Spintronics in the AI age

Lior Klein
Bar-Ilan University

Funded by
the European Union

*MultiSpin.AI has received funding from the
European Union under grant agreement 101130046*

MULTISPIN.AI
AI Co-Processor Research Project

AI risks leading humanity to 'extinction,' experts warn

It's the most recent in a series of alarms raised by experts in artificial intelligence – but also one that stoked growing pushback against a focus on what seem to be its overhyped hypothetical harms.

Google CEO: A.I. is more important than fire or electricity

Published Thu, Feb 1 2018•12:56 PM EST • Updated Thu, Feb 1 2018•12:56 PM EST

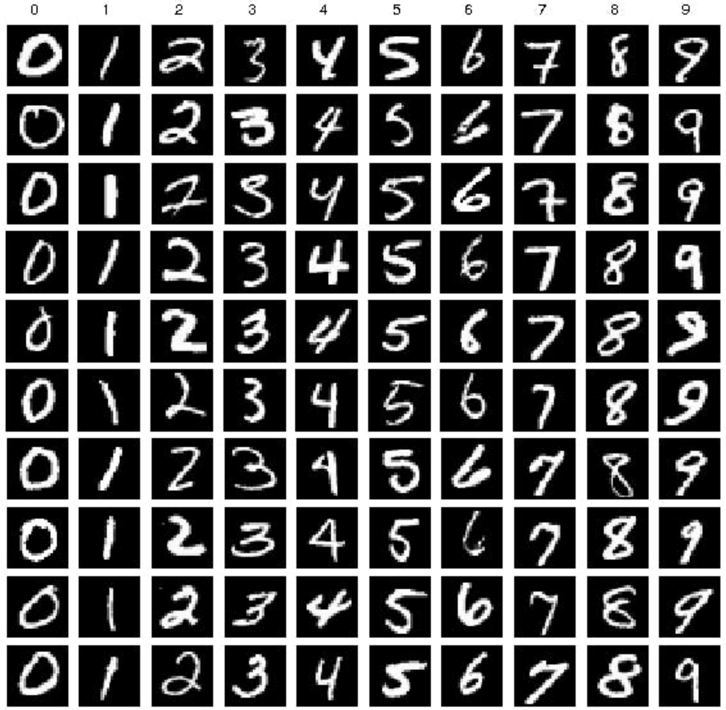
Catherine Clifford
@IN/CATCLIFFORD/
@CATCLIFFORD

What is AI?

The development of computer systems able to perform tasks that normally require human intelligence

natural language processing understanding

visual perception



What is AI good for?

AI - Healthcare

Medical diagnosis and treatment.

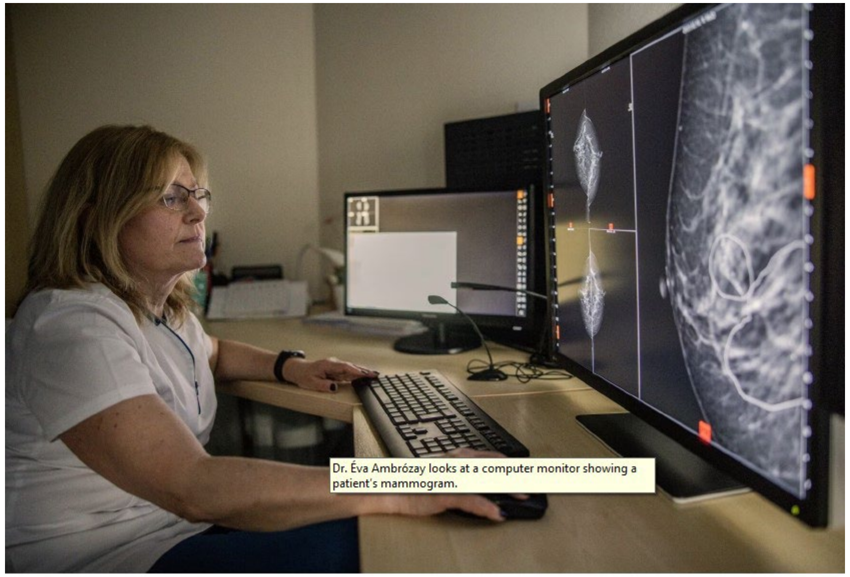
3/6/23, 9:52 AM How A.I. Is Being Used to Detect Cancer That Doctors Miss

The New York Times

Artificial Intelligence › | An Unsettling Chat With Bing | Read the Conversation | How Chatb

Using A.I. to Detect Breast Cancer That Doctors Miss

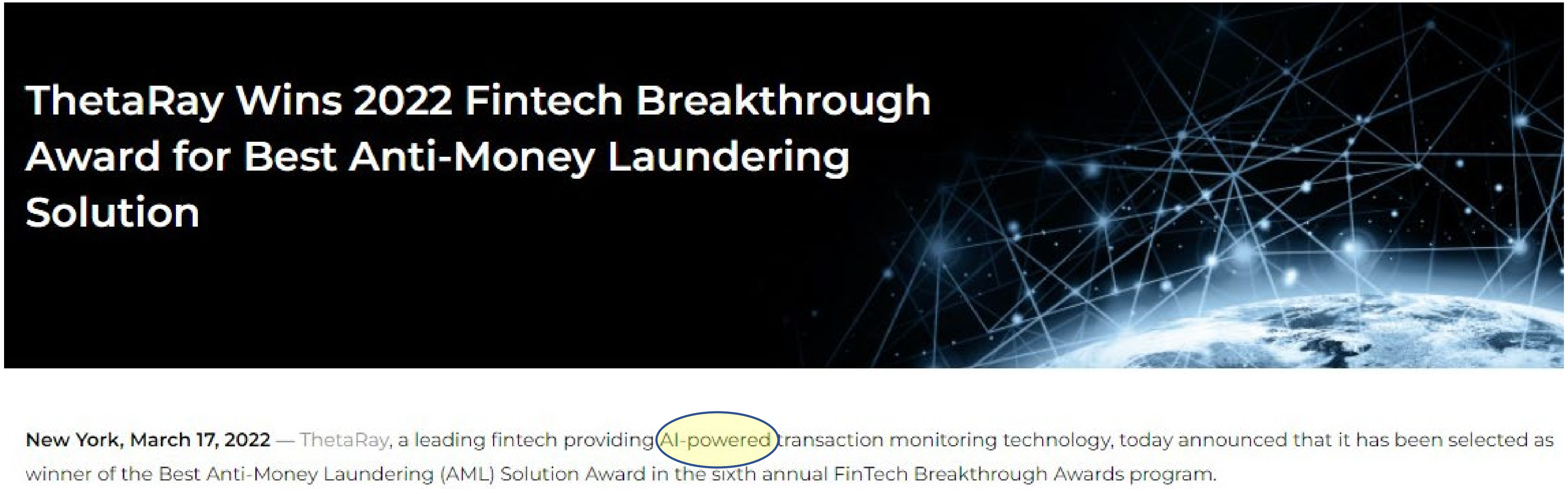
Hungary has become a major testing ground for A.I. software to spot cancer, as doctors debate whether the technology will replace them in medical jobs.



Dr. Éva Ambrózay looks at a computer monitor showing a patient's mammogram.

AI - Financial services

fraud detection
credit scoring
investment analysis



ThetaRay Wins 2022 Fintech Breakthrough Award for Best Anti-Money Laundering Solution

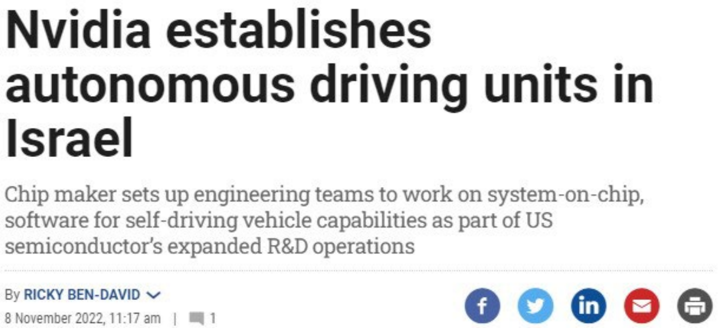
New York, March 17, 2022 — ThetaRay, a leading fintech providing AI-powered transaction monitoring technology, today announced that it has been selected as winner of the Best Anti-Money Laundering (AML) Solution Award in the sixth annual FinTech Breakthrough Awards program.

AI - Autonomous vehicles

AI algorithms are being used to develop self-driving cars and trucks,
which have the potential to reduce accidents and improve traffic flow.

Mercedes, Nvidia, and Google Are Creating Genuinely Smart Cars With AI

By Rob Enderle • February 27, 2023 4:00 AM PT • Email Article



Nvidia establishes autonomous driving units in Israel

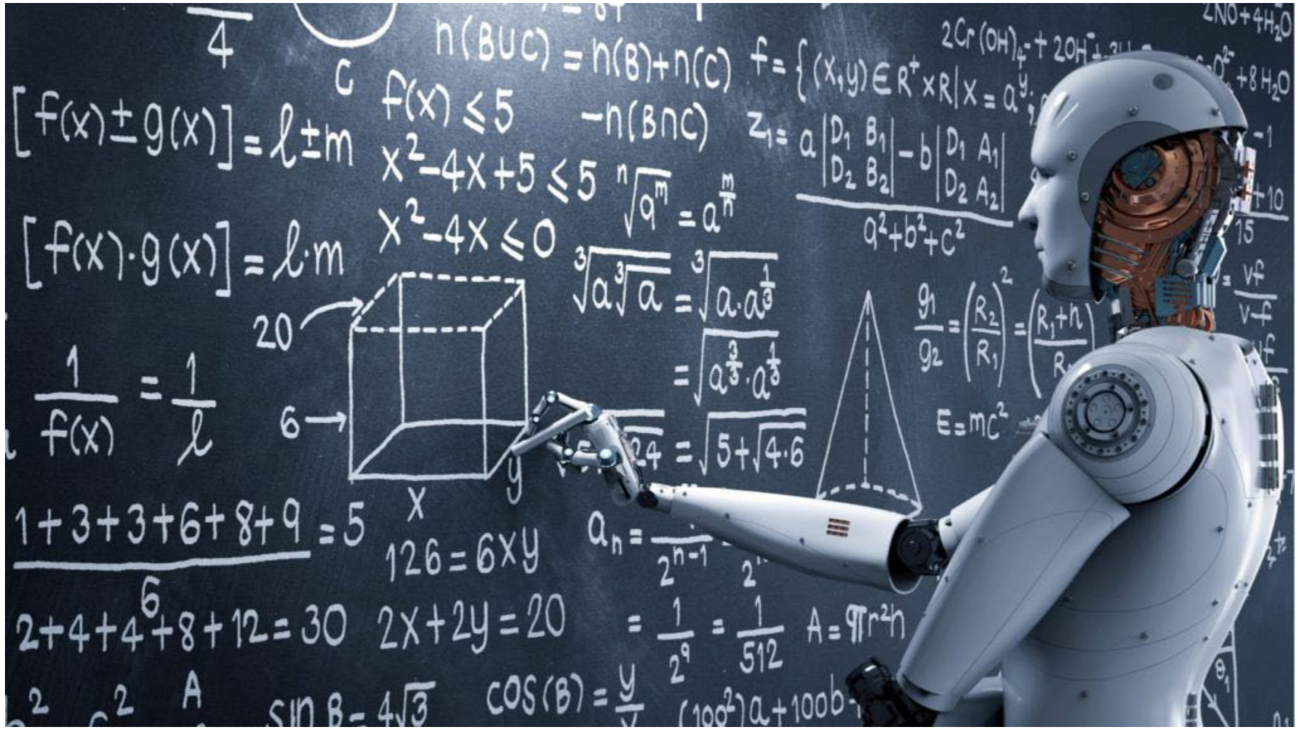
Chip maker sets up engineering teams to work on system-on-chip, software for self-driving vehicle capabilities as part of US semiconductor's expanded R&D operations

By RICKY BEN-DAVID

8 November 2022, 11:17 am | 1

AI - Education

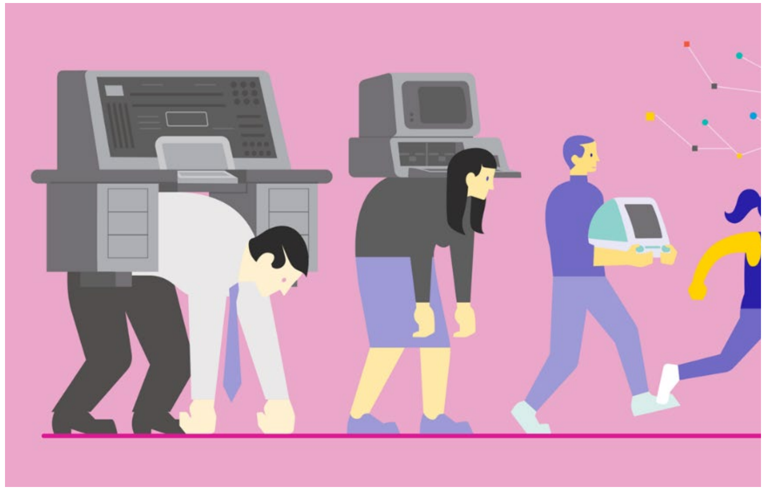
AI could be used to develop personalized learning experiences, allowing students to learn at their own pace and in their preferred style.



AI – computation challenges

The end of Moore’s law

Moore’s Law: the number of transistors in a dense integrated circuit doubles about every two years



‘Moore’s Law’s dead,’ Nvidia CEO Jensen Huang says in justifying gaming-card price hike

Last Updated: Sept. 22, 2022 at 7:43 a.m. ET
First Published: Sept. 21, 2022 at 6:16 p.m. ET

By Wallace Witkowski

Silicon wafers used to make chips ‘not a little bit more expensive, it is a ton more expensive,’ Huang says

The end of the Dennard scaling law

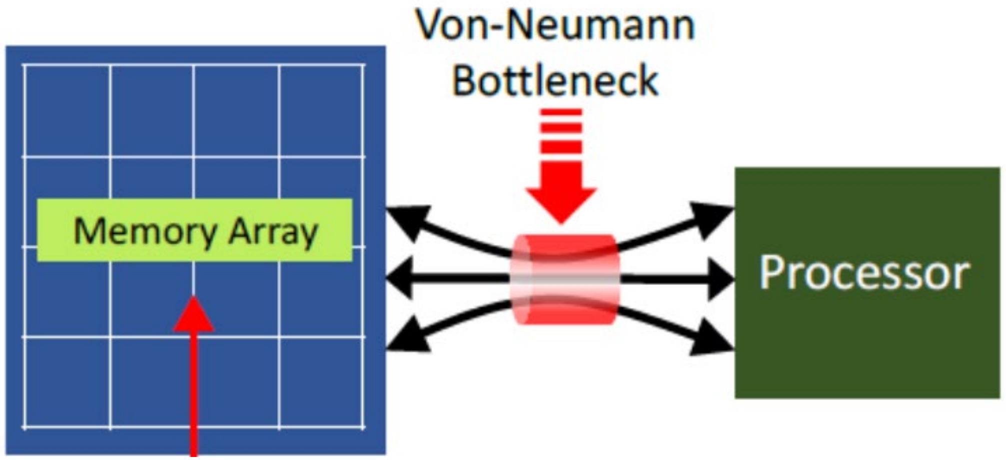
Dennard scaling law: as transistors get smaller, their power density stays constant, so that the power use stays in proportion with area

Since around 2005–2007 Dennard scaling appears to have broken down

Von Neumann bottleneck

While the processor can execute instructions very quickly, it must spend a significant amount of time waiting for data to be transferred to and from memory

CPU execution time - a few nanoseconds or less
memory access times - tens to hundreds of nanoseconds



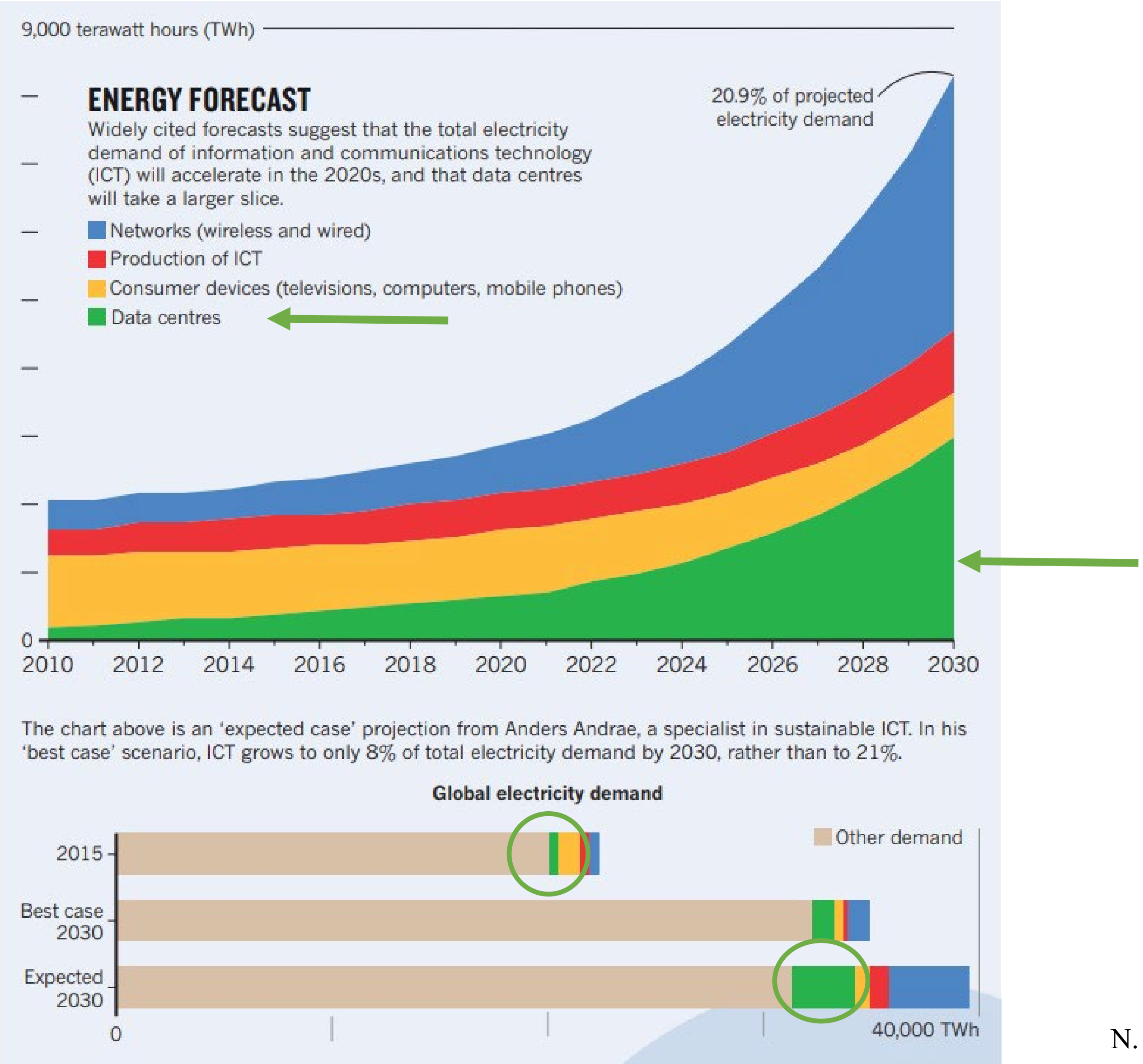
The need for Edge computing

Data processing and analysis are performed on the devices or systems that generate the data, rather than sending all the data to a centralized location, such as a cloud server, for processing

Applications for edge computing include autonomous vehicles, industrial automation, smart cities, and healthcare monitoring systems

AI and edge computing are making CCTV more effective than ever

The hunger for power



ENERGY FORECAST

Widely cited forecasts suggest that the total electricity demand of information and communications technology (ICT) will accelerate in the 2020s, and that data centres will take a larger slice.

The chart above is an 'expected case' projection from Anders Andrae, a specialist in sustainable ICT. In his 'best case' scenario, ICT grows to only 8% of total electricity demand by 2030, rather than to 21%.

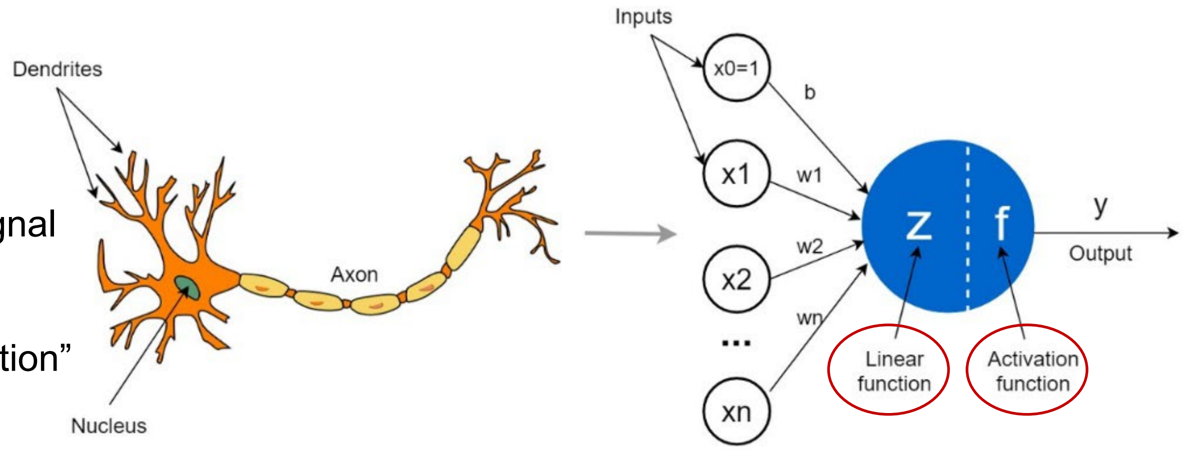
N. Jones Nature 2018

AI – Neuromorphic computing

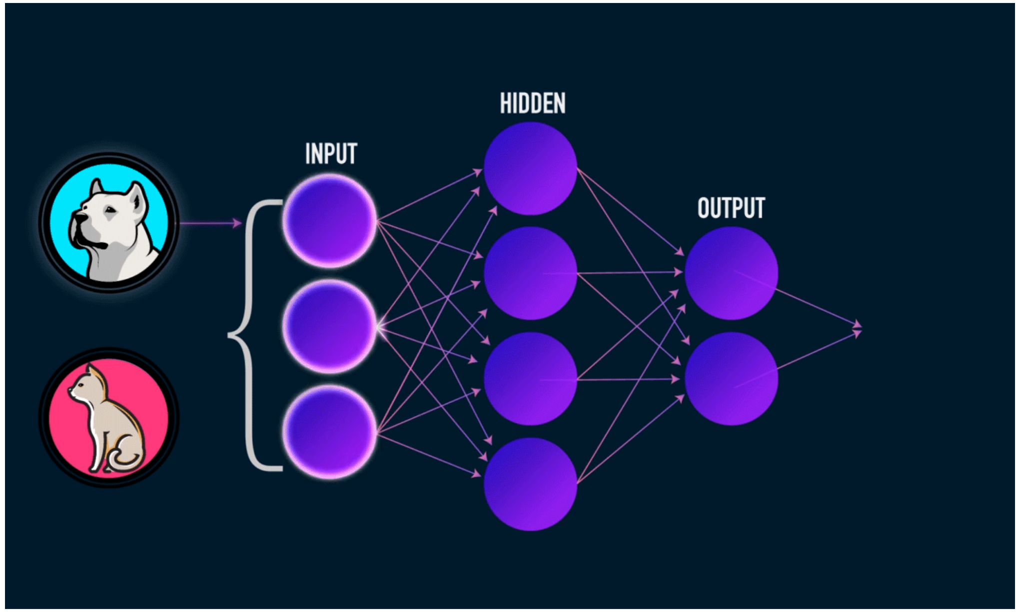
Visual perception

Stage 1:
Linear operation on the signal

Stage 2:
Non-linear “activation function”



Artificial neural network

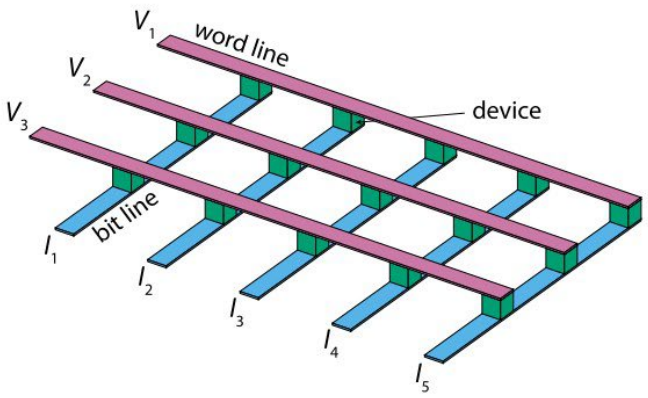


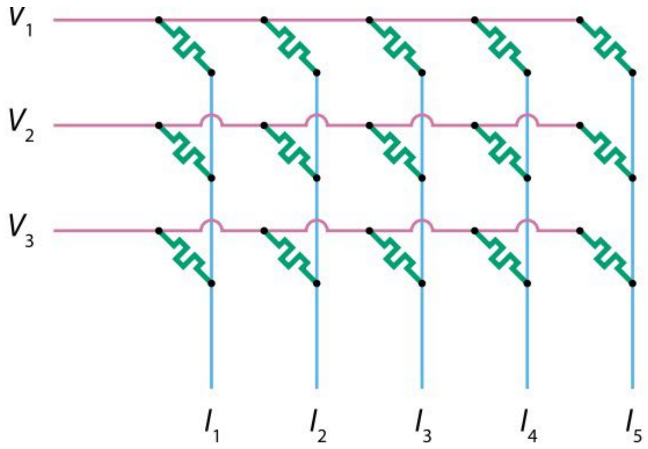
The crossbar – a central analogue AI hardware

Analogue execution of the linear operation

In memory processing:
fast and energy efficient

removes major computation barriers
particularly for AI at the edge





$$I_i = \sum_j G_{i,j} V_j$$

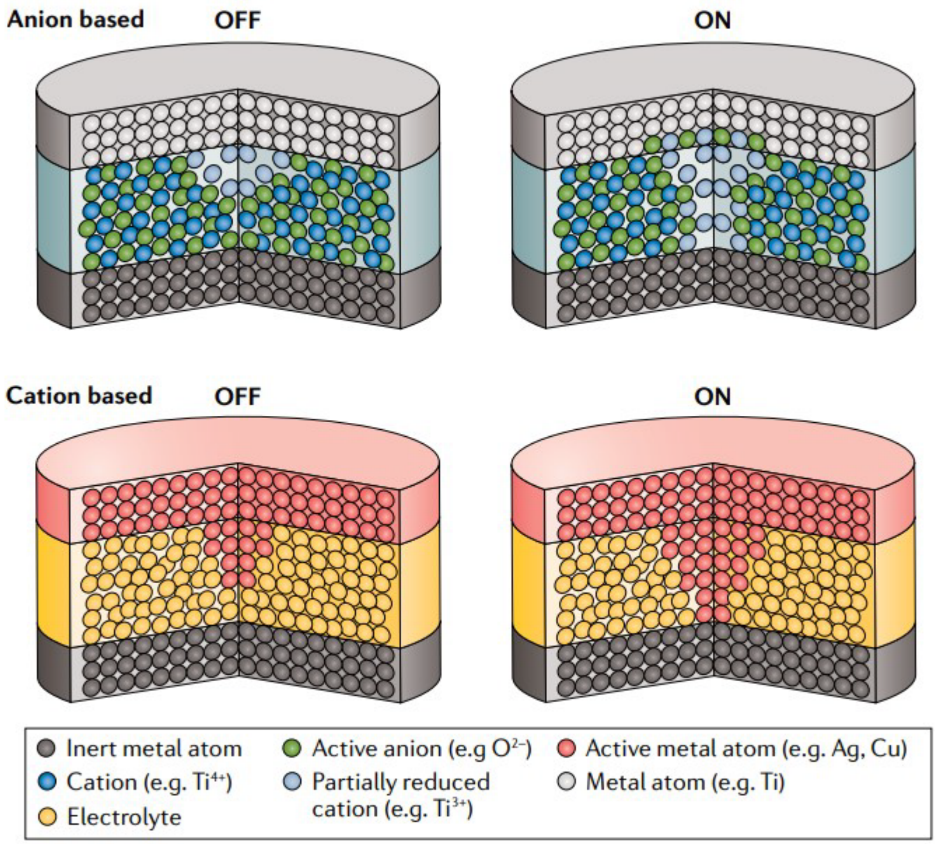
$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ \vdots \\ I_n \end{bmatrix} = \begin{bmatrix} G_{1,1} & G_{1,2} & G_{1,3} & \cdots & G_{1,m} \\ G_{2,1} & G_{2,2} & G_{2,3} & \cdots & G_{2,m} \\ G_{3,1} & G_{3,2} & G_{3,3} & \cdots & G_{3,m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ G_{n,1} & G_{n,2} & G_{n,3} & \cdots & G_{n,m} \end{bmatrix} \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ \vdots \\ V_m \end{bmatrix}$$

Ohm’s law for crosspoint multiplication
Kirchhoff's law for summation

Advantages in speed and energy efficiency

The crossbar – material options

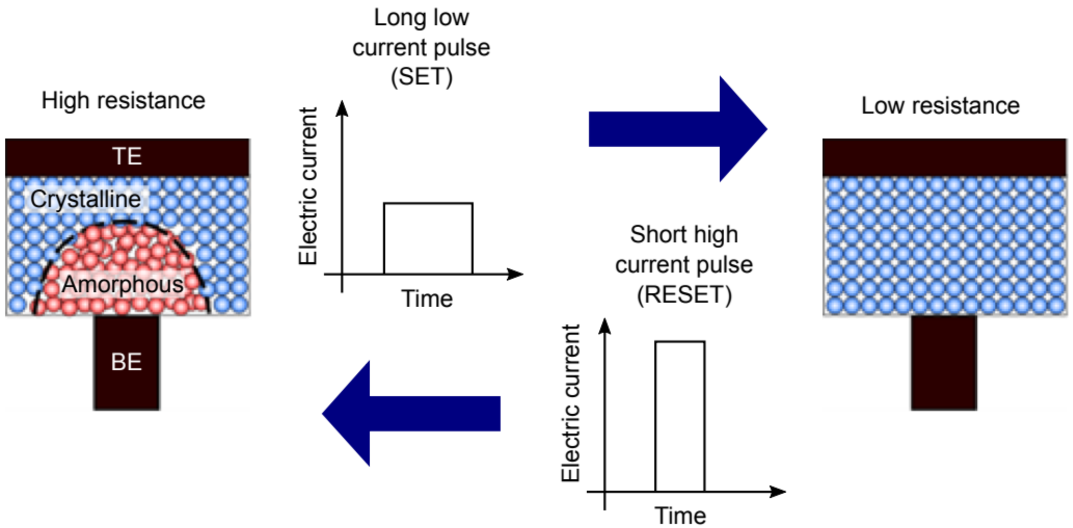
Redox (reduction oxidation) materials

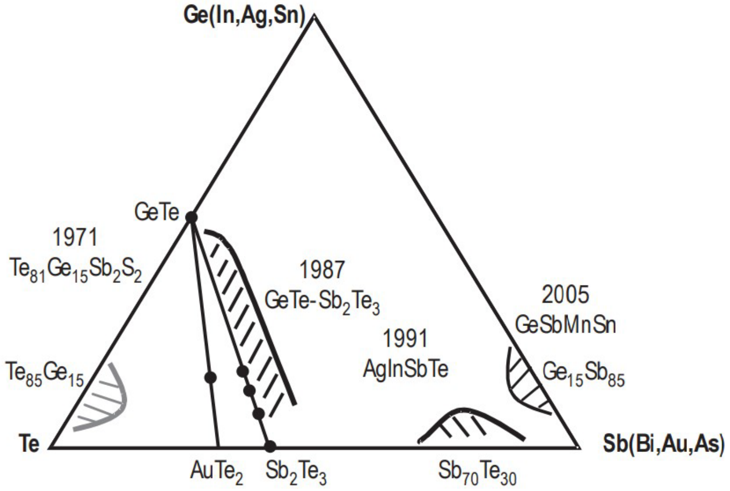


Wang et al. Nature Reviews 5, 173 (2020)

The crossbar – material options

Phase change materials

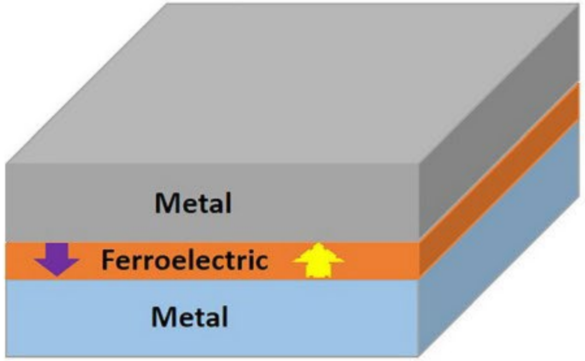


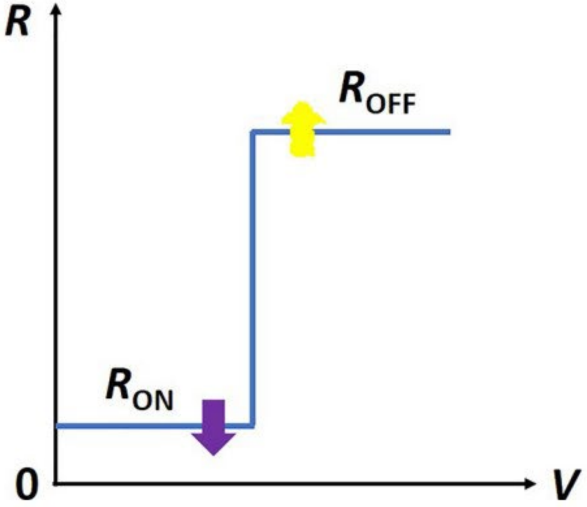


Wuttig and Yamada Nature Materials 6, 824 (2007)

The crossbar – material options

Ferroelectric tunnel junctions



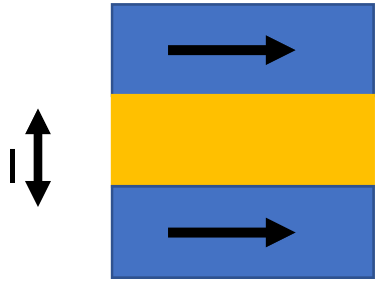


Shao et al. Nature Electronics 5, 67 (2022)

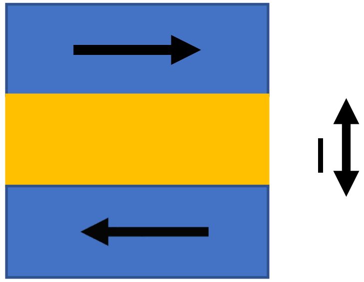
The crossbar – material options

Magnetoresistance (spintronics)

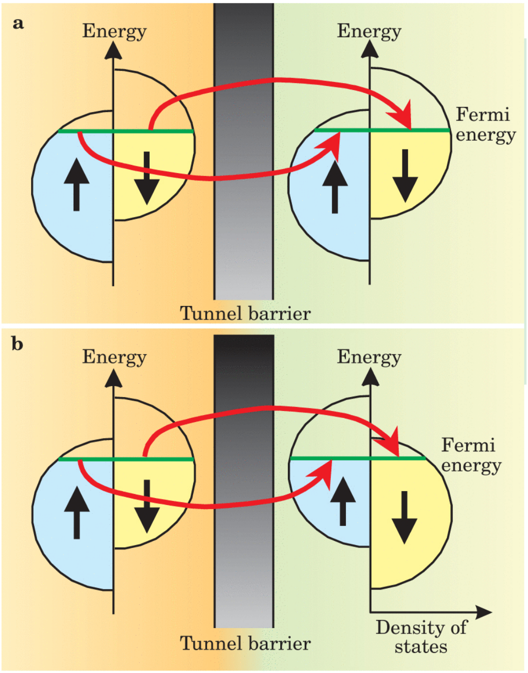
Magnetic tunnel junction (MTJ)



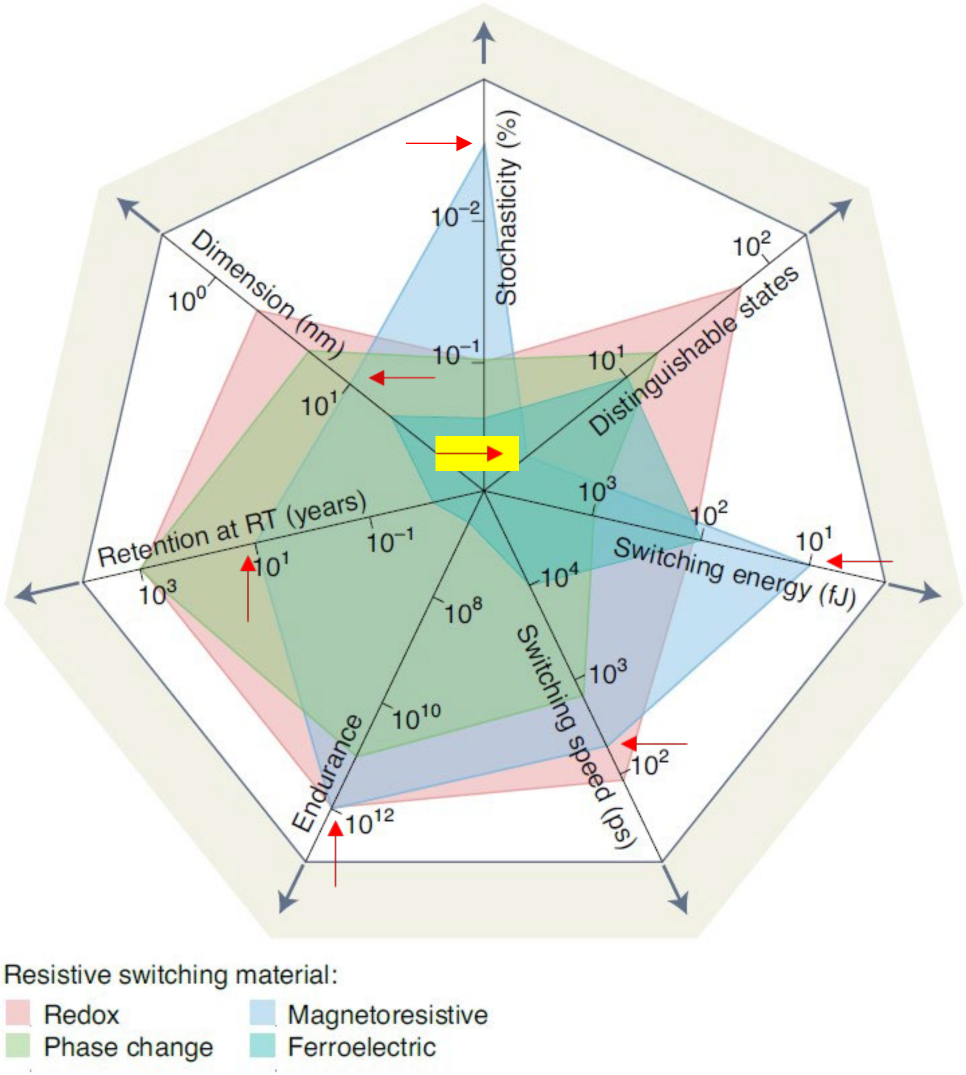
low resistance



high resistance



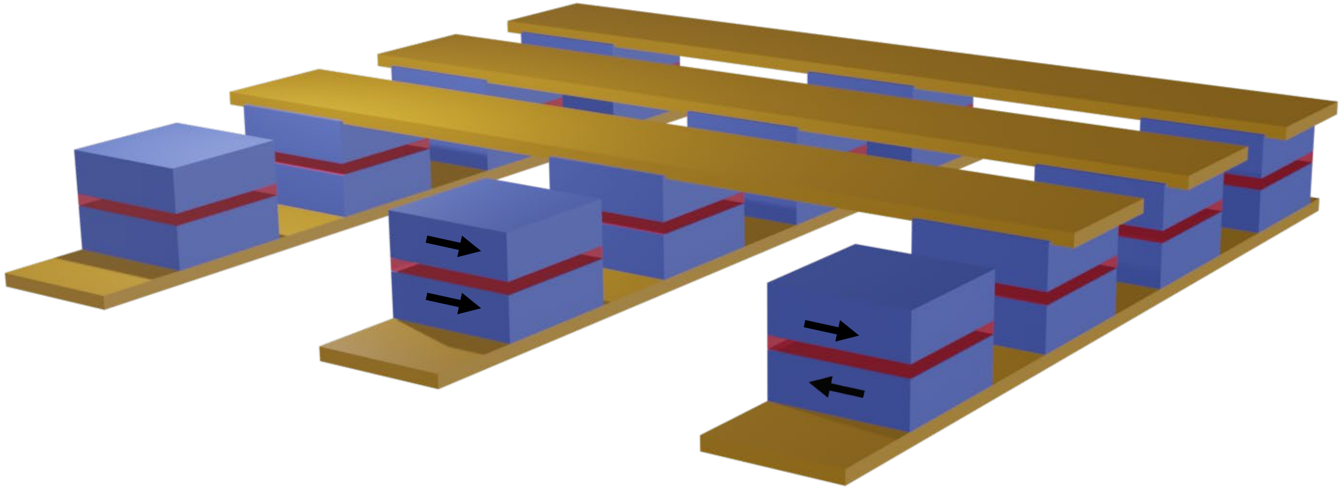
The crossbar – material options

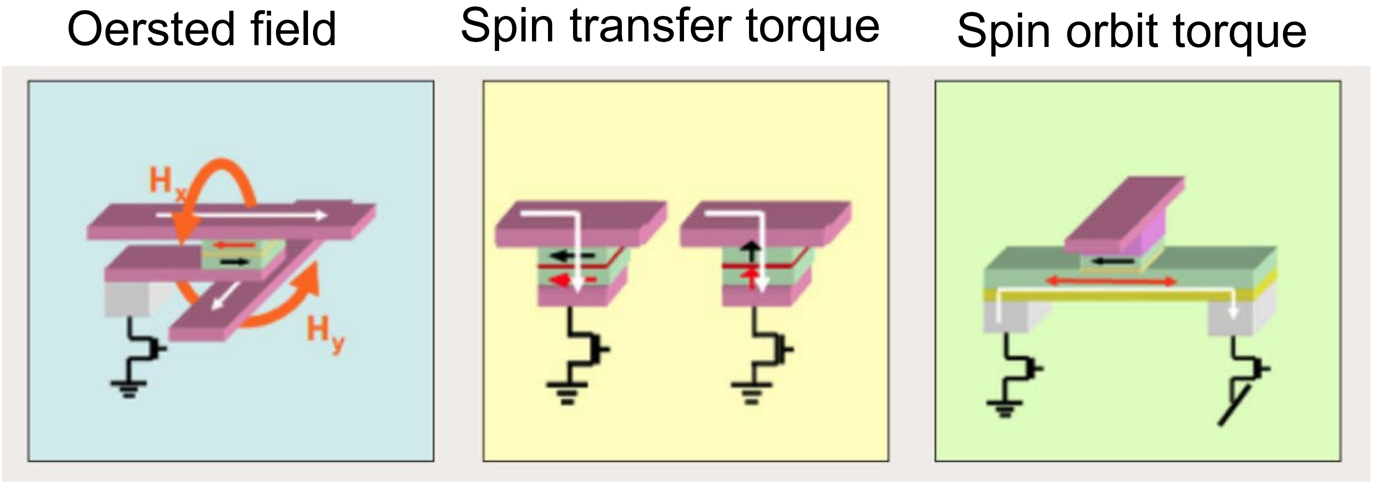


Wang et al. Nature Reviews 5, 173 (2020)

The spintronic crossbar

Spintronics – magnetic random access memory with MTJs

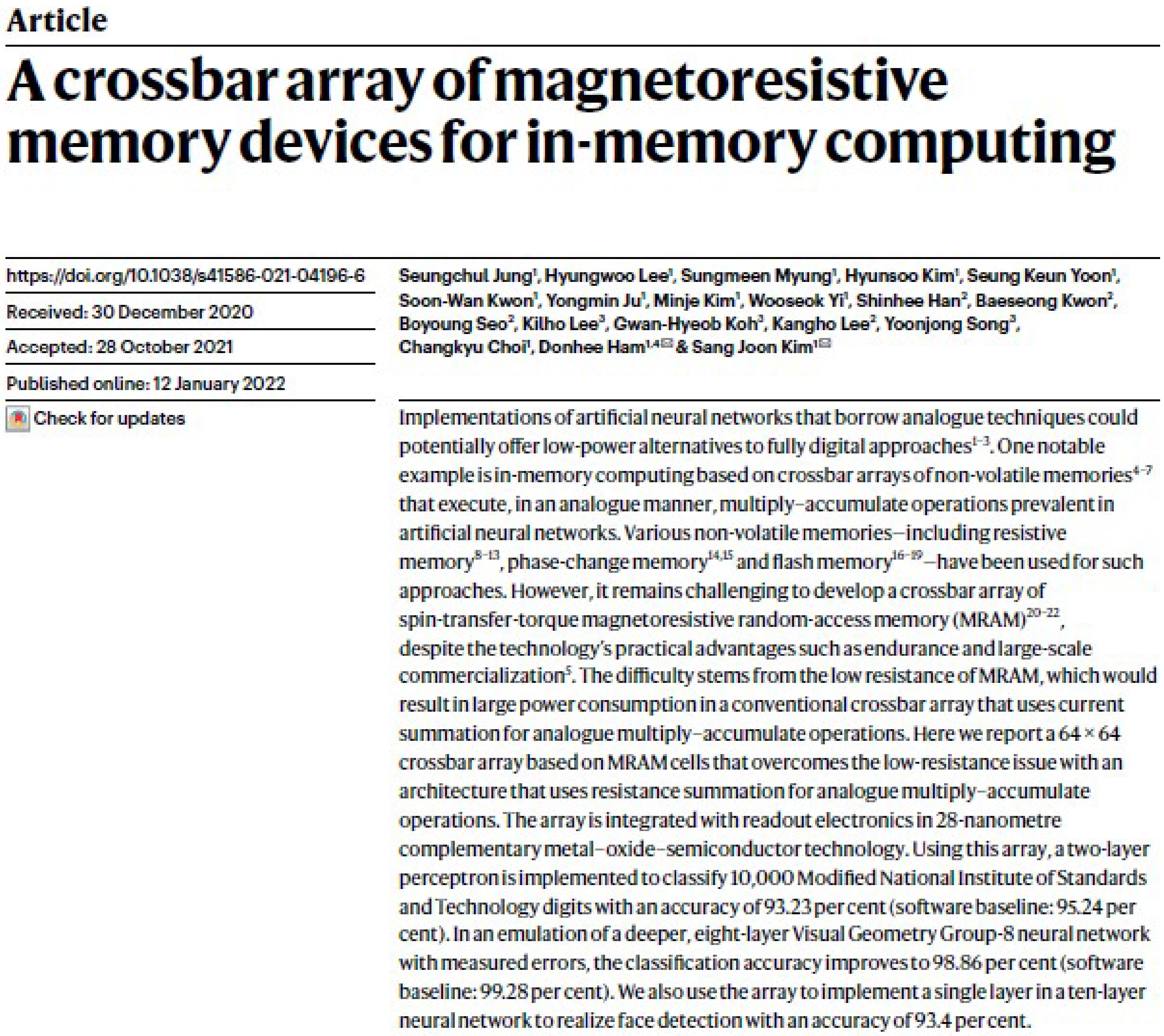




Oersted field

Spin transfer torque

Spin orbit torque



Article

A crossbar array of magnetoresistive memory devices for in-memory computing

https://doi.org/10.1038/s41586-021-04196-6

Received: 30 December 2020

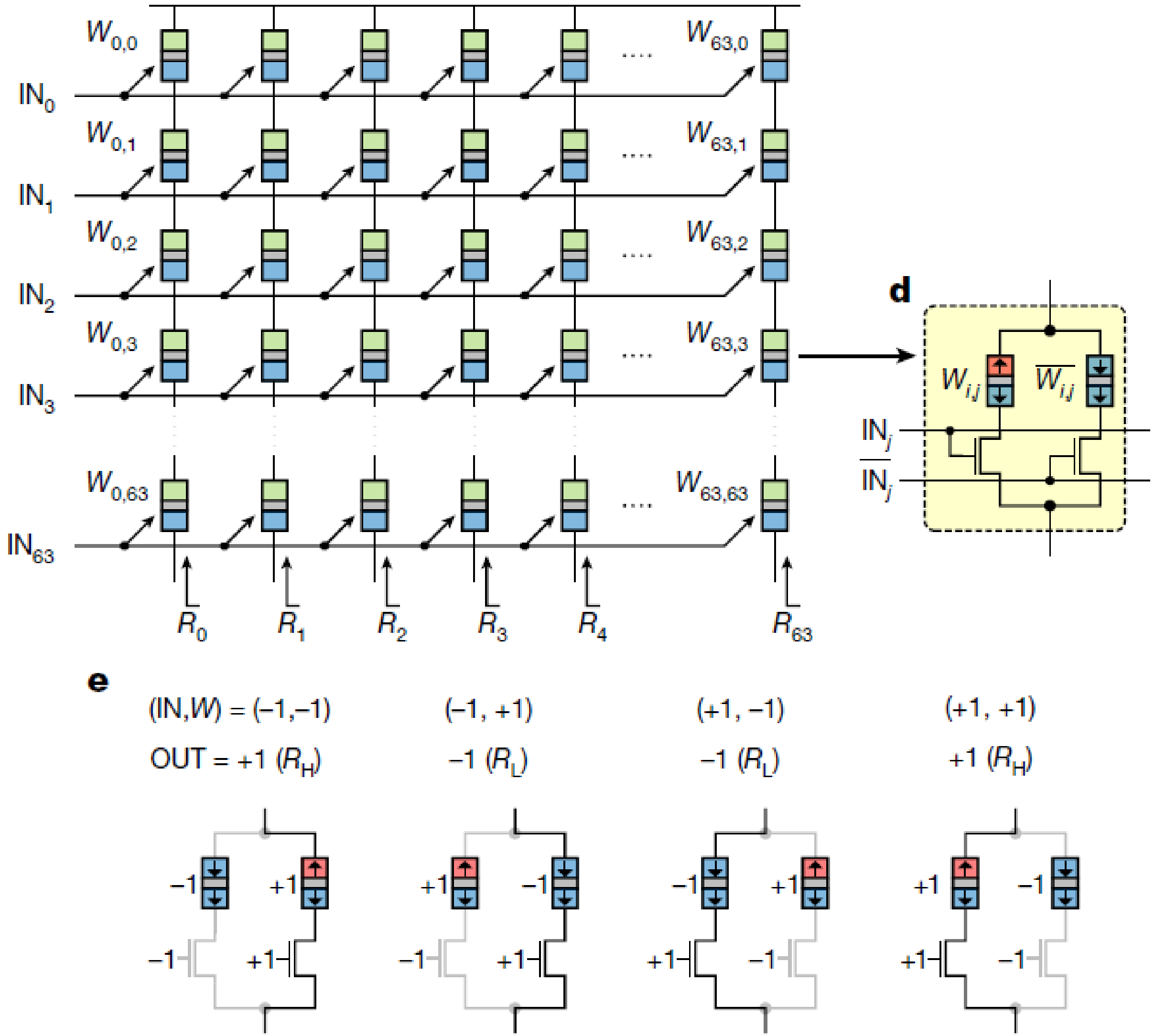
Accepted: 28 October 2021

Seungchul Jung[1], Hyungwoo Lee[1], Sungmeen Myung[1], Hyunsoo Kim[1], Seung Keun Yoon[1], Soon-Wan Kwon[1], Yongmin Ju[1], Minje Kim[1], Wooseok Yi[1], Shinhee Han[2], Baeseong Kwon[2], Boyoung Seo[2], Kilho Lee[3], Gwan-Hyeob Koh[3], Kangho Lee[2], Yoonjong Song[3], Changkyu Choi[1], Donhee Ham[1,4]✉ & Sang Joon Kim[1]✉

Implementations of artificial neural networks that borrow analogue techniques could potentially offer low-power alternatives to fully digital approaches[1–3]. One notable example is in-memory computing based on crossbar arrays of non-volatile memories[4–7] that execute, in an analogue manner, multiply–accumulate operations prevalent in artificial neural networks. Various non-volatile memories—including resistive memory[8–13], phase-change memory[14,15] and flash memory[16–19]—have been used for such approaches. However, it remains challenging to develop a crossbar array of spin-transfer-torque magnetoresistive random-access memory (MRAM)[20–22], despite the technology's practical advantages such as endurance and large-scale commercialization[5]. The difficulty stems from the low resistance of MRAM, which would result in large power consumption in a conventional crossbar array that uses current summation for analogue multiply–accumulate operations. Here we report a 64 × 64 crossbar array based on MRAM cells that overcomes the low-resistance issue with an architecture that uses resistance summation for analogue multiply–accumulate operations. The array is integrated with readout electronics in 28-nanometre complementary metal–oxide–semiconductor technology. Using this array, a two-layer perceptron is implemented to classify 10,000 Modified National Institute of Standards and Technology digits with an accuracy of 93.23 per cent (software baseline: 95.24 per cent). In an emulation of a deeper, eight-layer Visual Geometry Group-8 neural network with measured errors, the classification accuracy improves to 98.86 per cent (software baseline: 99.28 per cent). We also use the array to implement a single layer in a ten-layer neural network to realize face detection with an accuracy of 93.4 per cent.

Samsung’s Spintronic Crossbar with MTJs

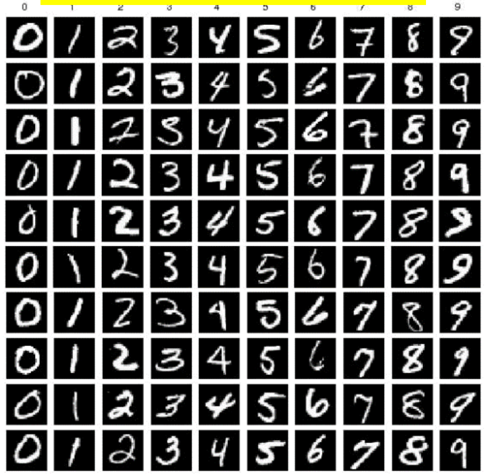
64x64 crossbar



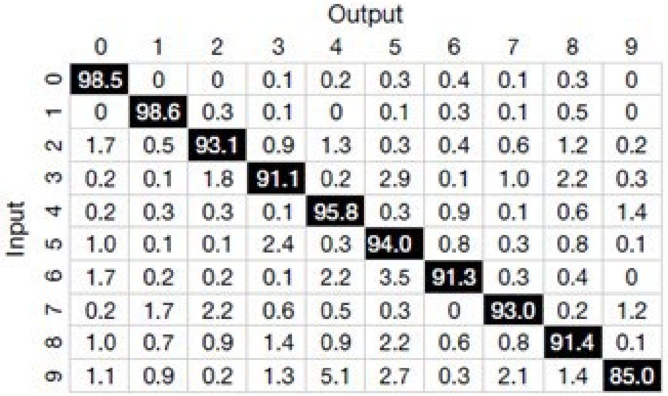
S. Jung et al. Nature 601, 2011 (2022)

The performance of MTJ-based Crossbar

28x28 pixel image

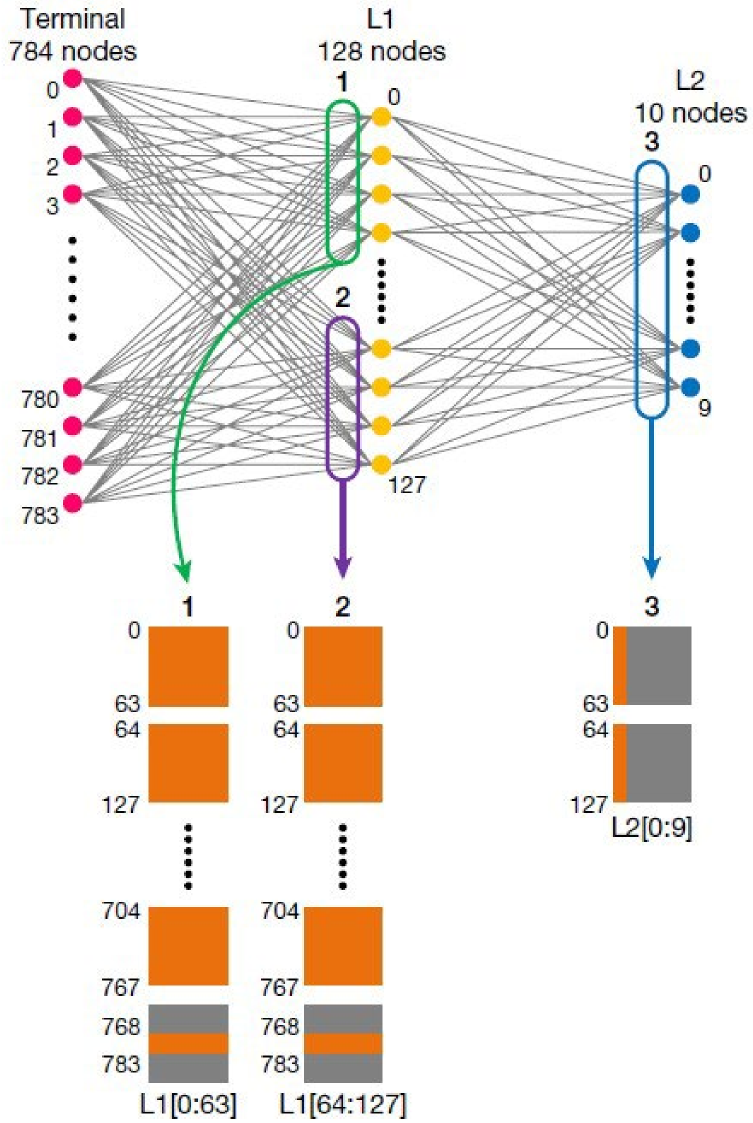


Limitation – only 2 weights per junction



Output

Input	0	1	2	3	4	5	6	7	8	9
0	98.5	0	0	0.1	0.2	0.3	0.4	0.1	0.3	0
1	0	98.6	0.3	0.1	0	0.1	0.3	0.1	0.5	0
2	1.7	0.5	93.1	0.9	1.3	0.3	0.4	0.6	1.2	0.2
3	0.2	0.1	1.8	91.1	0.2	2.9	0.1	1.0	2.2	0.3
4	0.2	0.3	0.3	0.1	95.8	0.3	0.9	0.1	0.6	1.4
5	1.0	0.1	0.1	2.4	0.3	94.0	0.8	0.3	0.8	0.1
6	1.7	0.2	0.2	0.1	2.2	3.5	91.3	0.3	0.4	0
7	0.2	1.7	2.2	0.6	0.5	0.3	0	93.0	0.2	1.2
8	1.0	0.7	0.9	1.4	0.9	2.2	0.6	0.8	91.4	0.1
9	1.1	0.9	0.2	1.3	5.1	2.7	0.3	2.1	1.4	85.0

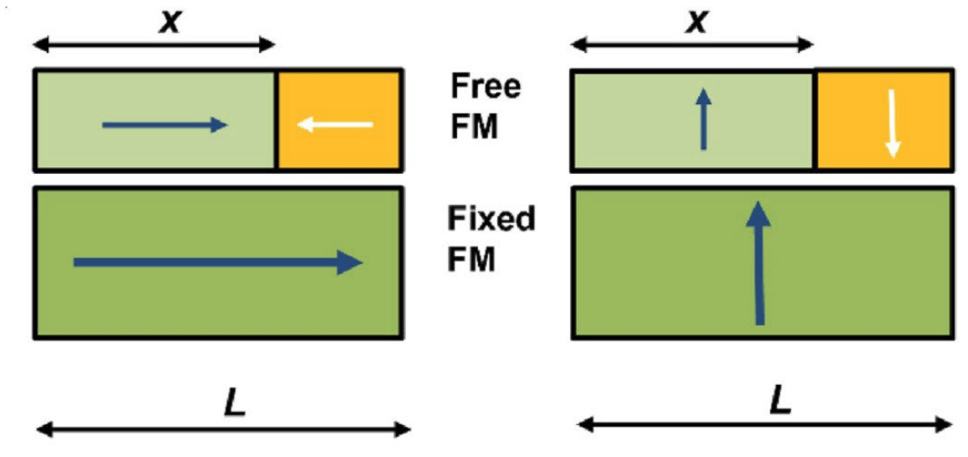


S. Jung et al. Nature 601, 2011 (2022)

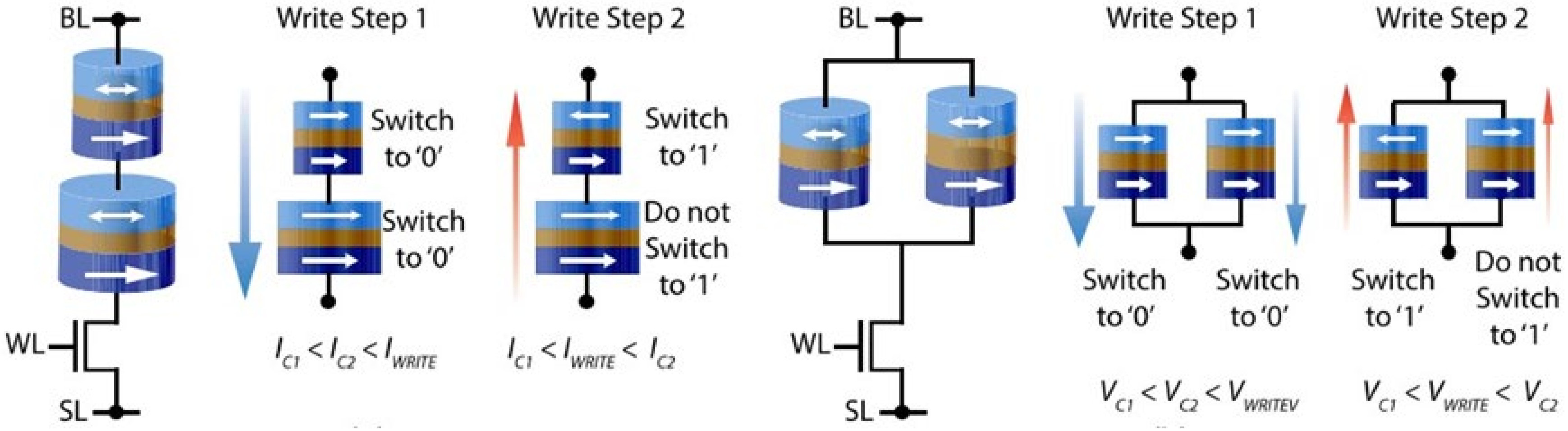
The spintronic crossbar

From binary to multi-state

Increasing the number of resistance states



Sci Rep 6, 31510 (2016)

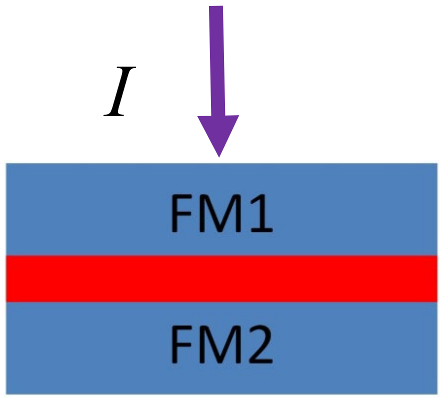


Proceedings of the IEEE, vol. 104, no. 7, 1449, July 2016

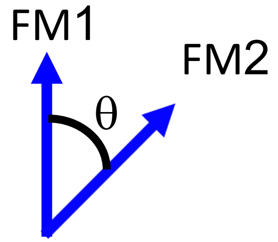
Using a new angle – the M^{2}TJ

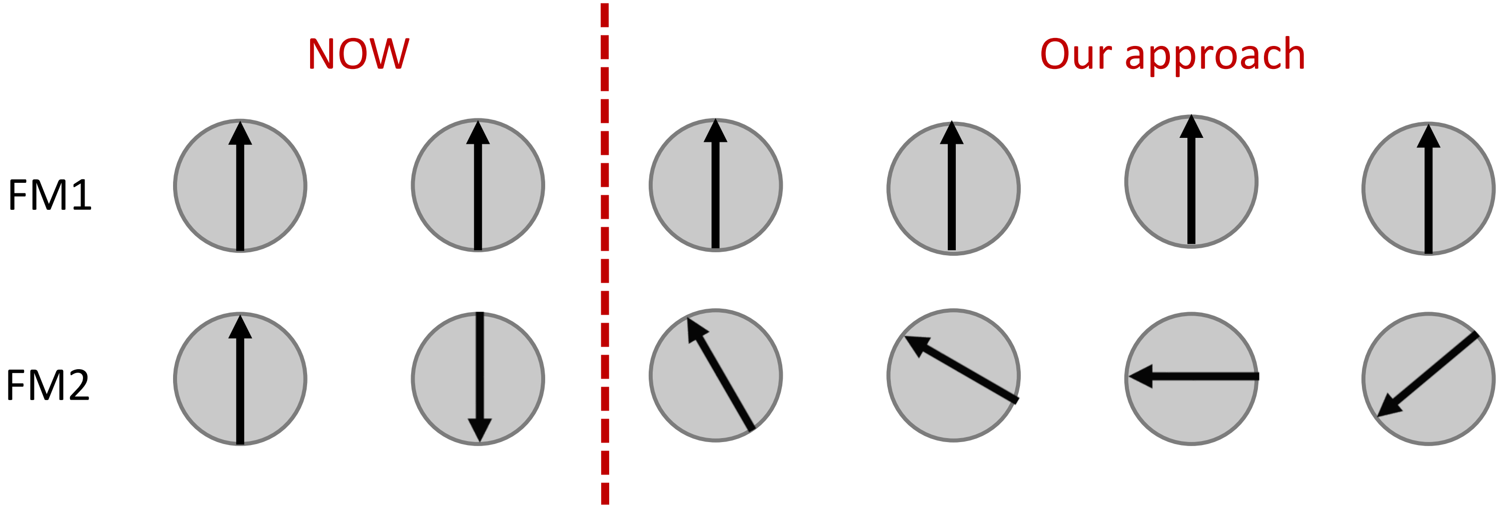
MULTISPIN.AI
AI Co-Processor Research Project

M^{2}TJ *the angular degree of freedom*



$$R_{junction} = R_{av} - \frac{1}{2}\Delta R \cos \vartheta$$

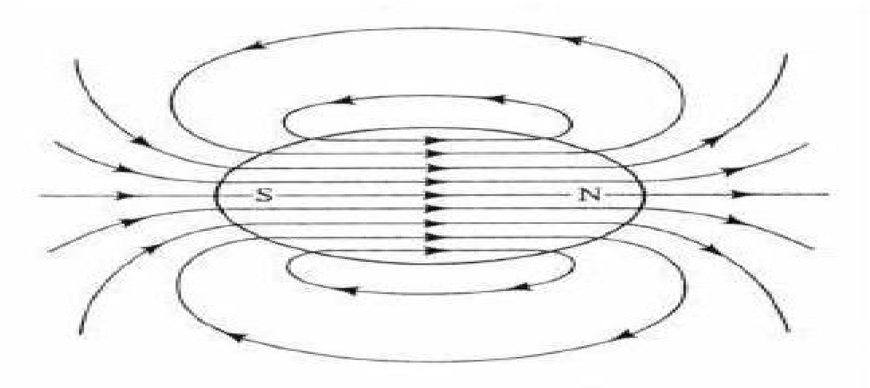


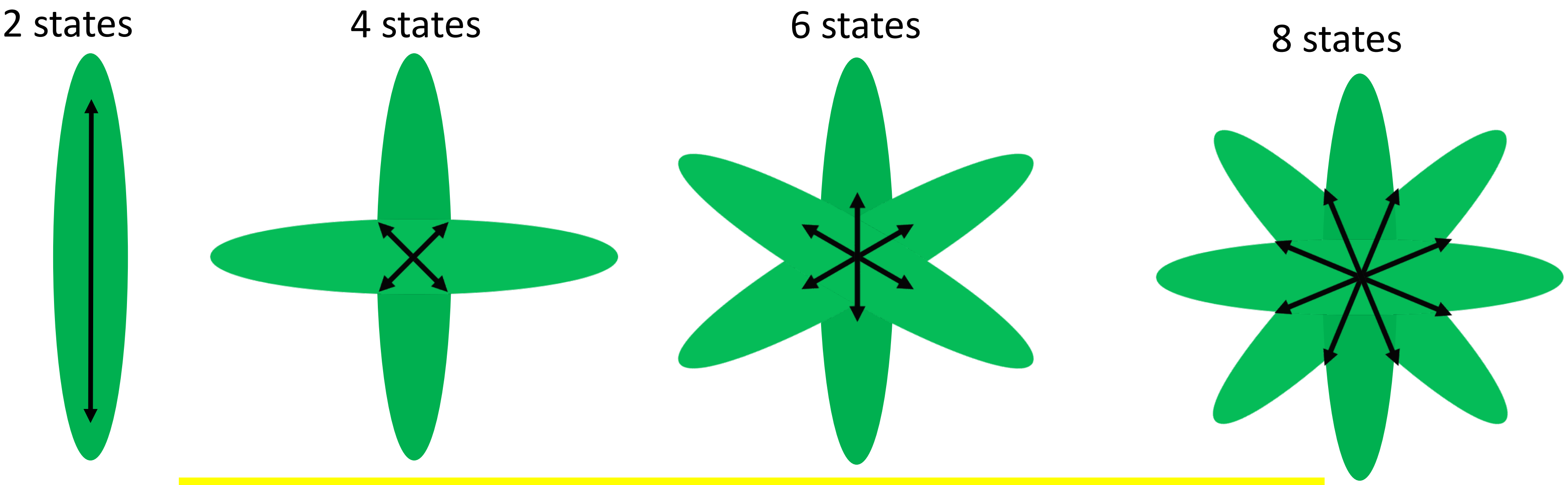


High-order magnetic anisotropy induced by shape

Basis for PHE sensors

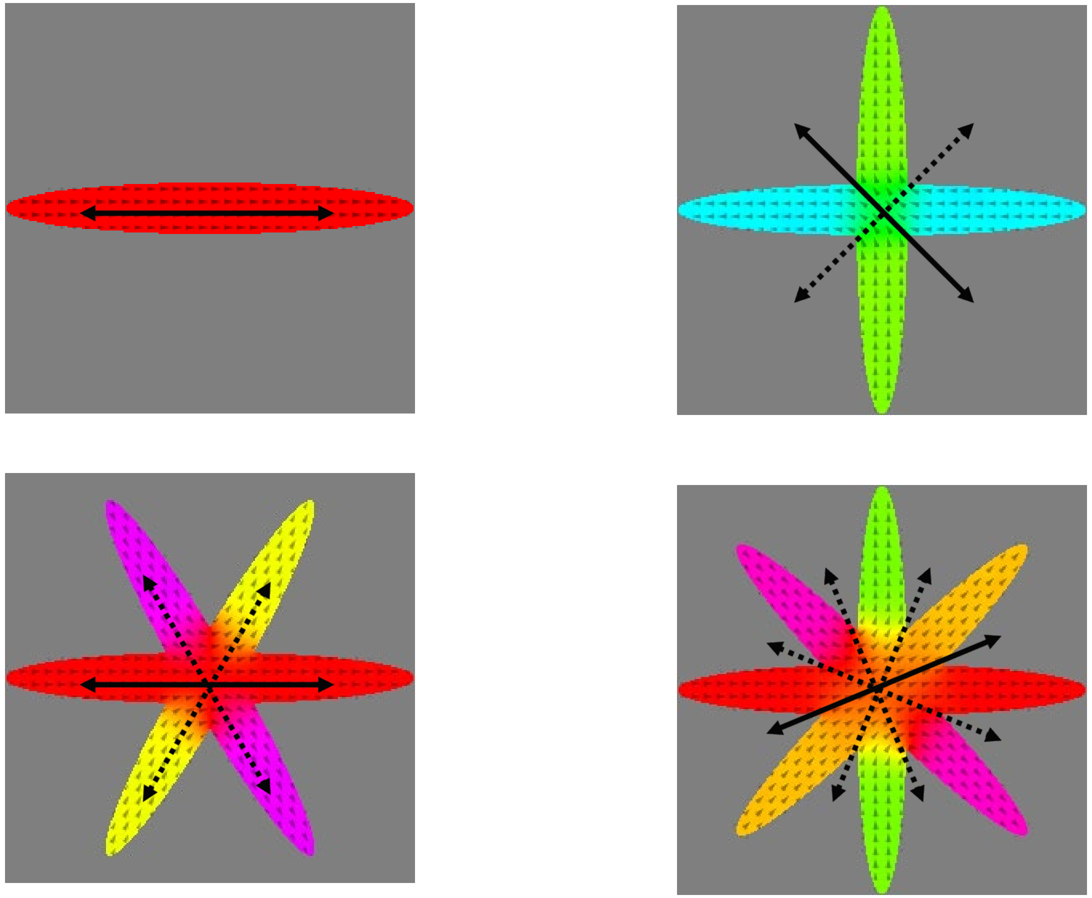
$$<10\ \frac{pT}{\sqrt{Hz}}@1\ Hz$$



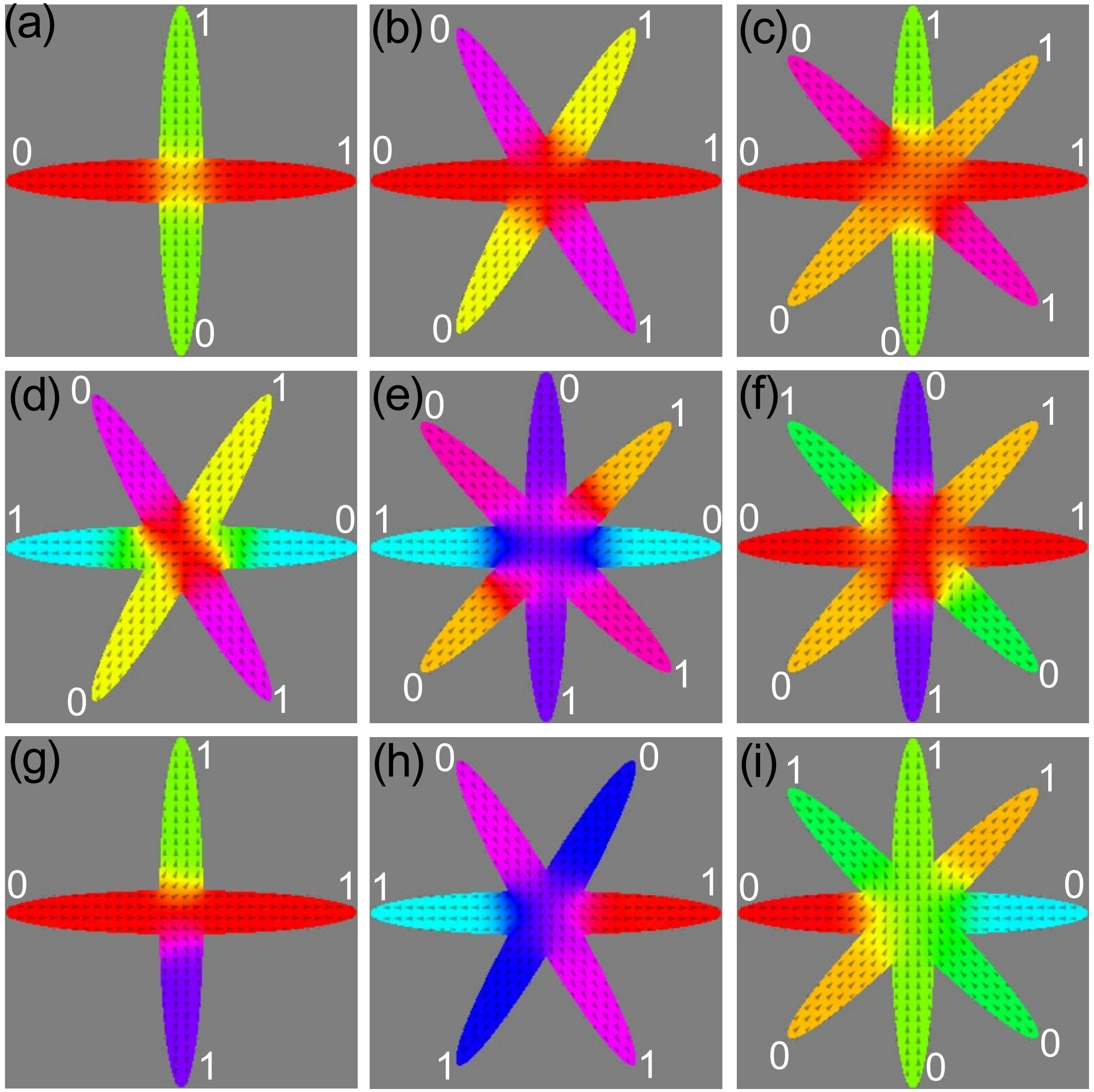


Using such structures for multi-level MRAM US Patent 10,204,678

The remanent magnetic configurations



Telepinsky et al., Appl. Phys. Lett. **108,** 182401 (2016)



2N states

Ordinary states

2^N states

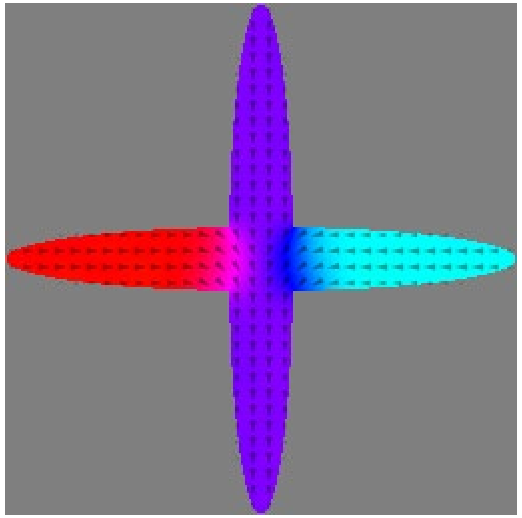
Staggered states

2^{2N} states

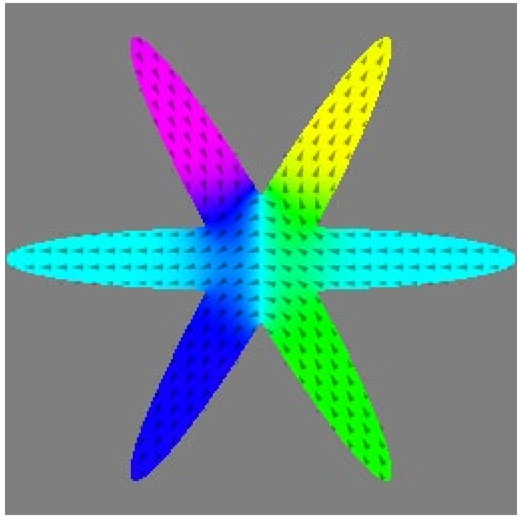
Neel states

S. Das et al., Appl. Phys. Lett. **116,** 262405 (2020).

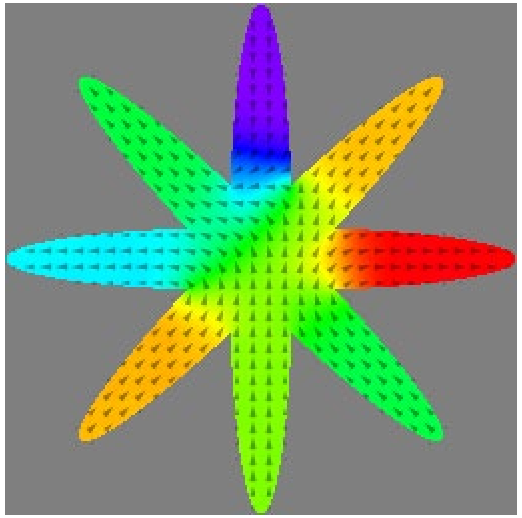
Exponential number of magnetic configurations



$2^{2*2} = 16$

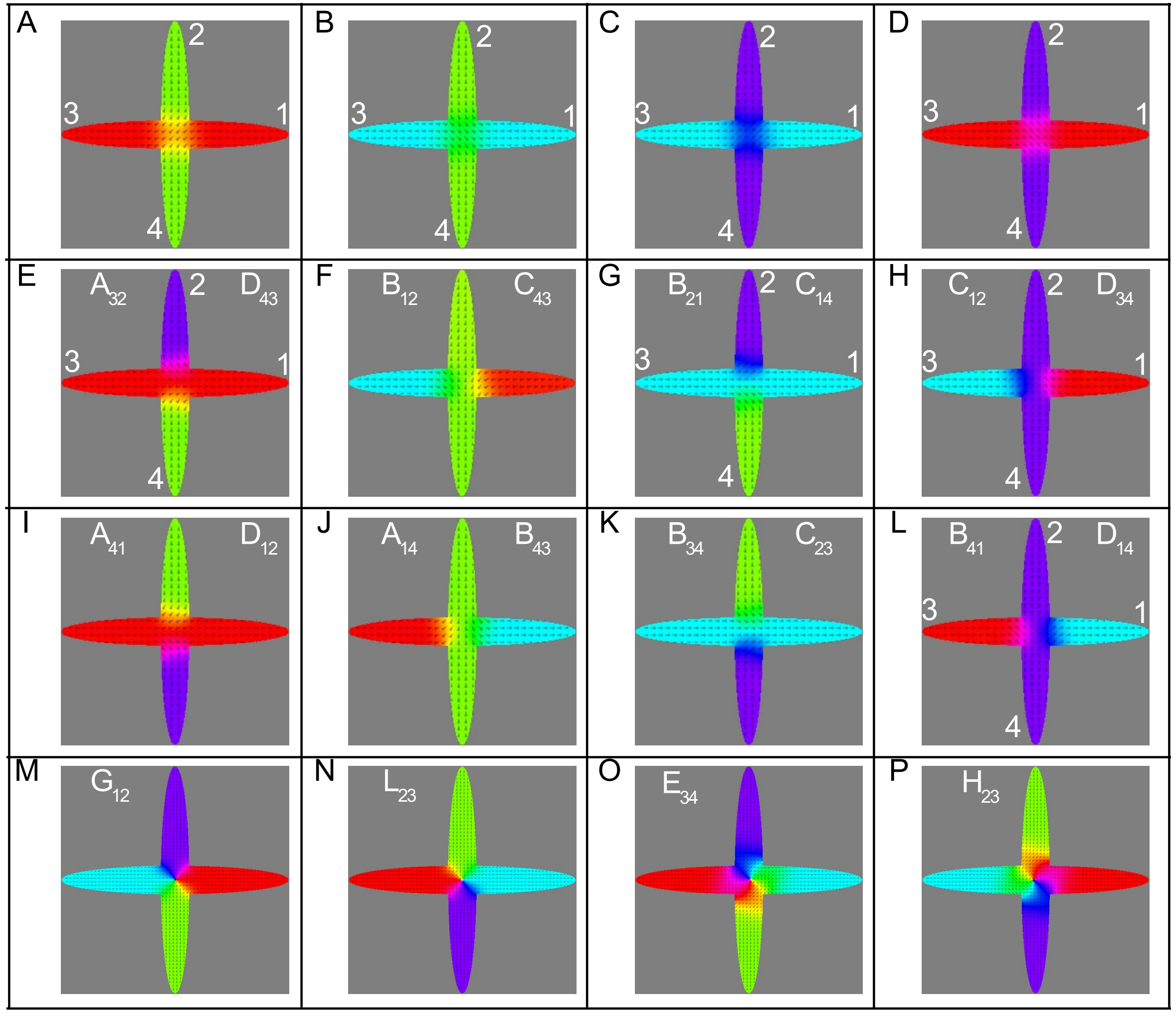


$2^{2*3} = 64$



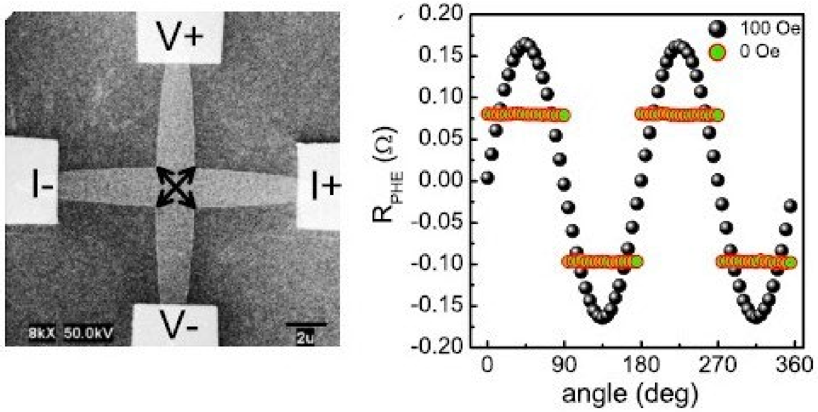
$2^{2*4} = 256$

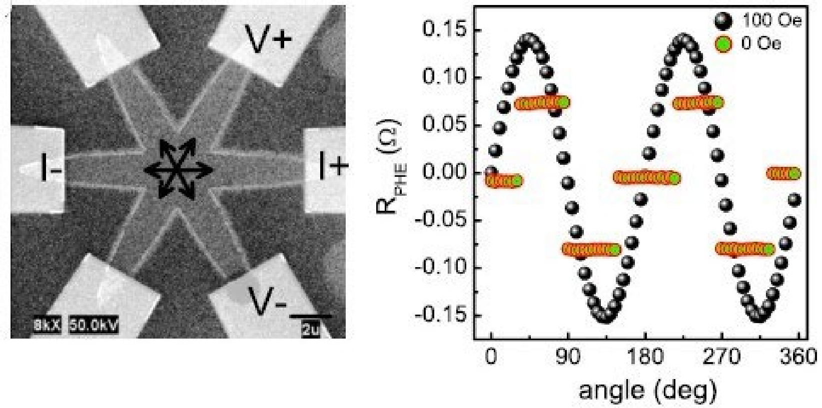
Stabilization of all 16 states

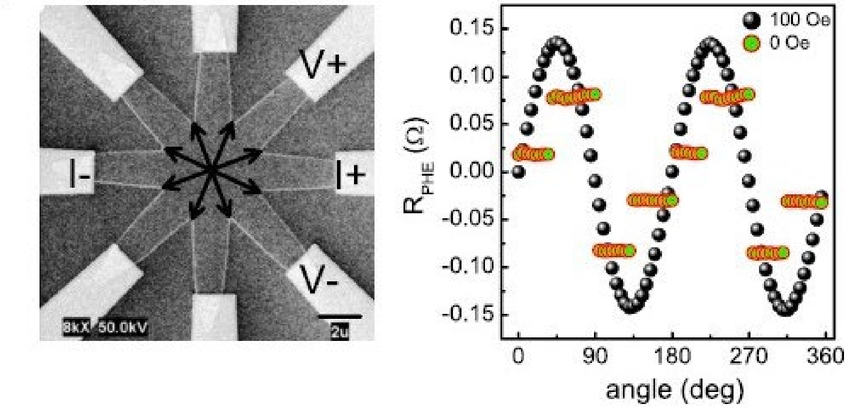


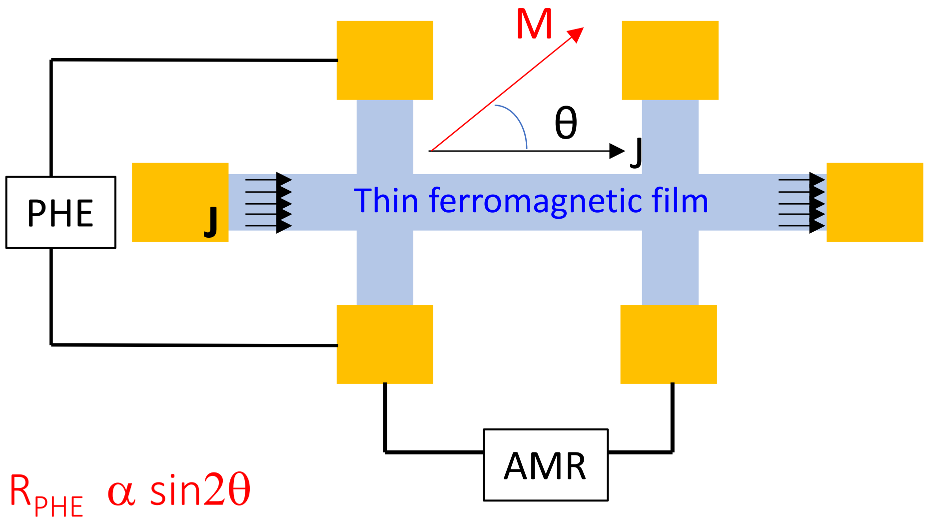
S. Das et al., Appl. Phys. Lett. **116,** 262405 (2020).

Experimental demonstration of high order magnetic anisotropy





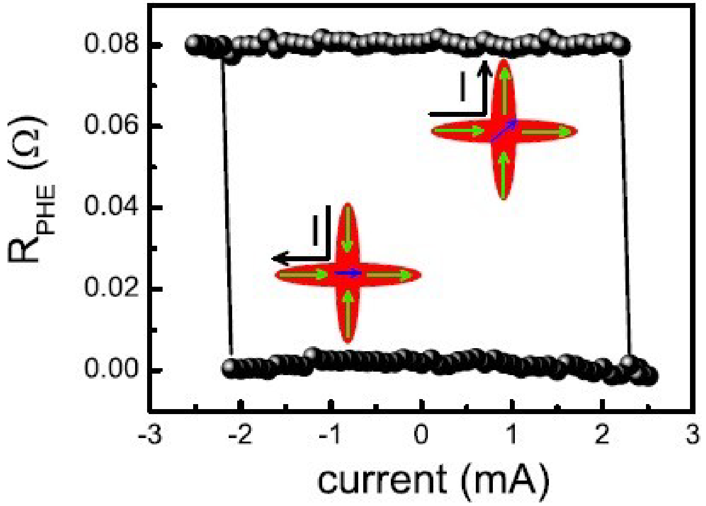


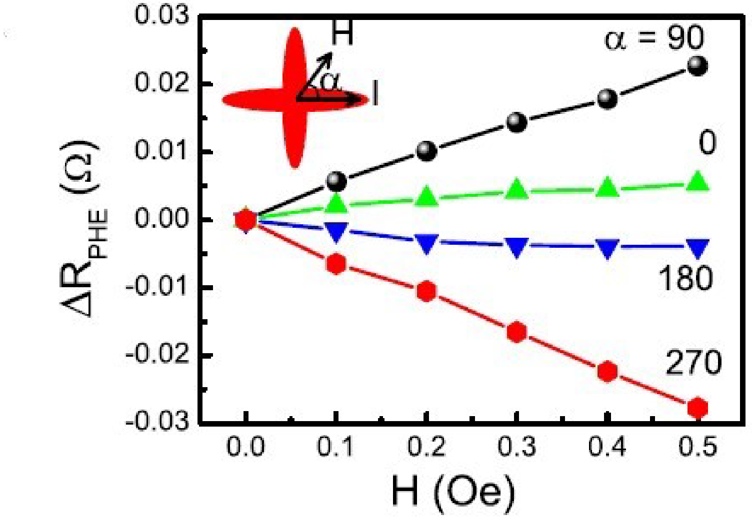


R_{PHE} α $\sin 2\theta$

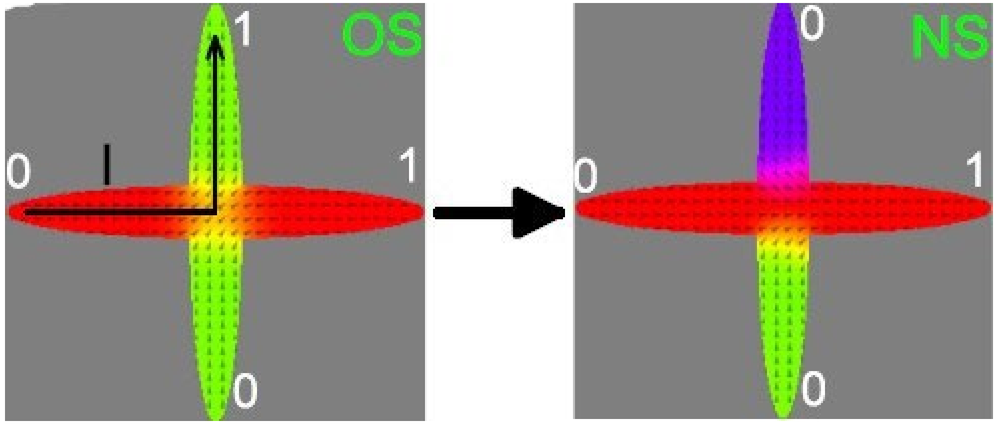
S. Das et al., Appl. Phys. Lett. **116,** 262405 (2020).

Magnetization manipulation with spin currents





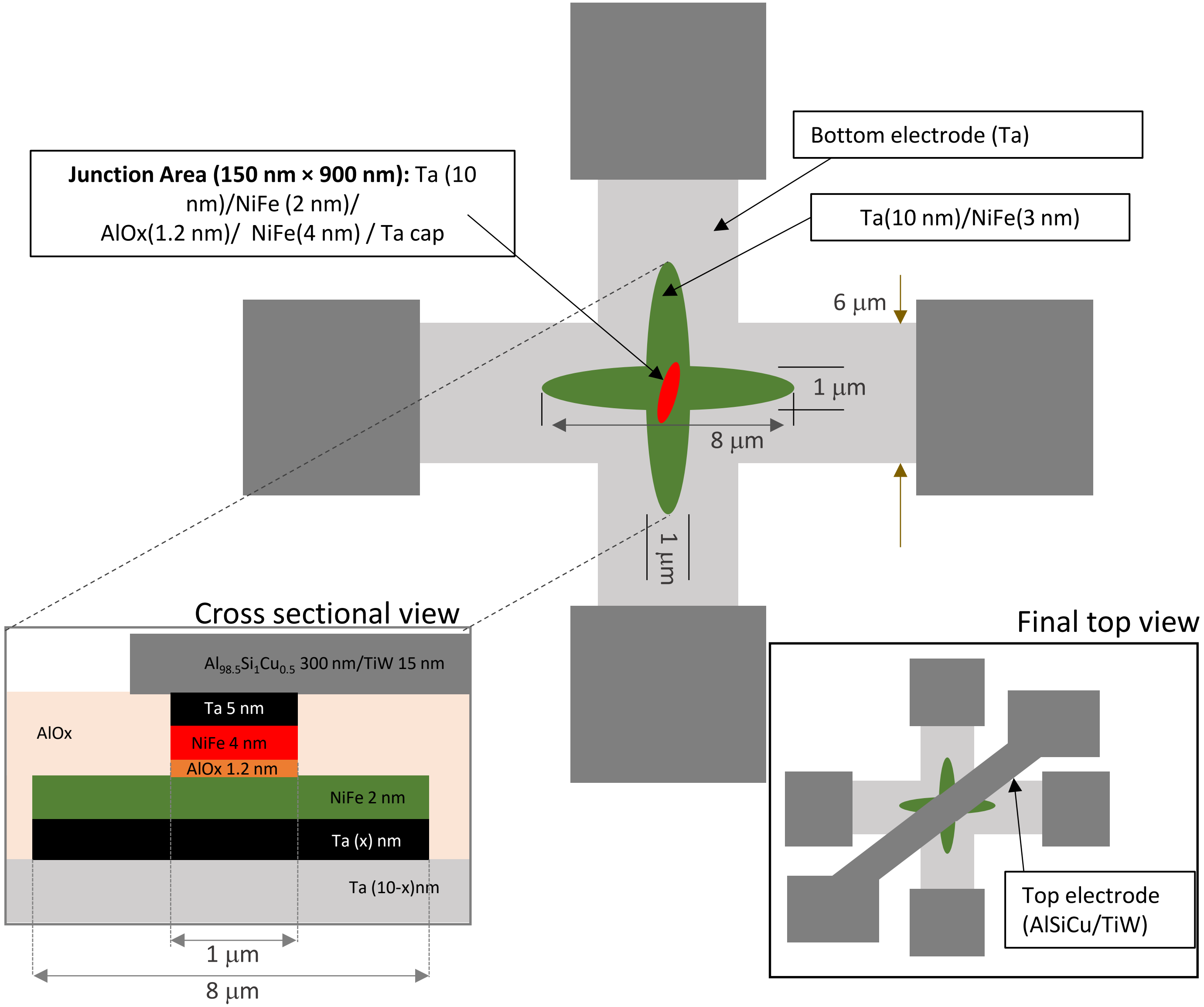
Switching with spin currents between an ordinary state and a Neel-like state



How to identify a Neel-like state?

S. Das et al., Appl. Phys. Lett. **116,** 262405 (2020).

Fabrication of M^2TJ



Junction Area (150 nm × 900 nm): Ta (10 nm)/NiFe (2 nm)/ AlOx(1.2 nm)/ NiFe(4 nm) / Ta cap

Bottom electrode (Ta)

Ta(10 nm)/NiFe(3 nm)

6 μm

1 μm

8 μm

1 μm

Cross sectional view

Al$_{98.5}$Si$_1$Cu$_{0.5}$ 300 nm/TiW 15 nm

Ta 5 nm

AlOx

NiFe 4 nm

AlOx 1.2 nm

NiFe 2 nm

Ta (x) nm

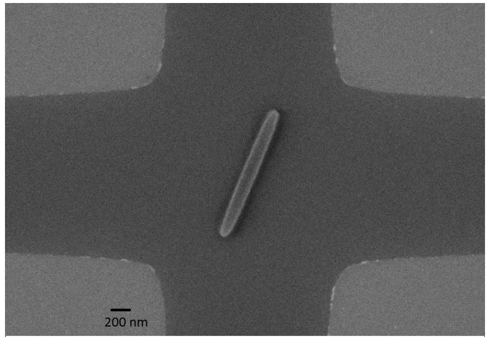
Ta (10-x)nm

1 μm

8 μm

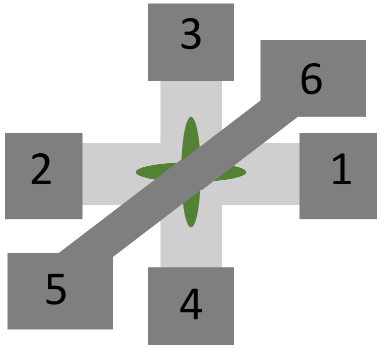
Final top view

Top electrode (AlSiCu/TiW)



S. Das et al., Appl. Phys. Lett. **117,** 072404 (2020)

Characterization of M^2TJ

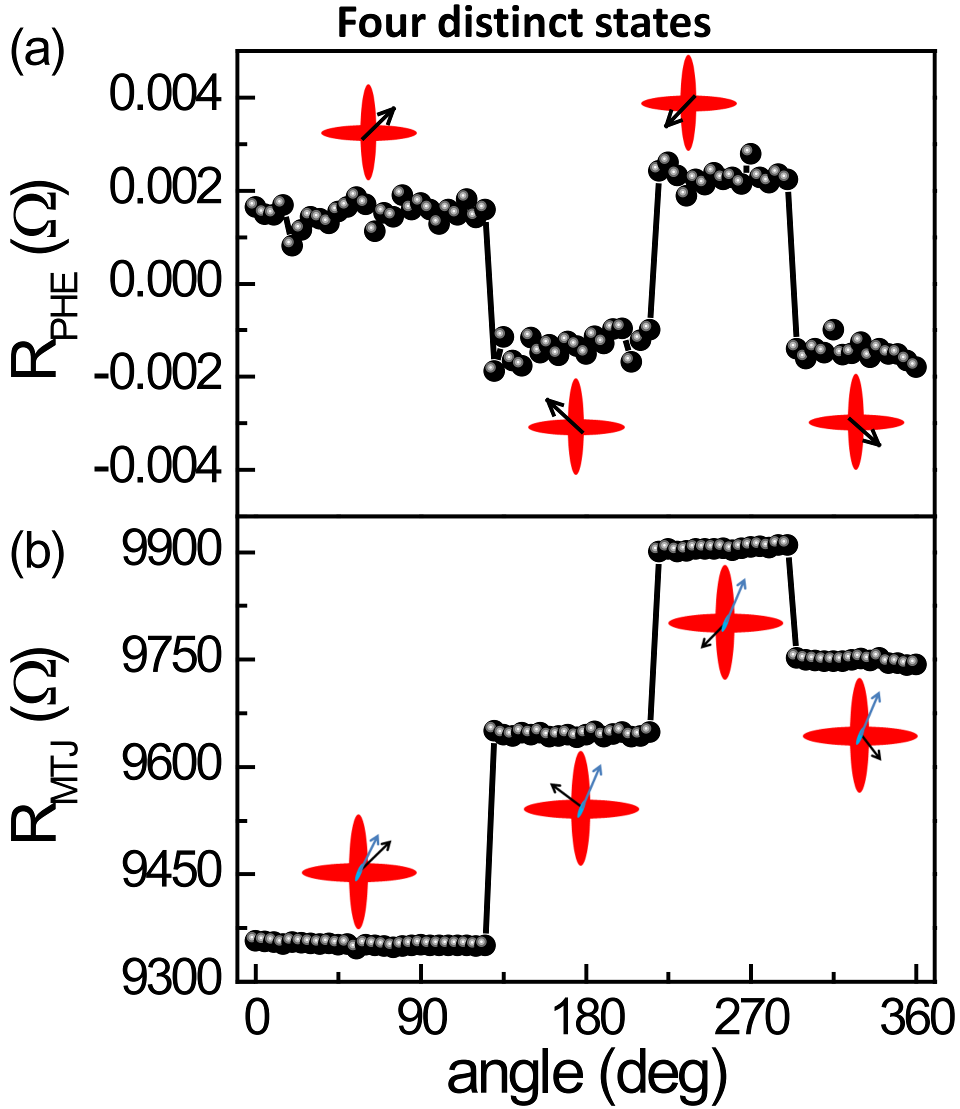


PHE: I+,I- : V+,V- = 1,2 : 3,4

MTJ: I+,I- : V+,V- = 6,1 : 5,2

$$R_{MTJ}(\beta) = R_{AV} - a\cos(\beta - \beta_0)$$

$$R_{PHE}(\beta) = b0 + b\sin(2\beta)$$

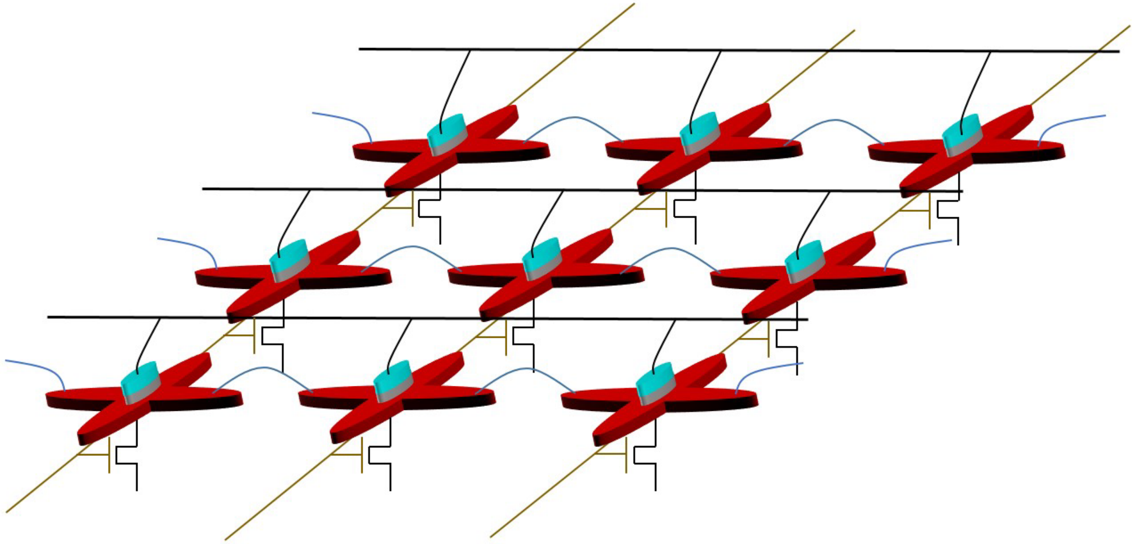


S. Das et al., Appl. Phys. Lett. **117,** 072404 (2020)

MultiSpin.AI

- EIC Pathfinder Open

The team, combining researchers from academia and industry, will develop innovative neuromorphic computer hardware to be used for AI computation that will be based, among other, on patents from Prof. Klein's laboratory in the field of Spintronics.



Lior Klein - coordinator

Susana cardoso de Freitas

Flavio Abreu Araujo

Bar-Ilan University

University

inesc mn

Private, non-for-profit research institute

UCLouvain

University

SpinEdge

SpinEdge AI accelerator – Spintronic analog instant AI

I-FEVS

Developing fully electric vehicles for sustainable urban mobility

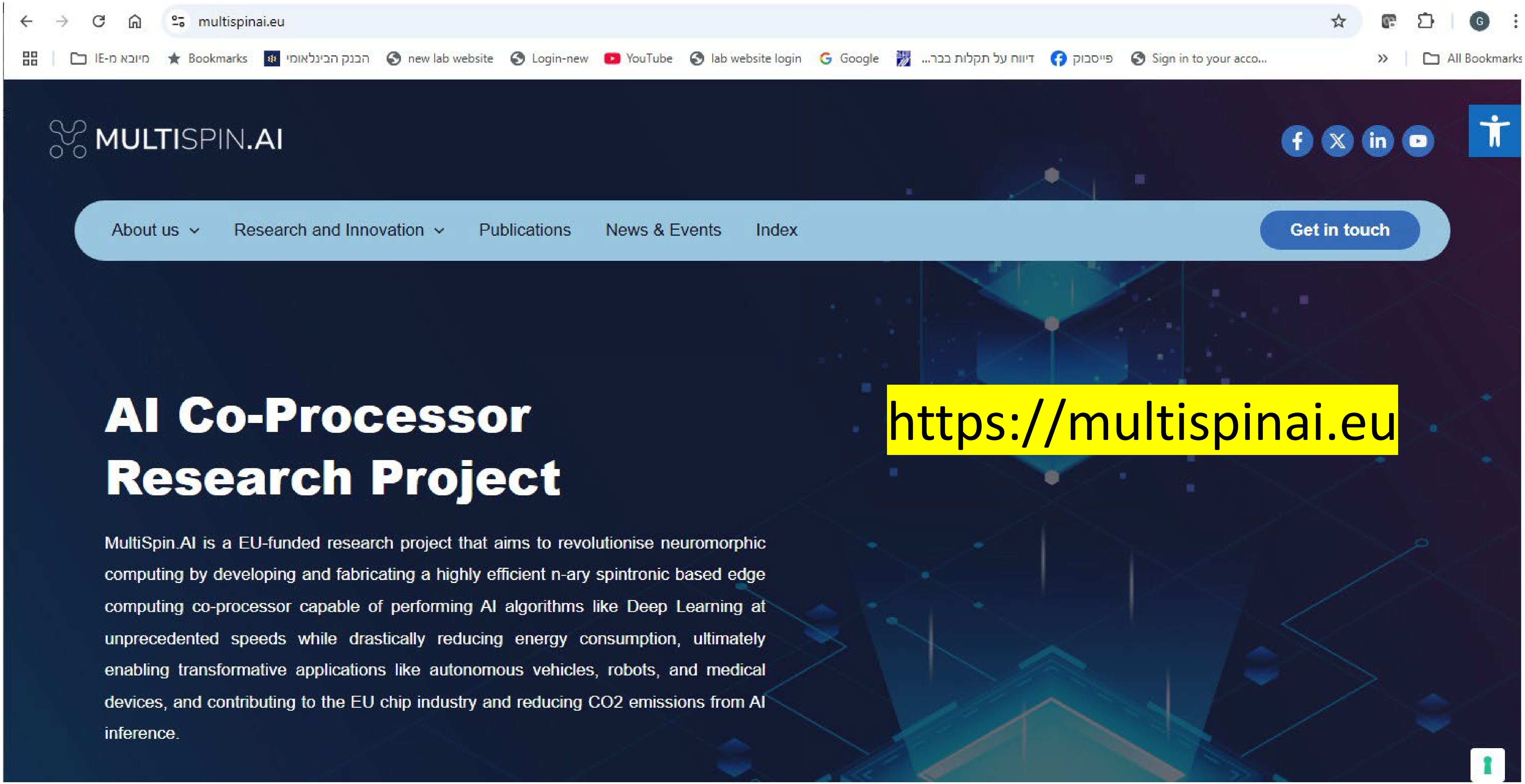
AMIRES

Supporting strategically oriented innovation for business and societal impact

Konstantin Zvezdin

Pietro Perlo

Marina de Souza Faria



AI Co-Processor Research Project

MultiSpin.AI is a EU-funded research project that aims to revolutionise neuromorphic computing by developing and fabricating a highly efficient n-ary spintronic based edge computing co-processor capable of performing AI algorithms like Deep Learning at unprecedented speeds while drastically reducing energy consumption, ultimately enabling transformative applications like autonomous vehicles, robots, and medical devices, and contributing to the EU chip industry and reducing CO2 emissions from AI inference.

https://multispinai.eu

Thank you!

Visit our website:

Follow us on social media:

- multispin.ai
- MultiSpin.AI
- MultiSpinAI
- MultiSpin.AI