Memory devices for neuromorphic computing

Fabien ALIBART

IEMN-CNRS, Lille

Trad: Toutes les questions que je me suis posé sur le neuromorphique sans jamais (oser) les poser

Introduction: Why ANNet

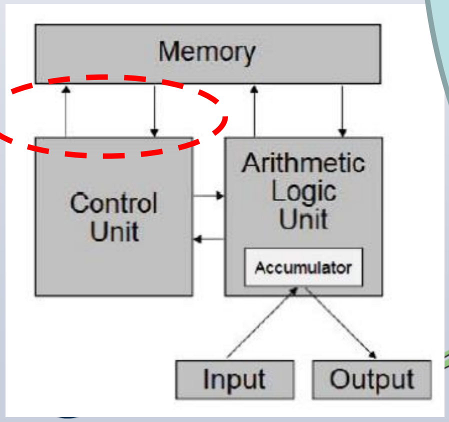
Increase of Fault (nanoscale engineering)

Saturation of clock frequency
+
Energy consumption

New needs for computing Recognition, Mining, Synthesis (Intel)

SEMICONDUCTOR TECHNOLOGY CHALLENGES

Von Neumann bottleneck



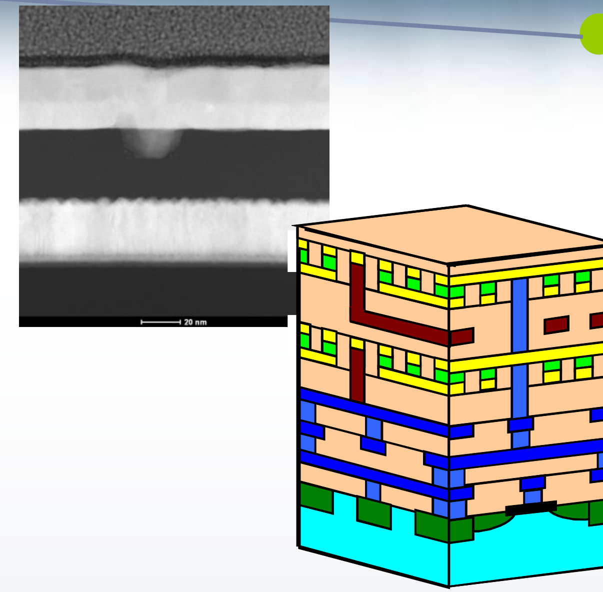
Shift toward a new paradigm for computation

BIO-INSPIRED COMPUTING to match the brain performances (low power consumption, fault tolerant, performances for RMS)

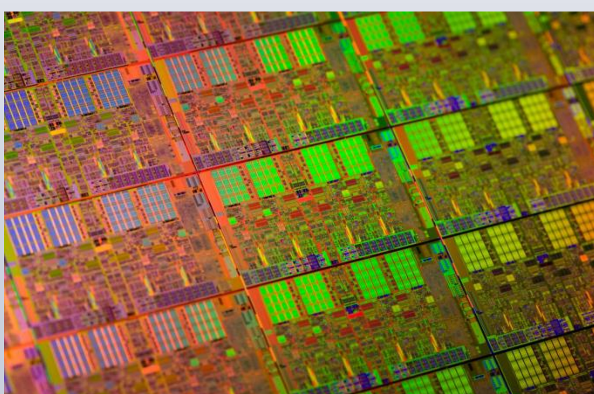
NNET directions

Supercomputer resources
Purely digital

10^{11} neurons
10^{15} synapses



Emerging nanotechnologies
New architecture concepts
and integration strategies



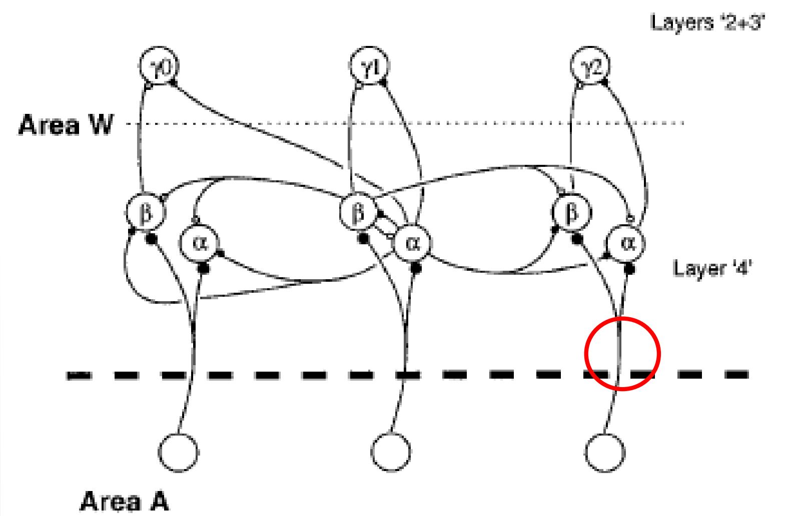
Custom IC, Mix analog/digital
Multichip approach,…
i.e. with conventional technologies

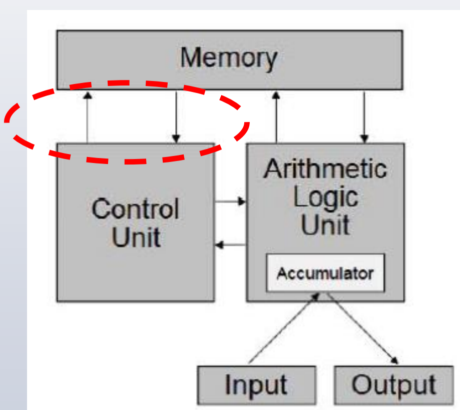
INSTITUT CARNOT IEMN

iemn UMR CNRS 8520 Recherche Formation Transfert

BNNs vs ANNs

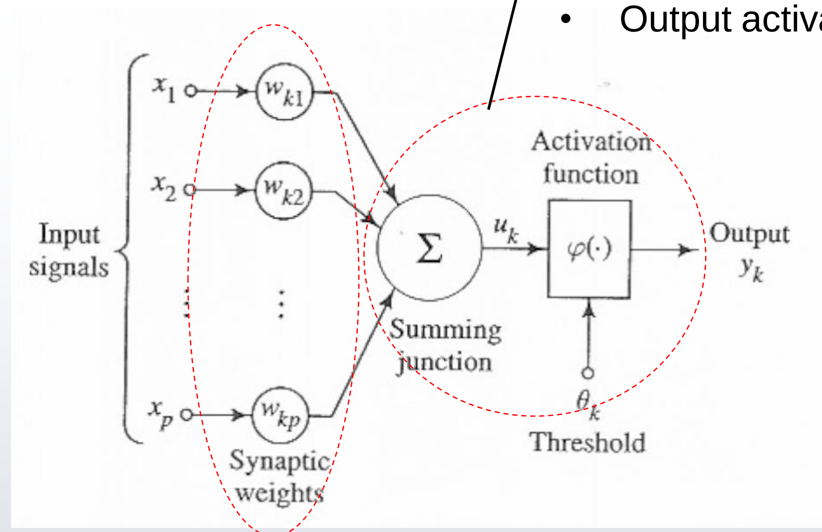
Biological neural network





Artificial neural network

NEURON:
- Summation
- Thresholding
- Output activation



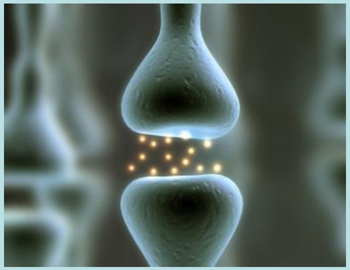
SYNAPSES:
- Input weighting
- Weight adaptation

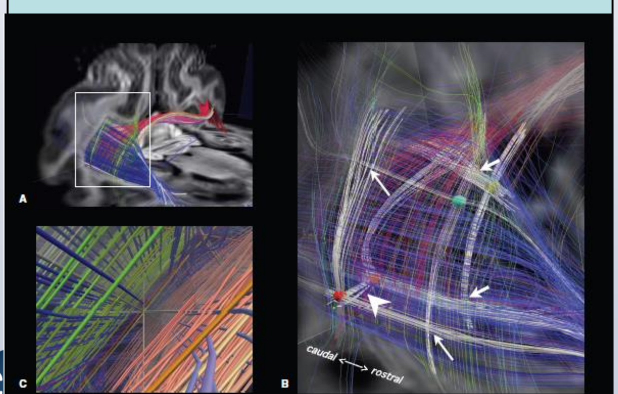
<u>The memory is in the processing unit</u>

(Direct solution to Von Neumann bottleneck!)

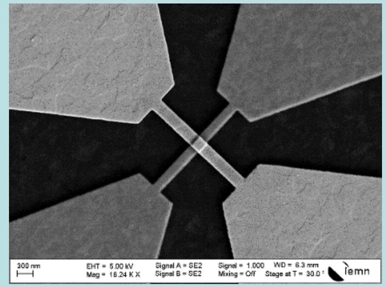
Neuromorphic in between

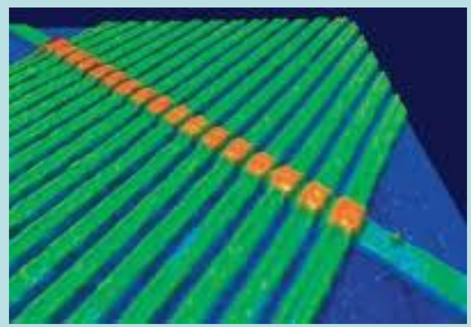
BNNs

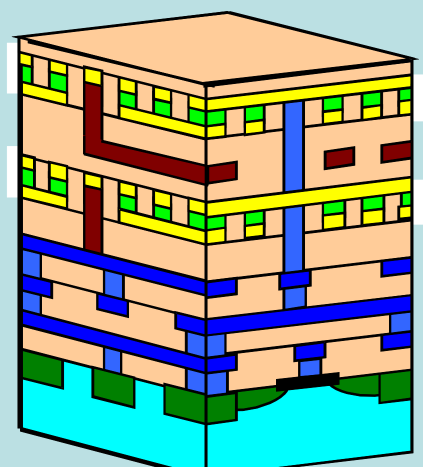




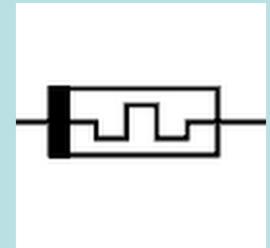
Neuromorphic Eng.

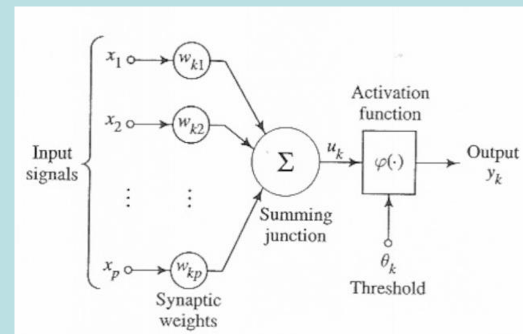


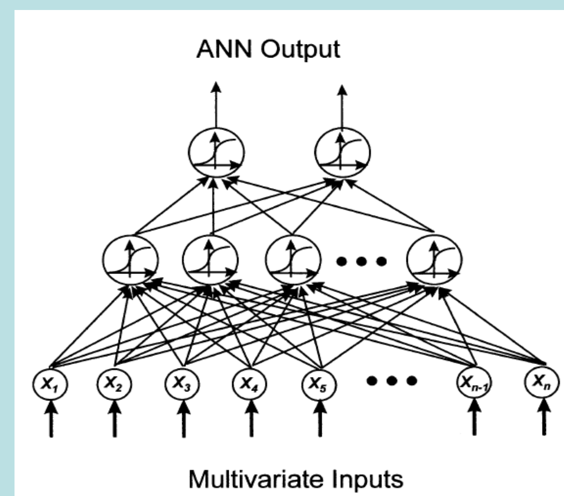




ANNs







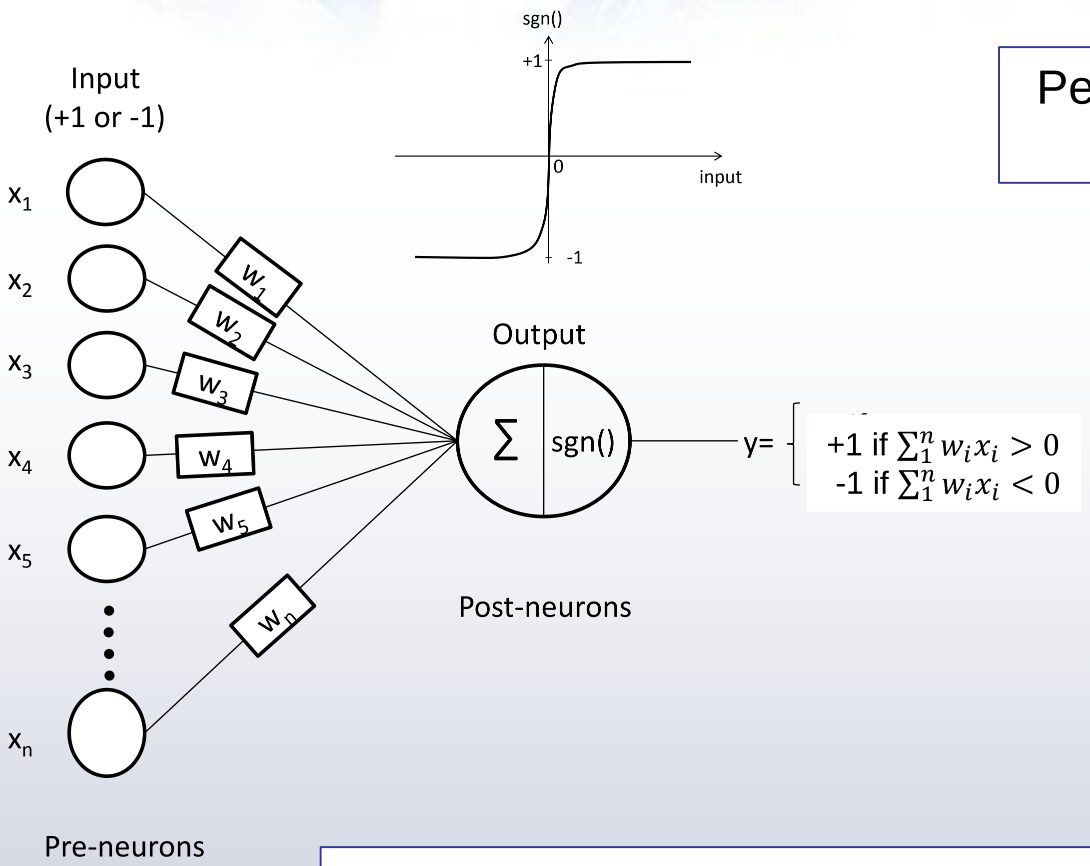
Outline

- Intro to Artificial Neural Network (ANNs)
- Intro to biological Neural Network (BNNs)
- Nanodevices for ANNs
- Nanodevices for BNNs
- STDP: somewhere in between

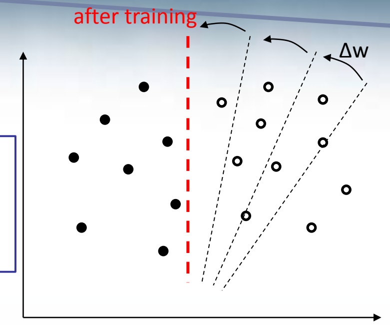
ANNs: basics

Rosenblatt, 1957

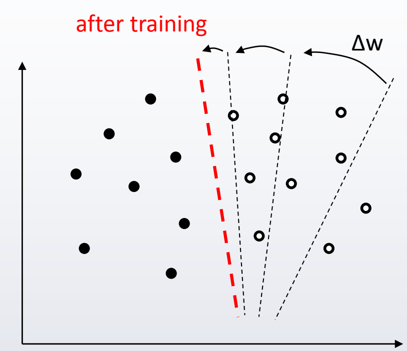
Classification of vectors (datas)



Perceptron rule



Delta rule



Backpropagation

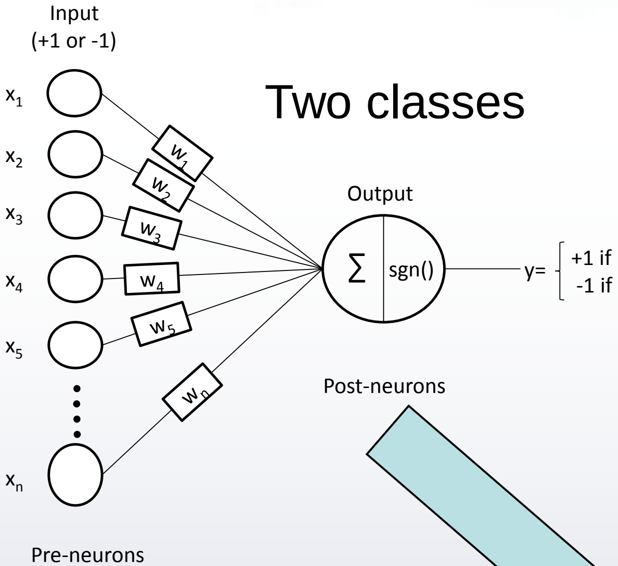
Training/learning: find the optimal weight

- No analytical solution
- Learning algorithms: iterative correction based on known data

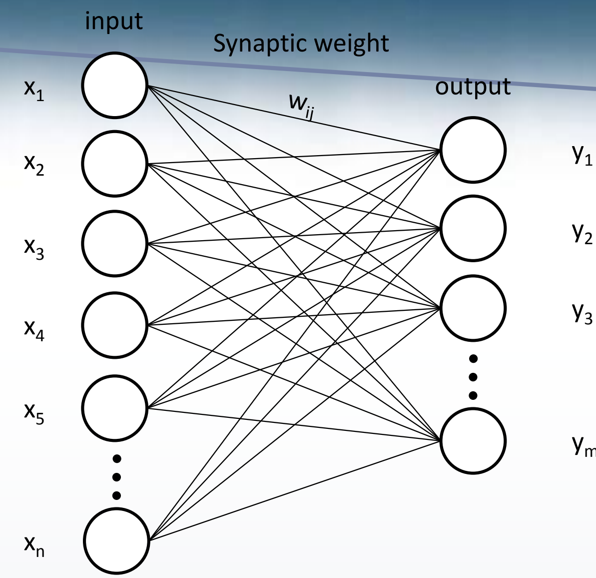
ANNs: basics

Practically, we can do: Pattern classification, Clustering, Prediction…

Two classes



Multiple classes



Non linearly separable ensembles

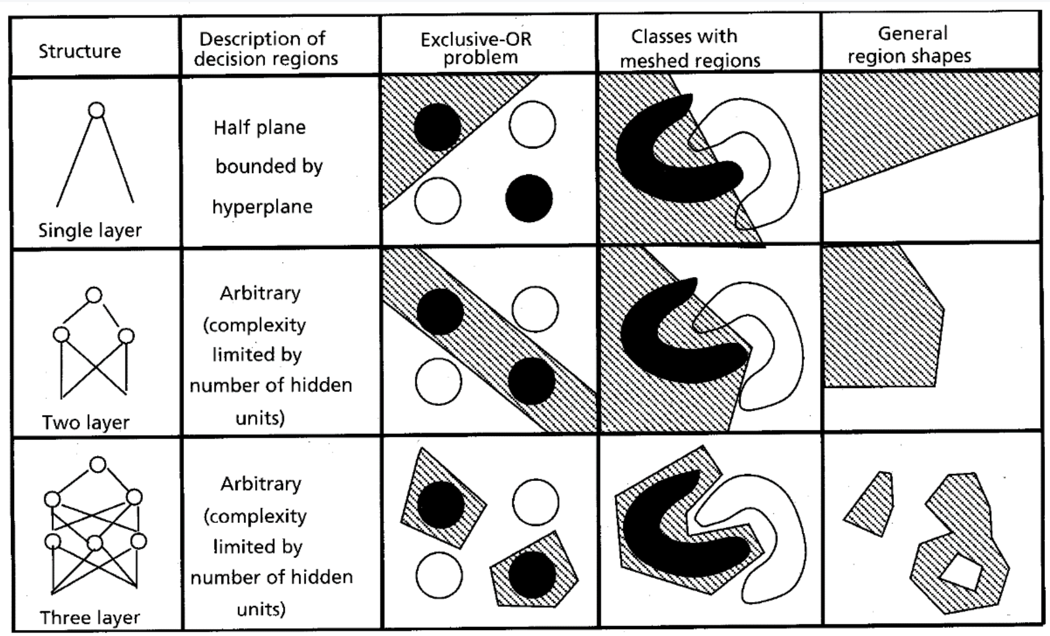
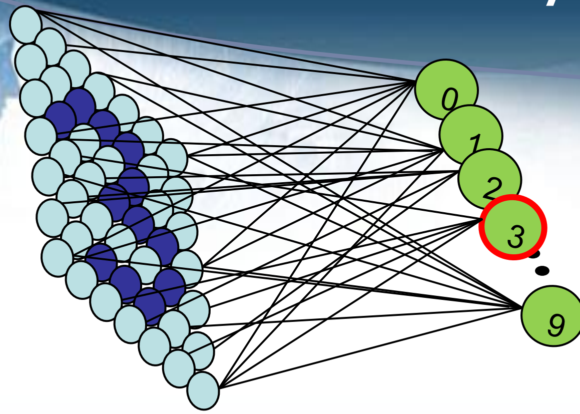
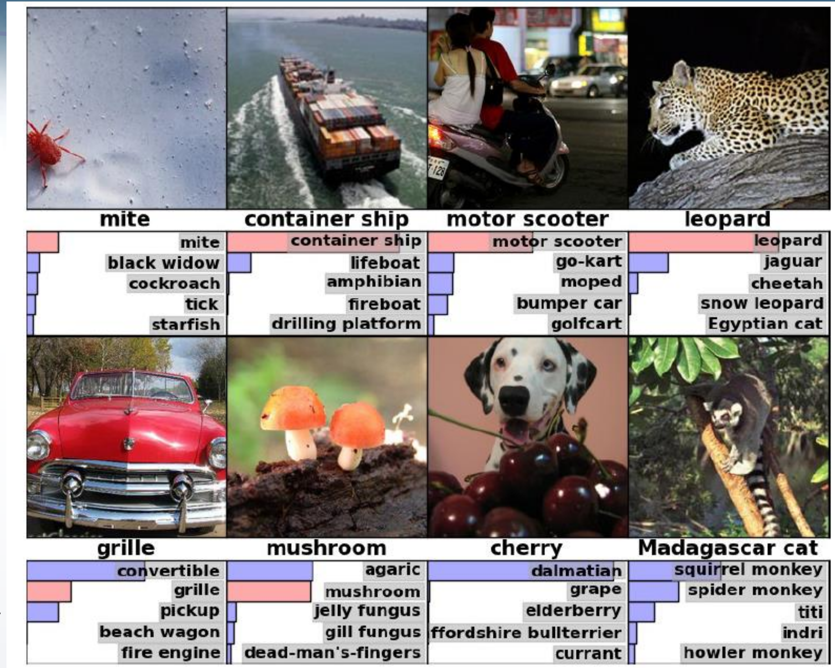


Figure 8. A geometric interpretation of the role of hidden unit in a two-dimensional input space.

ANNs: applications



- 28x28 pixel array
- 10 classes
- 7840 symapses



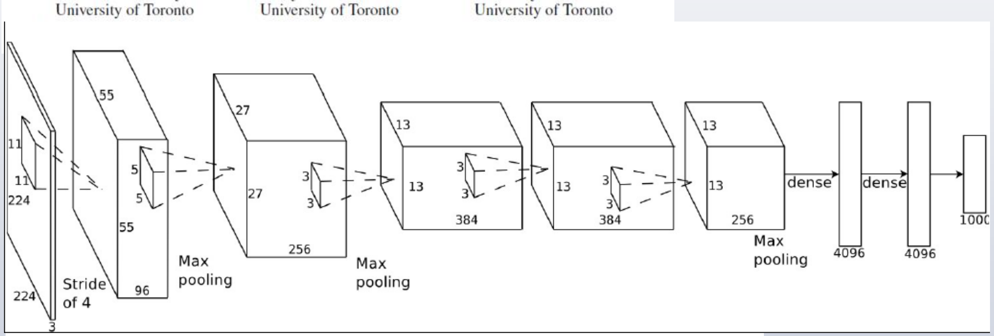
ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto

Ilya Sutskever
University of Toronto

Geoffrey E. Hinton
University of Toronto

- 256x256 pixel array
- 1000 classes



- Trained with stochastic gradient descent on two NVIDIA GPUs for about a week
- 650,000 neurons
- 60,000,000 parameters
- 630,000,000 connections

synapses

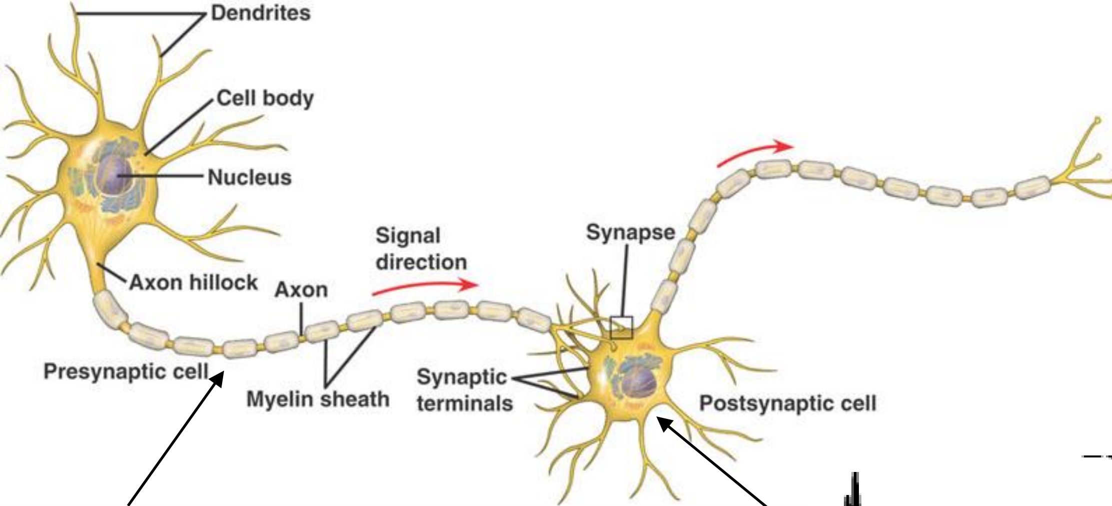
ANNs

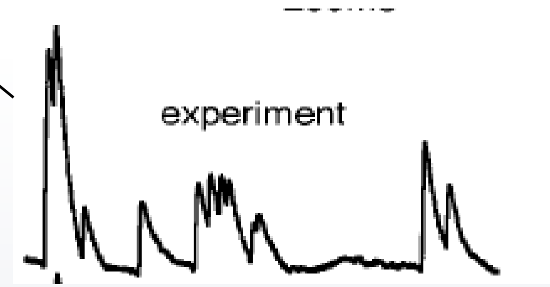
ANNs:

- Practical application will required ulra high density of nanodevices
- Main challenge: GPU or FPGA are serious challengers

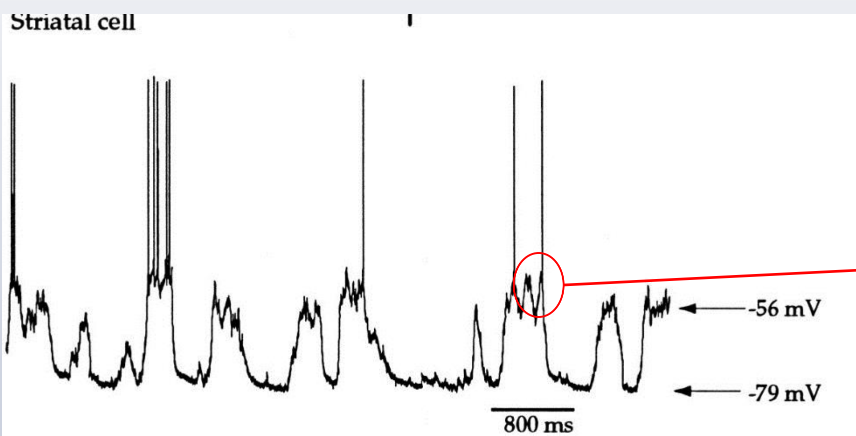
BNNs: neurons

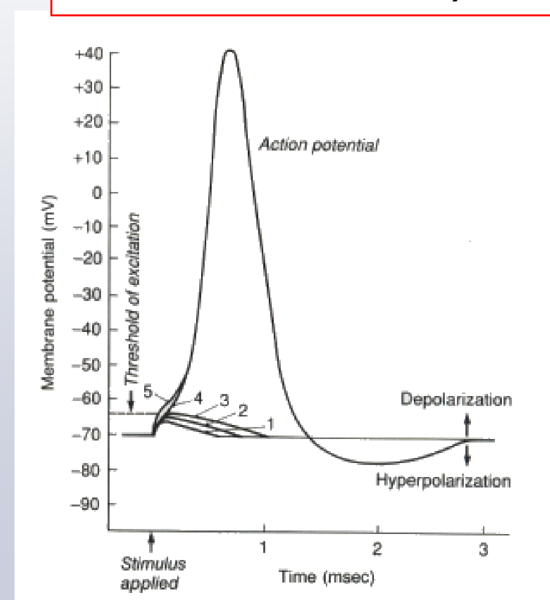
The neuron membrane can be seen as a transmission line (RC)
When the membrane potential reach a threshold, a spike is triggered



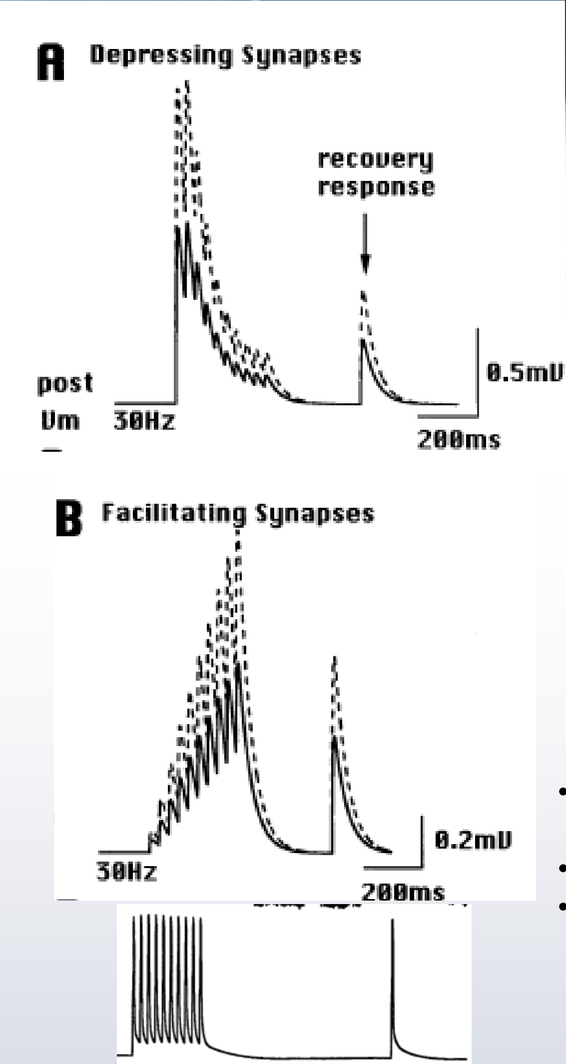


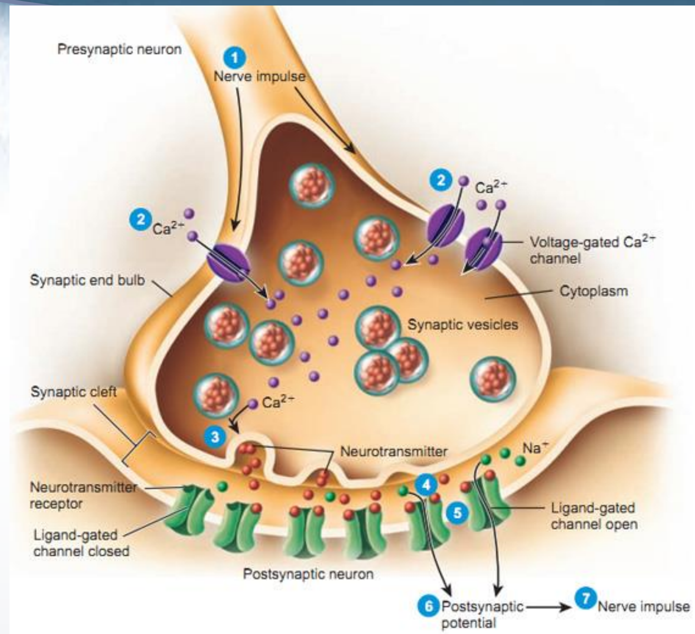
- Currents are ionic
- Low speed of propagation
- Active devices (soma and Ranvier's node)



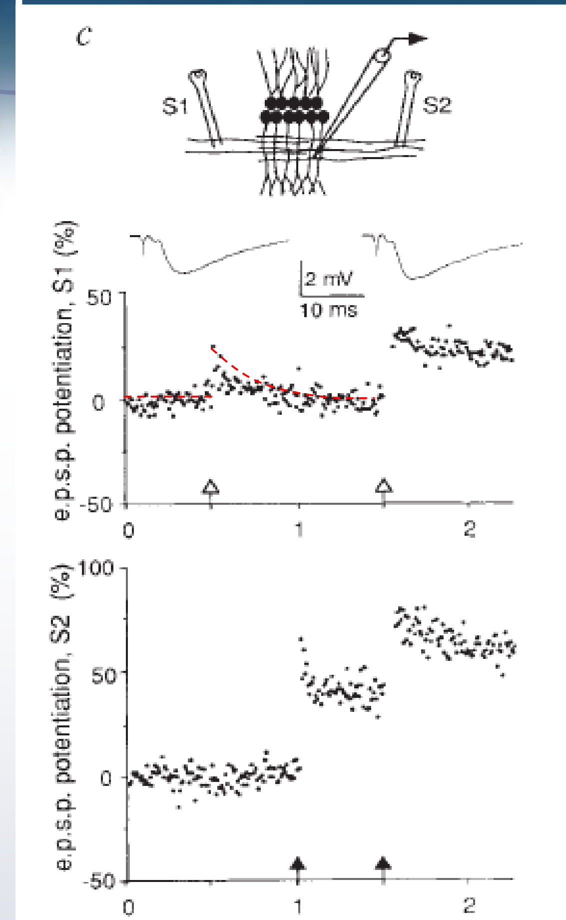


BNNs: synapses

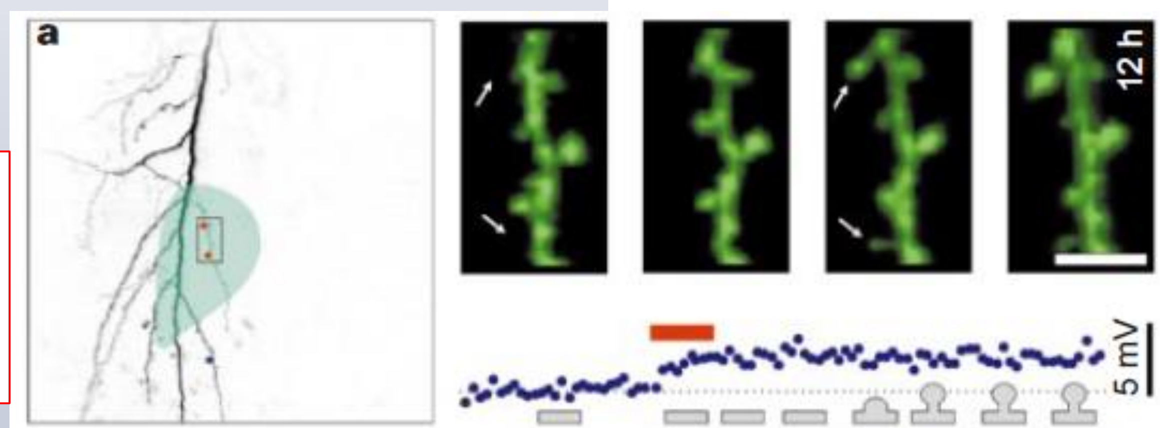




- The AP release neurotransmitters from the pre-neuron to the post neuron receptors
- Neurotransmitters open ionic channels
- Ionic concentration change the polarization of the post-neuron

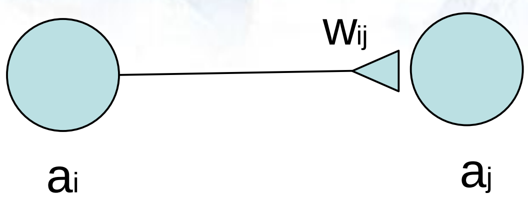


- Various time scale of plasticity
- Complex dynamics with many species (ions,NT,..)
- Unidirectionnal



BNNs: learning

Learning in BNNs:



Who fire together wire together, (Hebbs)

Synaptic learning

$$\frac{dw_{ij}}{dt} \propto a_i \cdot a_j$$

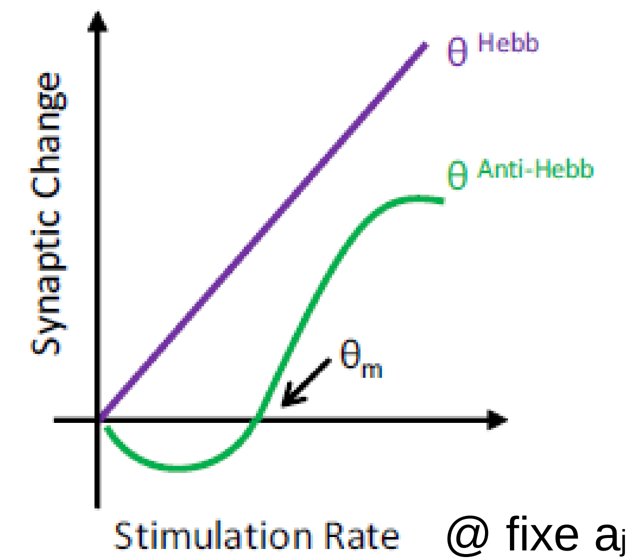
Synaptic adaptation

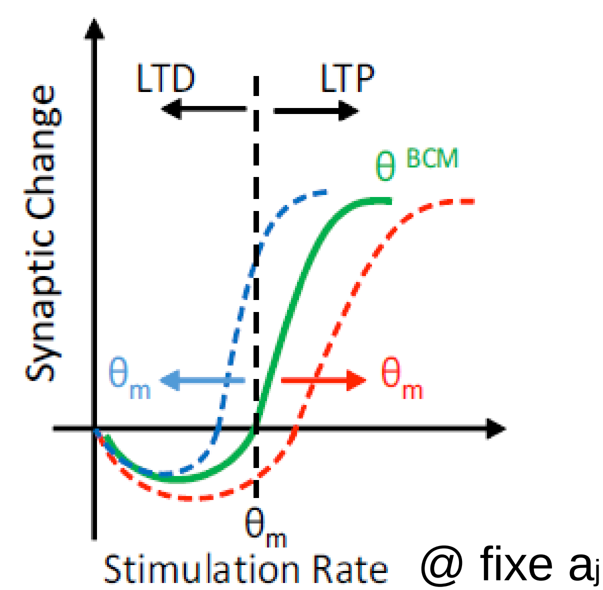
$$\frac{dw_{ij}}{dt} \propto a_i + a_j$$

Example, the BCM learning rule:

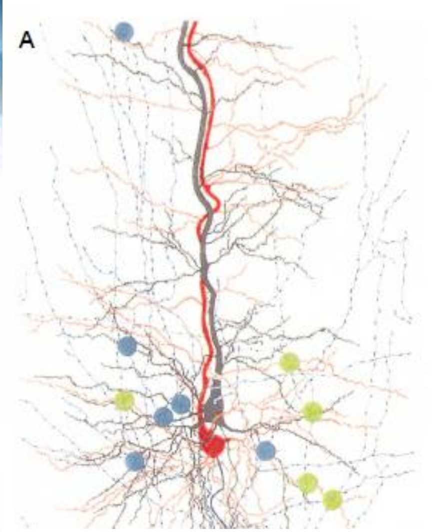
$$\frac{dw_{ij}}{dt} = \varphi(a_j(t)) \cdot a_i(t) - \varepsilon w_{ij}$$

$$\varphi(a_j) < 0 \quad for\ a_j < \theta_m \qquad \& \qquad \varphi(a_j) > 0 \quad for\ a_j > \theta_m$$

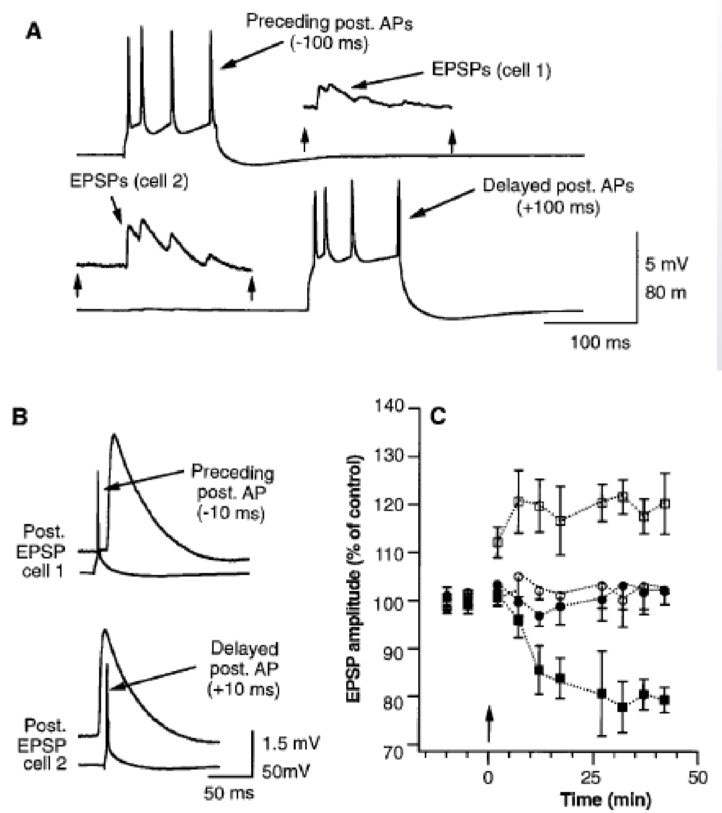


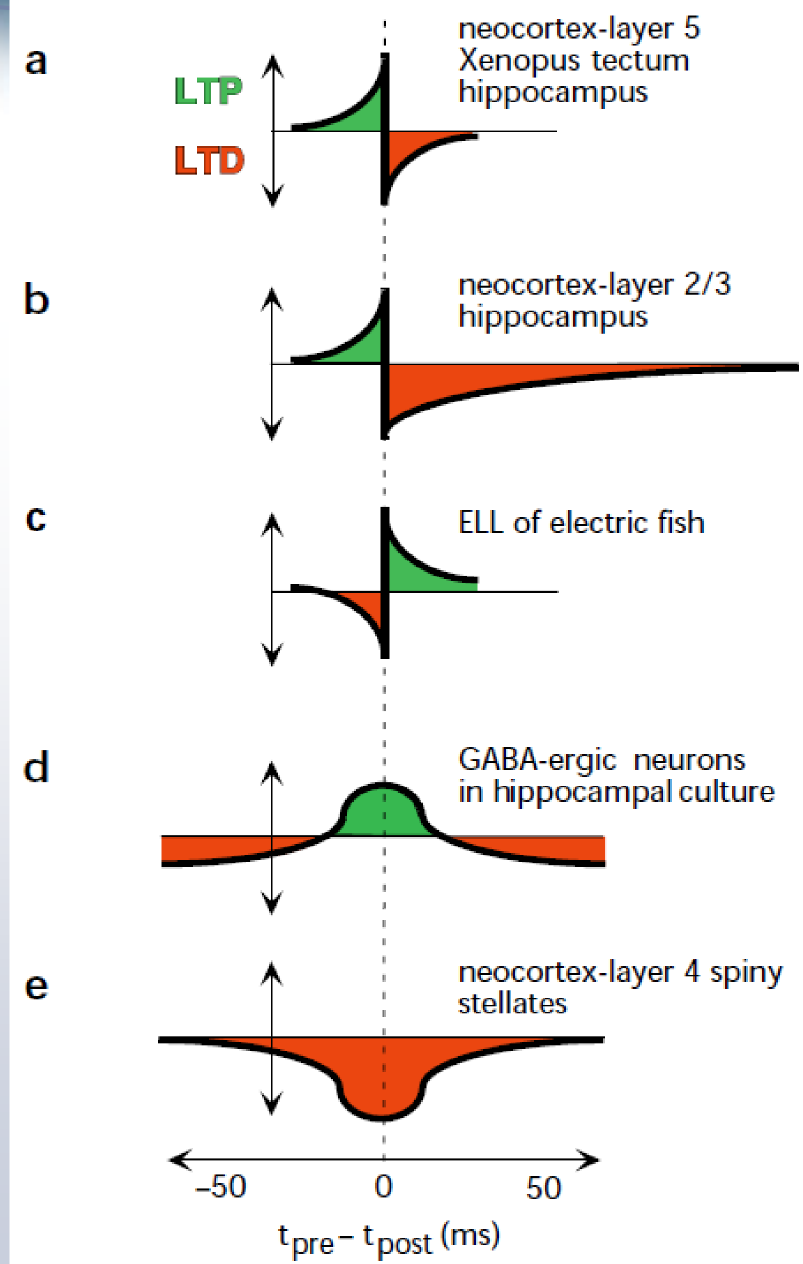


BNNs: learning with STDP



Variation of Hebbs rule
- Unsupervised
-



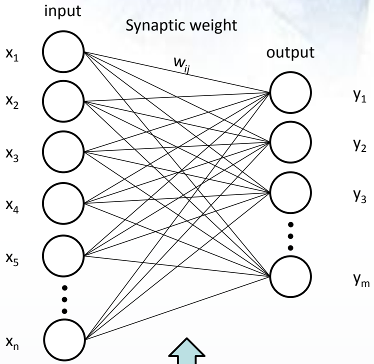


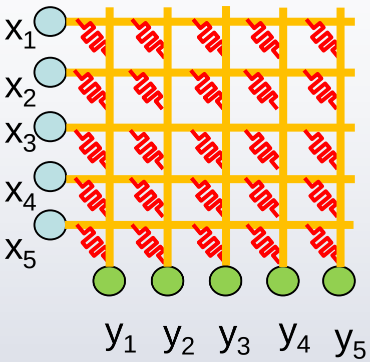
BNNs:

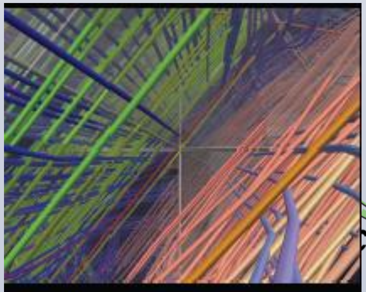
- No charges (i.e. electrons), only ions
- Slow, with rich dynamics
- Still unsolved issues
  - Basics of computing in the brain (coding,…)
  - What do we really need for computing (for practical applictions)

Neuromorphic: Main stream for nanodevices

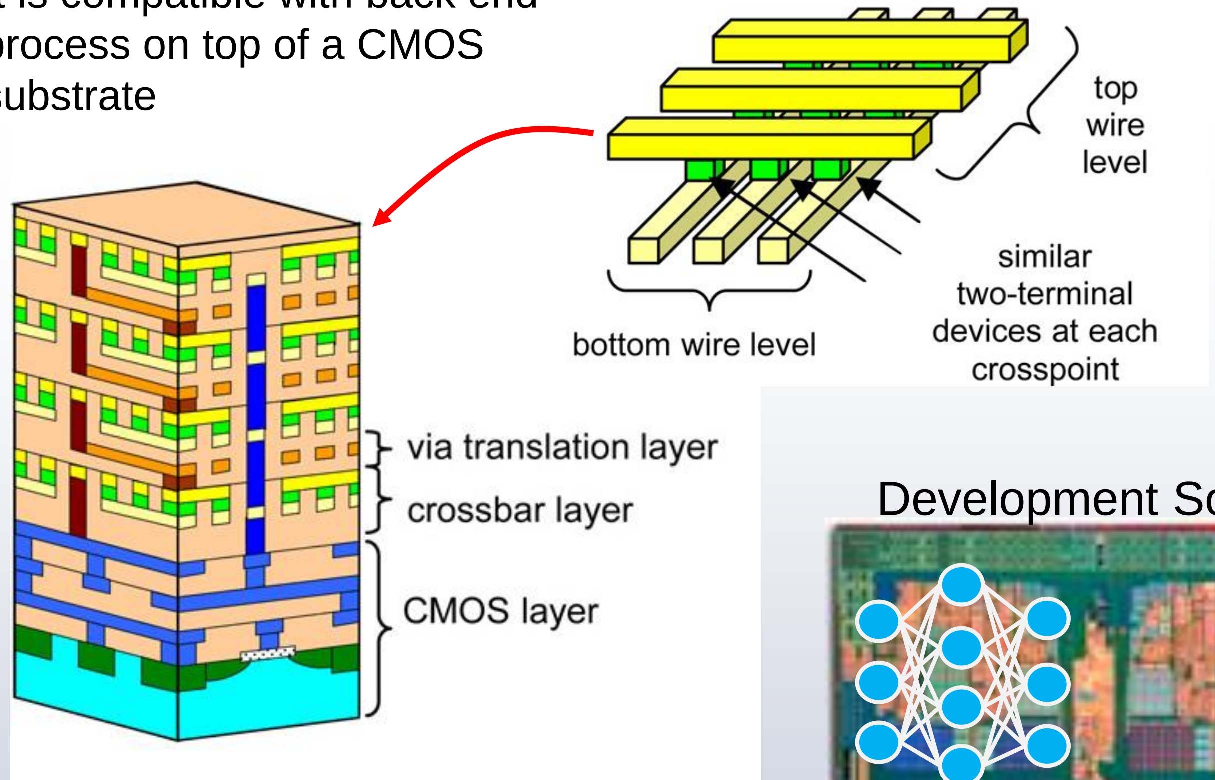
The crossbar structure is the perfect architecture for massively parallel processing





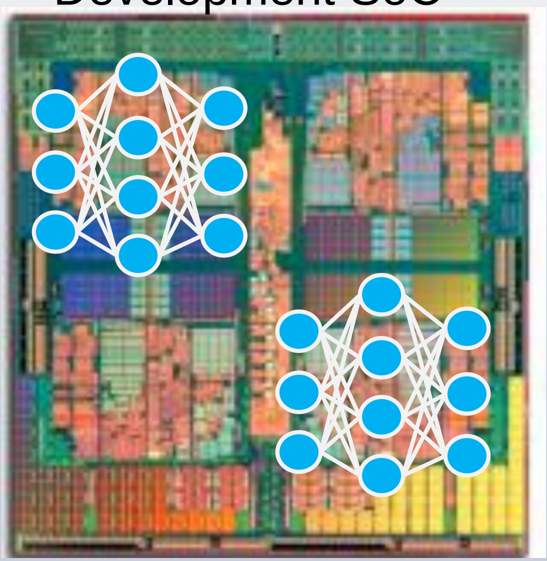


It is compatible with back end process on top of a CMOS substrate



Strukov, PNAS 2009

Development SoC



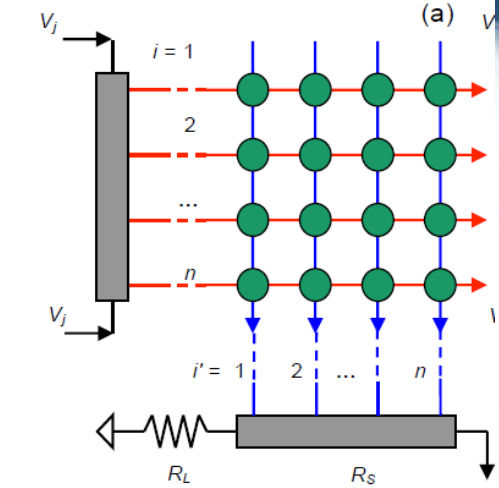
Footprint $4F^2$

10^{12} devices/cm^2

NANO FOR ANNs

ANNs: synapse, digital or analog?

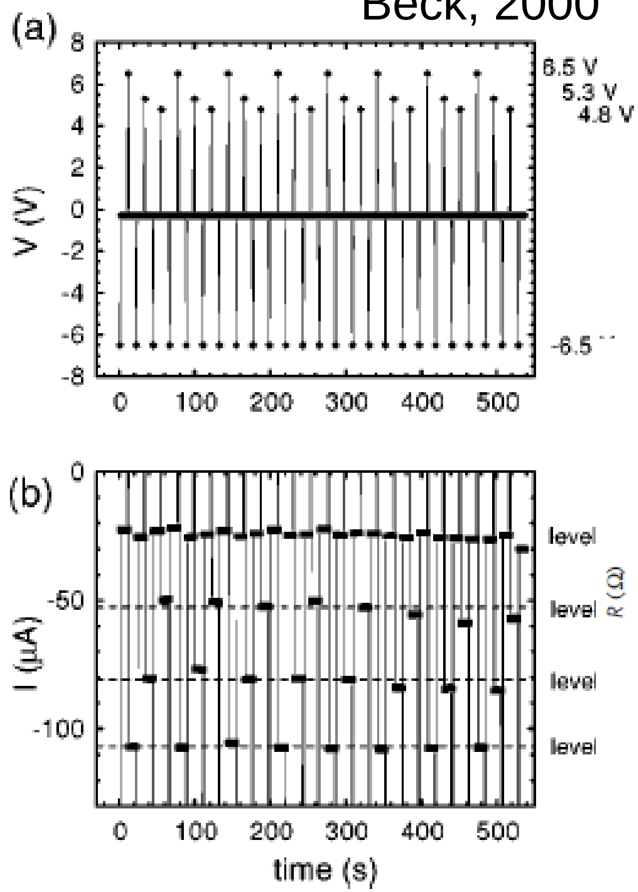
Crossnet



L= 2n² +1 states

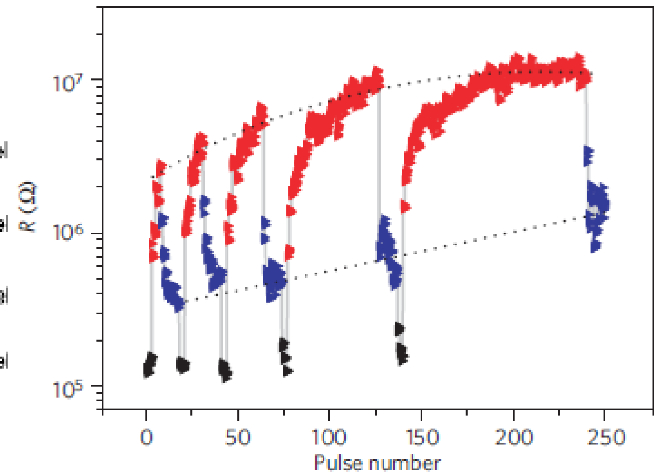
From binary…

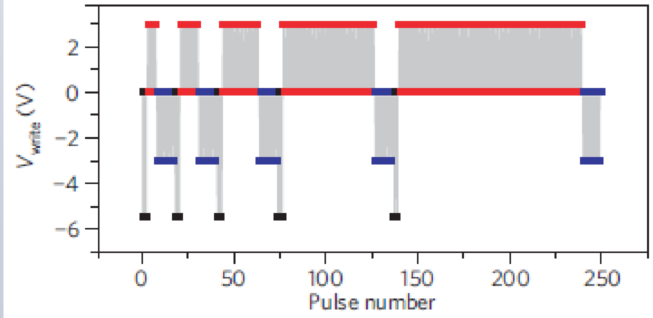
Beck, 2000



… to multilevel…

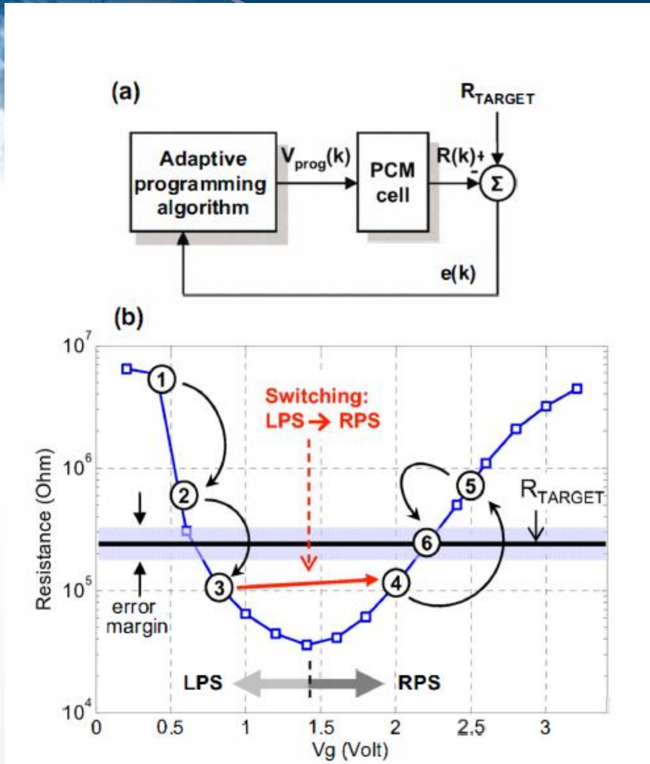
Chanthbouala, 2012



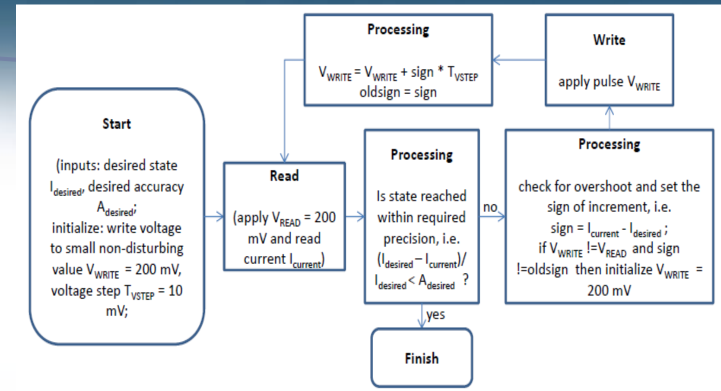


… to analog

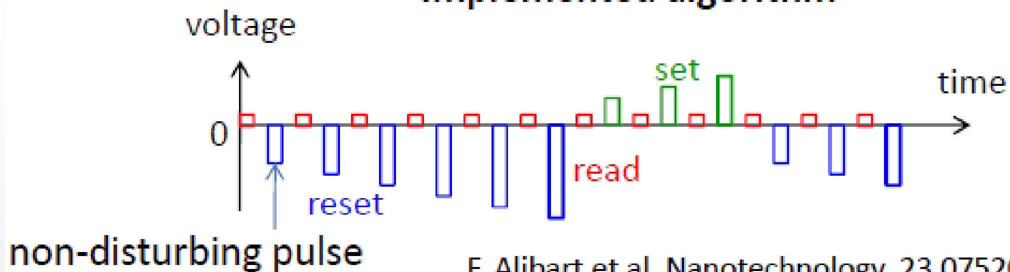
ANNs: the analog synapse



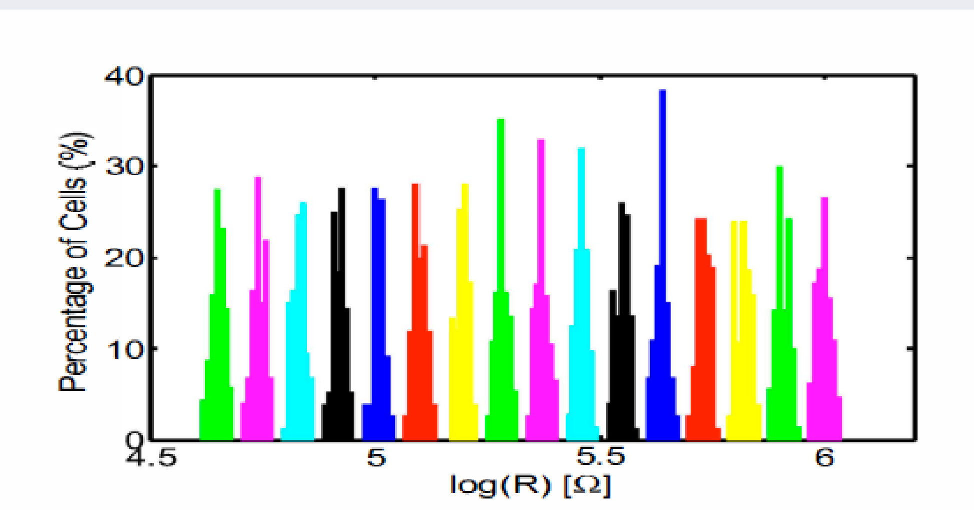
Papandreou, 2011

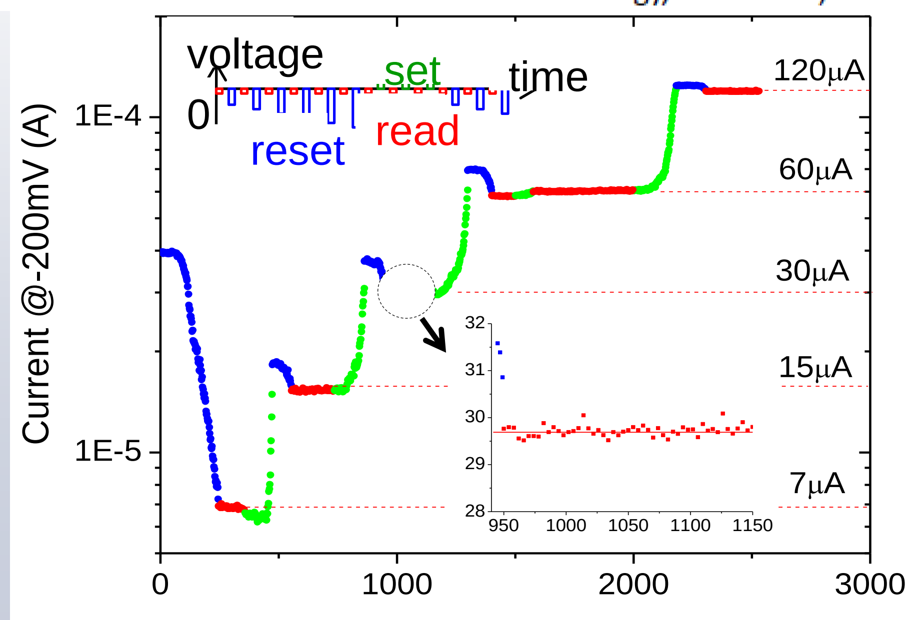


Implemented algorithm



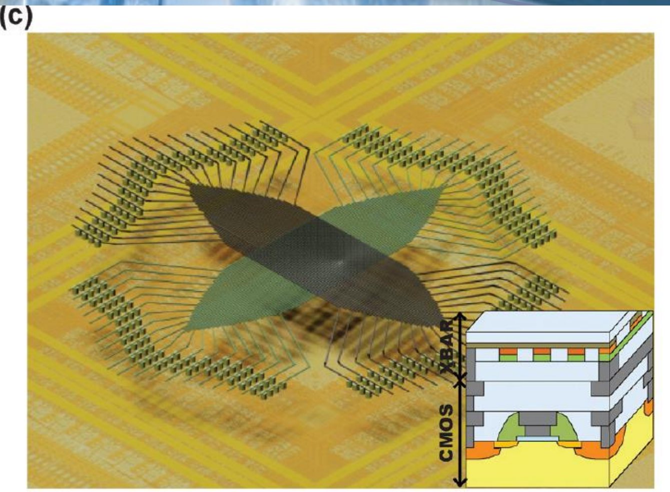
F. Alibart et al. Nanotechnology, 23 075201, 2012

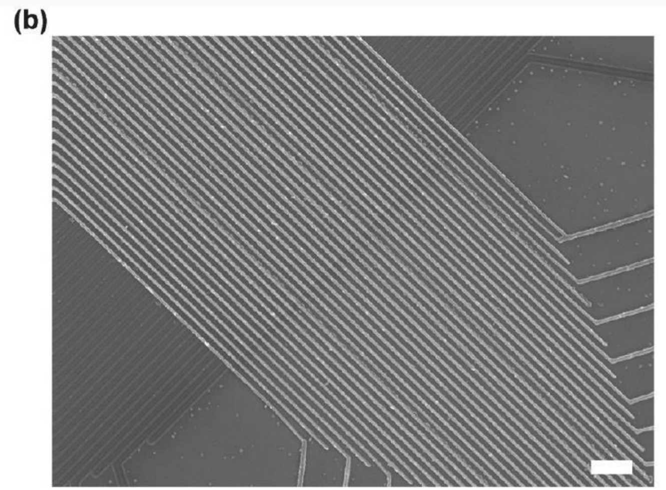


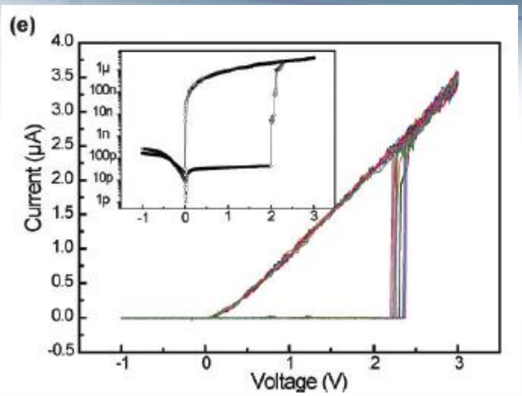


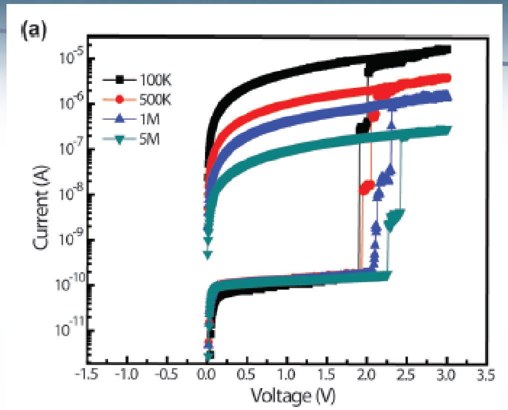
W. Lu, Nanoletters,2011

ANNs: implementations

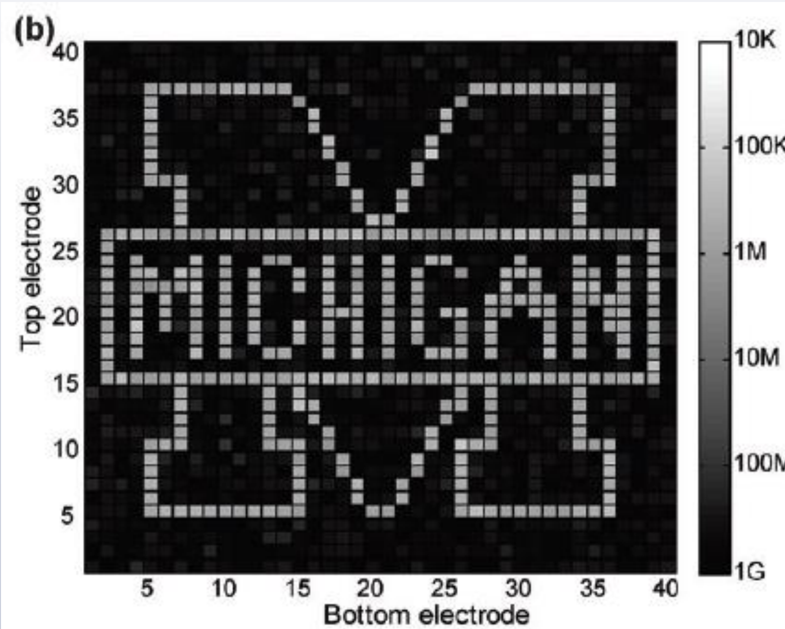






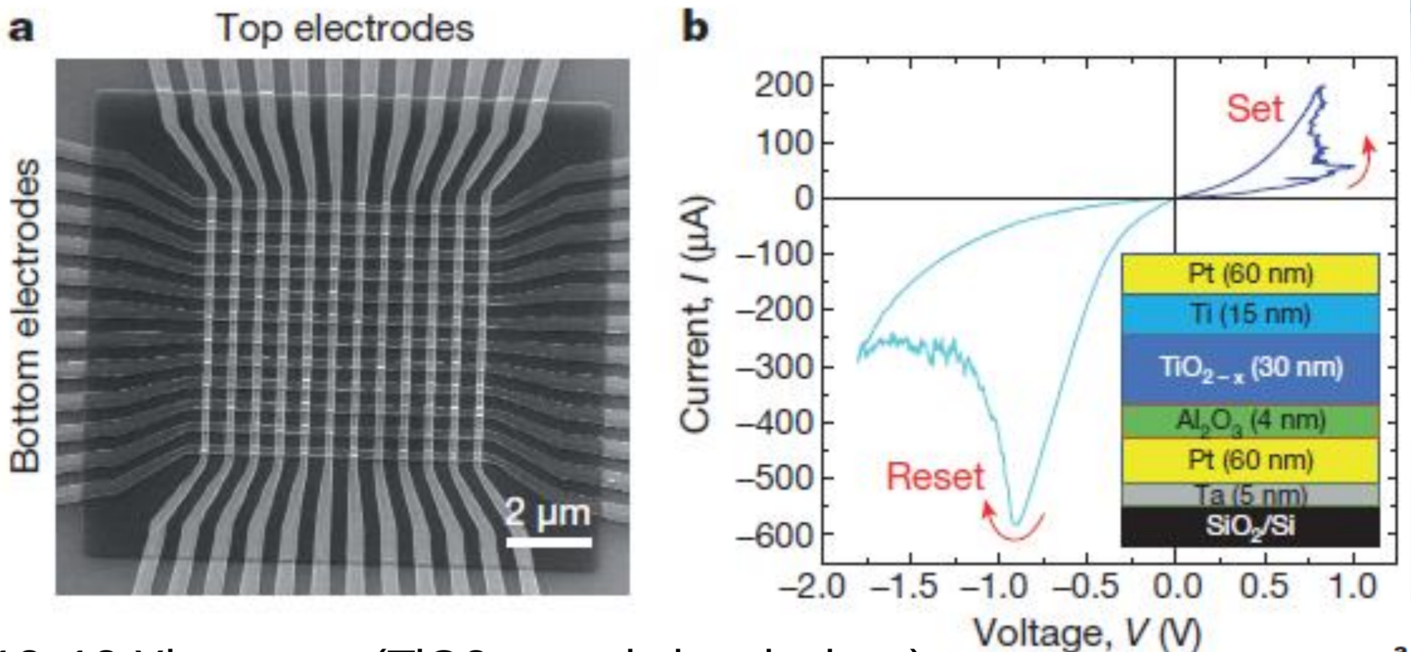


- 50x50 pasive crossbar array
- Binary operation (multilevel, more or less)
- CBRAM technology



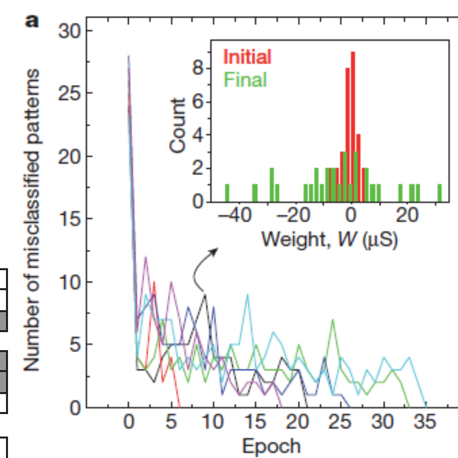
ANNs: implementations

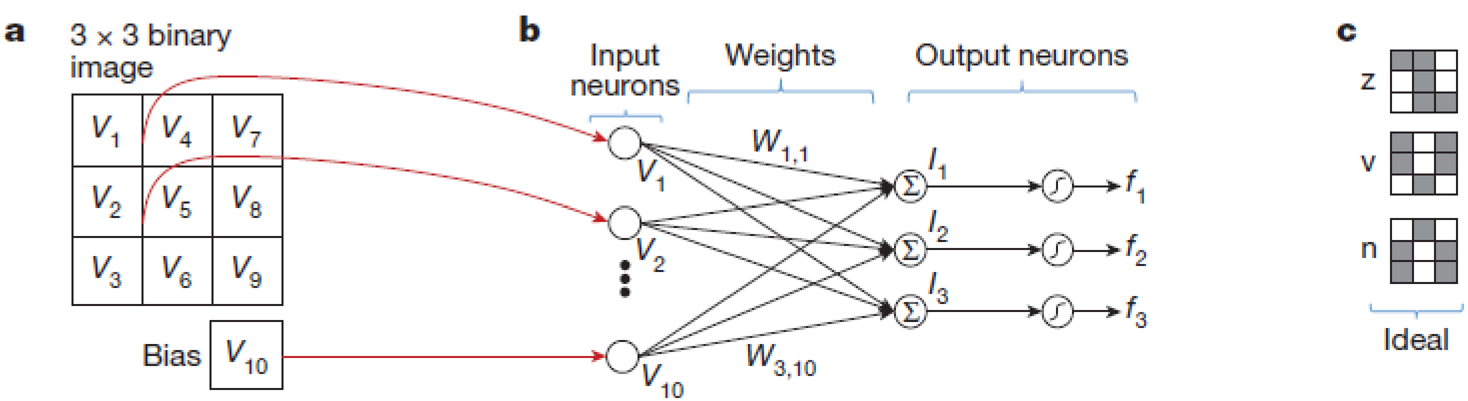
Strukov, 2015



12x12 Xbar array (TiO2 memristive devices)

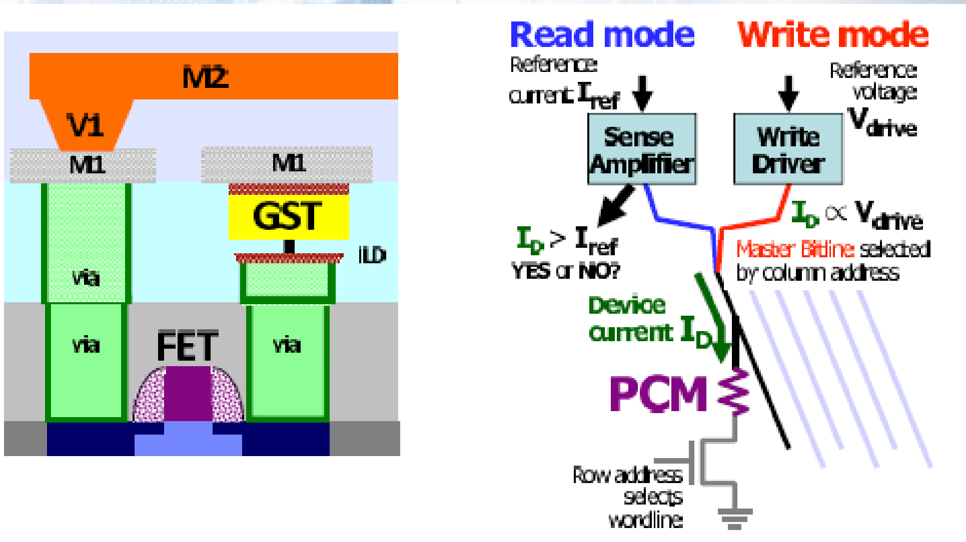
- Online training (variation of delta rule)
- Three classes, 60 synapses

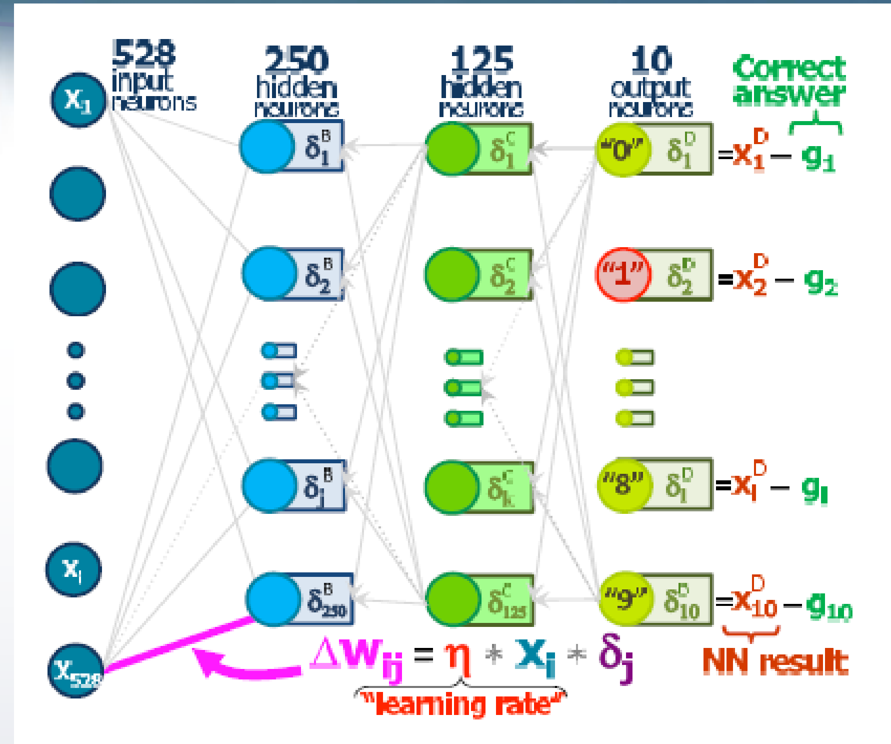


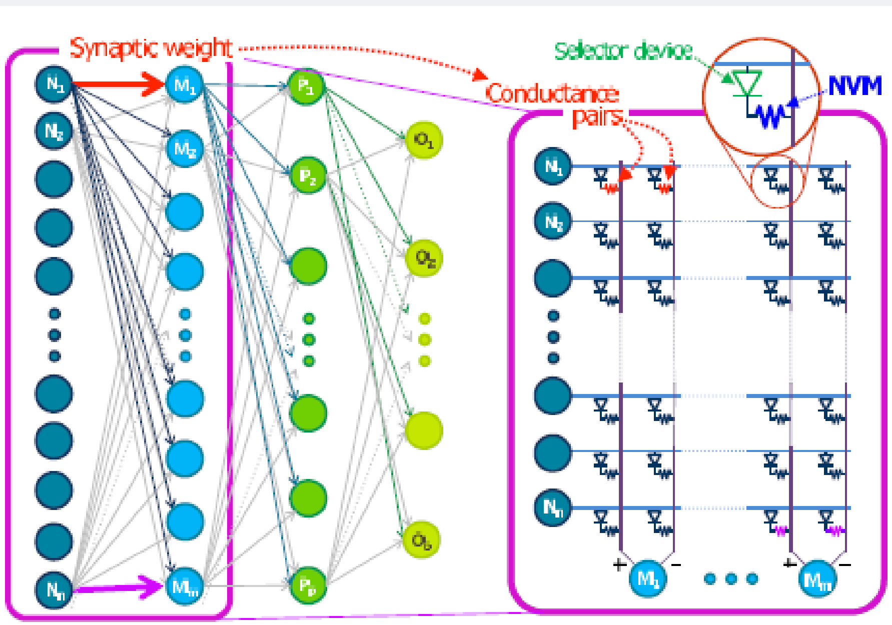


Burr, IEDM 2014

ANNs: implementations





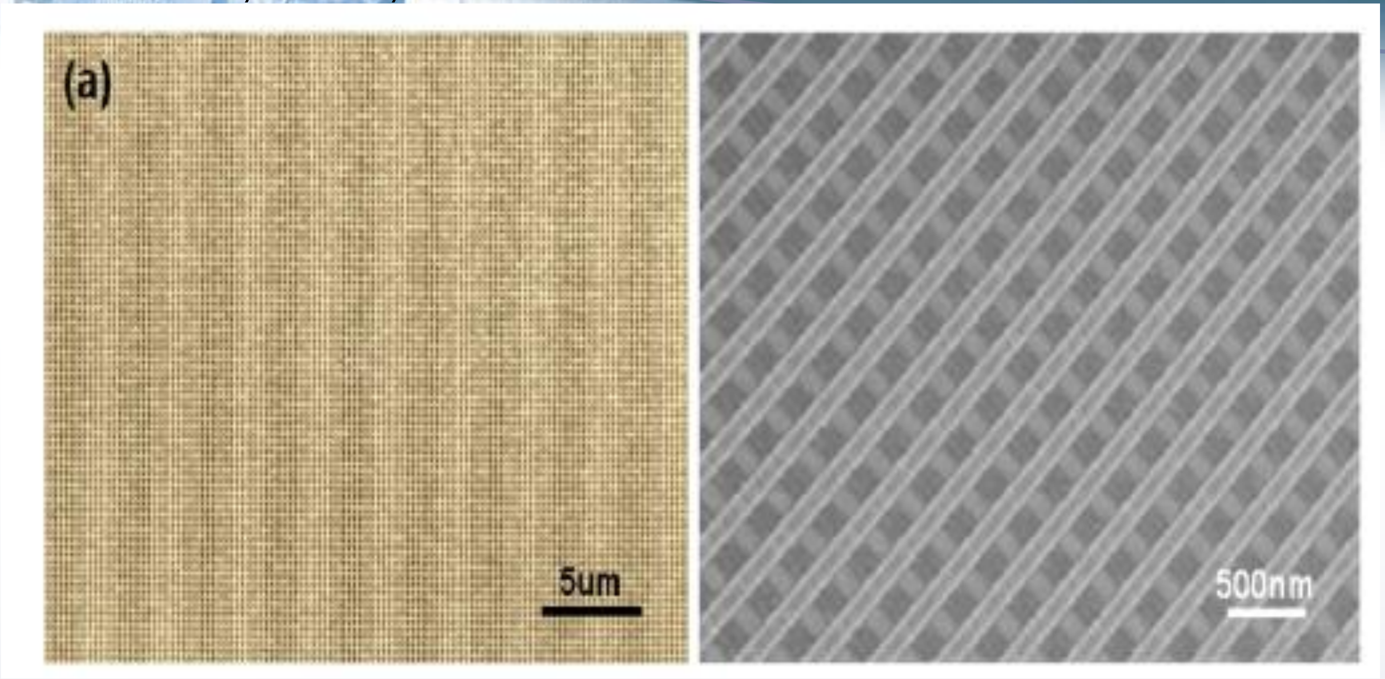


- Demo of multilayer perceptron (with backprop)
- 165000 PCM analog synapses

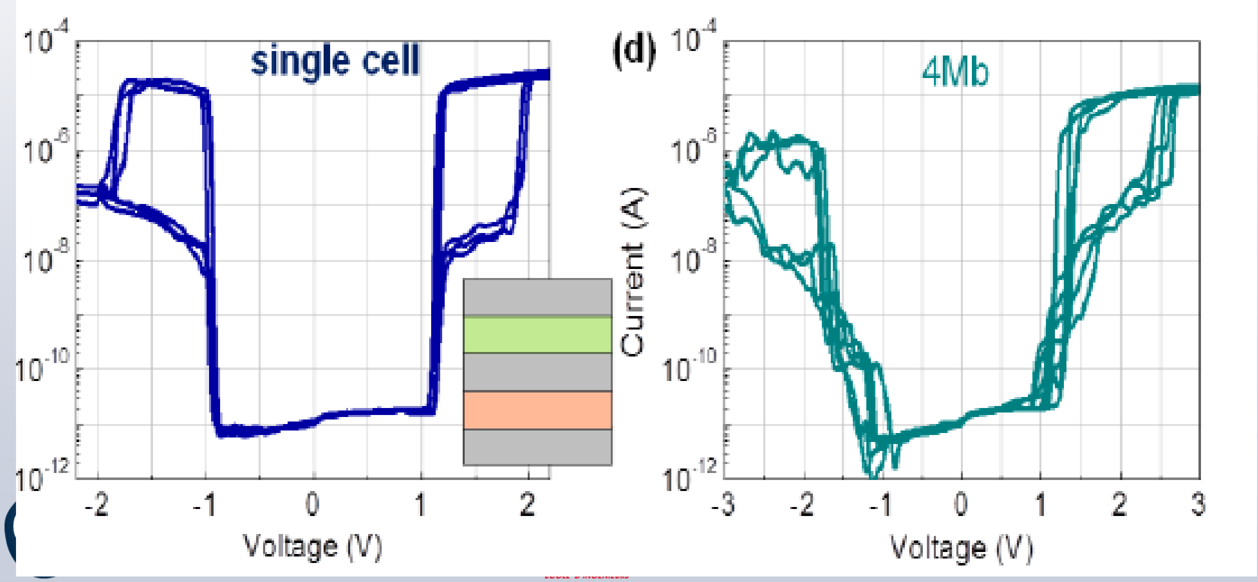
INSTITUT CARNOT IEMN
iemn UMR CNRS 8520 Recherche Formation Transfert

ANNs: implementations

Crossbar, IEDM, 2014



- Record density 4Mb
- Binary only
- 100nm half pitch
- Fully functionnal!



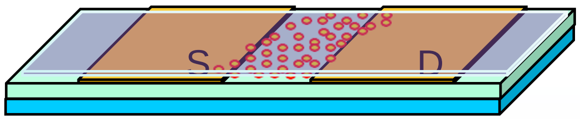
Nano for ANNs

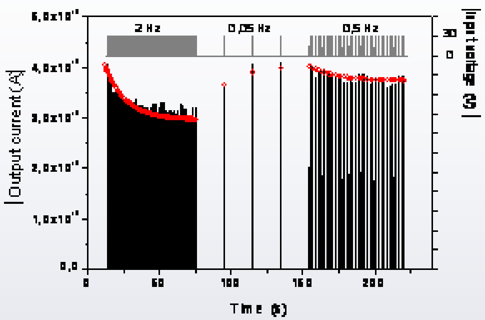
- Still a huge gap betwen requirements and what is available (from the memory perspectives)
- Do we need learning (i.e. smart memory) or pure memory (i.e. storage + analog)
- Co-integration still not demonstrated

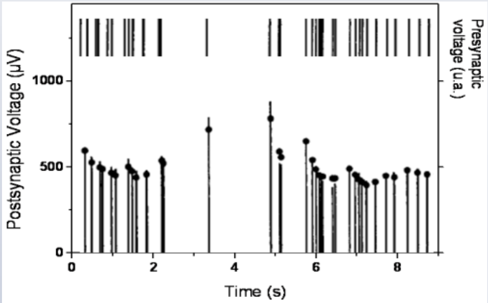
NANO FOR BNNs

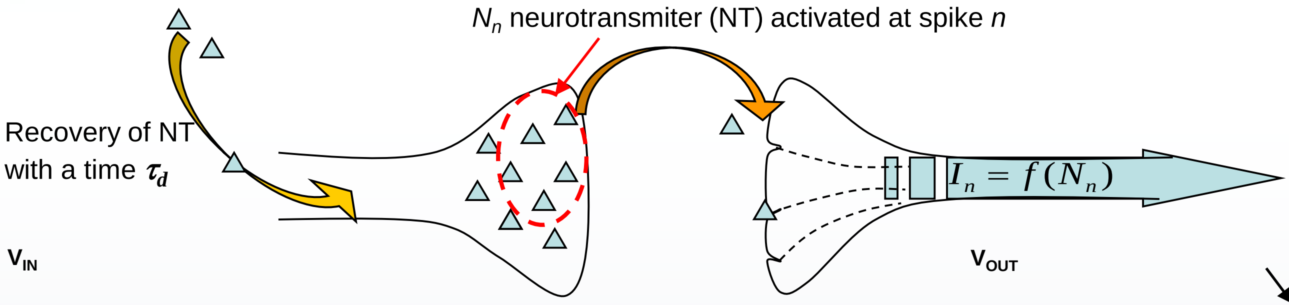
BNNs: synaptic plasticity (1)

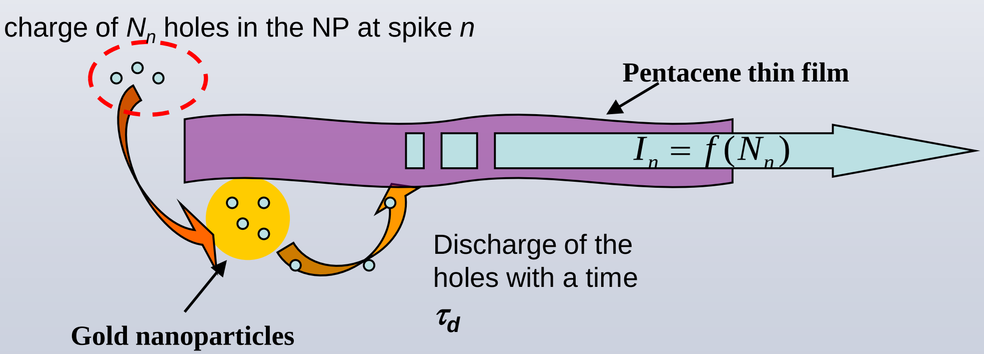
STP





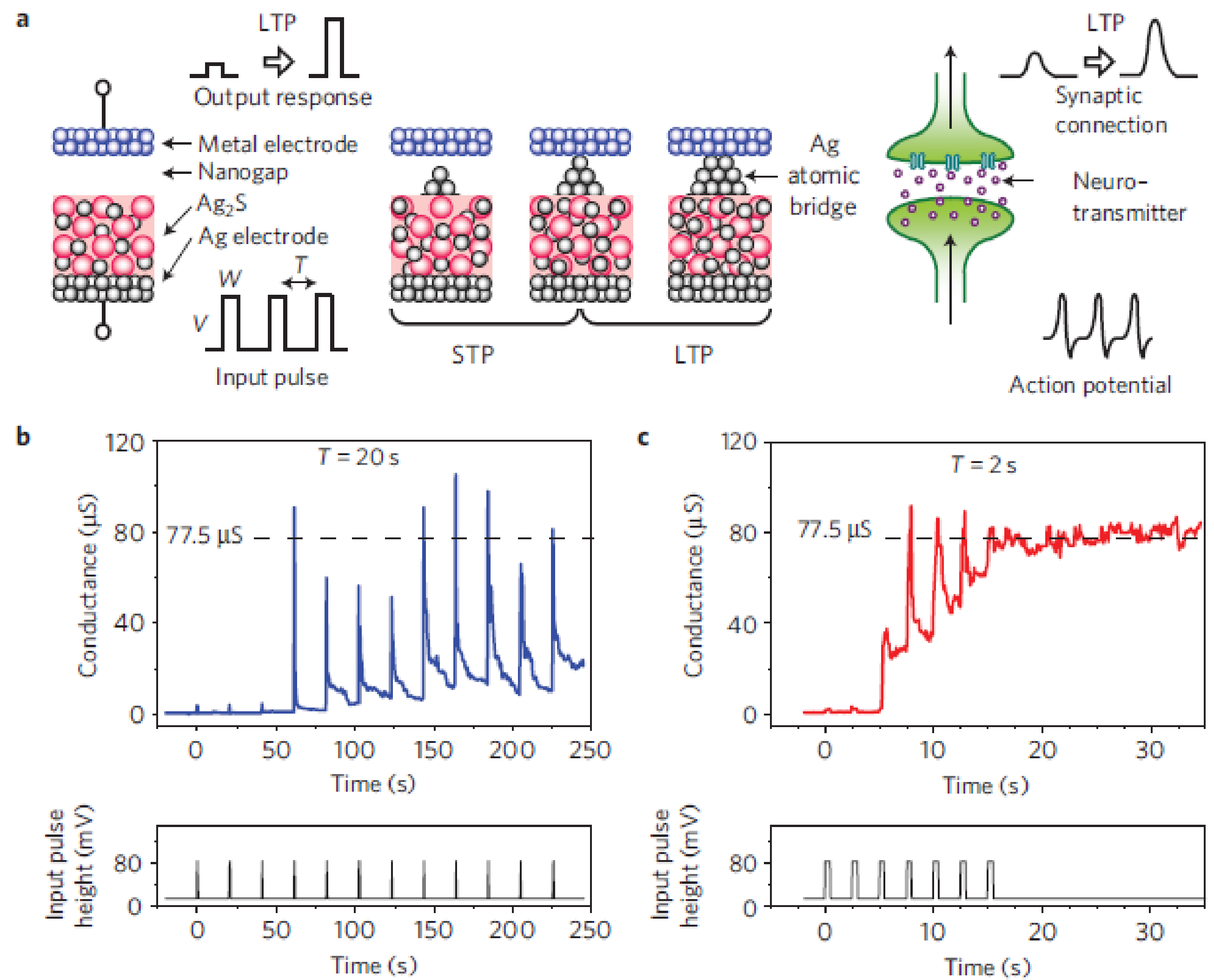




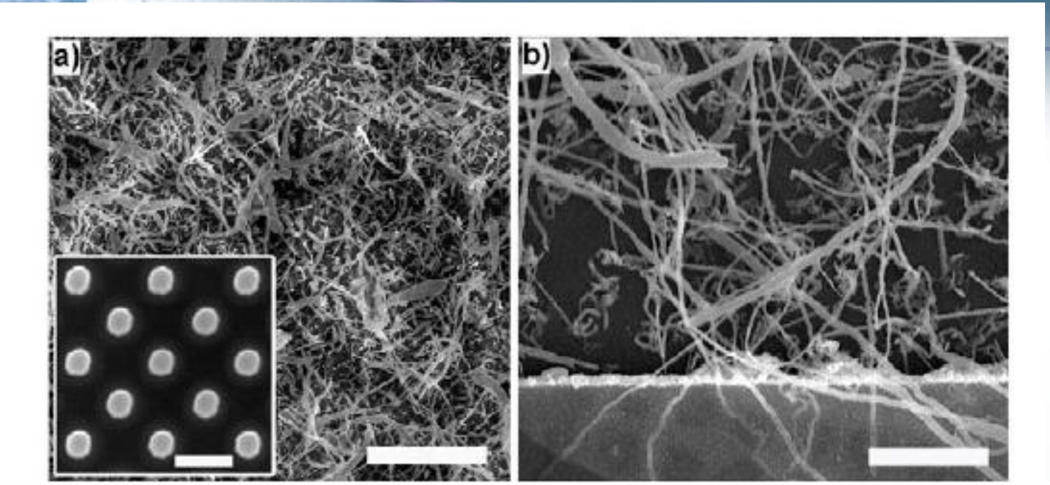


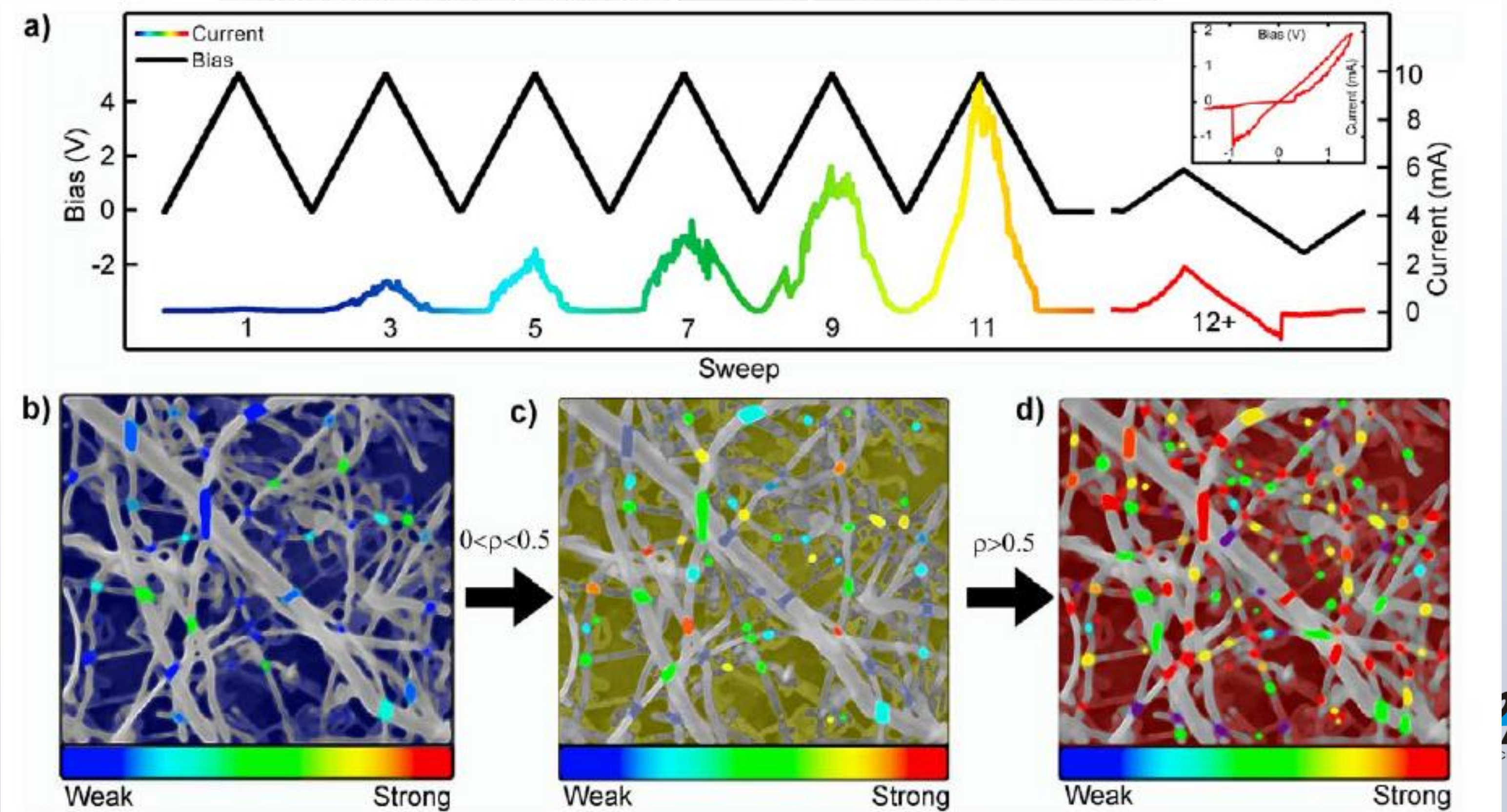
(Alibart, Adv. Funct. Mat, 2011

BNNs: synaptic plasticity (2)



BNNs: network scale implementations



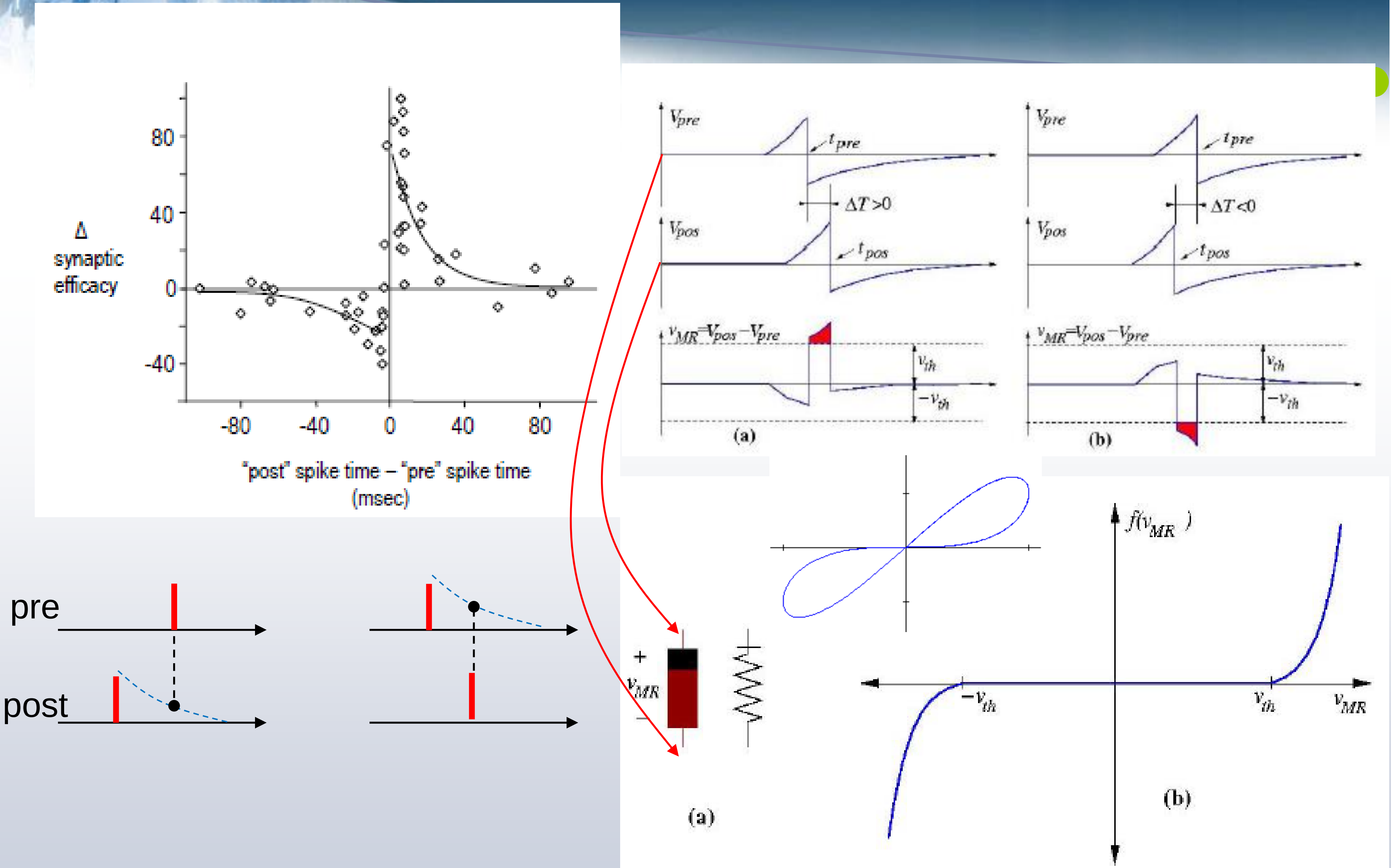


Nano for BNNs

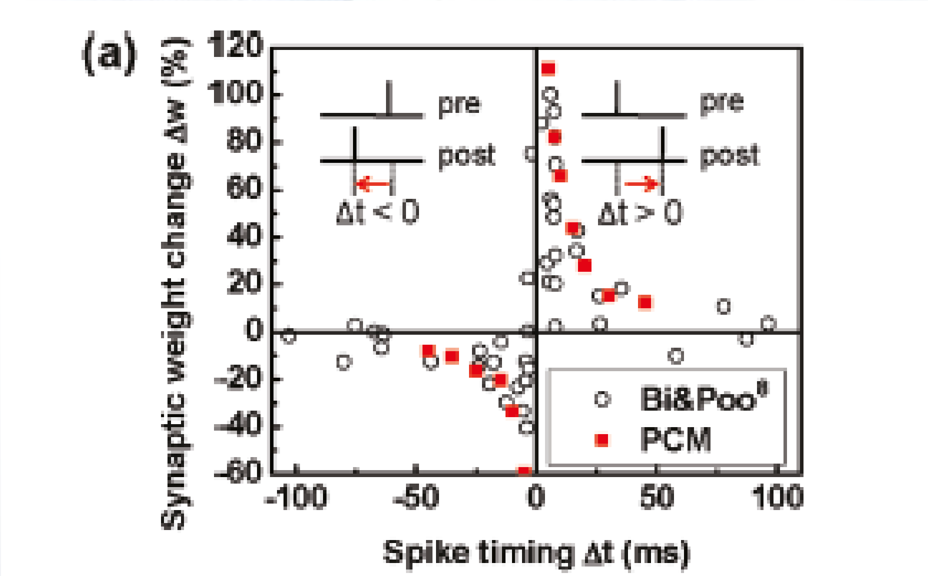
- Still very emerging
- No clear idea of applications but a very complex (and exciting!) field
- Can we map BNNs (ionic systems) with electronic devices (instead of ionic)

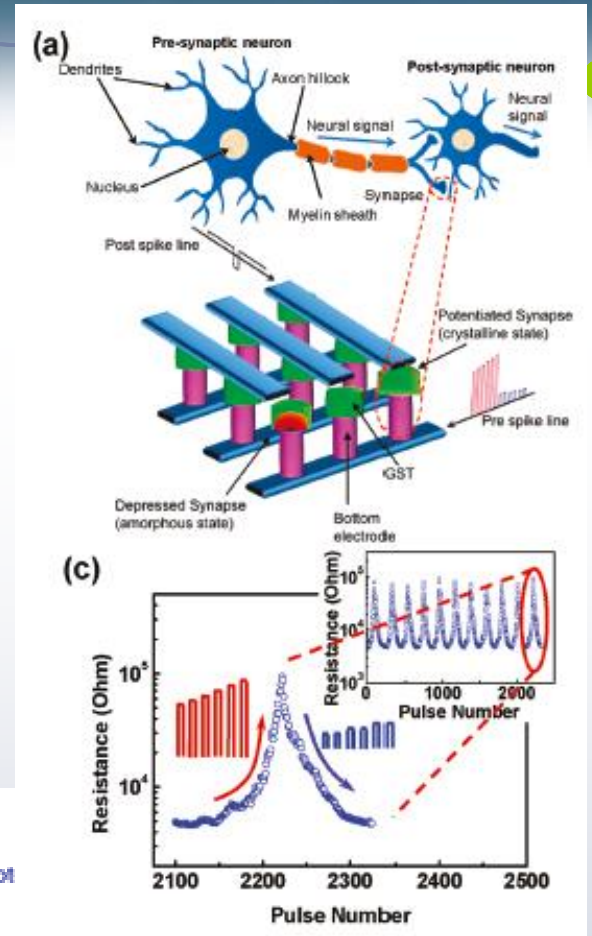
NANO IN BETWEEN: STDP

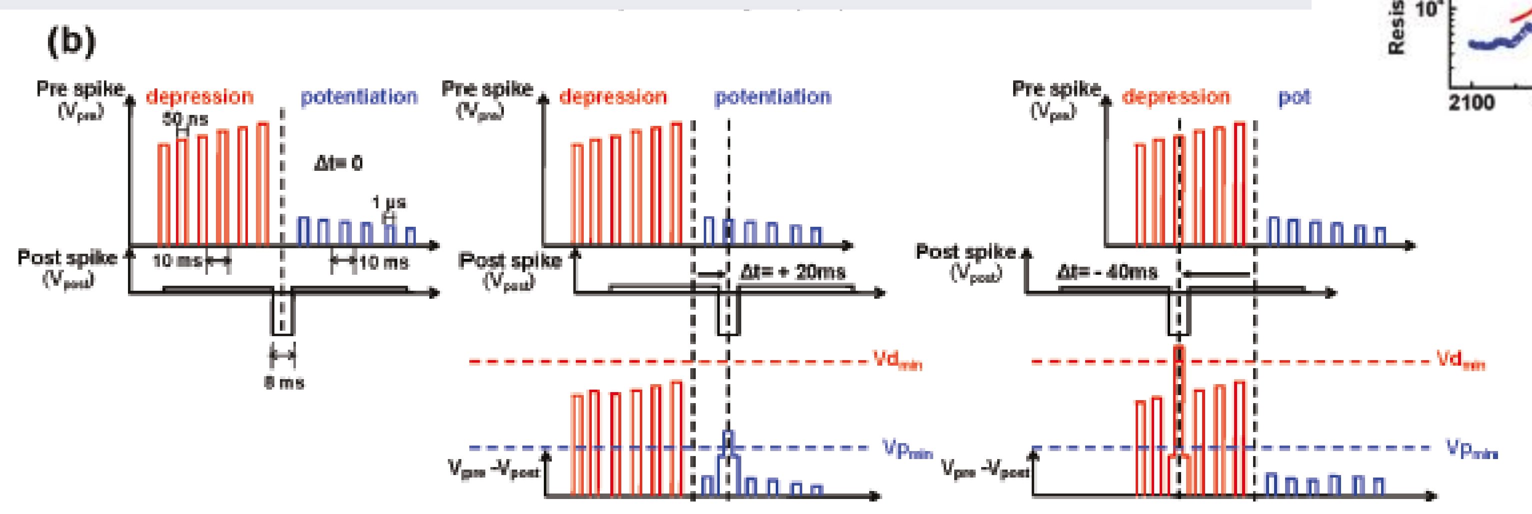
STDP with memristors



STDP: practical implementations

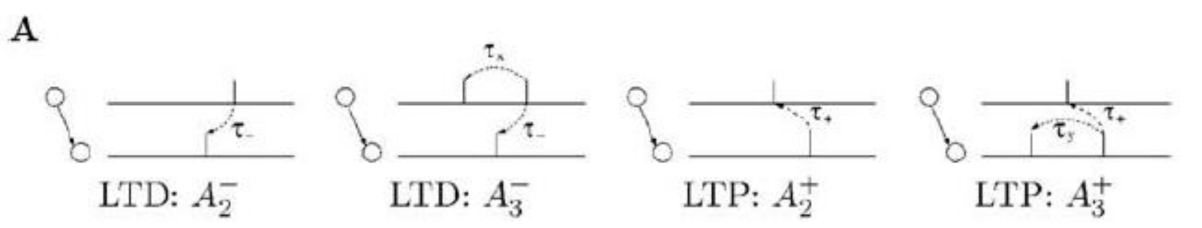


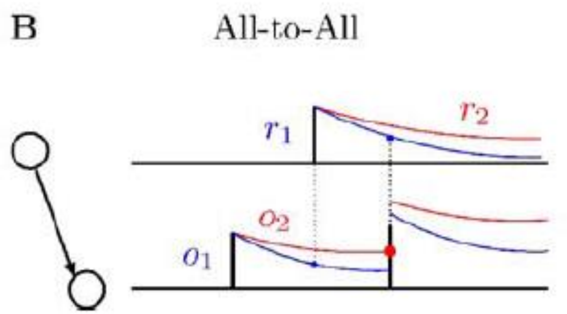


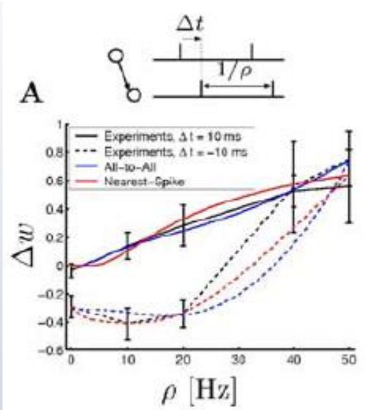


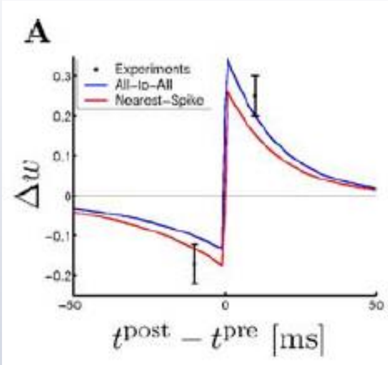
iemn
UMR CNRS 8520
Recherche
Formation
Transfert

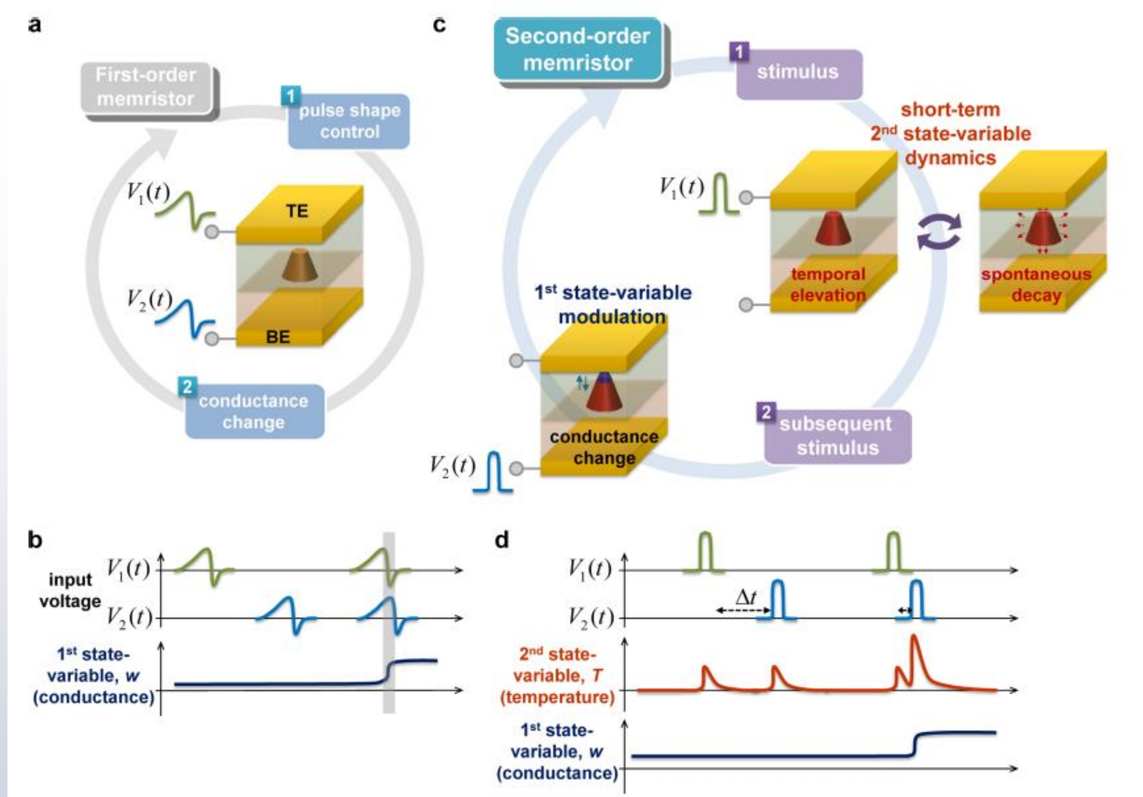
STDP: triplet rule











STDP

- No hardware demo at the network level
- One layer OK, what about multi-layers?
- STDP but what else?

Concluding remarks

- Neuromorphic in between BNNs and ANNs. Maybe an issue for visibility (what is our community?)
- Neuromorphic as a bridge between ANNs and BNNs?
- Doing more than identifying neuromorphic in nano, it is time to built it!

Thank you!

(One final tip: The hippocampus is not an animal)