature 558, 60 (2018)



- D. Querlioz group : M. Bocquet et al, IEDM (2018)
- CEA List

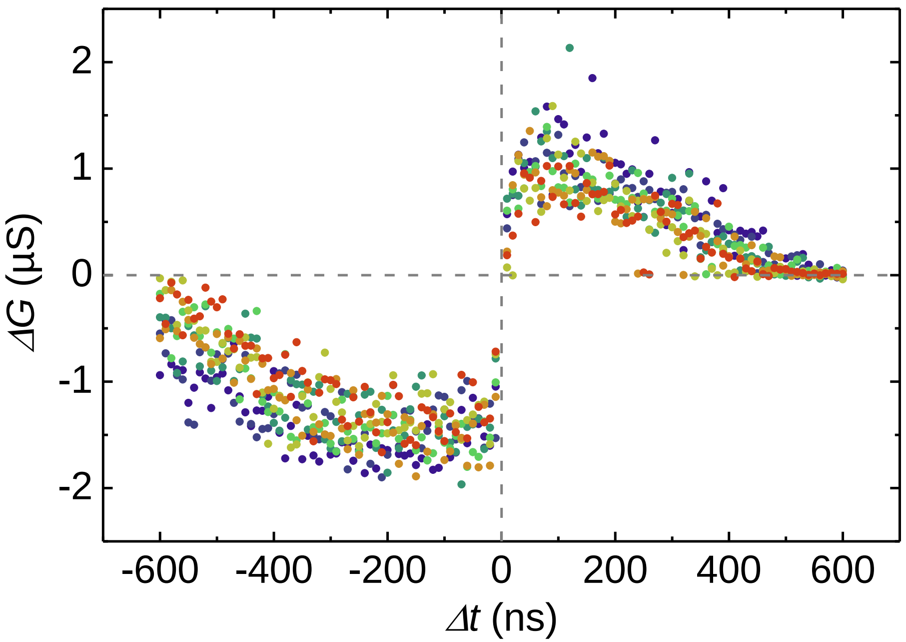
Memristors can implement bio-inspired learning rules

Biological synapse



Bi and Poo. Journal of Neuroscience 15 10464 (1998)

Ferroelectric memristor



Sören Boyn, **JG** et al, Nature Com. 8, 14736 (2017)

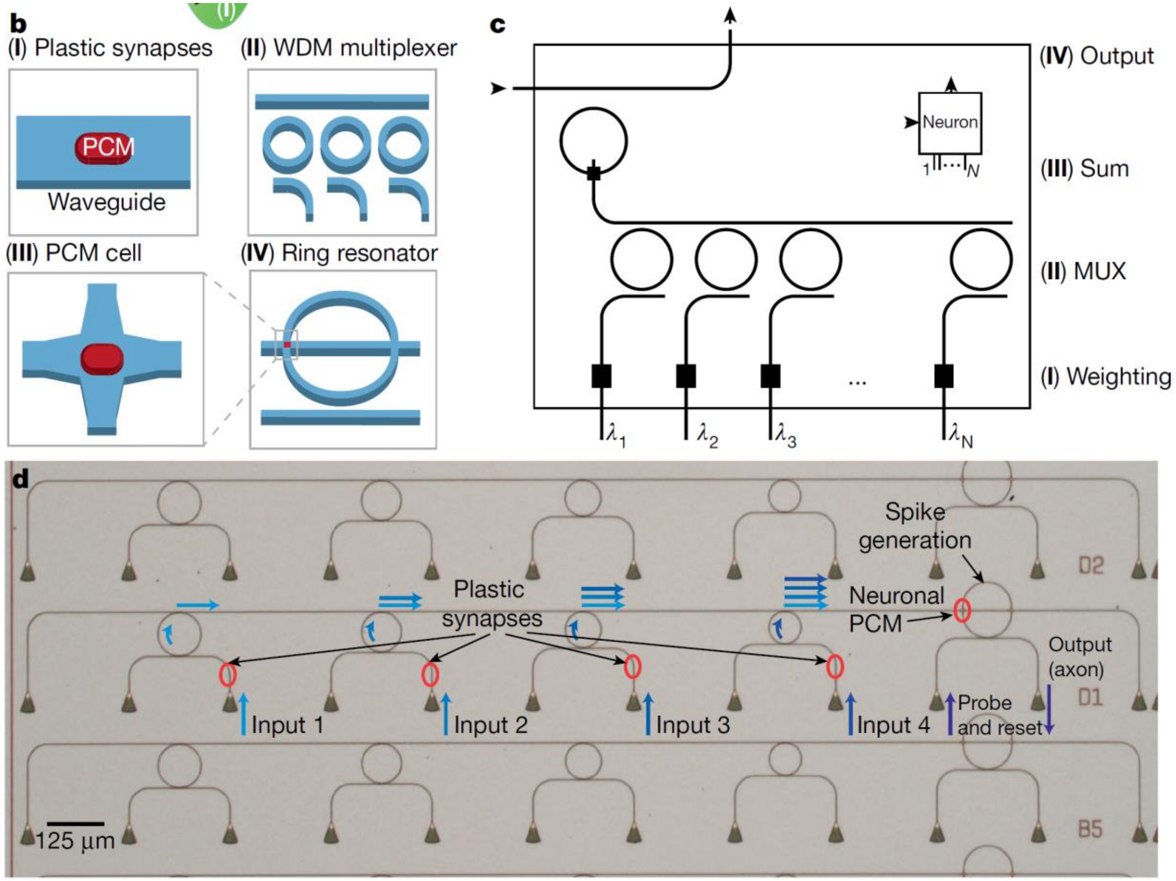
Spike Timing Dependent Plasticity enables unsupervised learning with memristors

Damien Querlioz et al, PIEEE 103, 1398 (2015)

Photonic neurons and synapses:

→ connecting through light

Scalable concept with photonic resonators as neurons and phase change synapses recently proposed



J. Feldmann et al, Nature 569, 208 (2019)

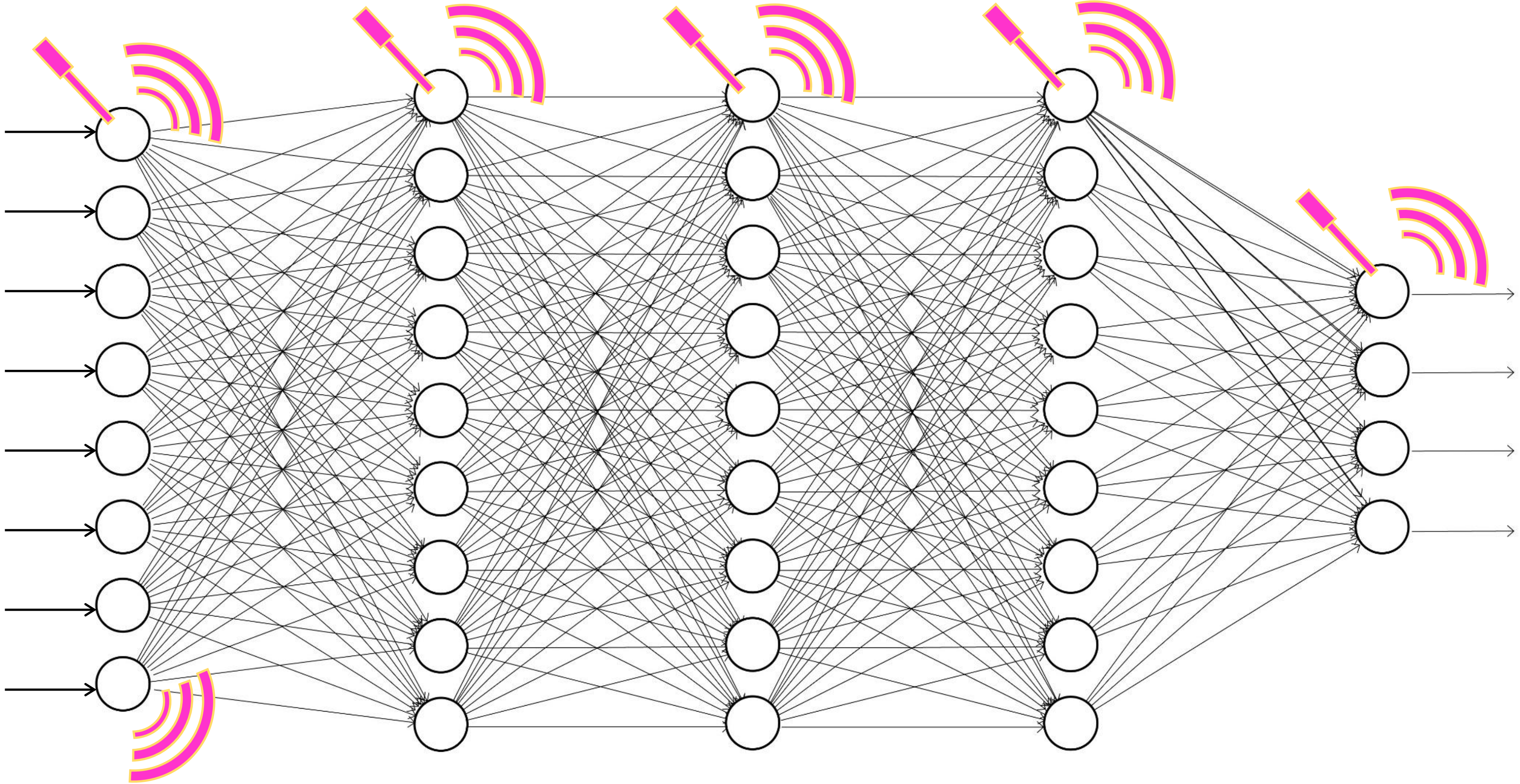
In Saclay :

C2N (S. Barbay, F. Raineri)
TRT (A. Brignon, A. de Rossi)
…

Magnetic nanoneurons and nanosynapses

→ connecting through radiowaves

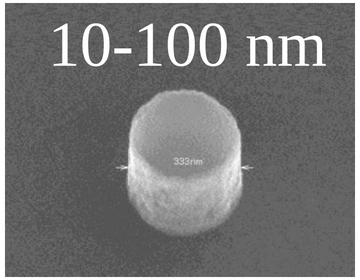
Wireless deep learning through RF communications



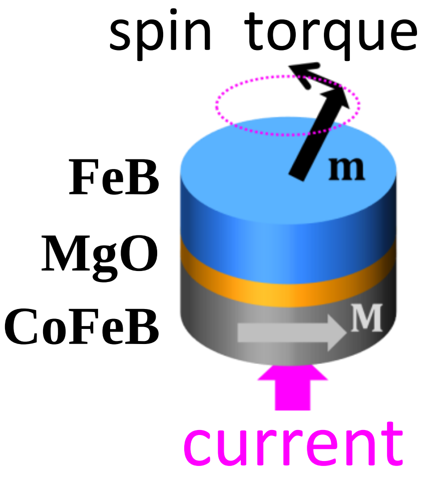
Magnetic nano-oscillators are non-linear nano-radios

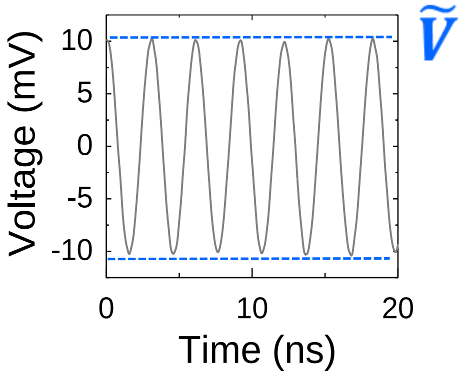
Nanoscale, fast (GHz), non-linear and easily measurable

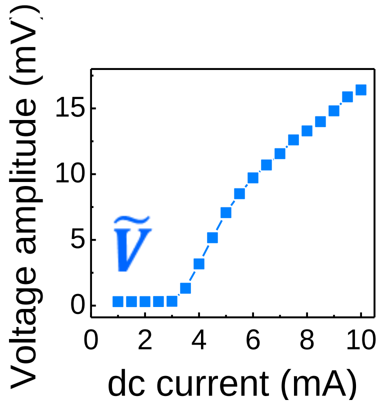
magnetic tunnel junction



compatible with CMOS







Same structure as magnetic memories

N. Locatelli, V. Cros and J. Grollier, Spin-torque building blocks, Nature Mat. 13, 11 (2014)

Collaborations

CNRS/Thales, France

Philippe Talatchian, Miguel Romera, Flavio Abreu Araujo, Mathieu Riou , Jacob Torrejon, Vincent Cros, Danijela Markovic, Nathan Leroux, Paolo Bortolotti, Juan Trastoy, Alice Mizrahi , Julie Grollier

C2N, France

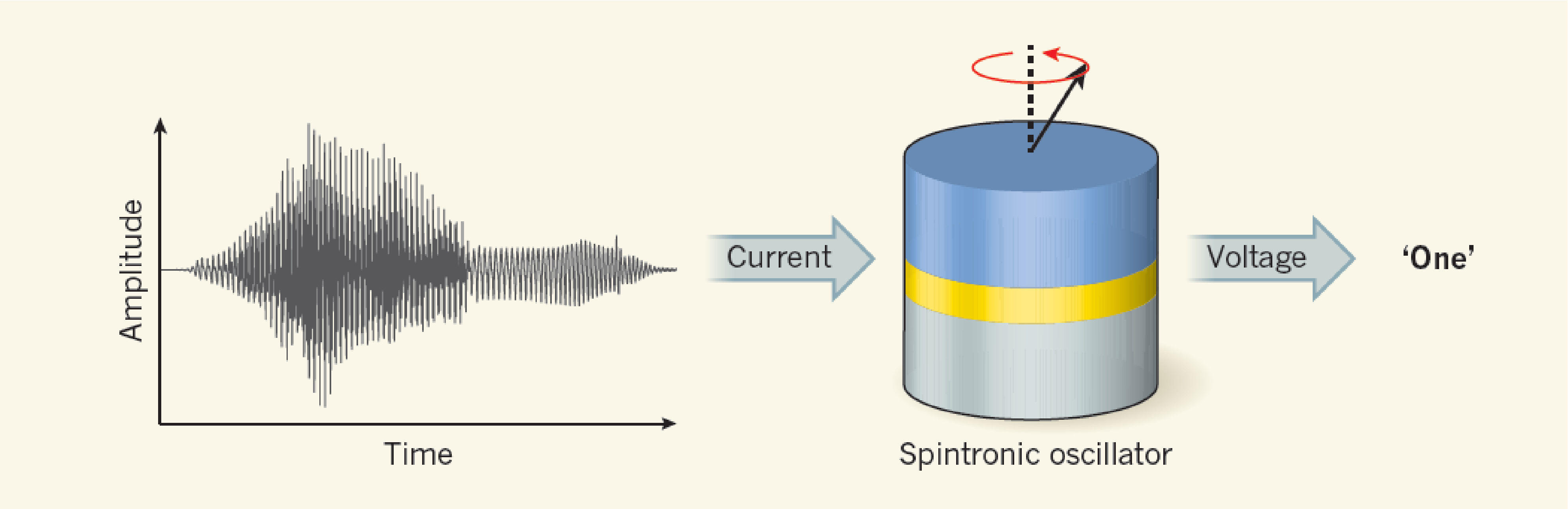
Maxence Ernoult, Damir Vodenicarevic, Tifenn Hirtzlin, Nicolas Locatelli, Damien Querlioz

AIST, Japan

Sumito Tsunegi, Kay Yakushiji, Akio Fukushima, Hitoshi Kubota, Shinji Yuasa

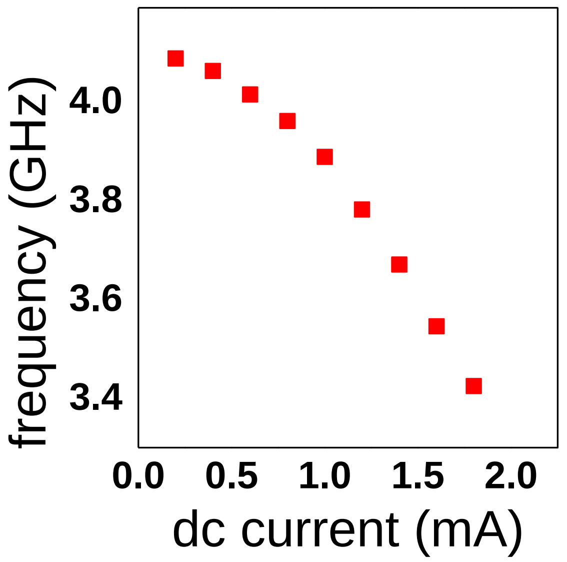
Due to its stability and non-linearity, a single magnetic oscillator can emulate an assembly of neurons and perform neuromorphic computing

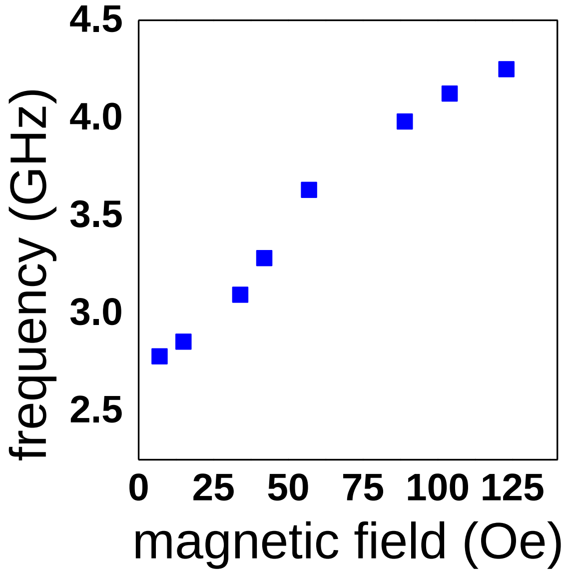
Spoken digit recognition through reservoir computing



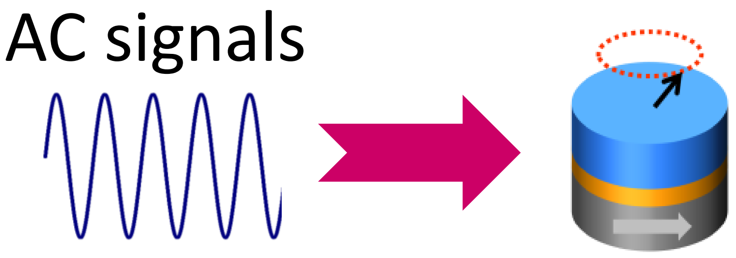
J. Torrejon, M. Riou, F. Abreu Araujo et al, Nature 547, 428 (2017)

Magnetic nano-oscillators have a high tunability : they are radio-receivers



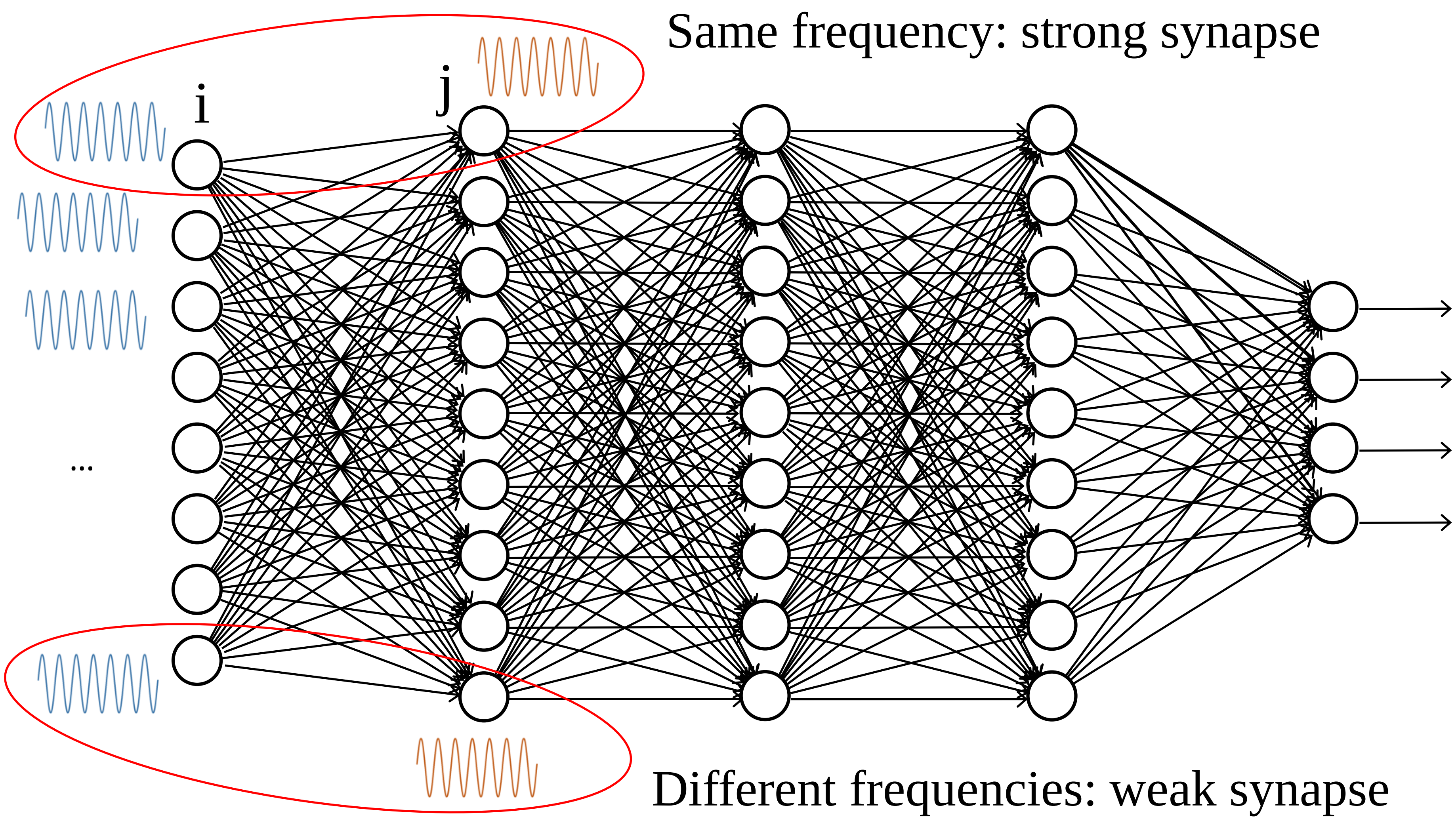


Enhanced sync ranges



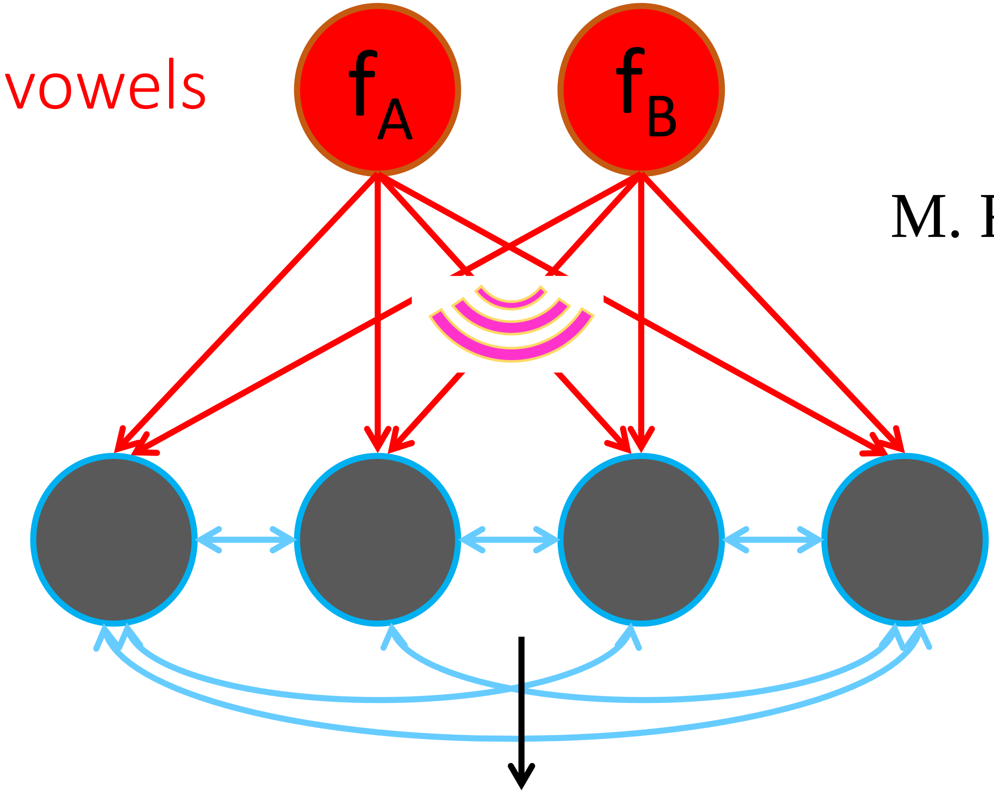
A. Slavin and V. Tiberkevich, IEEE TM 45, 1875 (2009)

The oscillators ability to mutually interact opens the path to RF on-chip communication between neuron layers



Vowels classification with spin-torque oscillator neural network

Inputs: vowels



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

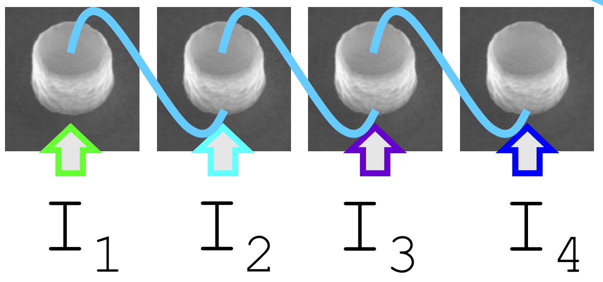
Magnetic Nano-Oscillators

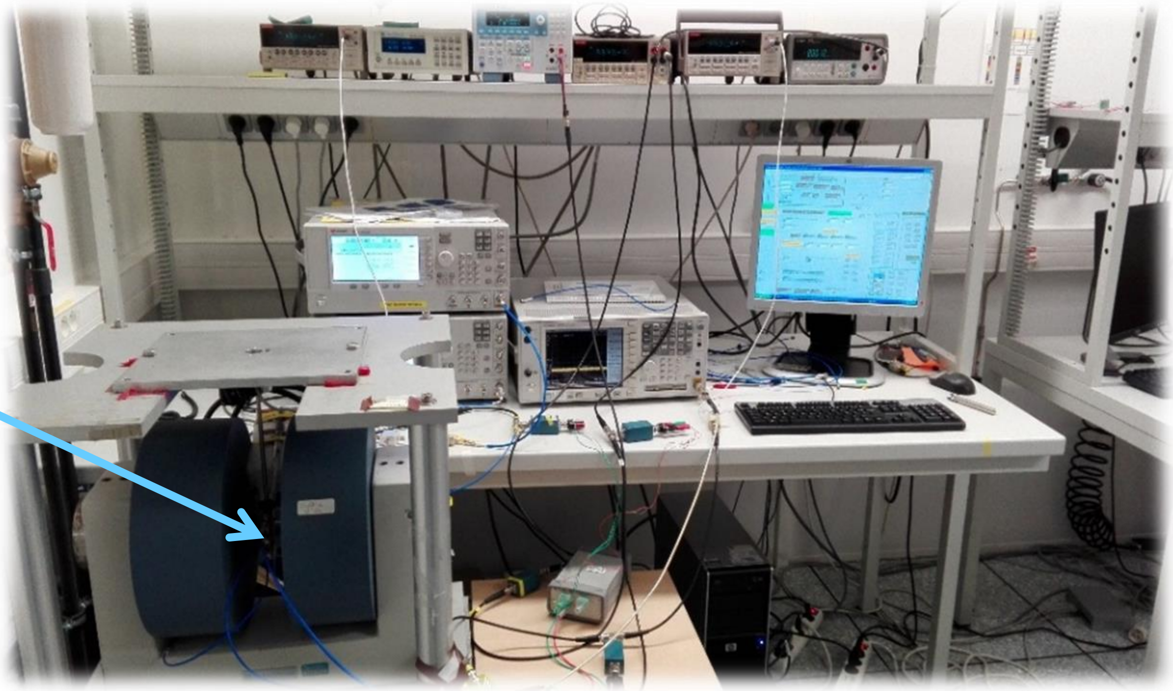
Outputs: synchronized states

We train the network by tuning the currents through the oscillators according to an online learning rule

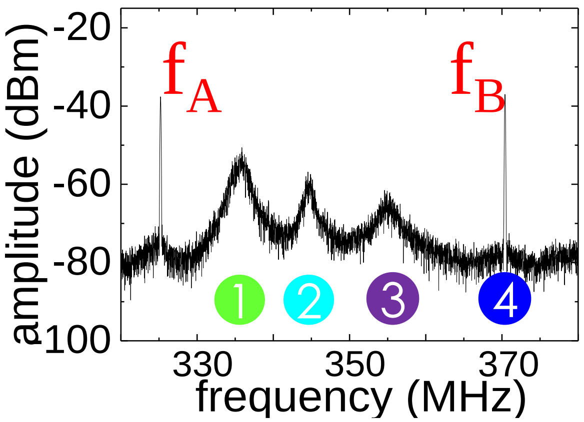
Experimental set-up ← → Computer (Learning algorithm)

Oscillators

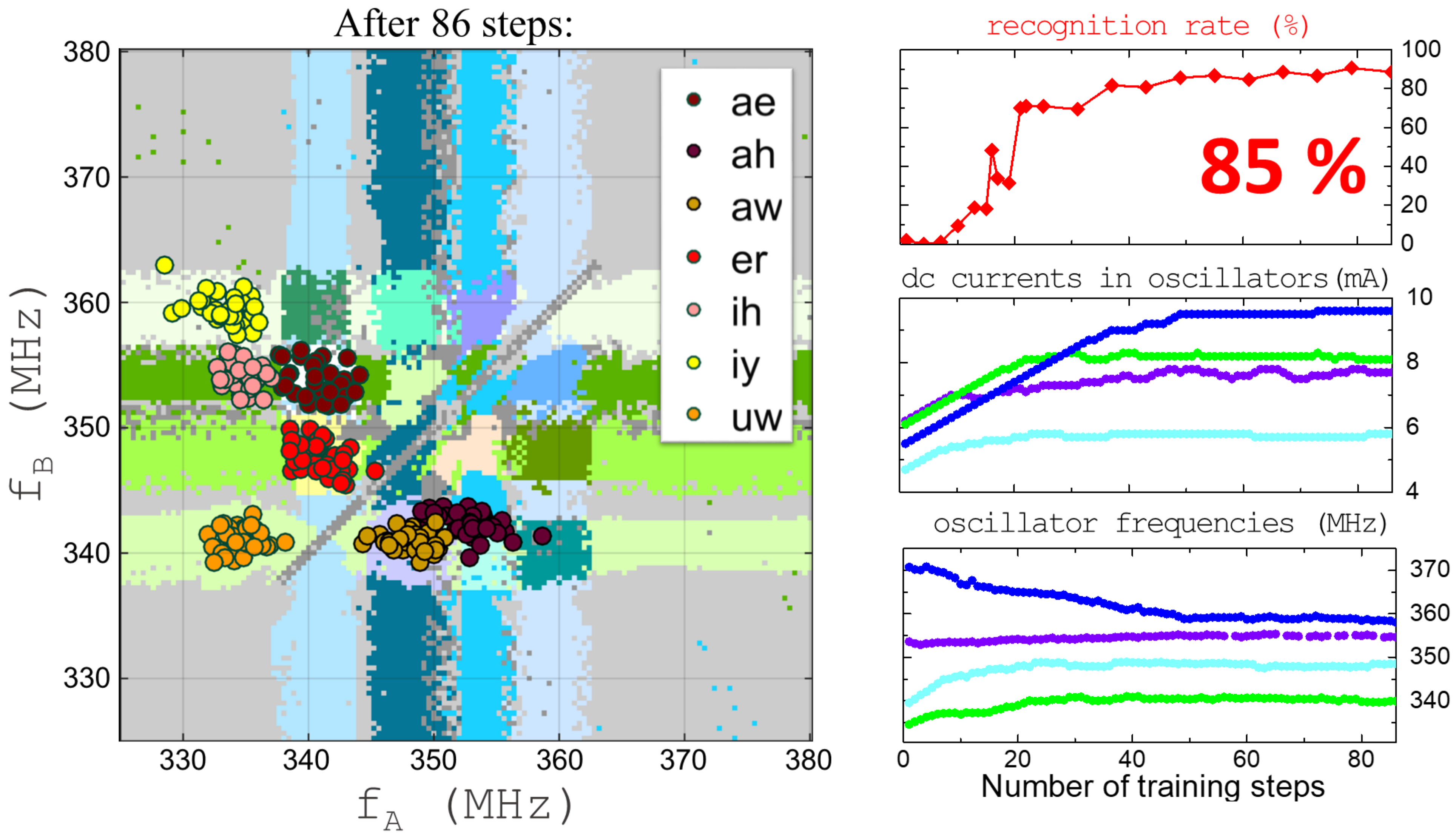




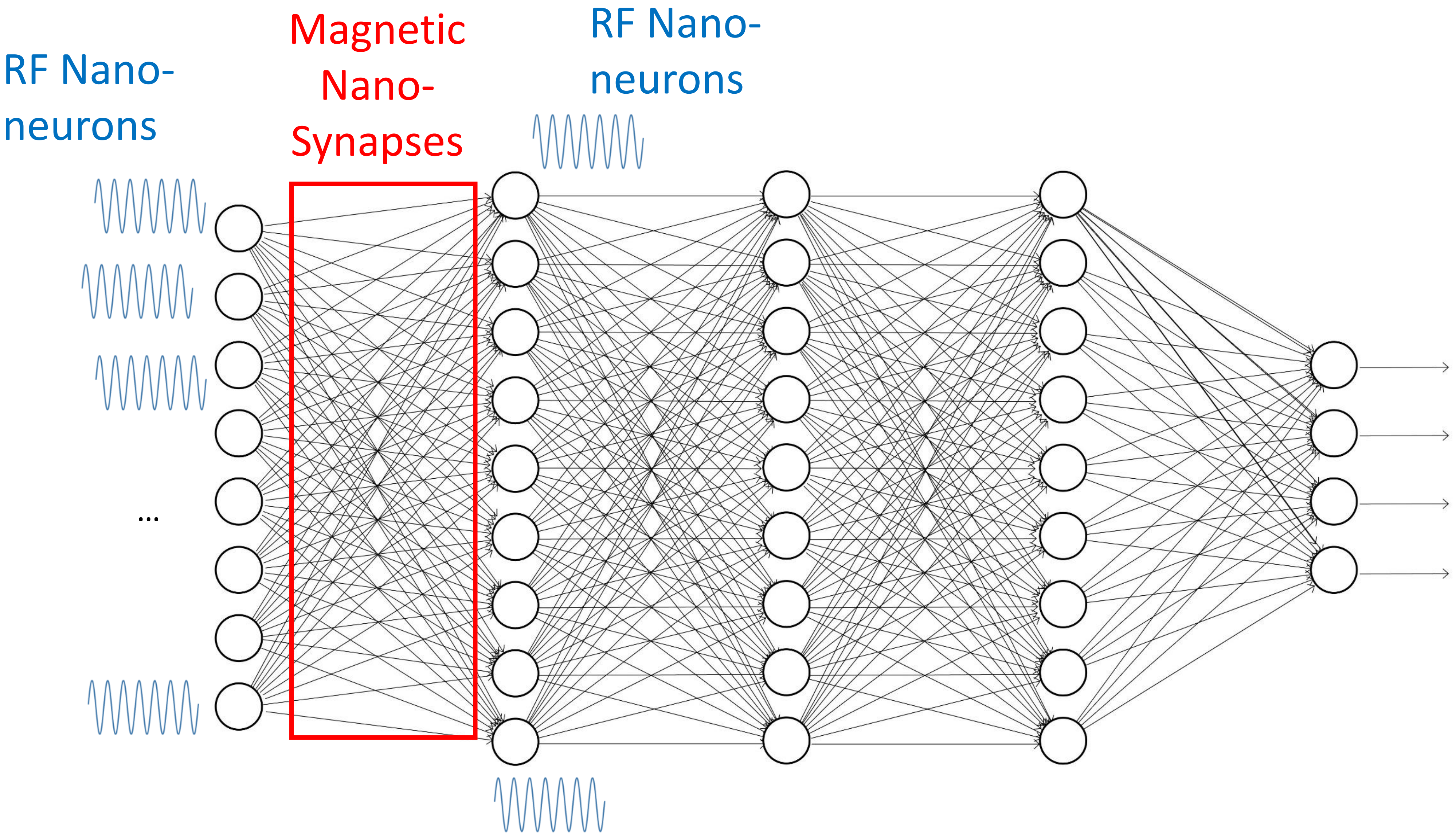
ex: « iy » 4B



We reach a recognition rate equivalent to a multilayer perceptron trained with the same number of parameters

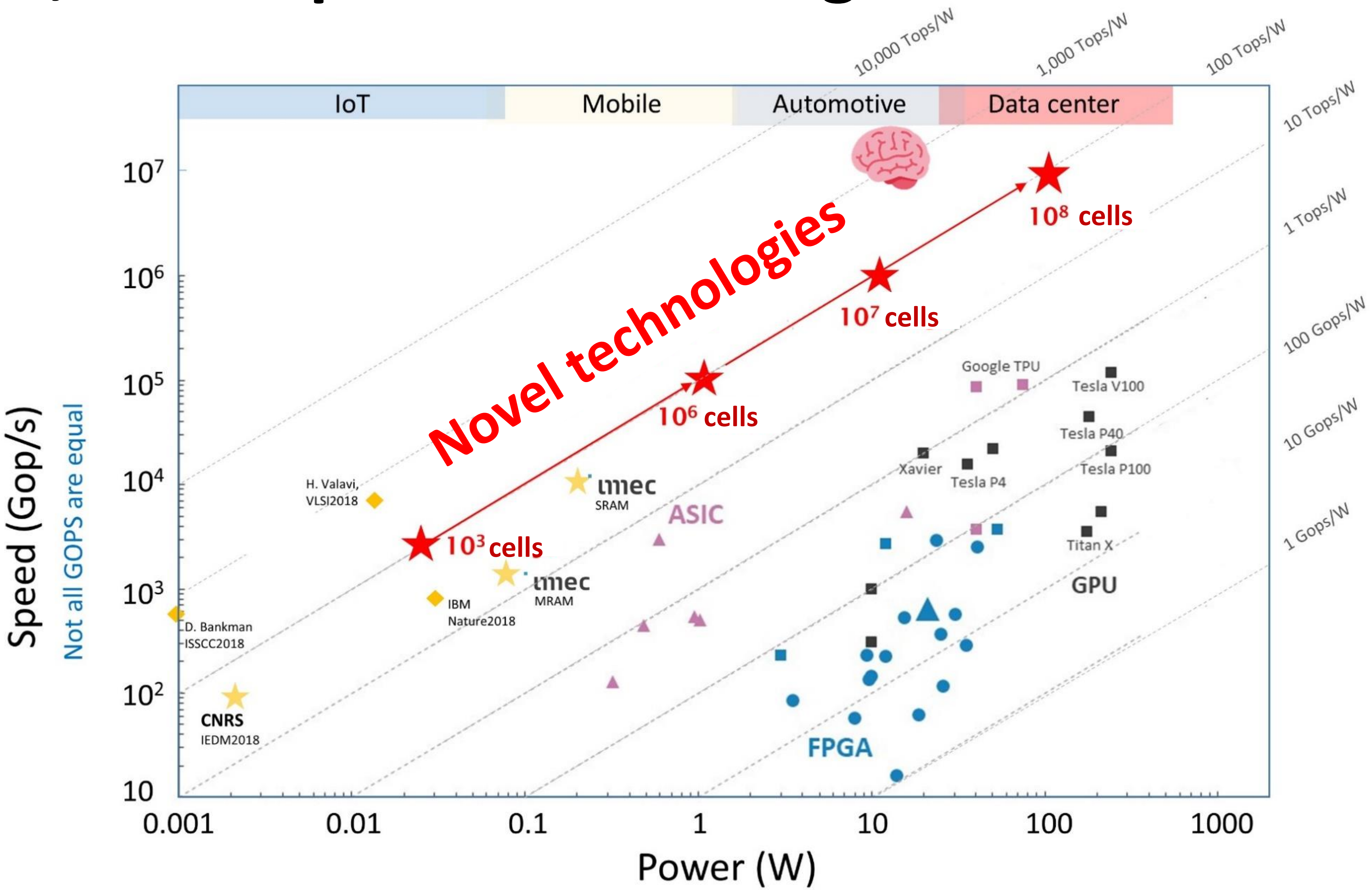


Perspectives: RF deep learning with spintronic nanodevices



Conclusion and perspectives

Novel technologies can speed up neural networks by 100, and implement learning



[Based on https://nicsefc.ee.tsinghua.edu.cn/projects/neural-network-accelerator/]

http://gdr-biocomp.fr/

	CMOS only	Memristor synapses with CMOS neurons	Photonic synapses and neurons	Magnetic synapses and neurons
Connections	wires	wires	light	radiowaves
Min Neuron lateral size	10 µm	10 µm	100 µm	10 nm
Min Synapse lateral size	10 µm	10 nm	1 µm ?	10 nm
Advantages	commercial	Nanoscale synapse, technology-ready	Can be totally passive (zero power consumption)	Nanoscale synapse and neurons, almost commercial technology
Disadvantages	Size of neurons and synapses No in-memory computing	Size of neurons Complex wiring	Size of neurons and synapses Dissipation due to lasers	Scalability to be demonstrated
Chips	Inference	Inference coming soon	no	no