Scalable neuromorphic in-memory computing with n-ary spintronic crossbars

Anatole MOUREAUX, **Anthony LOPES**,
Lindiomar BORGES DE AVILA, and **Flavio ABREU ARAUJO**

Institute of Condensed Matter and Nanosciences, **University of Louvain**, Louvain-la-Neuve, Belgium.

SPIE. OPTICS+ PHOTONICS

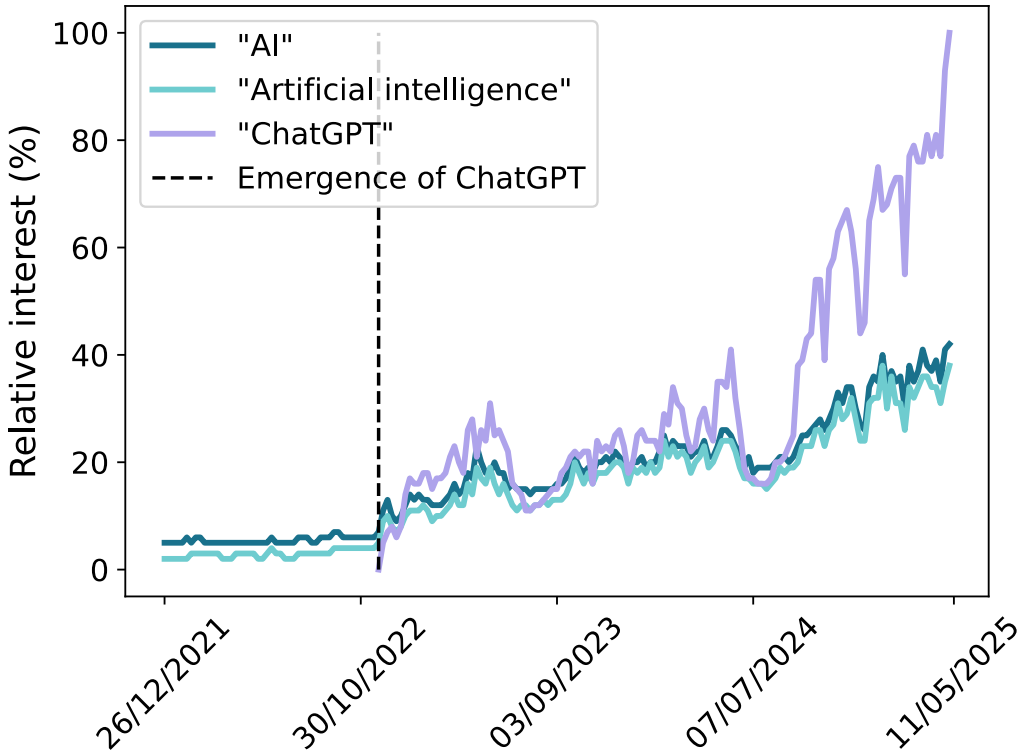
San Diego (CA, USA) | August 5, 2025

UCLouvain

IMCN

fnrs FREEDOM TO RESEARCH

Growing interest on the Internet



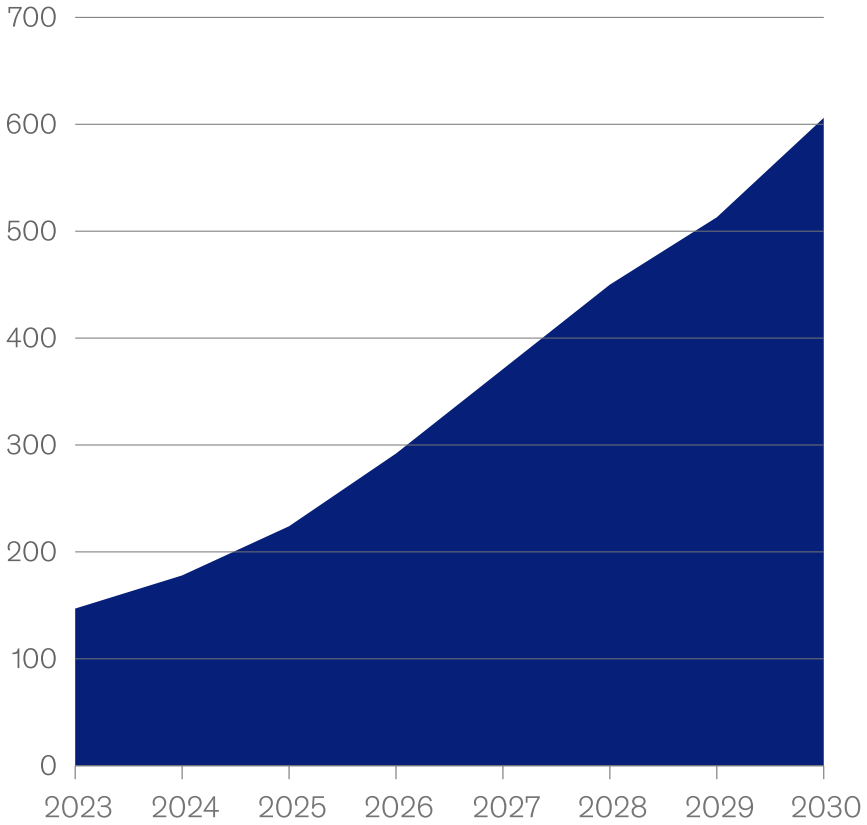
AI chatbots that feel human…



Google Trends (2025); *Global Energy Perspective 2023*, McKinsey, Oct 18, 2023; McKinsey analysis

Growing energy demand

Projected US datacenter energy consumption (TWh)



AI chatbots that feel human…

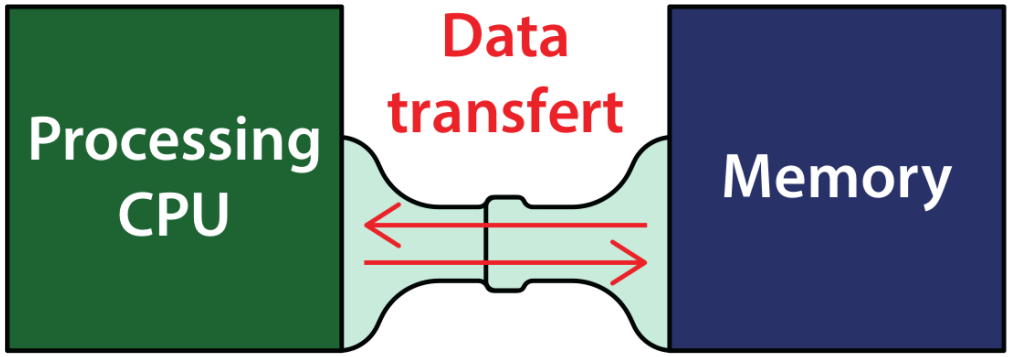
but are far less power-efficient than humans



Google Trends (2025); *Global Energy Perspective 2023*, McKinsey, Oct 18, 2023; McKinsey analysis

Digital computing systems are not flawless

von Neumann bottleneck

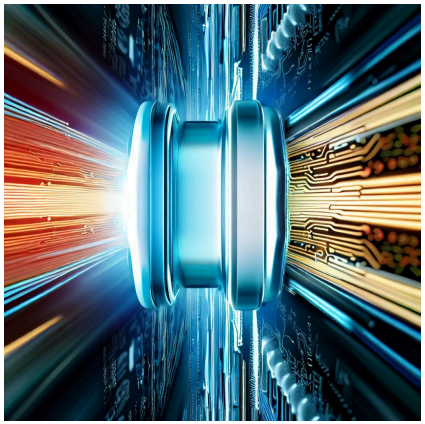


Time and energy consuming !

AI = Huge amount of simple operations…

…performed with high-precision hardware

DALL-E 3 prompt:
"A bottleneck representing limitations due to the separation between the CPU and the RAM"



130 kW !

⇒ slow and energy-hungry !

Digital computing systems



- **High-precision**
- **Separate from memory**
- **Limited parallelization**

VS

Mammalian brain



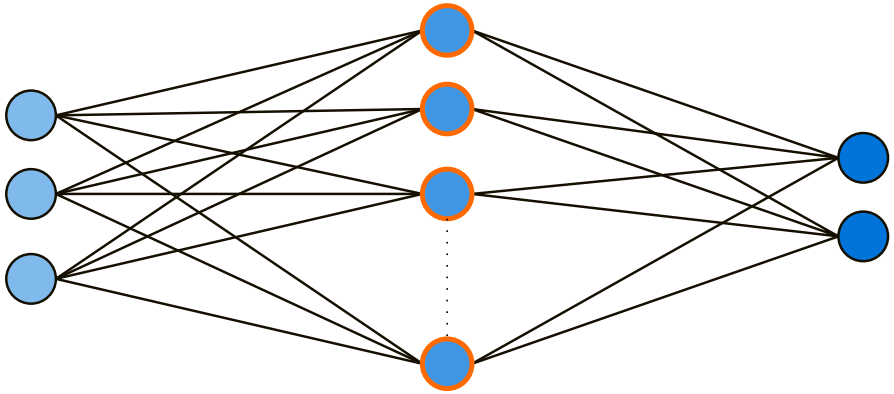
- **Low-precision**
- **Processing (neurons) and memory (synapses) intertwined**
- **Highly parallel**

Neuromorphic computing: Low power, analog, brain-inspired computing systems.

Neuromorphic spintronics: Neuromorphic computing systems based on **spintronic (nano)devices and phenomena**

Neural networks in a nutshell

Nonlinearity

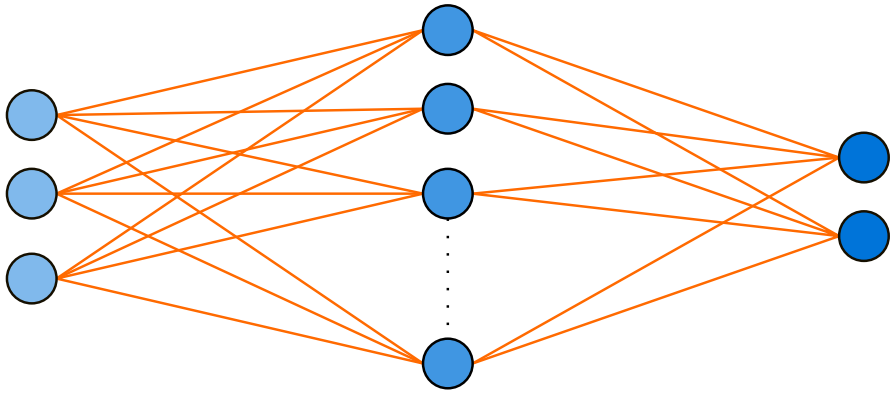


Nonlinear transformation of the input x in the neurons

$$y_{i,h} = \sigma\left(x_{i,h}\right)$$

ex: $\sigma \equiv$ sigmoid, tanh, reLU, …

Weighted sums



Weighted summation of the neurons output for computing the input of the next neurons.

The weights are learned during training.

$$x_{i,h} = \sum_{j} w_{ij} y_{j,(h-1)}$$

The Multiply-and-Accumulate (MAC) operation

The **MAC operation** is a **matrix-vector multiplication** written as

$$\boldsymbol{Ax} = \boldsymbol{b}$$

with:

- $\boldsymbol{A}$: $(m \times n)$ matrix
- $\boldsymbol{x}$: vector of length (n) corresponding to the **input**
- $\boldsymbol{b}$: vector of length (m) corresponding to the **ouput**

Each element b_i is the result of the weighted sum:

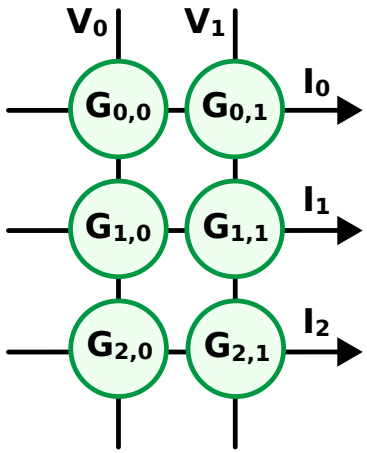
$$b_i = \sum_j A_{ij} x_j$$

Ex: $$\begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} A_{00} & A_{01} & A_{02} \\ A_{10} & A_{11} & A_{12} \\ A_{20} & A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix}$$

$$b_1 = A_{10}x_0 + A_{11}x_1 + A_{12}x_2$$

Hardware binary MAC implementation

Physical implementation



Multiplication
Ohm's law

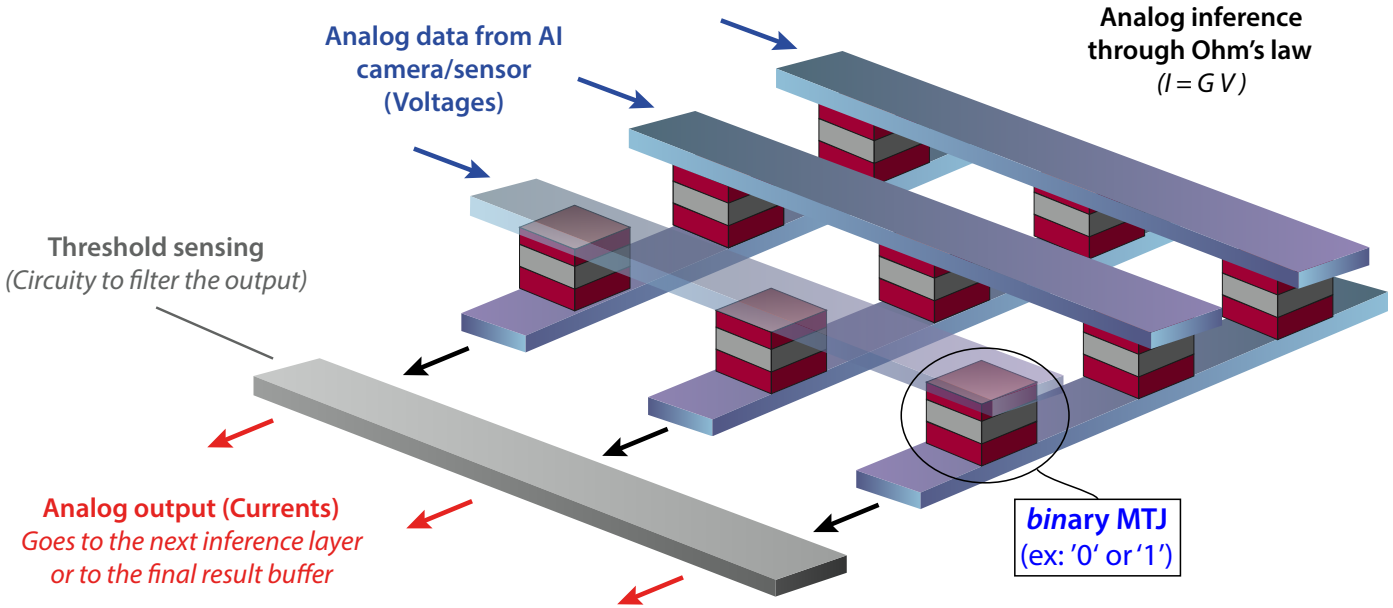
Accumulation
Kirchhoff's law

Input: Voltage V_j

Output: Current I_i

Weights:
Conductances G_{ij}

Spintronics crossbar architecture (MRAM compatible):
Neural weights (cell conductance)

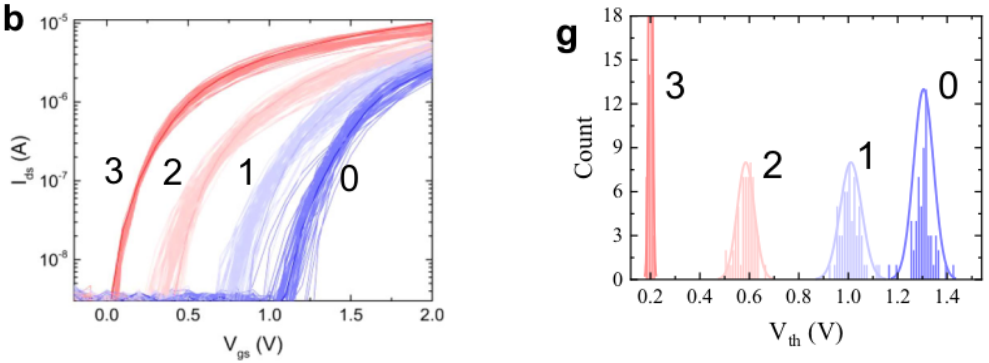


$$I_i = \sum_j I_{ij} = \sum_j G_{ij} V_j \iff \underline{I} = \underline{\underline{G}} \cdot \underline{V}$$

S. Jung *et al.*, *Nature* **601**, 211-216 (2022).

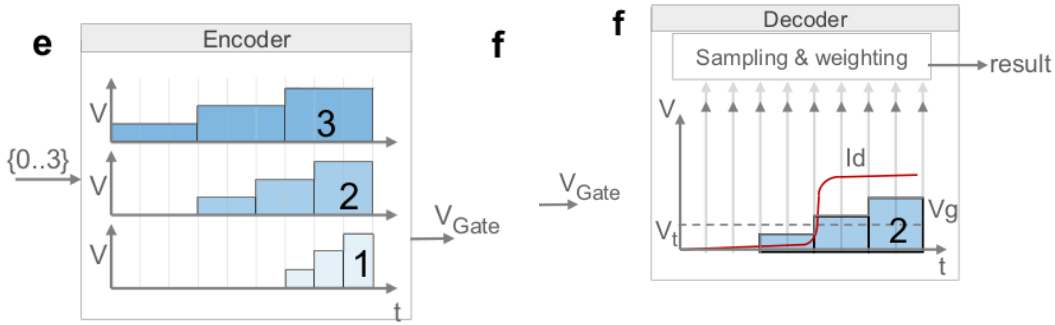
Hardware *n*-ary MAC operations

4 weights encoded in the V_{th} of the FET



Higher weight ⇒ FET triggering is faster

4 input levels in a **gate voltage** signal V_g



Higher input ⇒ V_g increases sooner

- **Multiplication**: time needed FET to trigger
- **Accumulation**: **capacitor** charged by the drains currents **along each line**

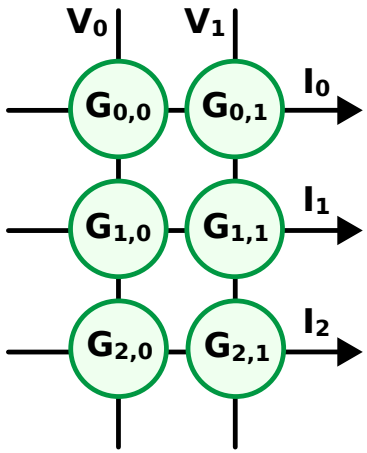
+ Overcomes computational limitations of binary crossbar

- Slower readout due to time needed to charge capacitor

T. Soliman *et al.*, *Nature Communications* **14**, 6348 (2023).

Hardware *n*-ary MAC implementation

Physical implementation



Multiplication
Ohm's law

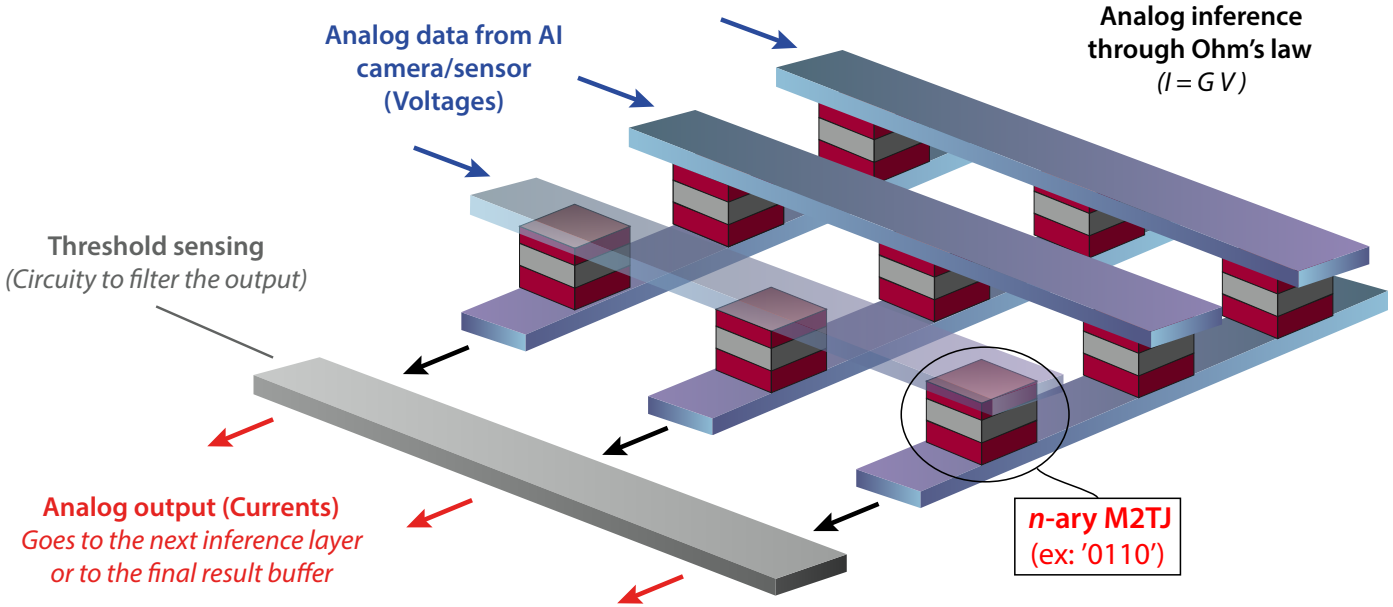
Accumulation
Kirchhoff's law

Input: Voltage V_j

Output: Current I_i

Weights: Conductances G_{ij}

Spintronics crossbar architecture (MRAM compatible):
Neural weights (cell conductance)



$$I_i = \sum_j I_{ij} = \sum_j G_{ij} V_j \iff \underline{I} = \underline{\underline{G}} \cdot \underline{V}$$

S. Jung *et al.*, *Nature* **601**, 211-216 (2022).

Mathematical abstraction to physical measurements

Legend:	What we know	What we measure	What we want to know

- We want to compute the following MAC operation $b_m = \sum_n A_{mn} x_n$, or in vectorial form $\underline{b} = \underline{\underline{A}}\,\underline{x}$.
- To do so, we use the crossbar array and Ohm's and Kirchhoff's laws: $i_m = \sum_n G_{mn} v_n$ (or $\underline{i} = \underline{\underline{G}}\,\underline{v}$).

1. We first assume that the input $\underline{x}$ is linearly encoded into a voltage $\underline{v}$:

$$\underline{v}(\underline{x}) = a_v \underline{x} + b_v$$

2. The conductance G of the cell (m, n) is a **known arbitrary function** of the state A at position (m, n):

$$\underline{\underline{G}}(\underline{\underline{A}})$$

3. We can write

$$\frac{i_m - b_v \sum_n G_{mn}}{a_v} = \sum_n G_{mn} x_n = y_m$$

Mathematical abstraction to physical measurements

However:

$$b_m = \sum_n A_{mn} x_n \neq y_m = \sum_n G_{mn} x_n = \text{what we measure}$$

We have to work with G to retrieve b

MAC operation = linear combination ⇒ G has to provide a linear weighting of A

The linearity of G:

- Guarantees uniqueness of solutions **(bijection)**
- Conservation of the topological properties of the input space **(homeomorphism)**

What if $G(A)$ is linear?

If $\underline{\underline{G}}\left(\underline{\underline{A}}\right)$ **provides a linear weighting of** $\underline{\underline{A}}$, $\underline{b}$ can be retrieved **algebraically**:

1. We first assume that the input $\underline{x}$ is linearly encoded into a voltage $\underline{v}$:

$$\underline{v}\left(\underline{x}\right) = a_v \underline{x} + b_v$$

2. And consider a linear $\underline{\underline{G}}\left(\underline{\underline{A}}\right)$:

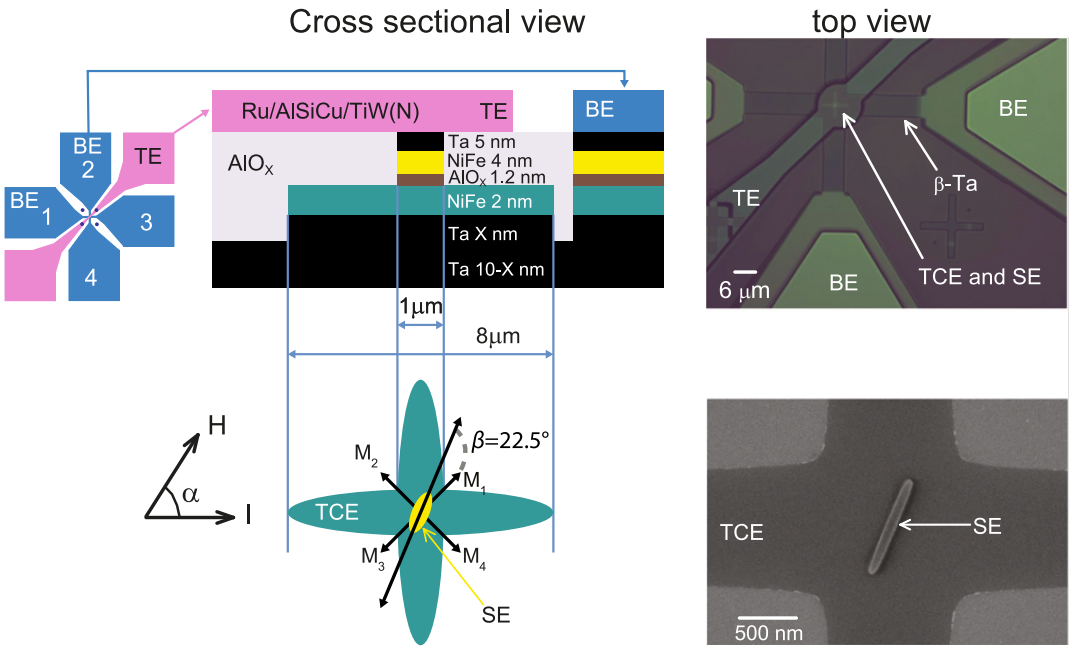
$$\underline{\underline{G}}\left(\underline{\underline{A}}\right) = a_G \underline{\underline{A}} + b_G$$

3. Under vectorial form writes:

$$\boxed{\underline{b} = \frac{\underline{i} - b_G\left(n b_v + a_v\left(\underline{x}^T \cdot \underline{1}\right)\right)\underline{1} - a_G b_v \underline{\underline{A}} \cdot \underline{1}}{a_G a_v}}$$

where all the **green** terms are known a priori and $\underline{i}$ is measured currents vector.

M^2TJs as multistate memristive cells



Key points:

- Use **four states** (*or more*) states for encoding the weights
- **M^2TJs versatility**: large range of different implementations (*current/resistance summation, encoding method, etc*)

⇒ **Improved computing power**

Conductance/resistance of the cell is tuned via the free layer magnetization

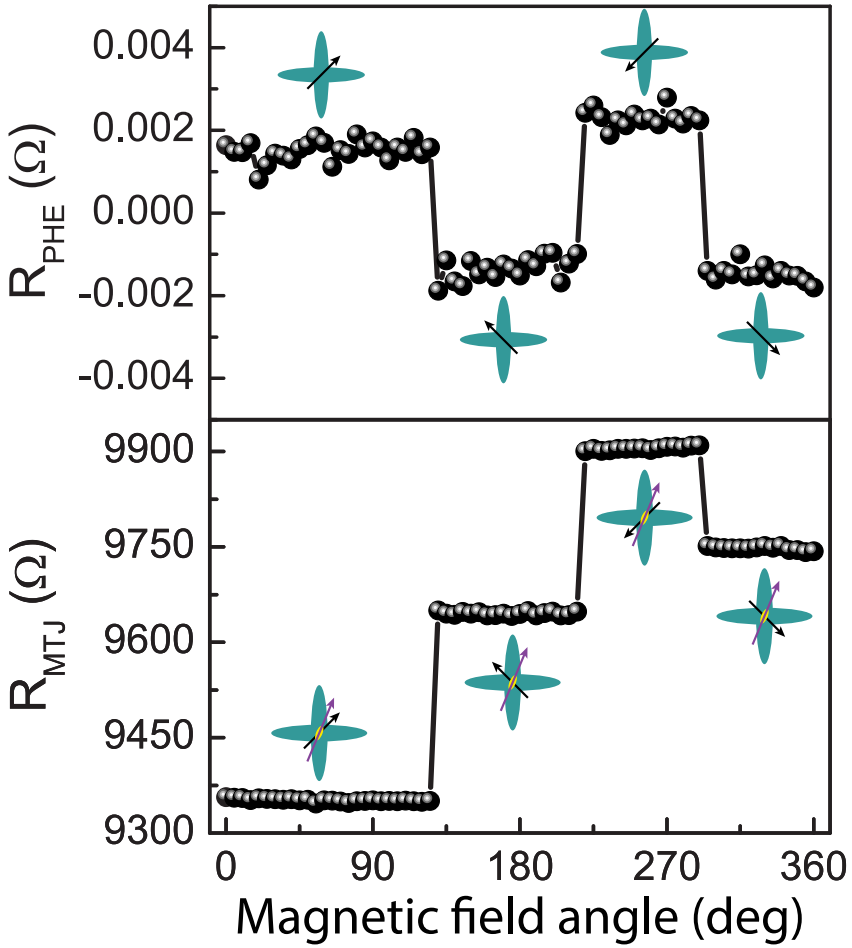
The references made here to S. Das *et al.* belong to our partners from BIU & INESC MN.

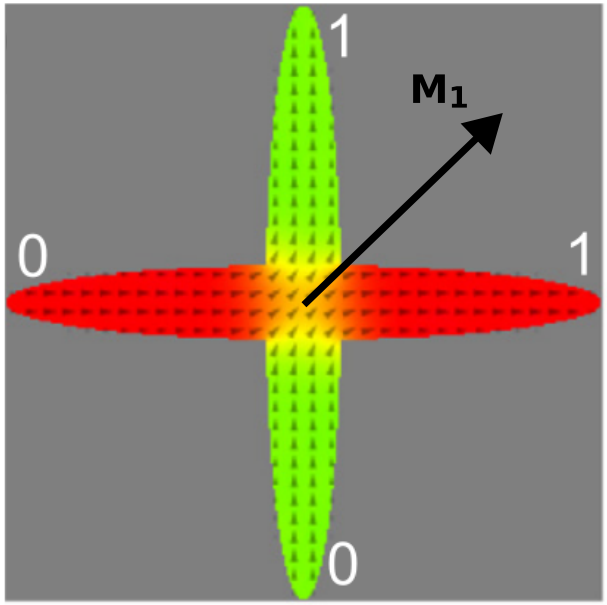
אוניברסיטת בר-אילן
Bar-Ilan University

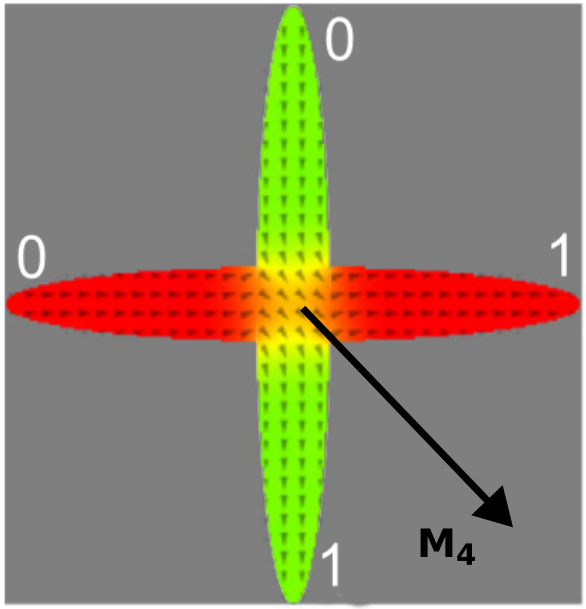
inesc mn
Microssistemas e Nanotecnologias

S. Das *et al.*, *Applied Physics Letters* **116**, 26 (2020). | S. Das *et al.*, *Applied Physics Letters* **117**, 7 (2020).

M^2TJs as multistate memristive cells: writing the states







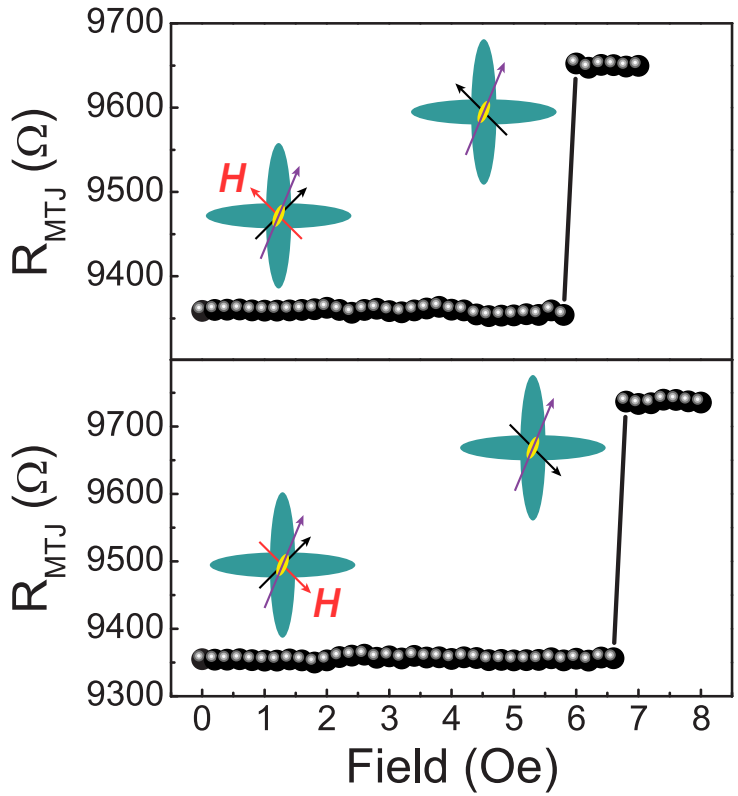
External magnetic field selects a magnetization state *M* in the free layer, hence controlling the misalignment with the magnetization of the fixed layer.

$$G_{ij}(\text{state}) = \frac{1}{R_{ij}(\text{state})} = \frac{1}{R_{ij}(\text{angle}, \text{noise})}$$

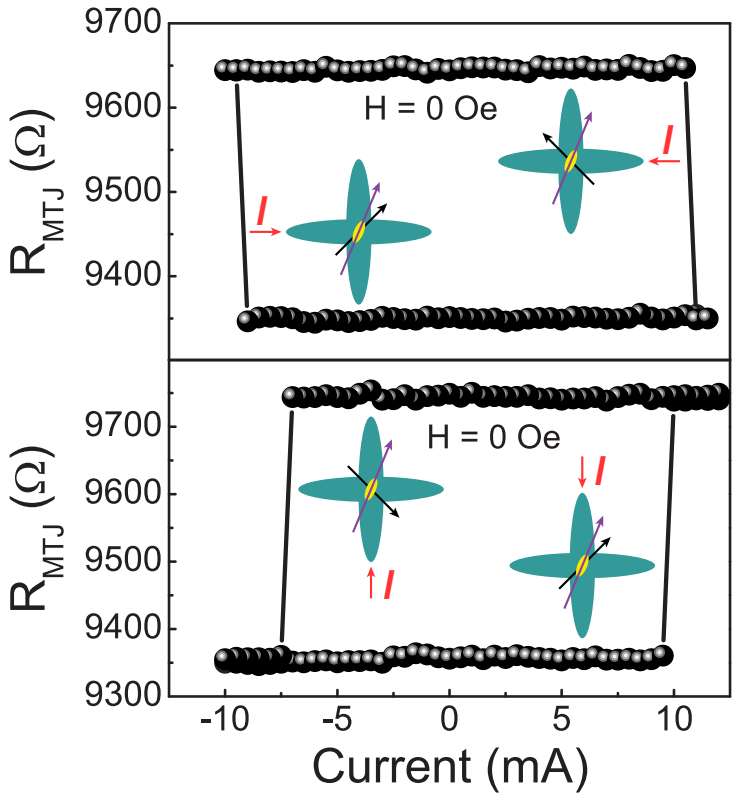
S. Das *et al.*, *Applied Physics Letters* **116**, 26 (2020). | S. Das *et al.*, *Applied Physics Letters* **117**, 7 (2020).

M^2TJs as multistate memristive cells: writing the states

Magnetic field switching



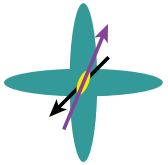
Spin-orbit torque (SOT) switching

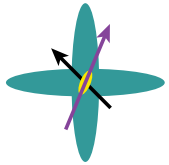


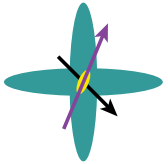
S. Das *et al.*, *Applied Physics Letters* **116**, 26 (2020). | S. Das *et al.*, *Applied Physics Letters* **117**, 7 (2020).

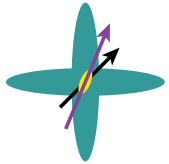
Encoding 4 states (2 bits)

⇒ 4-states junctions: R_{MTJ} =


~101.01 μS (state 1)


~103.41 μS (state 3)

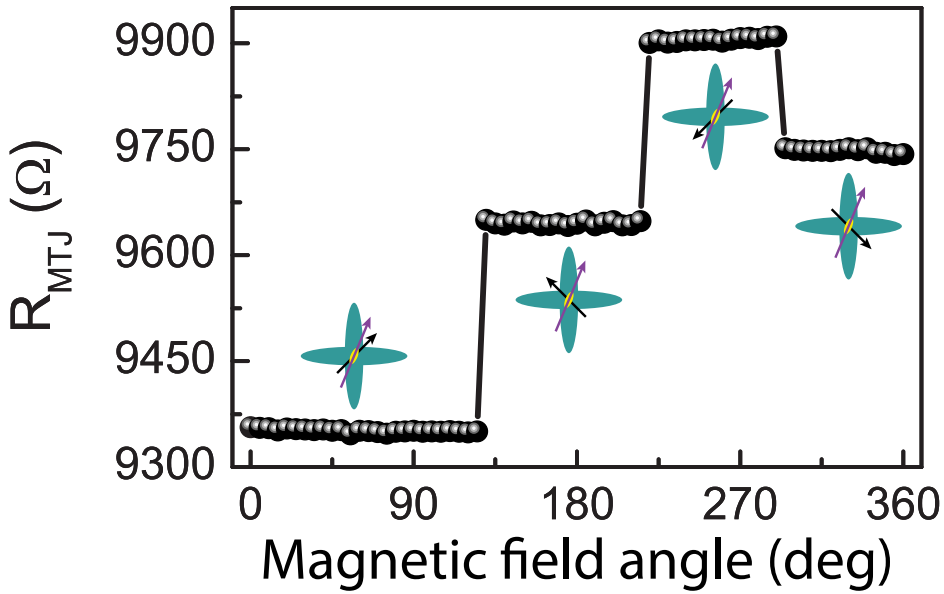

~102.56 μS (state 2)


~106.72 μS (state 4)

Magnetic field switching

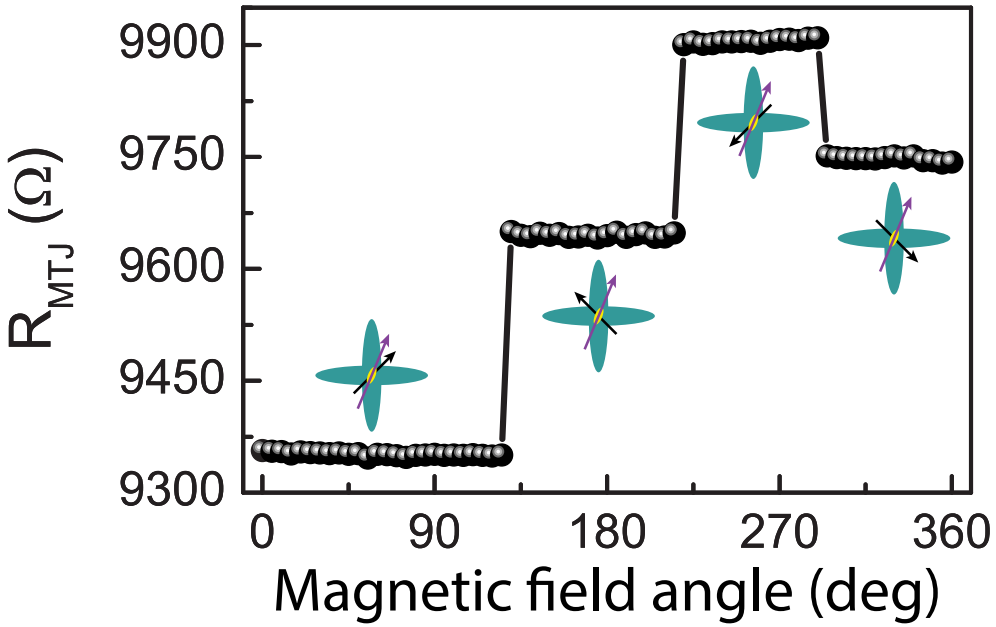
or

Spin-orbit torque switching



S. Das *et al.*, *Applied Physics Letters* **116**, 26 (2020). | S. Das *et al.*, *Applied Physics Letters* **117**, 7 (2020).

M²TJs non-idealities



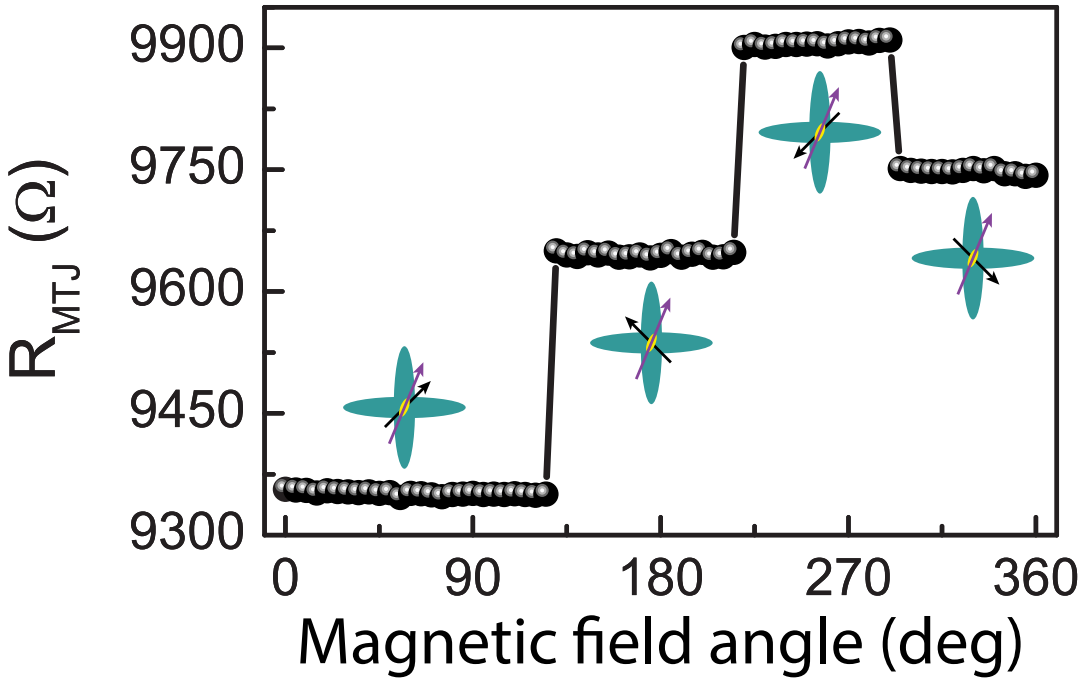
2 Challenges:

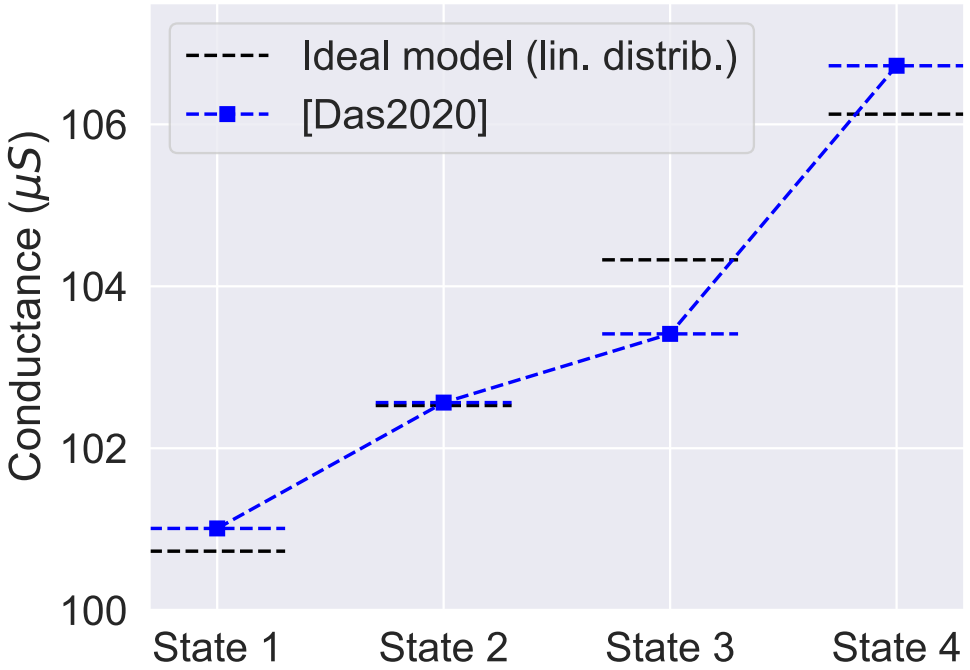
1. States are not increasingly ordered
2. States are not equidistant (not linearly distributed)

S. Das *et al.*, *Applied Physics Letters* **116**, 26 (2020). | S. Das *et al.*, *Applied Physics Letters* **117**, 7 (2020).

Linear approximation of the conductance states distribution

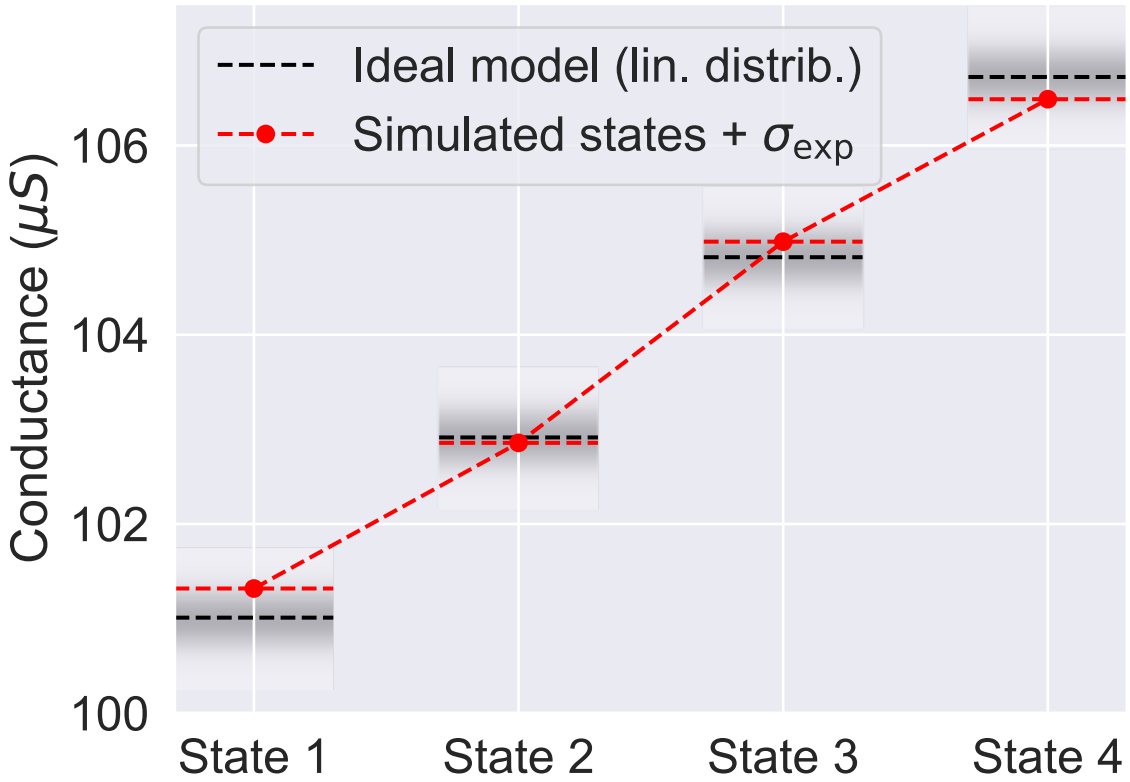
- The states of the **full crossbar** are modeled with a **linear regression**.
- 4 (generalized to ***n***) theoretical and **ideal states** are obtained.
- **Normally distributed** variabilities tend to vanish for large crossbar.
- **Cell-to-cell variability** and **measurement noise** can be included.





S. Das *et al.*, *Applied Physics Letters* **116**, 26 (2020). ; S. Das *et al.*, *Applied Physics Letters* **117**, 7 (2020).

Experimental (measurement) noise



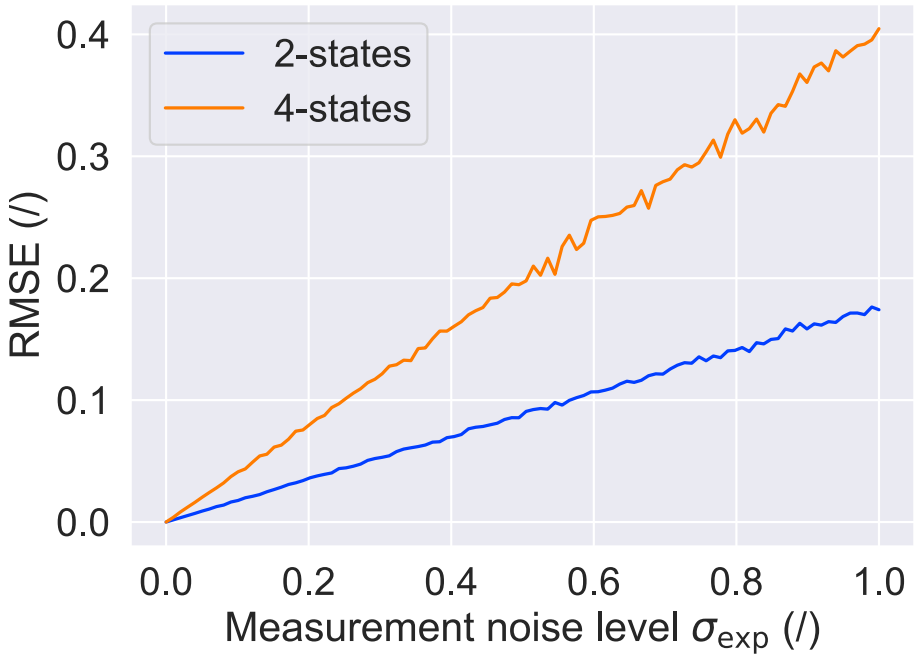
Example when σ_{exp} = 0.25μS

In the **linear distribution** case:

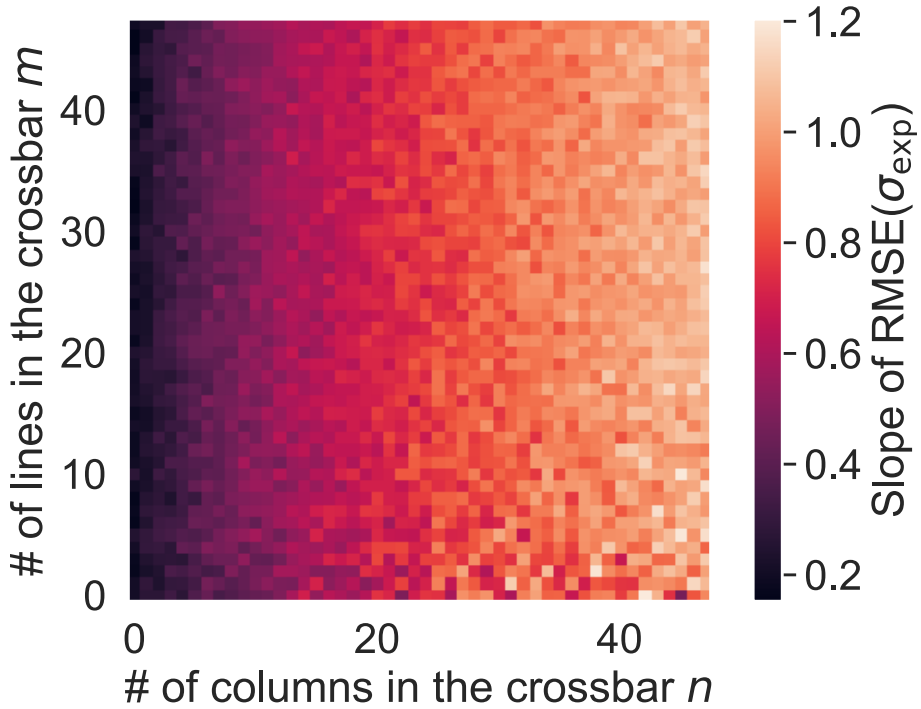
- Assessment of **noise** affecting the **measurement of the current**
- Additive **Gaussian white noise** (μ = 0 and $\sigma = \sigma_{exp}$)

Scaling of the measurement error

RMSE of the MAC operation result from σ_{exp} = 0μS to 1μS



Slope of RMSE $\left(\sigma_{exp}\right)$ for various crossbar geometries

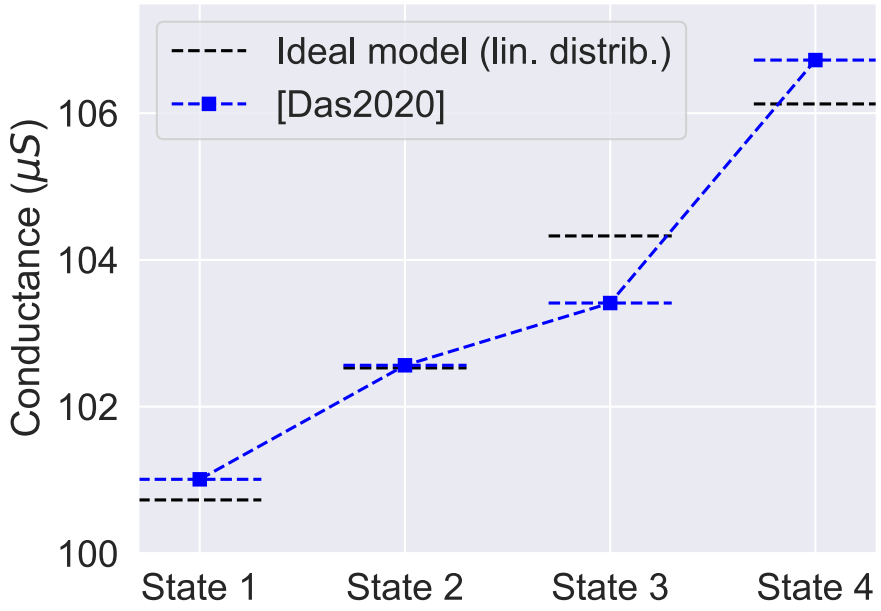


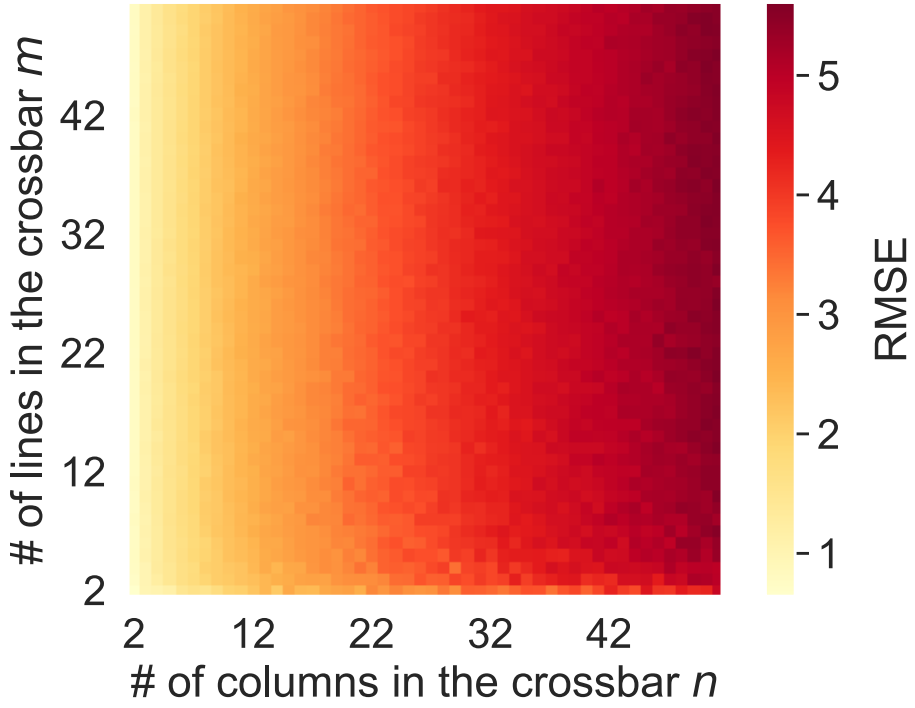
Slope of RMSE $\left(\sigma_{exp}\right)$ is constant with the number of lines m in the crossbar array, but **increases as** $\sqrt{n}$ (n being the number of columns).

Error scaling due to states non-ideality

Experimentally measured states are **nonlinearly distributed**, **even in the absence of noise**.

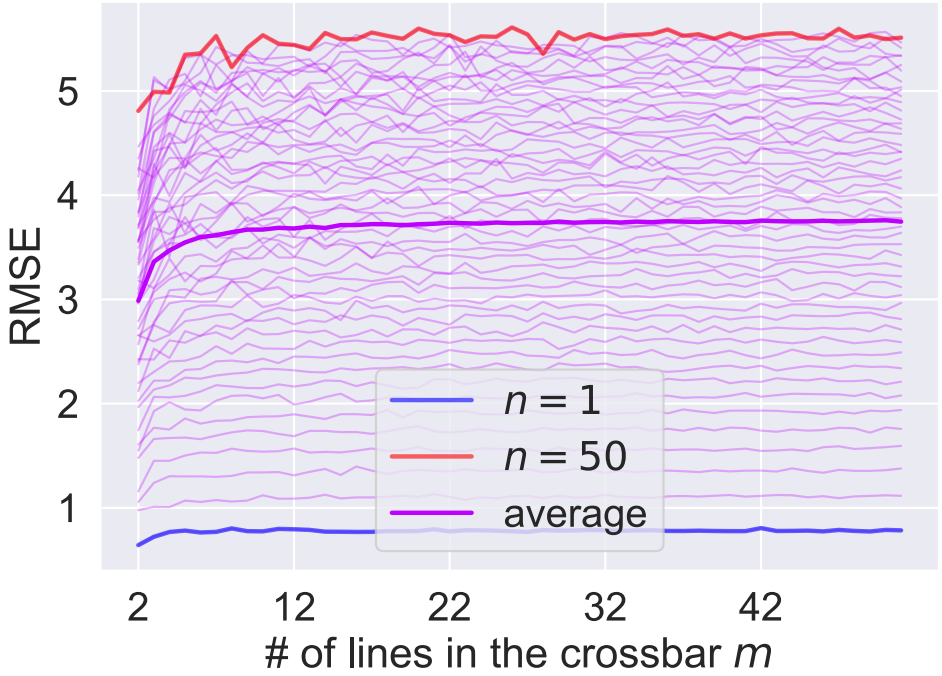
⇒ Modeling using a **linear regression** ⇒ deviations are **normally distributed**



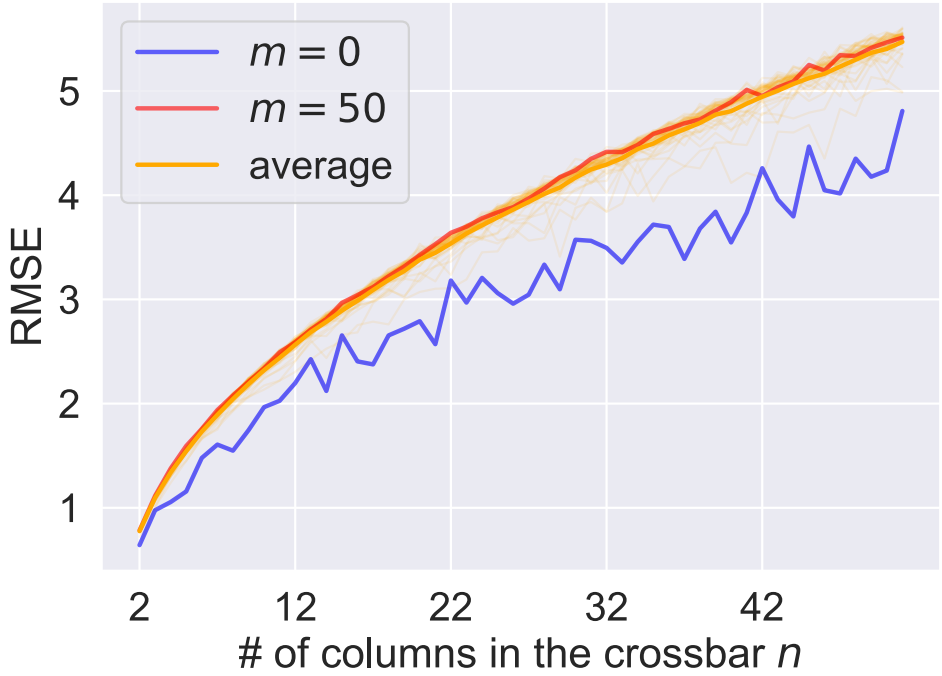


Similar scaling as the experimental measurement noise (σ_{exp})

Error scaling due to states non-ideality



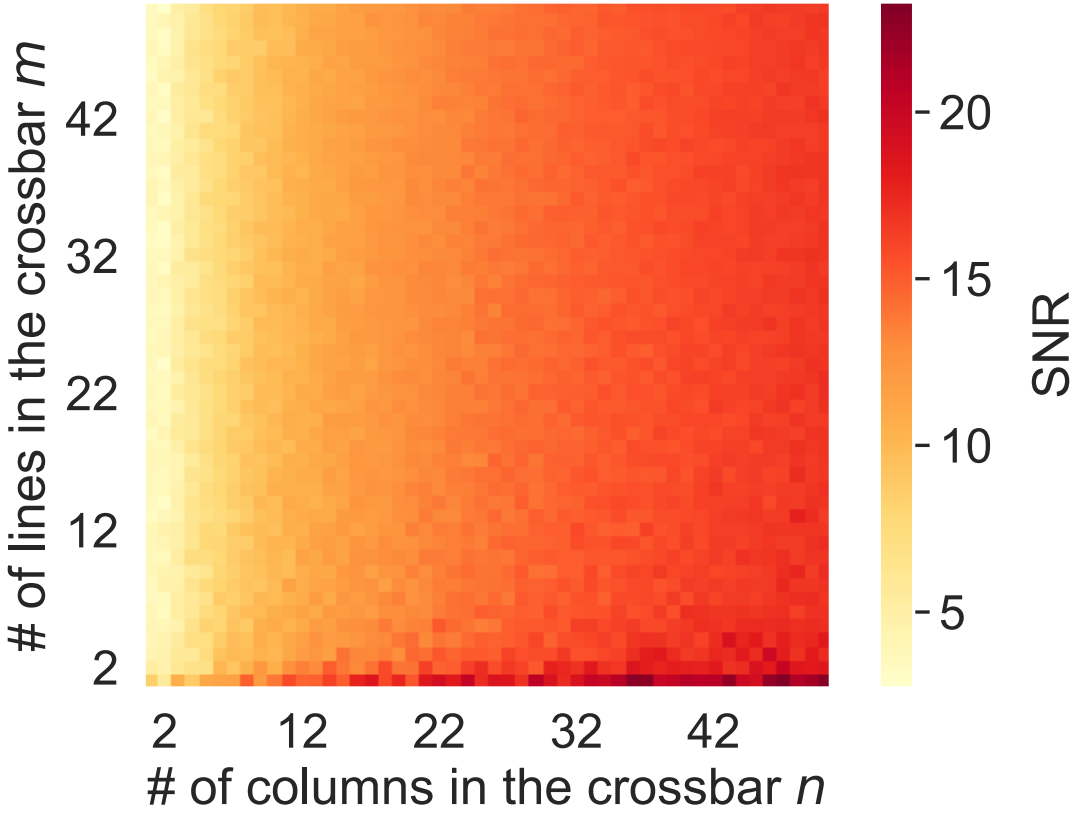
Constant RMSE when the number of lines *m* increases (except at low *m*).



$\sqrt{n}$**-like** increase of the RMSE with the number of columns *n*.

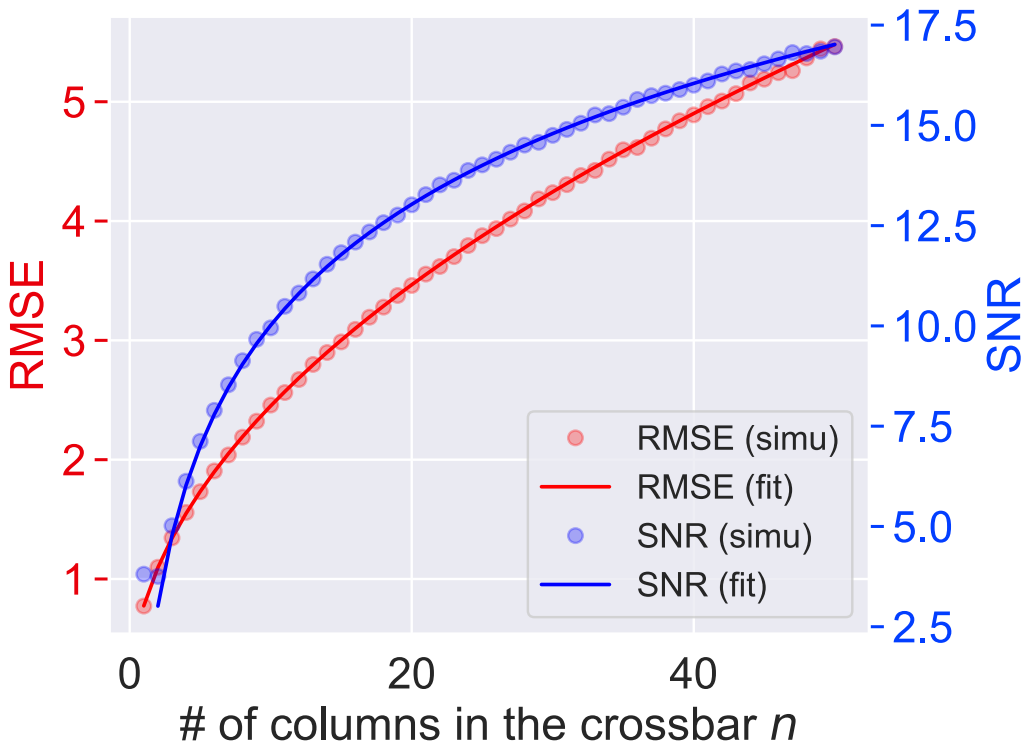
Signal-to-noise ratio (SNR) scaling due to states non-ideality

- RMSE increases as $\sqrt{n}$, with n the number of columns in the crossbar array
- Amplitude of the output current **scales linearly** with n
- SNR increases as $\log(n)$ and is constant with the number of lines m



RMSE and SNR scaling due to state non-ideality

$\mathrm{RMSE}(n) \approx 0.78\sqrt{n}$ and $\mathrm{SNR} \approx 4.35 \log(n)$

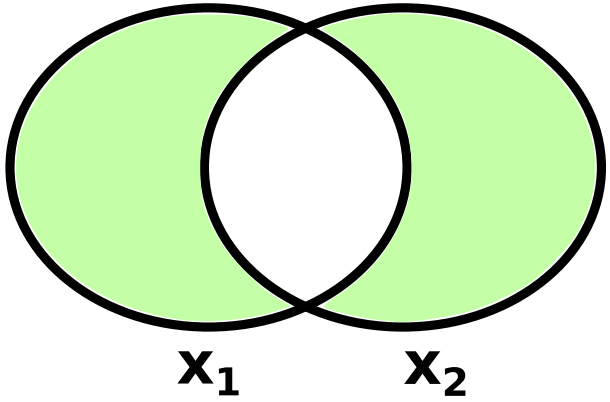


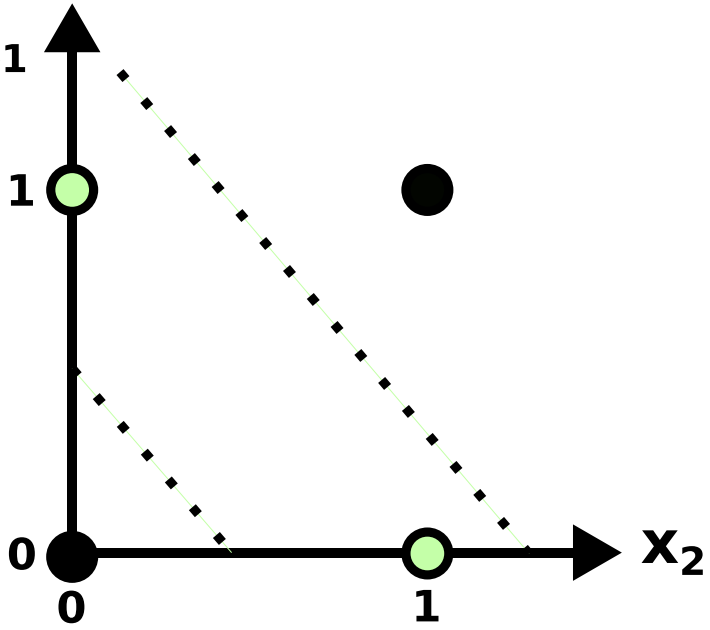
# columns (n)	RMSE	SNR (dB)
2	1.04	3.25
4	1.50	6.20
8	2.15	9.15
16	3.06	12.01
32	4.35	15.05
64	6.18	18.00
128	8.76	20.95
256	12.42	23.90
512	17.59	26.85

Toy Case: XOR

Simple machine learning task: the **XOR problem**

XOR		
x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	0

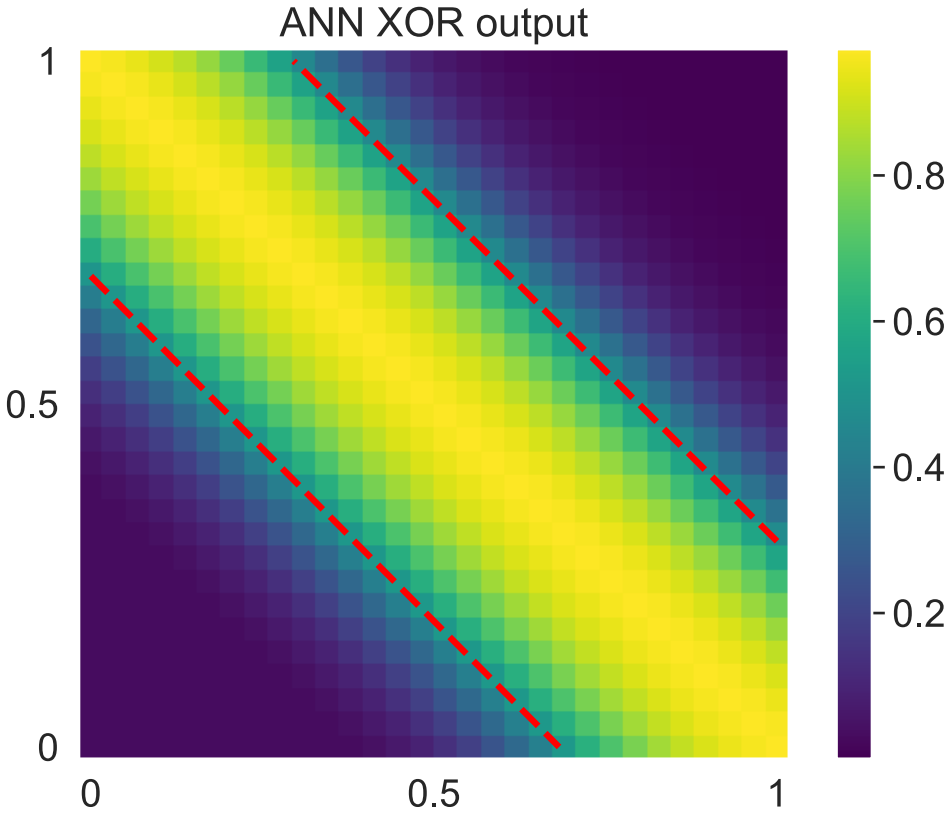


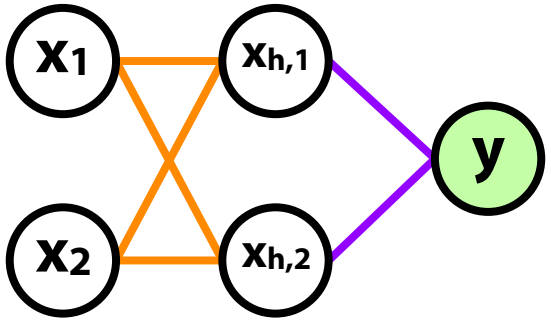


Goal: approximate the (nonlinear) XOR function by respecting the truth table.

XOR problem: software solution

Effectively solved using a **software** artificial neural network:

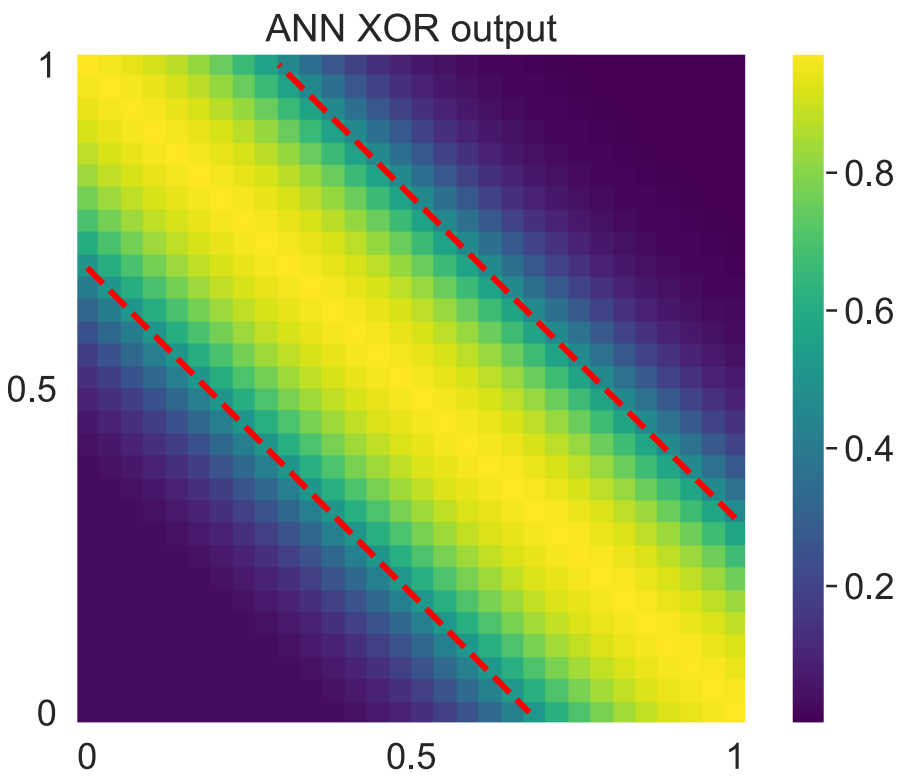


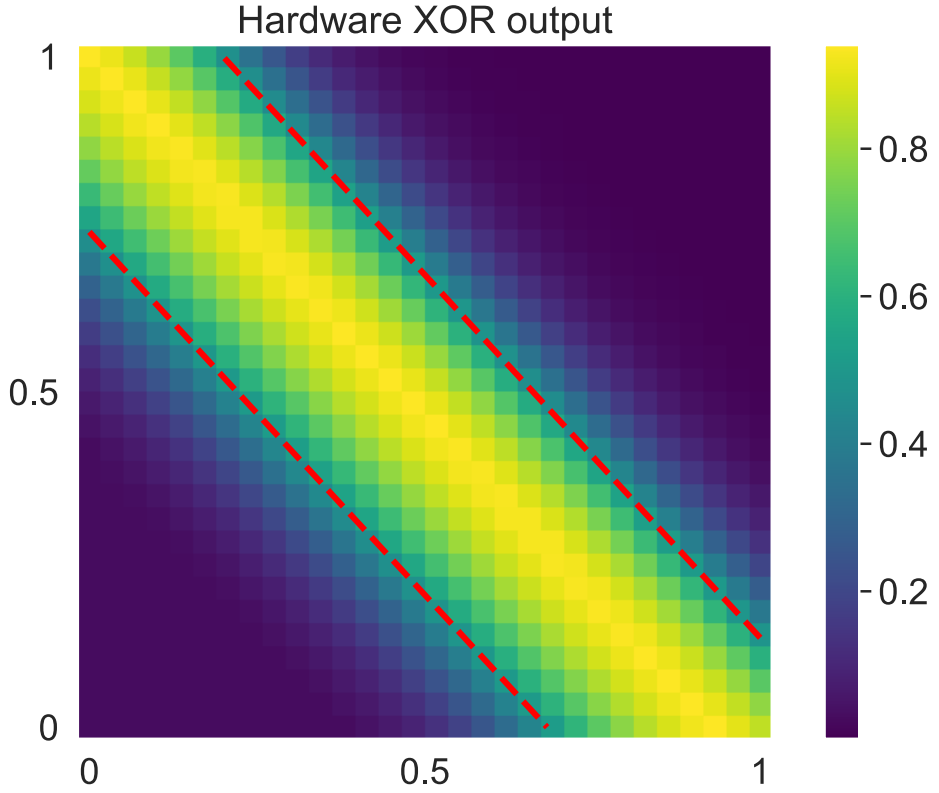


- **1 hidden layer**: **2 neurons** activated by **reLU**
- **Output node**: **sigmoid** activated
- **Training**: gradient descent and error backpropagation.

XOR problem: hardware solution

Inference is performed using the **simulated 2-by-2 crossbar** array with **linearly distributed states**





Small discrepancies due to the **limited bit width** for writing the states.

Still solve the XOR problem

Conclusion & take-home message

- The **deviation from perfectly linearly distributed states** induces an **intrinsic error** in the MAC operation.
- The **SNR however increases** when the **number of columns in the matrix** increases. Hence the array is optimized for multiplying vectors and **wide matrices**.
- The SNR should also increase with the number of columns when considering the **measurement noise**.
- We developed a framework for **dealing with the non-idealities** of M^2TJs crossbar arrays for performing the MAC operation.
- The framework allows to assess the **impact of experimental noise and device non-idealities** on the quality of the MAC operation.
- Simulated M^2TJs crossbar arrays were successfully used to **solve the XOR problem** using a simple neural network.

Aknowledgements

Neuromorphic Engineering Group @ UCLouvain



Prof. Flavio ABREU ARAUJO
Group leader



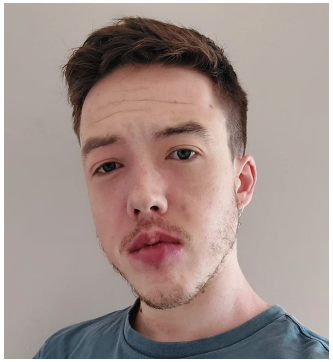
Dr. Lindiomar BORGES DE AVILA Jr
Post-doctoral researcher



Dr. Tristan DA CÂMARA SANTA CLARA GOMES
Post-doctoral researcher
(currently at INESC-MN)



Simon DE WERGIFOSSE
Doctoral researcher



Anatole MOUREAUX
Doctoral researcher



Anthony LOPES TEMPORAO
Doctoral researcher

MULTISPIN.AI
AI Co-Processor Research Project

Funded by the European Union

Related presentation:

Paper **13586-97** | Thursday 9:30 AM | Conv. Ctr. Room 5B
To be presented by Dr. **Dmirty Leshchiner**