FROM RESEARCH TO INDUSTRY

cea tech

www.cea.fr

leti & list

Advanced Technologies for Brain-Inspired Computing

Fabien CLERMIDY*

Rodolphe Heliot*, Alexandre Valentian*, Christian Gamrat**, Olivier Bichler**, Marc Duranton**, Bilel Blehadj# and Olivier Temam#

.

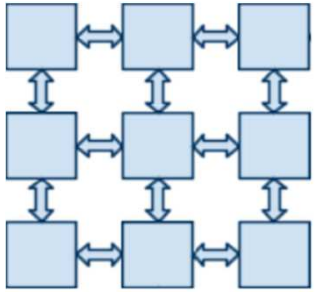
***CEA-Leti, Grenoble, France**

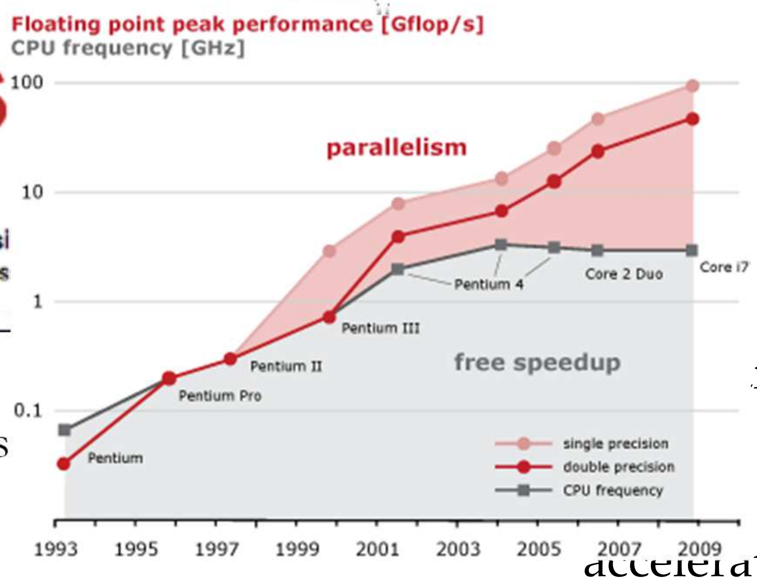
****CEA-List, Paris, France**

 # INRIA, Paris, France

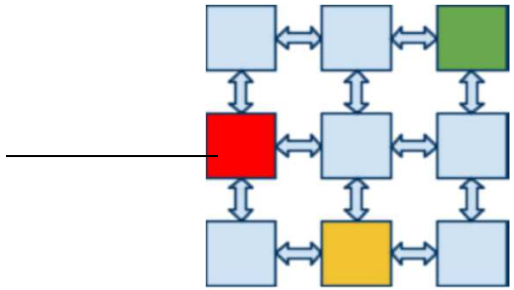
Context: dark silicon…

- Dramatic changes in the area of computing architectures
- Power has become the main limiting factor for the scalability of microprocessors
  - Multi-cores architectures





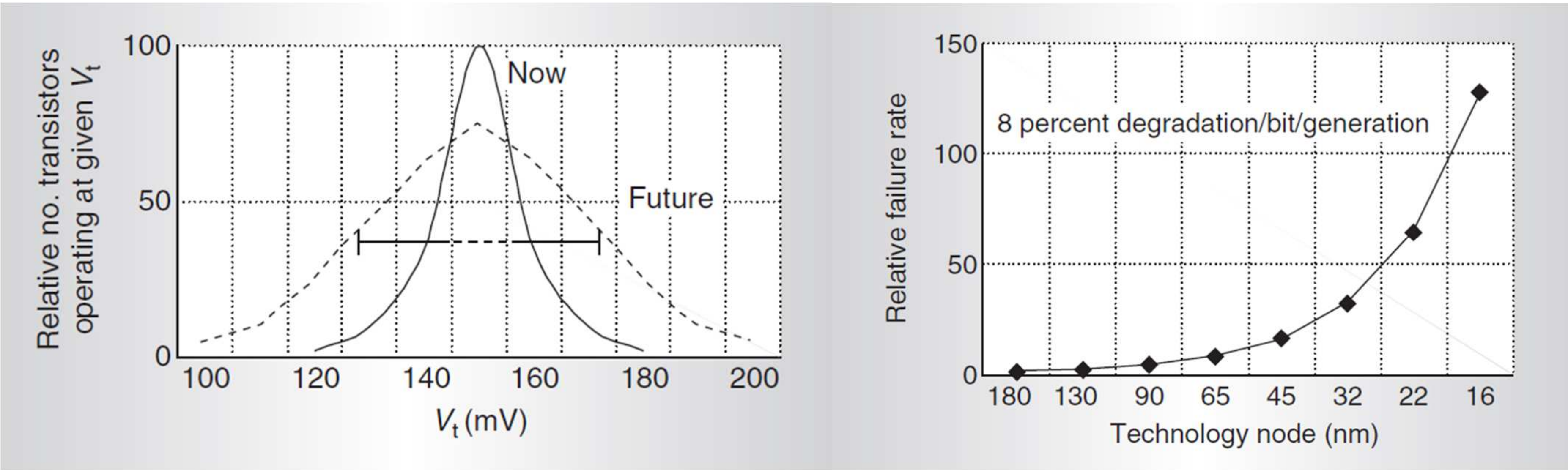
- Simultaneous … nsistors
  - Heterogenous … accelerator



➢ Need for versatile hardware accelerators

...and variability

- Dramatic changes in the area of computing architectures
- Robustness issues

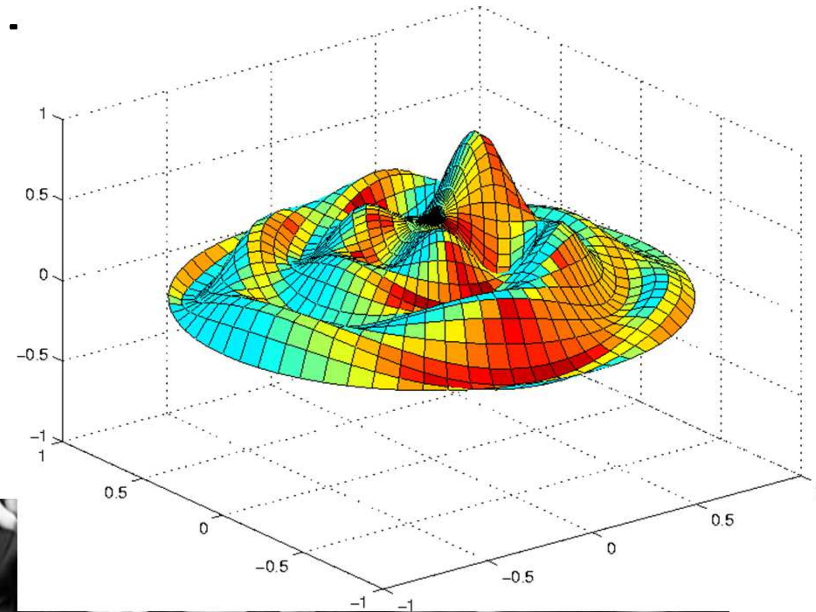


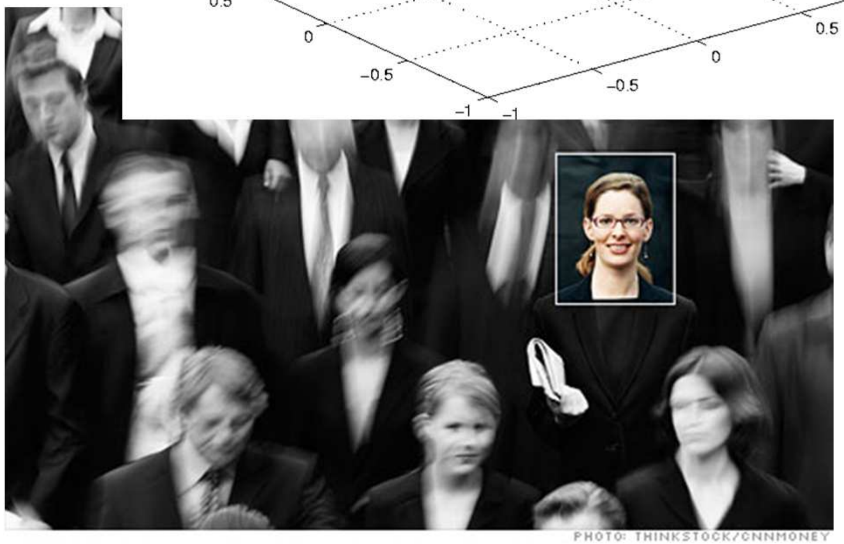
(Borkar 2005)

➢ Robust accelerators

for new applications

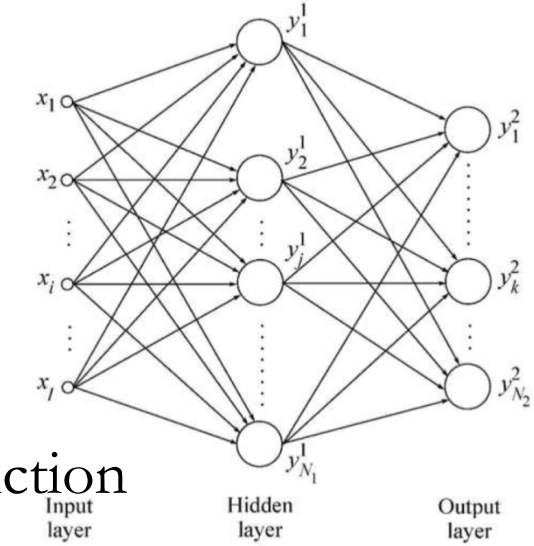
- Today : scientific computing, cryptography, …
- ***on, Mining, Synthesis***





A potential answer : Neural networks

- Neural Networks are good candidates
  - They provide Robustness



Signal processing

Approximation

  - Good application scope

Prediction

Optimization

Clustering

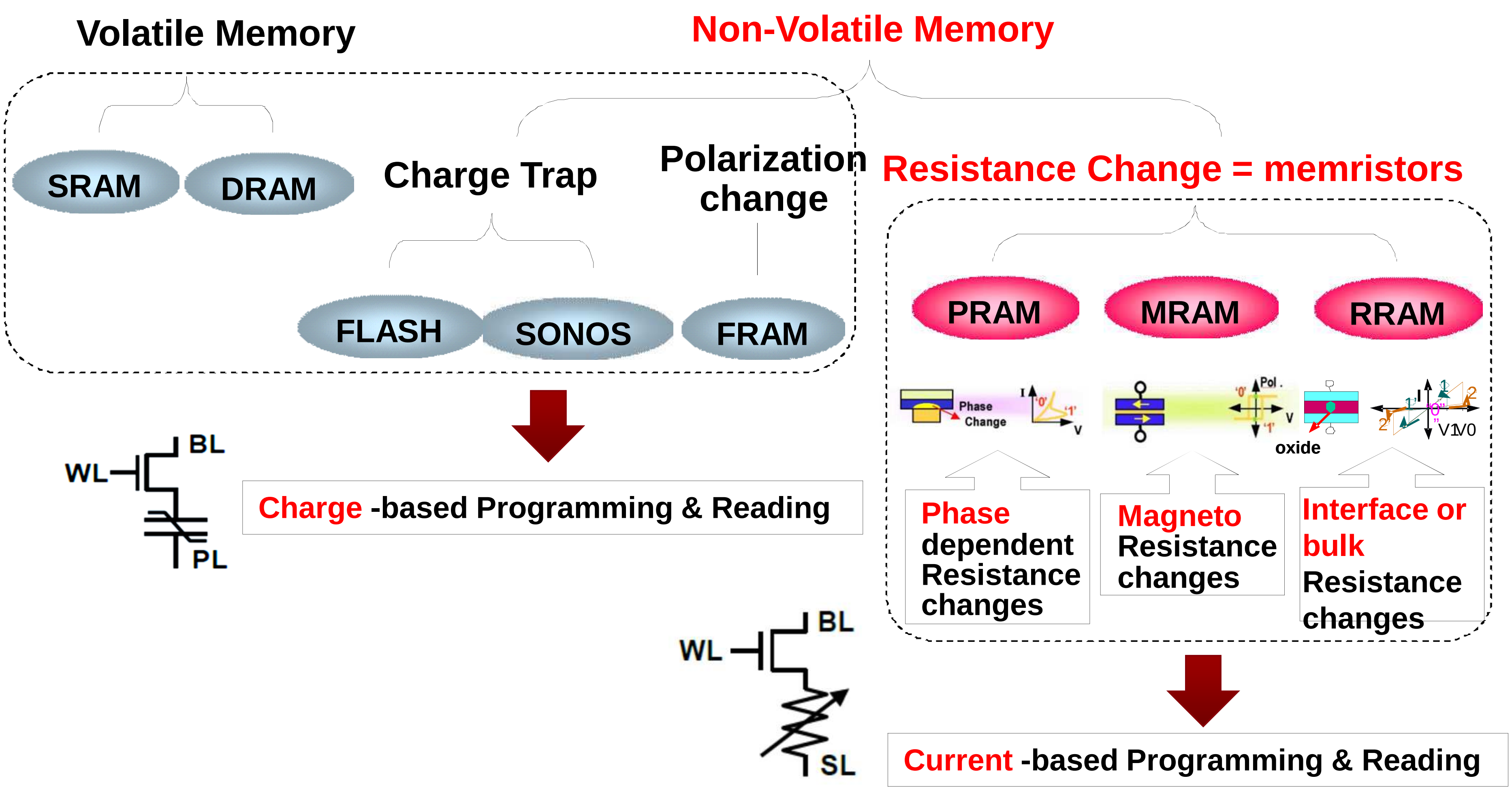
Classification

  - But classically limited by 2D hardware solutions
    - Limited interconnections, costly long range ones
    - But the HW situation has changed !

Outline

- Context
- **Exploiting resistive memories**
- Analogue neurons and 3D-TSV
- Towards monolithic 3D…
- Conclusion

New memory technologies

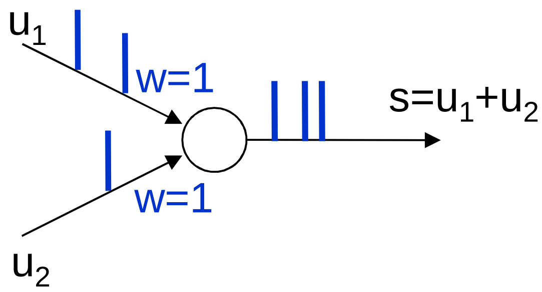


Memristors

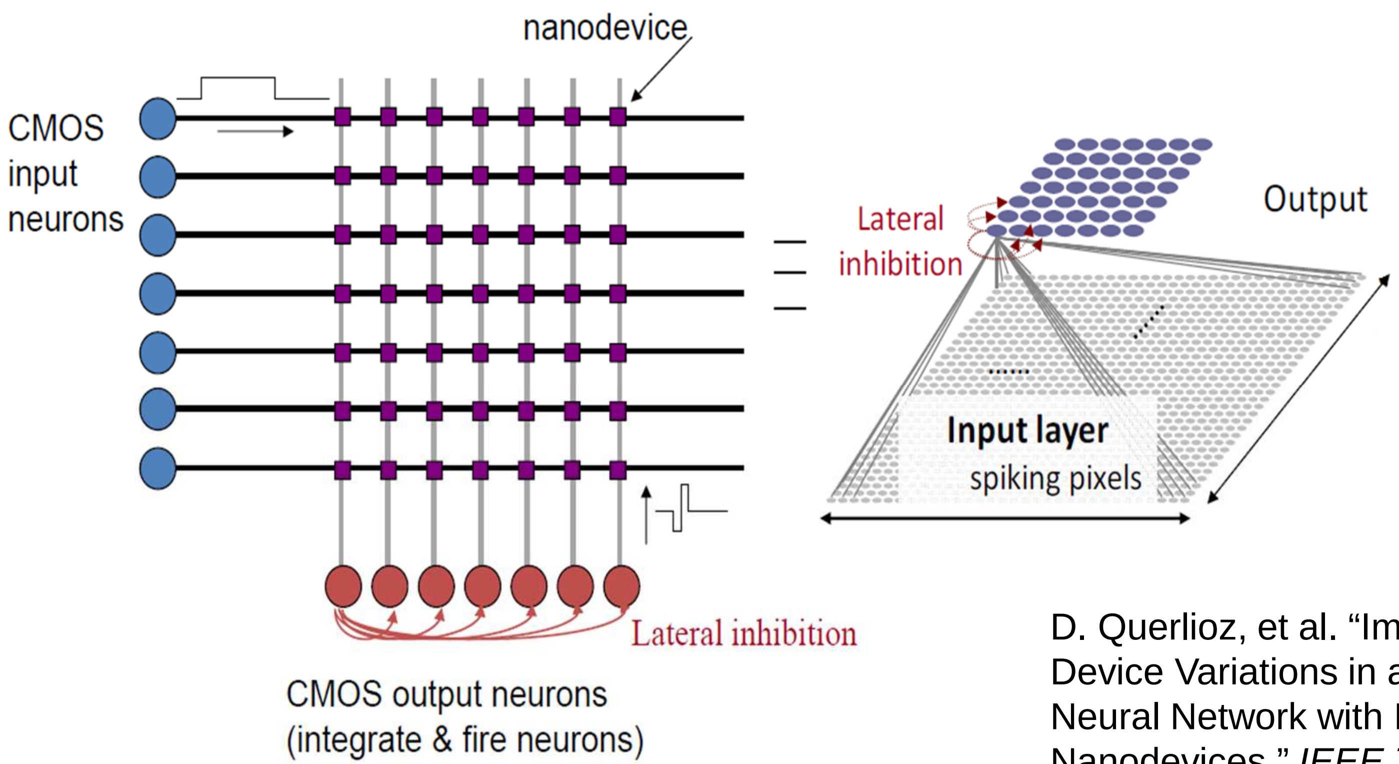
	DRAM	FLASH	MRAM	PRAM	RRAM	
		Nand	STT		OxRAM	CBRAM
Compatibility with CMOS	- -	-	-	+	+ +	+ +
Scalability	32nm	15nm	20-30nm	10-20 nm	10 nm	10-20 nm
Density	6-8f²	4f²	35-40f²	4-6f²	4-6f²	4-6f²
Maturity	+ +	+ +	-	-	- -	- -
Byte @	YES	NO	YES	YES	YES	YES
Writing cycle	50ns	0.1 ms	20ns	100 ns	150 ns	1-10 us
Endurance	1,00E+16	1,00E+05	1,00E+15	1,00E+09	→1,00E+06	1,00E+06
Plus	**Perf.**	**NV**	**Perf. + NV + Endurance**	**Perf. + NV+ density**	**Perf. + NV+ density**	**Vitesse lect+ NV+ densité**
Minus	**Scalability**	**Endurance Perf - -**	**Costly tech. Delay wrt CMOS nodes**	**consumptio n**	**Endurance, Forming**	**Endurance, Perf.**

Memristor as a Synapse

- Concept introduced by Chua
  - "Memristor-The missing circuit element" - 1971
- Strukov & Snider developed it
  - "The missing memristor found" - Nature 2008
  - "Spike-timing-dependent learning in memristive nanodevices" – Nanoarch 2008
- Concept "memristor = synapse weight"
  - So easy ?



Crossbar of Synapses



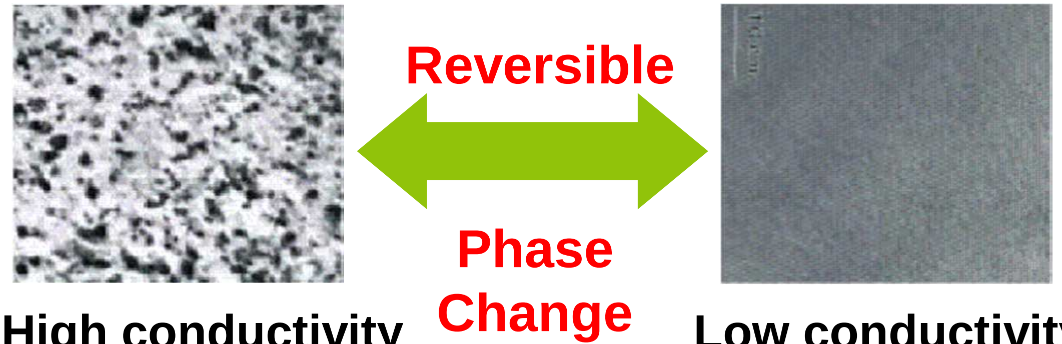
D. Querlioz, et al. “Immunity to Device Variations in a Spiking Neural Network with Memristive Nanodevices,” *IEEE TNANO 2013*

- Require “gradually” programming memristors
  - Difficulty is to obtain equivalent potentiation / depression phases

Phase-Change Memories

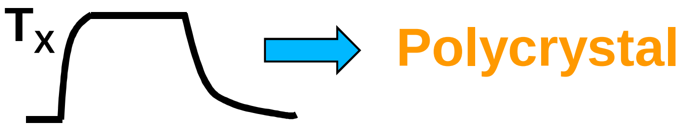
- Material with 2 stable phases
  - Chalcogenide alloys : GST, GeTe, …
  - Hysteresis cycle between 2 states

Polycrystal **Amorphous**

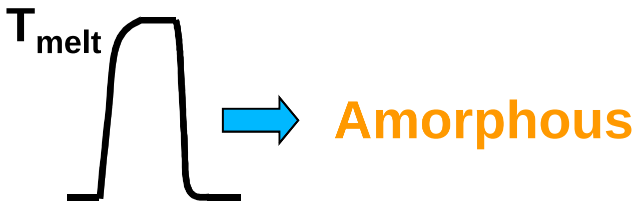


High conductivity **Low conductivity**

- Transition by Joule heating



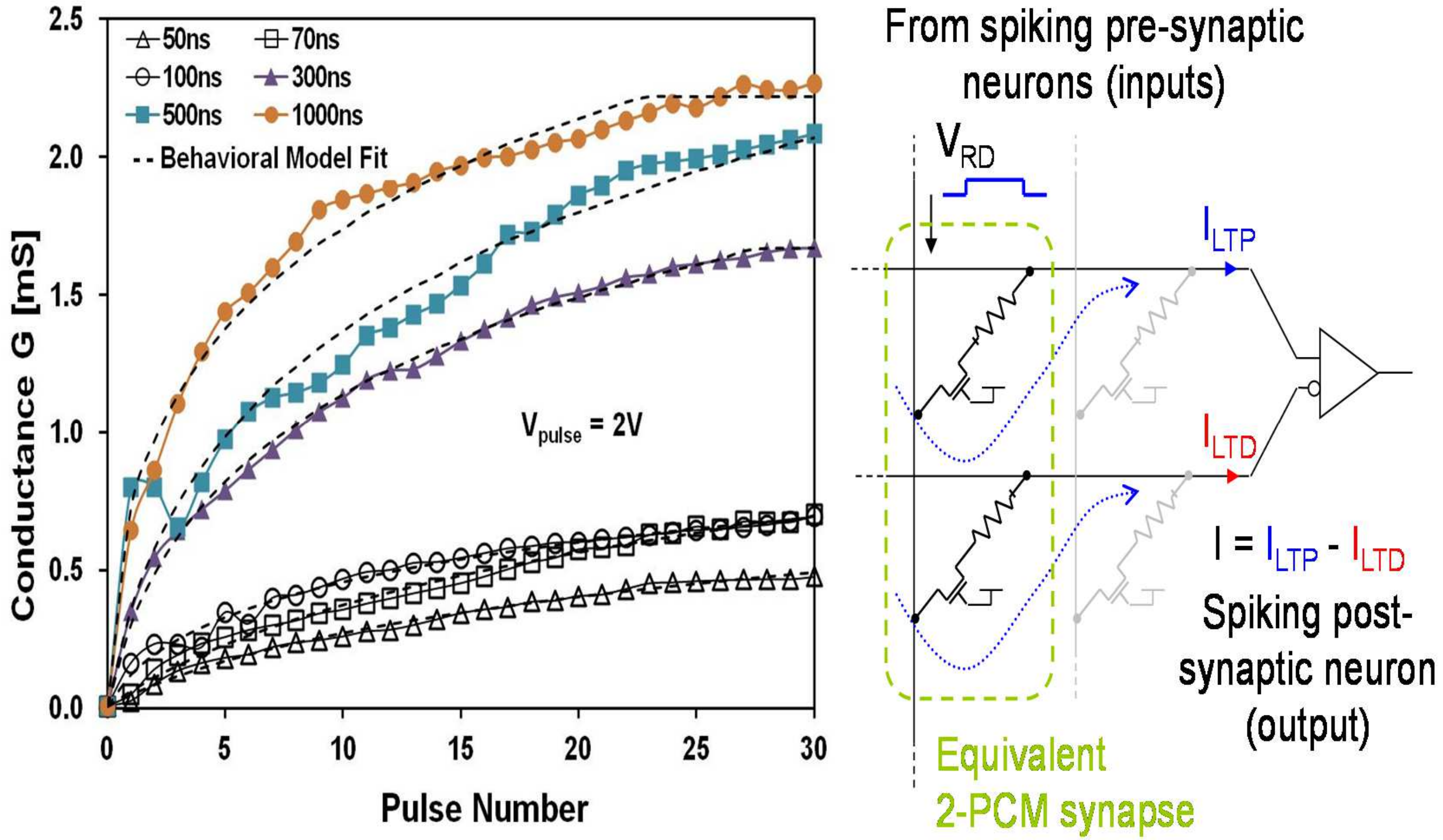
t_{QUENCH} > 100 ns



t_{QUENCH} < 10 ns

A solution: 2-PCM Synapse

- Symmetric Long-Term-Potentiation and Long-Term depression behavior by using pulsed crystallization.



Outline

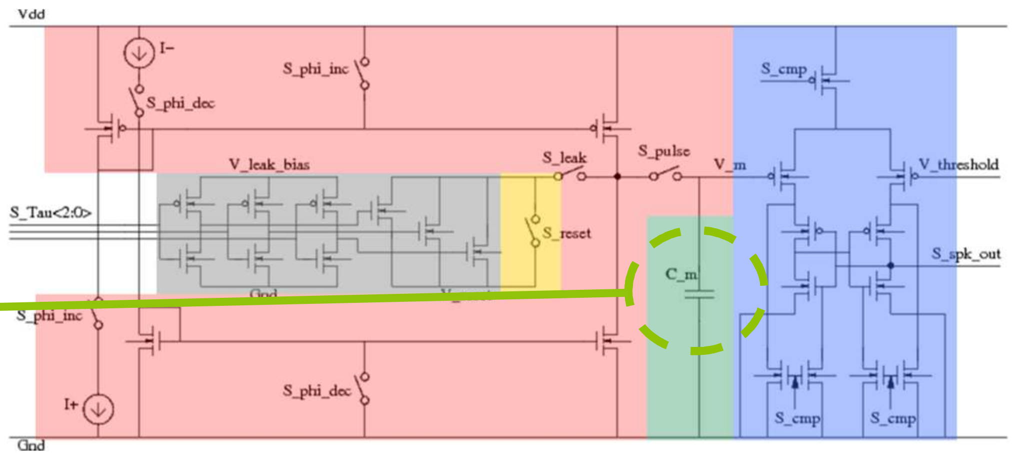
- Context
- Exploiting resistive memories
- **Analogue neurons and 3D-TSV**
- Towards monolithic 3D…
- Conclusion

Analogue neurons

- Compact design (5x gain vs digital)
- Low power consumption (20x gain vs digital)
- Ability to interface sensors directly with the processing part
- Computational efficiency

A. Joubert et al. “A robust and compact 65 nm LIF analog neuron for computational purposes,” IEEE NEWCAS 2011

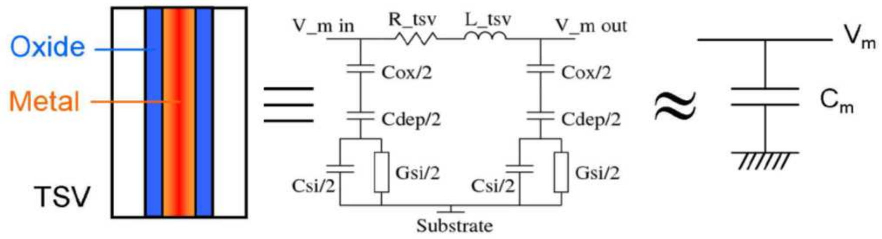
Drawback:
Large capacitance required (0,5-1pF)
$\Rightarrow$ MIM
$\Rightarrow$ Large area (not scalable !)



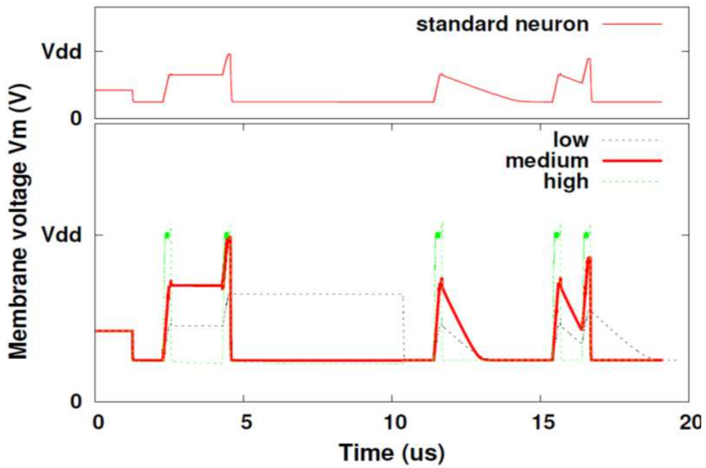
Leaky Integrate and Fire Neuron in CMOS 65nm accept spikes in 1kHz-1 Mhz range

Through Silicon Vias

- Used for crossing circuits in 3D stacking
- Present large "parasitic" capacitance

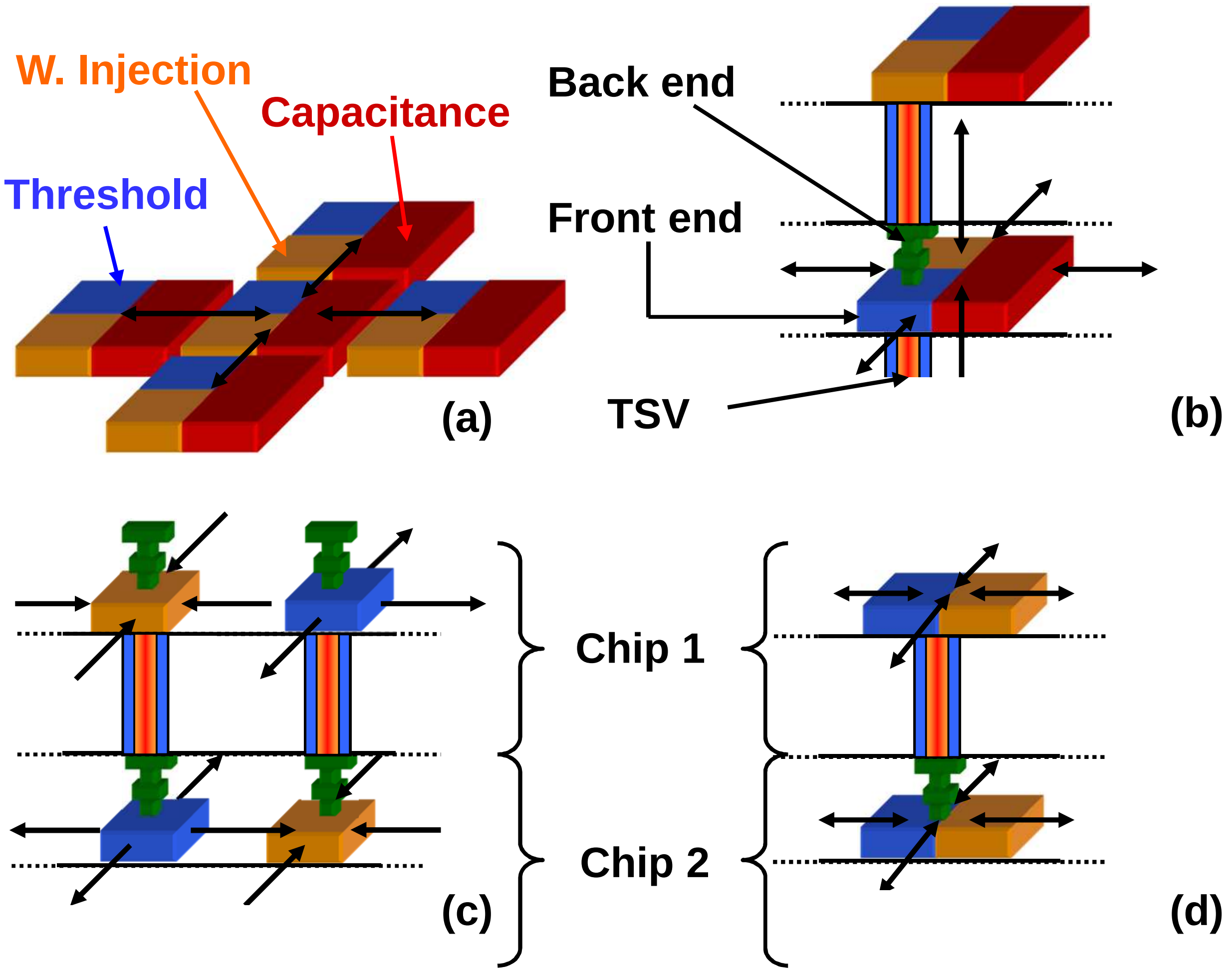


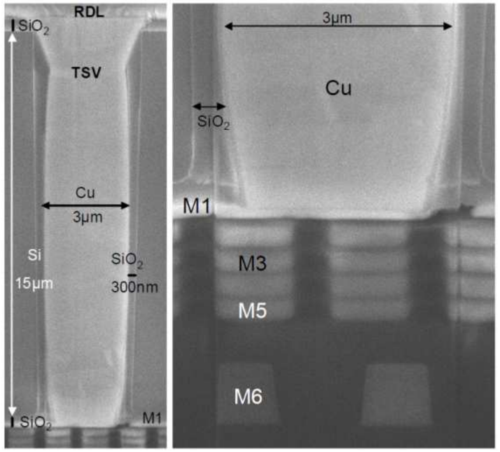
- Demonstrated feasibility



A. Joubert et al., "Capacitance of TSVs in 3-D stacked chips a problem? Not for neuromorphic systems", DAC'12

3D integration using TSV





Chaabouni et. al. (2010)

(a) Standard 2D neuromorphic architecture

(b) 3D stacking with standard 2D neurons

(c) et (d) 3D stacking with TSV-based neurons

Outline

- Context
- Exploiting resistive memories
- Analogue neurons and 3D-TSV
- **Towards monolithic 3D…**
- Conclusion

Brain is 3D !

- **3D technologies**
  - **Parallel Integration**

    Dies fabricated separately then vertically stacked, such as:

    ➔ Through-Silicon-Via (TSV)

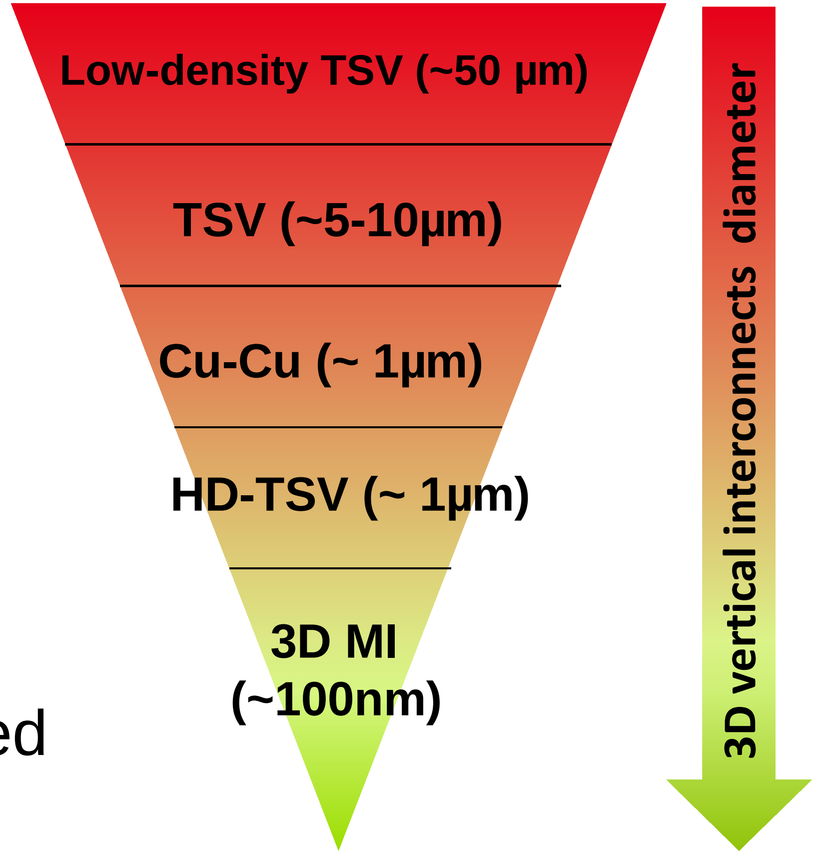
    ➔ Copper-to-Copper (Cu-Cu) contact

    ➔ High-Density TSVs

  - **Sequential Integration**

    Second transistor layer is fabricated directly on the top of first one. (Cold process)

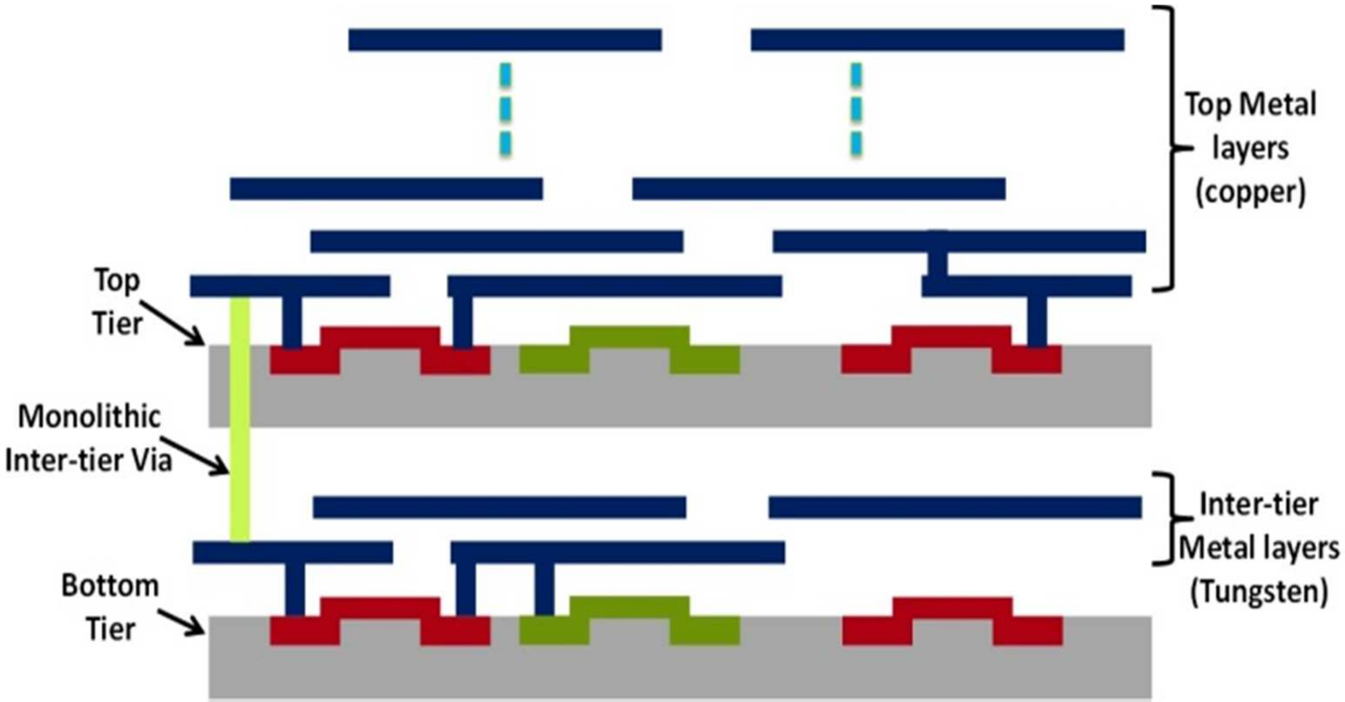
    ➔ **3D Monolithic Integration (3DMI)**



Monolithic 3D

- Integration is monolithic using a "cold" process
- Allows 100 nm 3D vias pitch

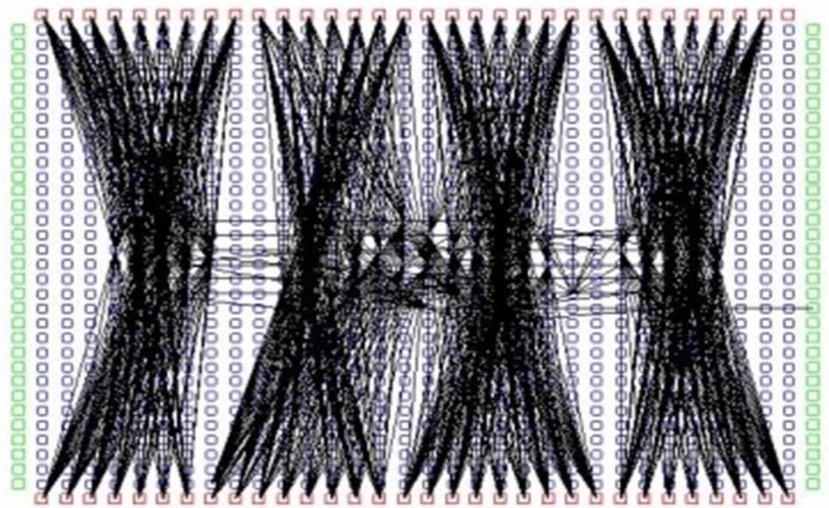
=> Open perspectives for real 3D implementation



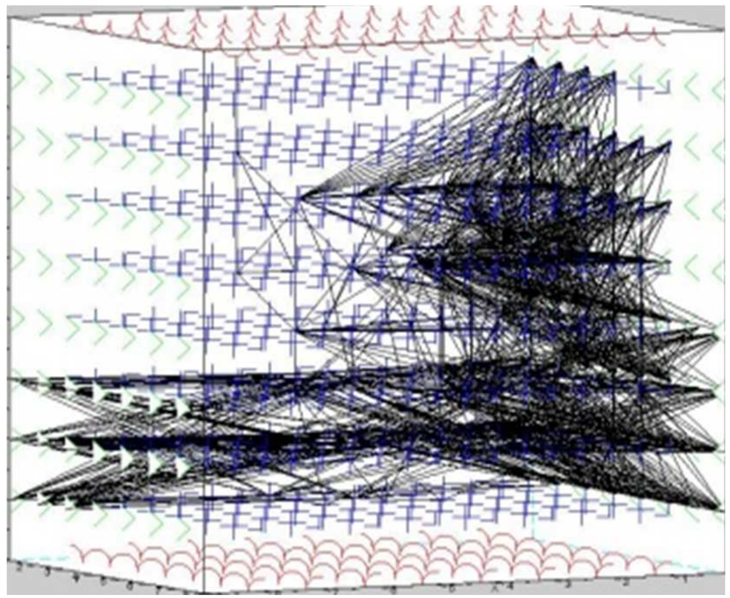
P. Batude et al., "GeOI and SOI 3D Monolithic Cell integrations for High Density Applications" VLSIT 2009

2D versus 3D mapping on NN

- Using 3D TSV allows a “cube” of neurons
- Drastically reduce interconnects
- Example shown here is 3x more efficient in 3D



2-D



3-D

184 neurons with 1181 connections

Conclusion

- Neuromorphic architectures are appealing
  - Solve dark silicon issues
  - Adapted to modern applications
- And can take benefit of advances technologies
  - Synapse => Memristors = RRAM, PCM
  - Neurons => Analogue neurons + high-capacitance TSV
  - Interconnections => high-density 3D
- The path is open for the next neuromorphic revolution!

Questions ?

Centre de Grenoble
17 rue des Martyrs
38054 Grenoble Cedex

Centre de Saclay
Nano-Innov PC 172
91191 Gif sur Yvette Cedex