Magnetic memories from a broad IT, materials, and physics perspectives

T. Jungwirth
Institute of Physics, Czech Academy of Sciences
University of Nottingham, United Kingdom
jungw@fzu.cz

1. Recording & computers
2. Conventional & neuromorphic computing

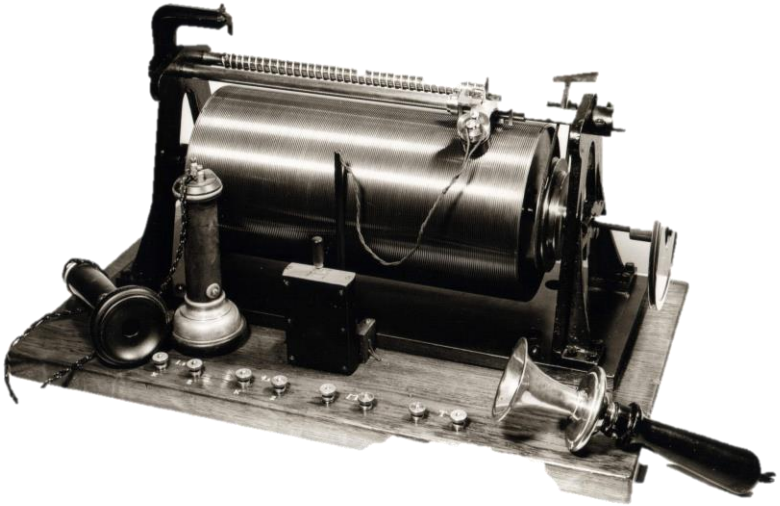
Lecture I

3. Non-CMOS materials and devices
4. Physical principles of operation of magnetic devices

Lecture II

Recording

Mechanical gramophone *1870‘s*



Electro-magnetic wire recorder *1880‘s*

Boom, bust, boom,…

Recording & computers

sound & video

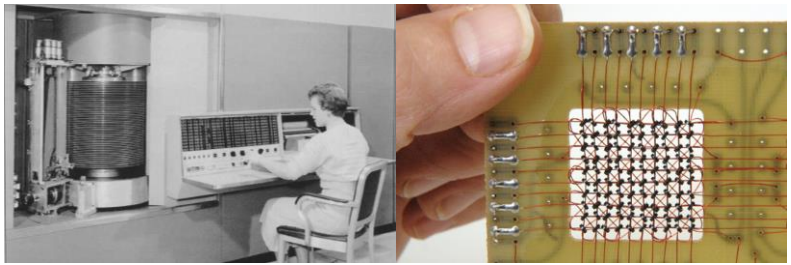
1930's Tape recorder

1980's Compact disk
(100sMB-GBs)

- **Analog to digital**
- **Phase-change**
- **Optical**

1990's - **Spintronic**

data

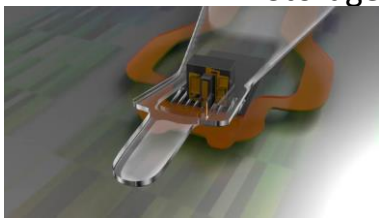


1950's Magnetic hard disk (MBs) & core memory (kbs)

Back-up

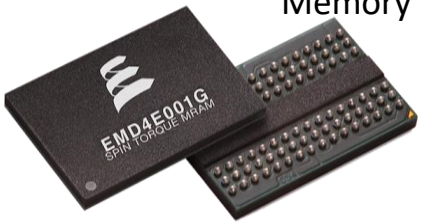
Sony/IBM tape (330TB)

Storage



Seagate HDD (16TB)

Memory



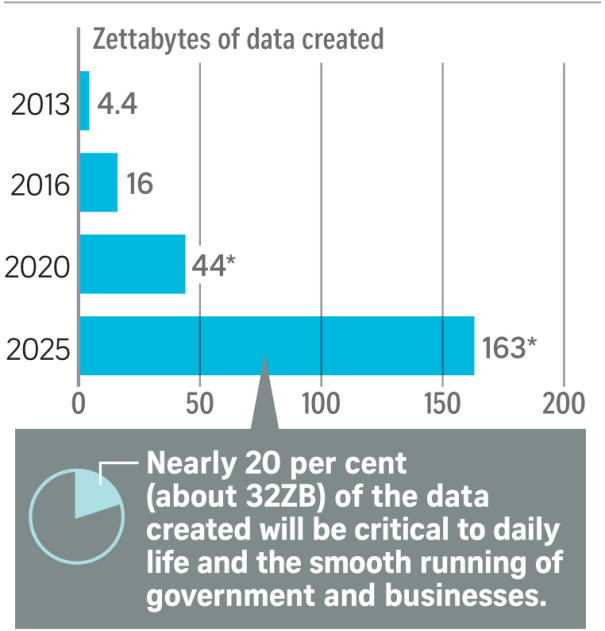
Everspin MRAM (1Gb)

Recording & computers

- Internet (PC & cloud IT)

- Internet of Things (edge IT)

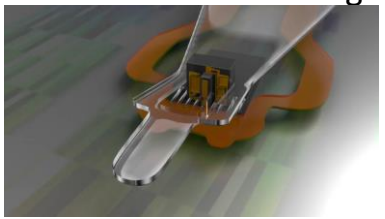
- Big data



Back-up

Sony/IBM tape (330TB)

Storage



Seagate HDD (16TB)

ZB = 10^9 TB = 10^9 people x 1TB mobile phone

Computer: PC & cloud IT

Microcontroller: edge IoT

Microprocessor

volatile

Non-volatile

CPU
(ALU,FPU)
SRAM
cache

1ns — Mb

- von Neumann

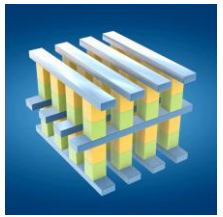
Main memory

10ns — DRAM — Gb

Storage

100µs — HDD, Flash-SSD — 10TB

Everspin
MRAM
10ns & 1Gb



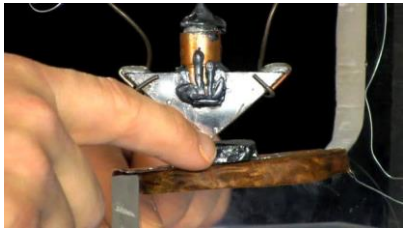
Intel/Micron
XPoint-PCRAM
1µs & 128Gb

MPU

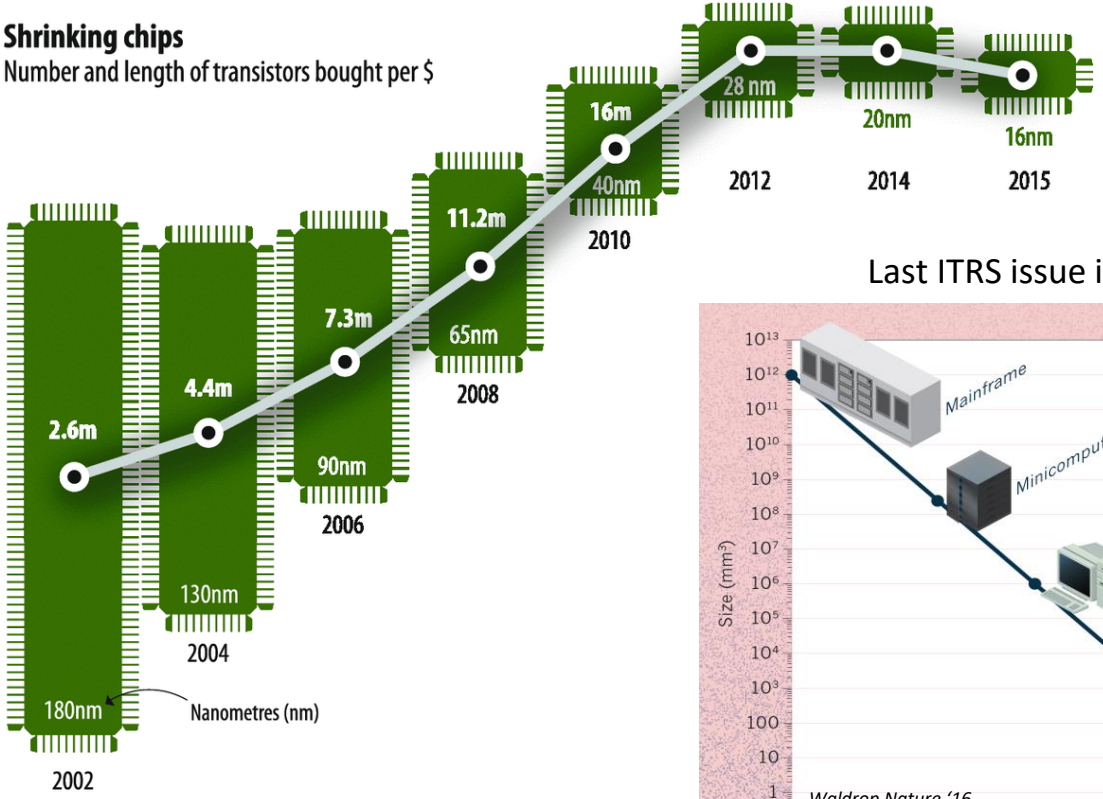
eFlash, eMRAM

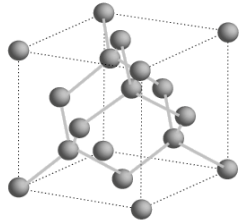
EEPROM	Adesto CBRAM **1µs & 512kb**	Panasonic/Fujitsu RRAM **10ms & 8Mb**	Fujitsu FRAM **100ns & 8Mb**

International technology roadmap for semiconductors

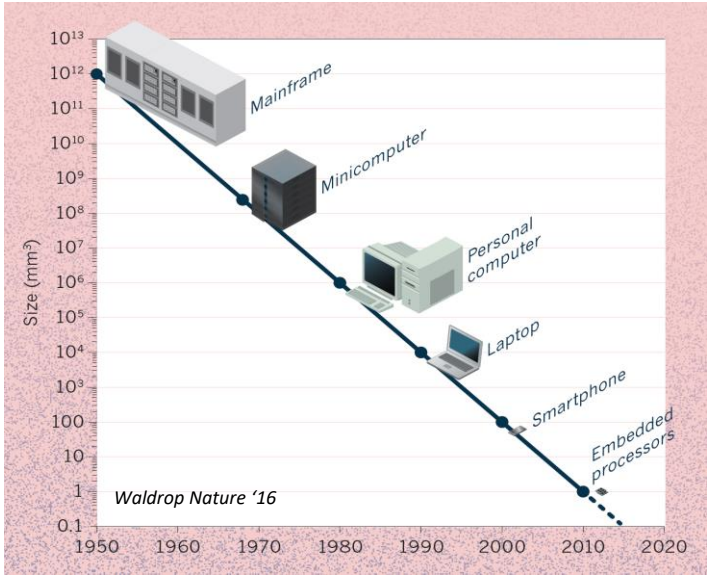


Transistor *1947*
Integrated circuit *1958*

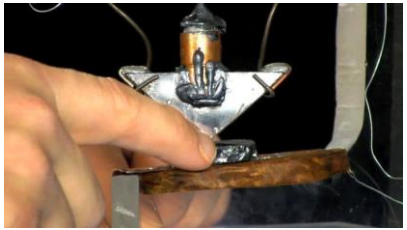




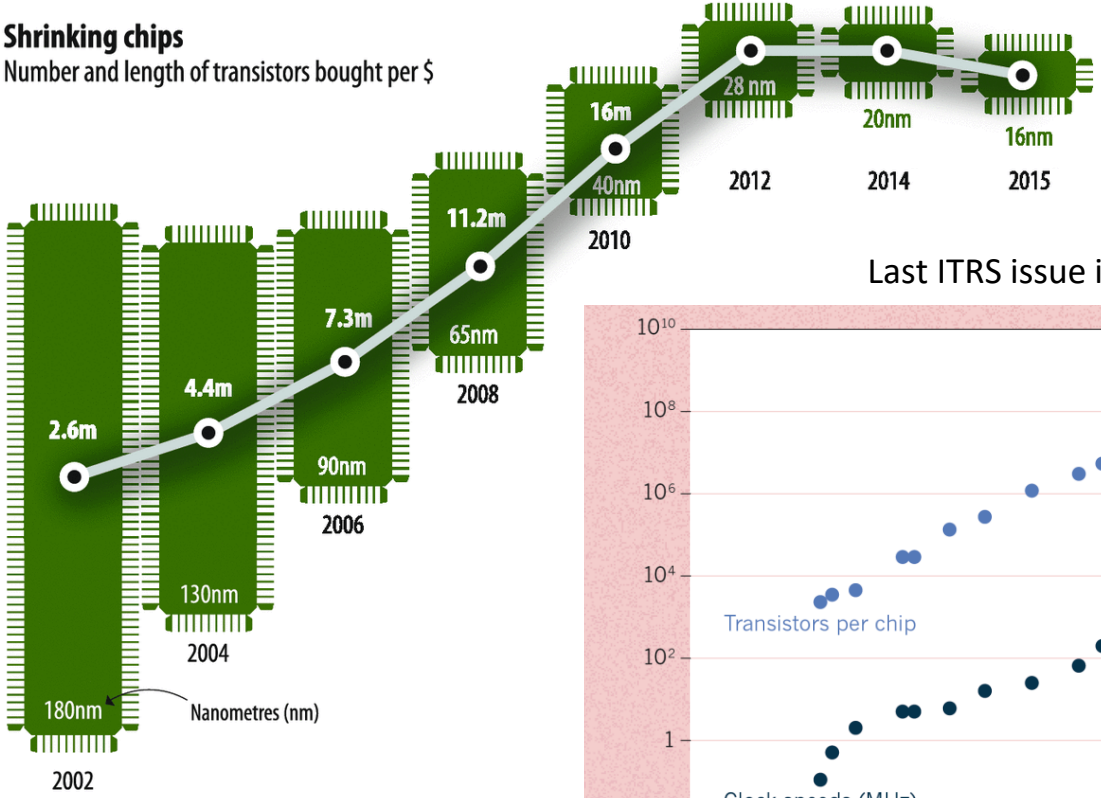
Last ITRS issue in 2016 - **No Moore**

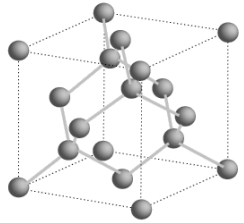


International technology roadmap for semiconductors



Transistor *1947*
Integrated circuit *1958*





Last ITRS issue in 2016 **- No Moore**

problem

- **Internet**
- **Internet of Things**
- **Big data**

- **No Moore**
  Speed
  Energy
  CMOS scaling

Let's racap

1. Recording & computers

2. Conventional & neuromorphic computing

3. Non-CMOS devices and materials

4. Physical principles of operation of magnetic devices

do differently

- **von Neumann**
  Revisit the architecture to tackle the bottleneck

- **Analog to digital**
  Revisit the noise vs. complexity trade-off

do more

- **Spintronic**
- **Phase-change**
- **….**
  Exploit full potential of non-CMOS devices

- **Optical**
  Explore speed and energy efficiency limits

problem

- **Internet**
- **Internet of Things**
- **Big data**

- **No Moore**
  Speed
  Energy
  CMOS scaling

Let's racap

1. Recording & computers
2. Conventional & neuromorphic computing
3. Non-CMOS devices and materials
4. Physical principles of operation of magnetic devices

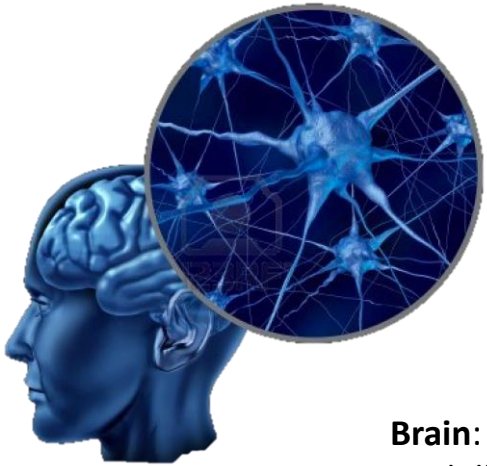
do differently

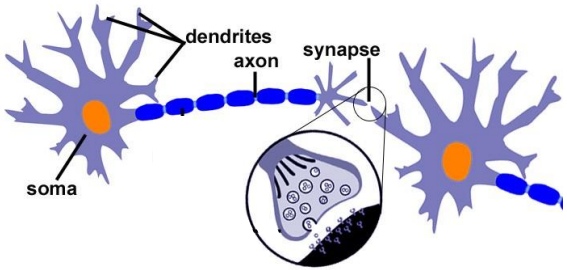
- **von Neumann**
  Revisit the architecture to tackle the bottleneck
- **Analog to digital**
  Revisit the noise vs. complexity trade-off

do more

- **Spintronic**
- **Phase-change**
- **….**
  Exploit full potential of non-CMOS devices
- **Optical**
  Explore speed and energy efficiency limits

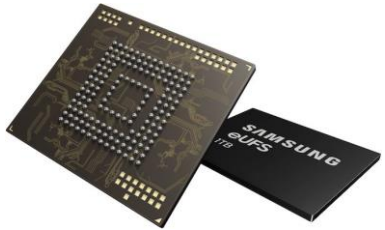
Neuromorphic architecture





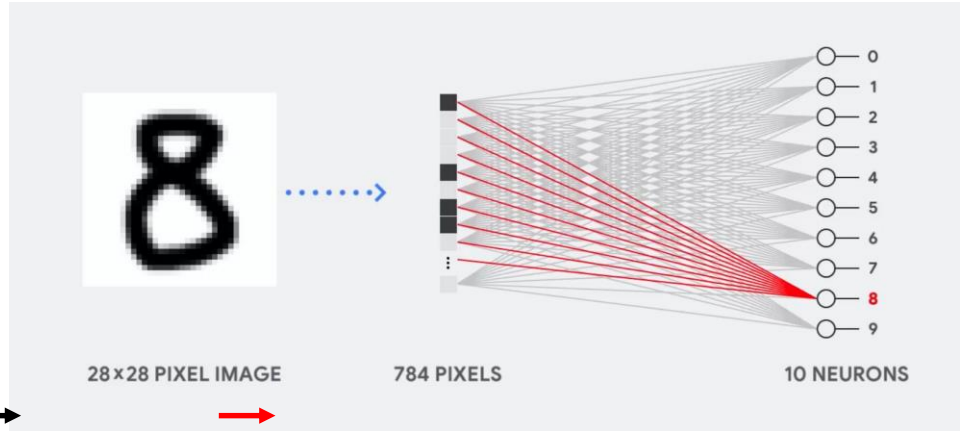
Brain: Massively **parallel** neural network architecture connecting 100billion low-power **computing and memory** elements

cf. Samsung 1TB Flash-SSD for smart phones with 2 trillion transistors on a chip



1.Synchronous: All components run under global clock
Artificial Neural Networks: Input/output/internal variables coded in real numbers

Mass applications – Google Brain (2012 – image recognition, 2016 – language translation)

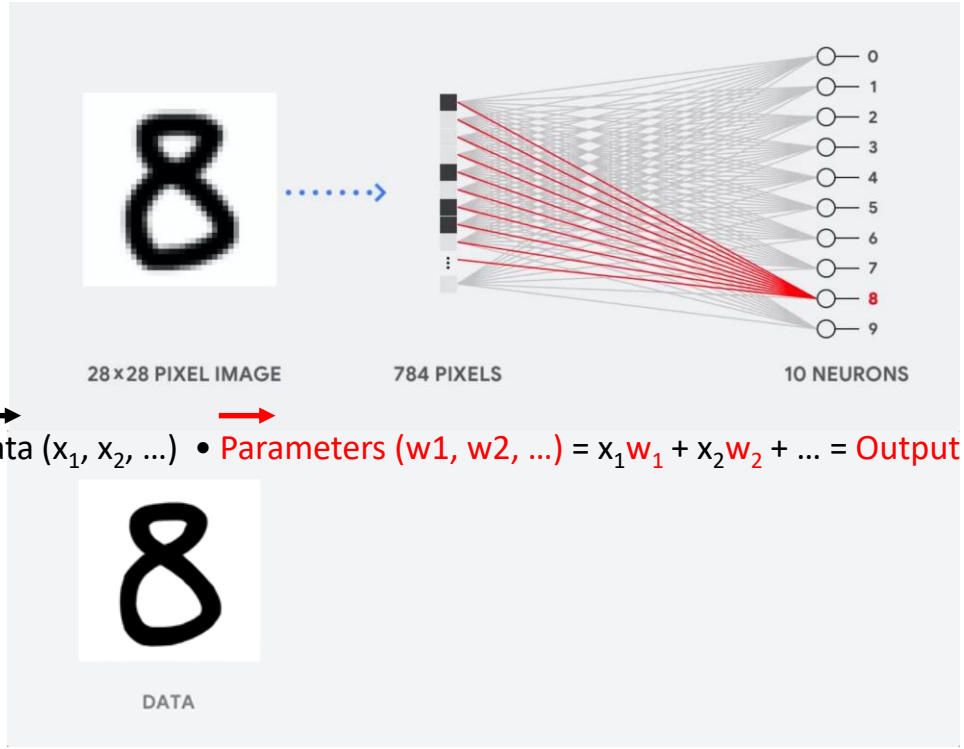


Data (x_1, x_2, …) • Parameters (w1, w2, …) = $x_1w_1 + x_2w_2 + \ldots$ = Output (y)

https://cloud.google.com/blog/products/ai-machine-learning/what-makes-tpus-fine-tuned-for-deep-learning

1.Synchronous: All components run under global clock
Artificial Neural Networks: Input/output/internal variables coded in real numbers

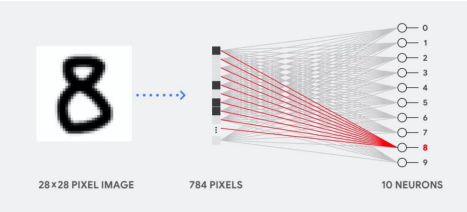
Mass applications – Google Brain (2012 – image recognition, 2016 – language translation)



Data (x_1, x_2, …) • Parameters (w1, w2, …) = $x_1w_1 + x_2w_2 + \ldots$ = Output (y)

1.Synchronous: All components run under global clock
Artificial Neural Networks: Input/output/internal variables coded in real numbers

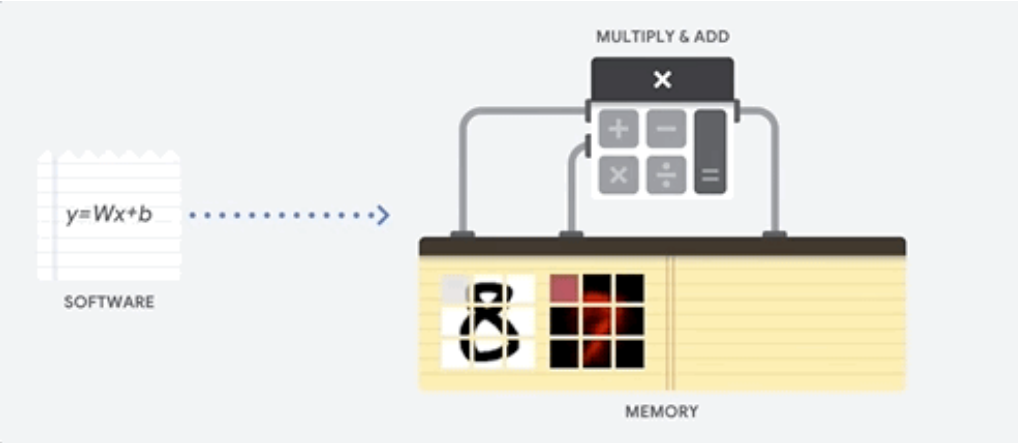
Mass applications – Google Brain (2012 – image recognition, 2016 – language translation)



1.1 Off-shelf

General purpose **CPU** (Intel,…)

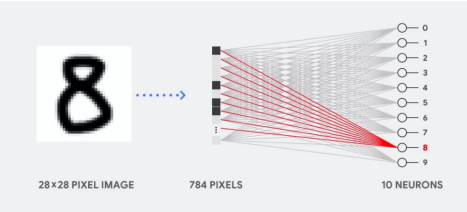
1 or few big cores



Serial & von Neumann bottleneck

1.Synchronous: All components run under global clock
Artificial Neural Networks: Input/output/internal variables coded in real numbers

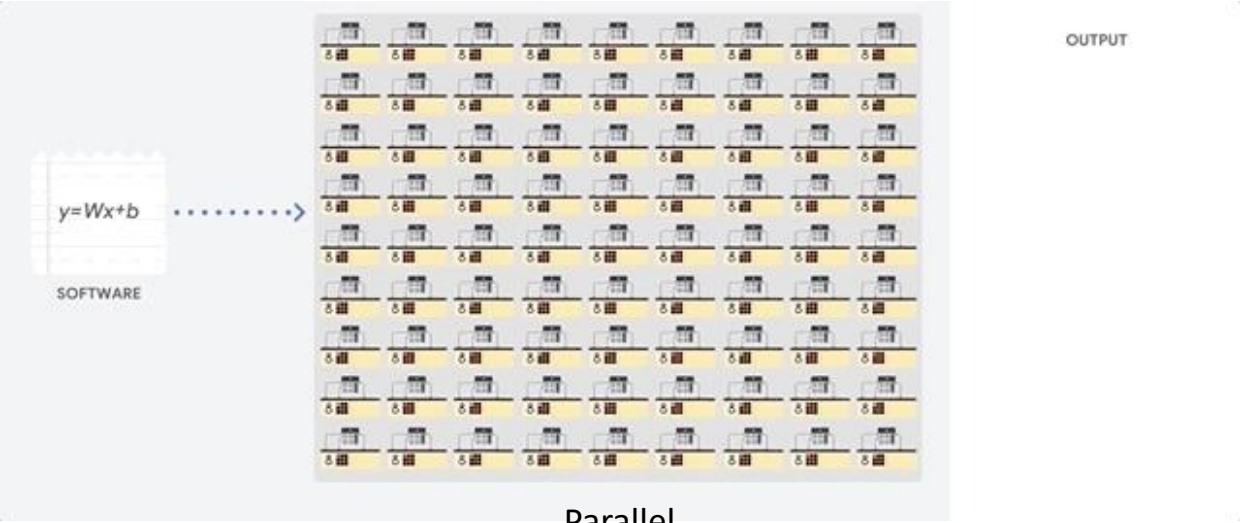
Mass applications – Google Brain (2012 – image recognition, 2016 – language translation)



1.1 Off-shelf

General purpose **GPU** (NVIDIA, AMD)

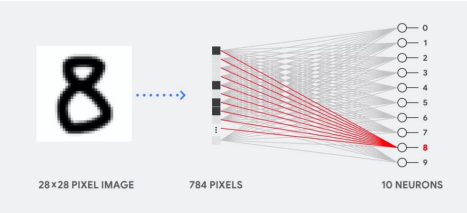
2,000 medium cores



Parallel

1.Synchronous: All components run under global clock
Artificial Neural Networks: Input/output/internal variables coded in real numbers

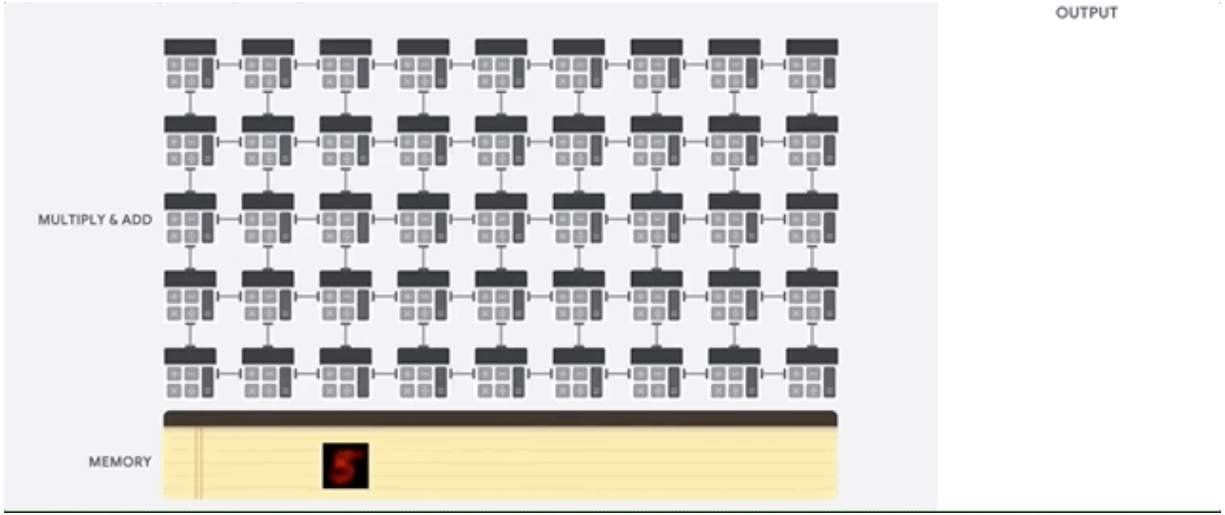
Mass applications – Google Brain (2012 – image recognition, 2016 – language translation)



1.2 Custom-designed

Neuromorphic **TPU** (Google)

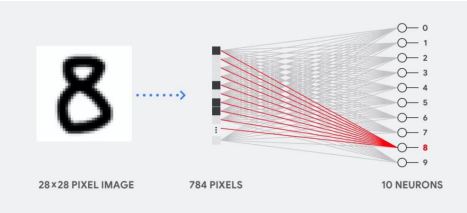
30,000 small cores



Highly parallel & tackles von Neumann bottleneck

1.Synchronous: All components run under global clock
Artificial Neural Networks: Input/output/internal variables coded in real numbers

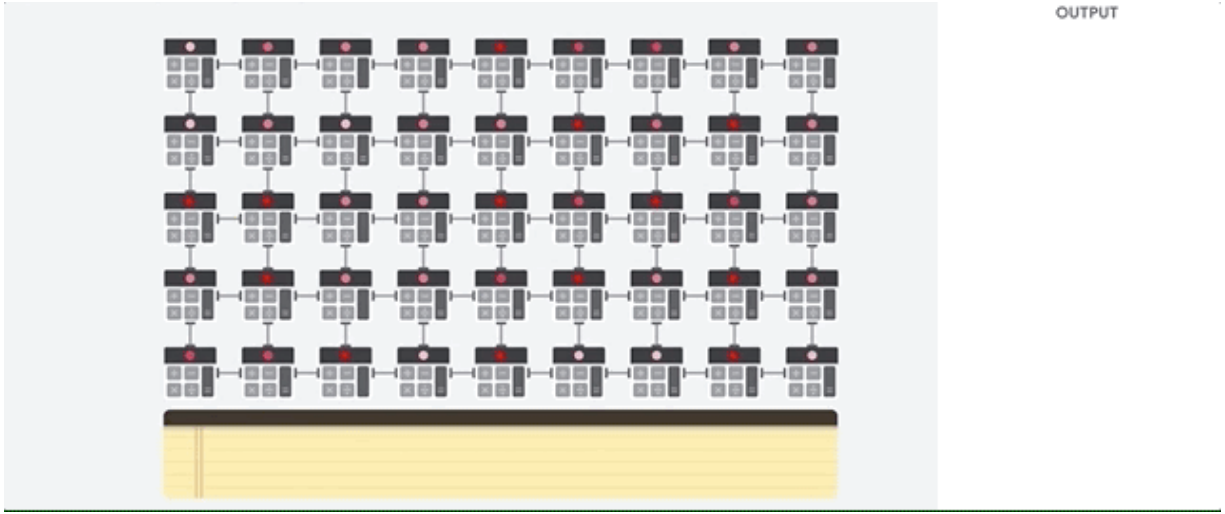
Mass applications – Google Brain (2012 – image recognition, 2016 – language translation)



1.2 Custom-designed

Neuromorphic **TPU** (Google)

30,000 small cores



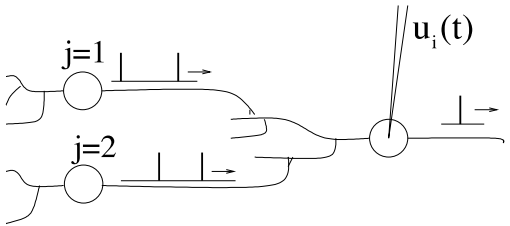
Highly parallel & tackles von Neumann bottleneck

2. Asynchronous: Individual components have local clocks

Spiking Neural Networks: Input/output/internal variables coded in spikes and their timing

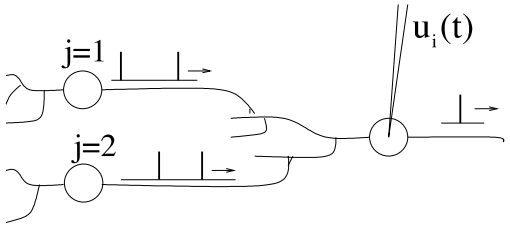
Still mostly in research phase

- Short-term: Save bandwidth & energy

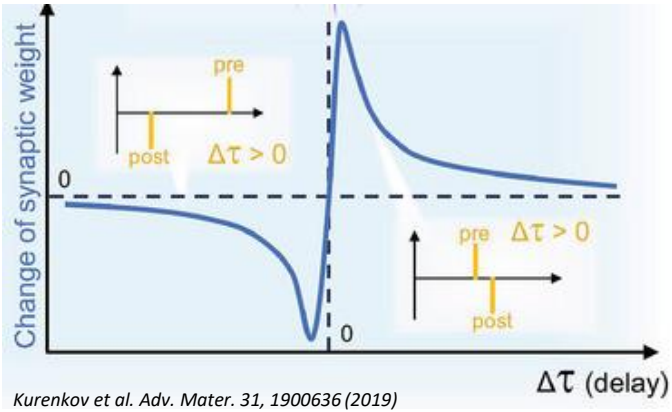


2. Asynchronous: Individual components have local clocks
Spiking Neural Networks: Input/output/internal variables coded in spikes and their timing

Still mostly in research phase
- Long term: Help understand neuroscience, develop General Artificial Intelligence

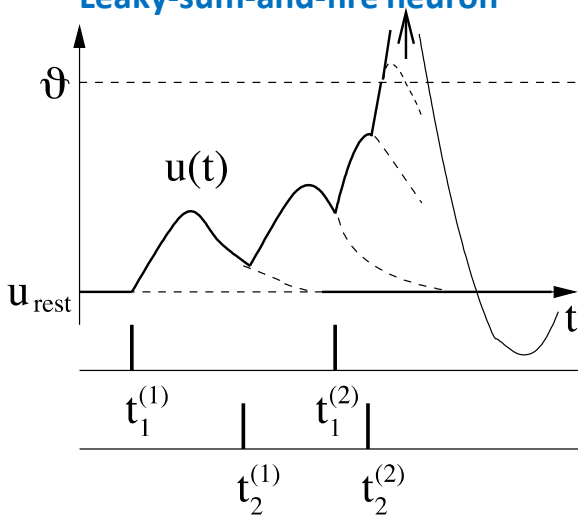


Spiking time dependent plasticity of synapse
("neurons that fire together wire together")



Kurenkov et al. Adv. Mater. 31, 1900636 (2019)

Leaky-sum-and-fire neuron



Gerstner & Kistler, Spiking Neuron Models, Cambridge University Press (2002)

2. Asynchronous: Individual components have local clocks

Spiking Neural Networks: Input/output/internal variables coded in spikes and their timing

2.1 Off-shelf based

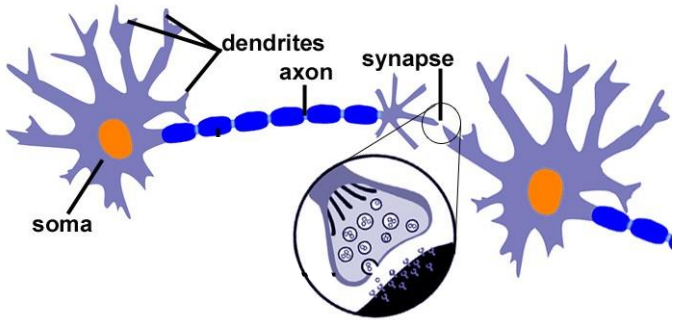
FPGA DeepSouth (Sydney Univ.)

Mobile CPU SpiNNaker (Manchester Univ.)

Thakur et al. Frontiers in Neuroscience 12, 891(2018)

2.2 Custom-designed CMOS

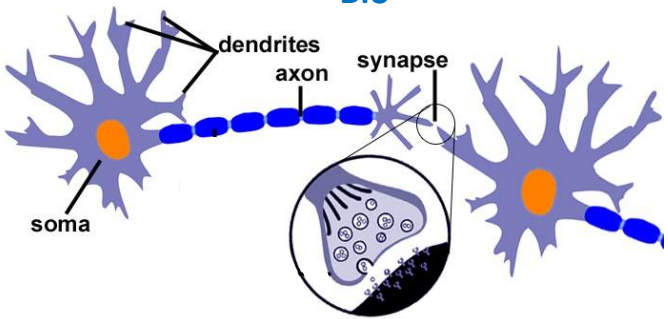
Individual circuit components mimic bio-neuron structure and functions



2. Asynchronous: Individual components have local clocks

Spiking Neural Networks: Input/output/internal variables coded in spikes and their timing

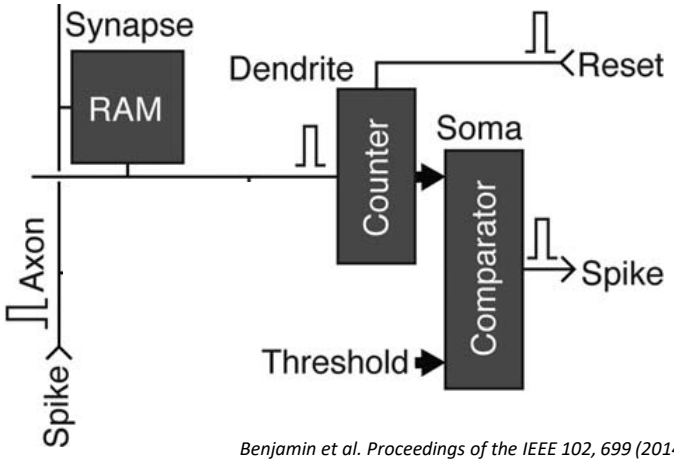
Bio



2.2 Custom-designed CMOS

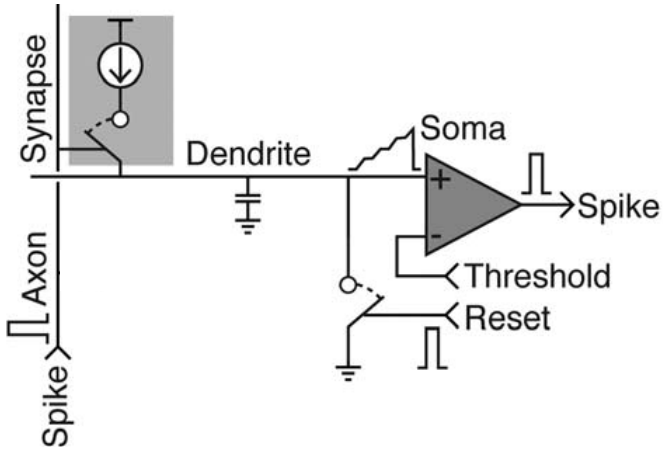
Individual circuit components mimic bio-neuron structure and functions

CMOS digital



Benjamin et al. Proceedings of the IEEE 102, 699 (2014)

CMOS analog



2. Asynchronous: Individual components have local clocks

Spiking Neural Networks: Input/output/internal variables coded in spikes and their timing

2.2.1 CMOS digital

TrueNorth (IBM) – 1M neurons

Low-power execution

Learning done externally

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

Loihi (Intel) – 100k neurons

Includes learning

2.2.2 CMOS mixed digital/analog

Neurogrid (Stanford) – 60k neurons

Dynap-SEL (Zurich Univ.) – 1000 neurons

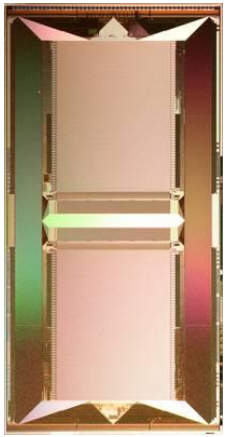
HICANN (Heidelberg Univ.) – 500 neurons

Benjamin et al. Proceedings of the IEEE 102, 699 (2014)

Digital communication

Analog neuron

Analog synapse with weights stored in digital RAM



Reviews:

Thakur et al. Frontiers in Neuroscience 12, 891(2018)

Yu (ed.), Neuro-inspired Computing Using Resistive Synaptic Devices, Springer (2017)

Burr et al. Adv. Phys. X 2, 89 (2017)

3. Mixed CMOS/non-CMOS

3.1 Analog memristive synapse

CBRAM (Michigan Univ.)

Jo et al. Nano Lett., 10, 1297 (2010)

RRAM (Pohang Univ.)

Moon et al. Nanotechnology 25, 495204 (2014)

PCRAM (IBM)

Eryilmaz et al. Frontiers in Neuroscience 8, 205(2014)

FRAM (Panasonic)

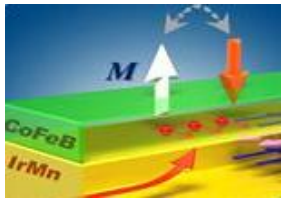
Ueda et al. PLOS ONE 9, e112659 (2014)

MRAM (Tohoku Univ.)

Borders et al. Appl. Phys. Exp. 10, 013007 (2017)

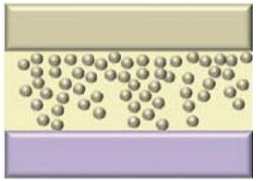
with SW or CMOS HW neurons

Ferromagnetic domains



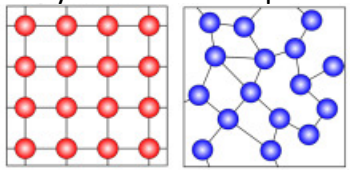
MRAM

Defects in insulator



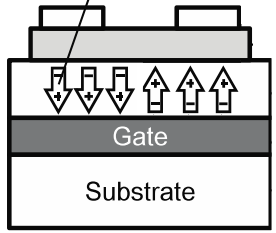
CBRAM/RRAM

Crystalline Amorphous



PCRAM

Ferroelectric domains



FRAM

3.2 Analog memristive synapse & neuron

Analog PCRAM (IBM)

Pantazi et al. Nanotechnology 27, 355205 (2016)

Spiking NN

Analog MRAM (Tohoku Univ.)

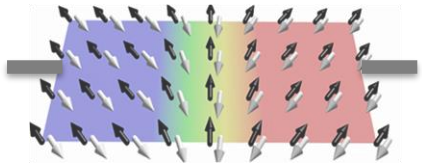
Kurenkov et al. Adv. Mater. 31, 1900636 (2019)

Analog AFMEM (Prague/Nottingham/Mainz/…)

Discrete synapse or neuron

Kaspar et al. preprint (2019)

Antiferromagnetic domains



AFMEM

Reviews:

Thakur et al. Frontiers in Neuroscience 12, 891(2018)

Yu (ed.), Neuro-inspired Computing Using Resistive Synaptic Devices, Springer (2017)

Burr et al. Adv. Phys. X 2, 89 (2017)

3. Mixed CMOS/non-CMOS

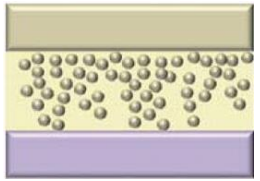
3.3 Analog memristive weighted-sum (dot product) array

RRAM passive array (UCSB)

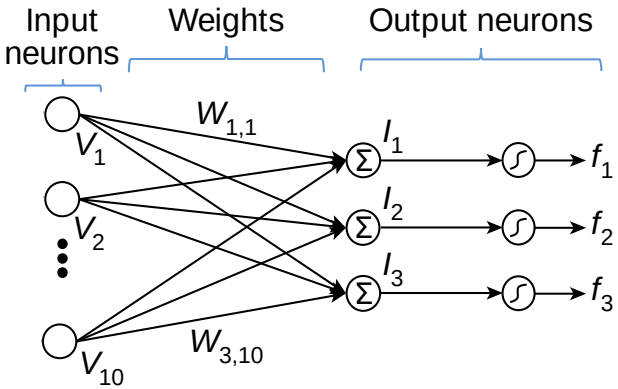
Prezioso et al. Nature 521, 61 (2015)

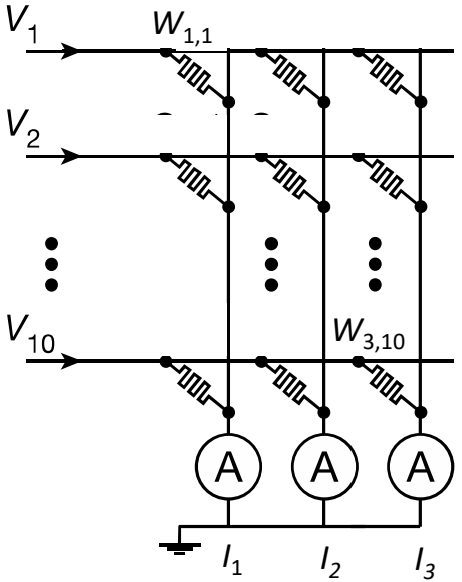
RRAM 1T1R array (Mass. Univ., HP)

Hu et al. Nature Elec. 1, 52 (2018)



RRAM

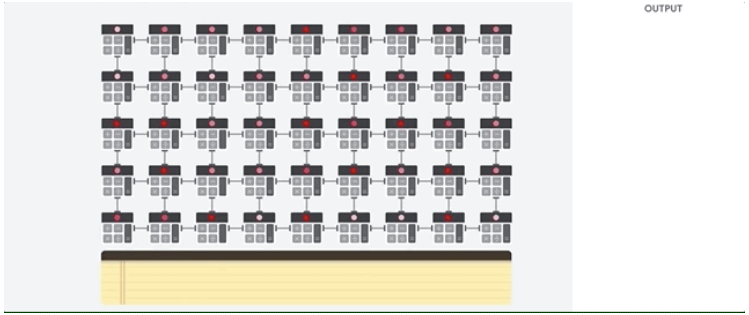




Dot product (weighted sum) → Kirchhoff's rule

Non-CMOS vs. CMOS for neuromorphics

Digital artificial neural networks for cloud IT
Compete with Google



Analog spiking neural devices for edge IoT
More realistic R&D start

problem

- **Internet**
- **Internet of Things**
- **Big data**

- **No Moore**
  - Speed
  - Energy
  - CMOS scaling

Let's racap

1. Recording & computers
2. Conventional & neuromorphic computing
3. Non-CMOS devices and materials
4. Physical principles of operation of magnetic devices

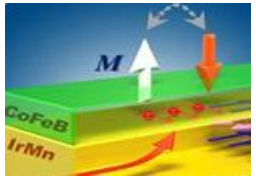
do differently

- **von Neumann**
  - Revisit the architecture to tackle the bottleneck
- **Analog to digital**
  - Revisit the noise vs. complexity trade-off

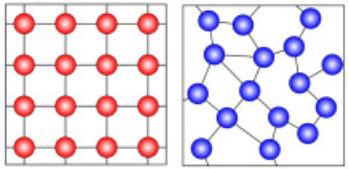
do more

- **Spintronic**
- **Phase-change**
- **….**
  - Exploit full potential of non-CMOS devices
- **Optical**
  - Explore speed and energy efficiency limits

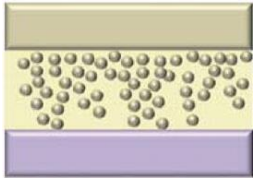
Non-CMOS memristive materials



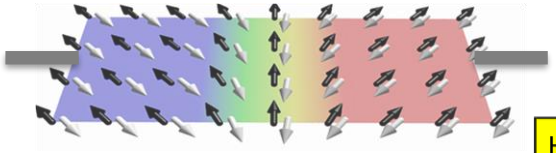
MRAM Metal



PCRAM Metalloid



CBRAM, RRAM Nonmetal



AFMEM

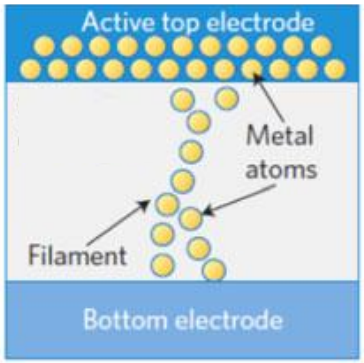
H																	He
Li	Be											B	C	N	O	F	Ne
Na	Mg											Al	Si	P	S	Cl	Ar
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
Cs	Ba	La-Lu	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
Fr	Ra	Ac-Lr															

La	Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
Ac	Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr

Conductive-Bridging RAM

EEPROM	Adesto CBRAM **1µs & 512kb**

Filamentary

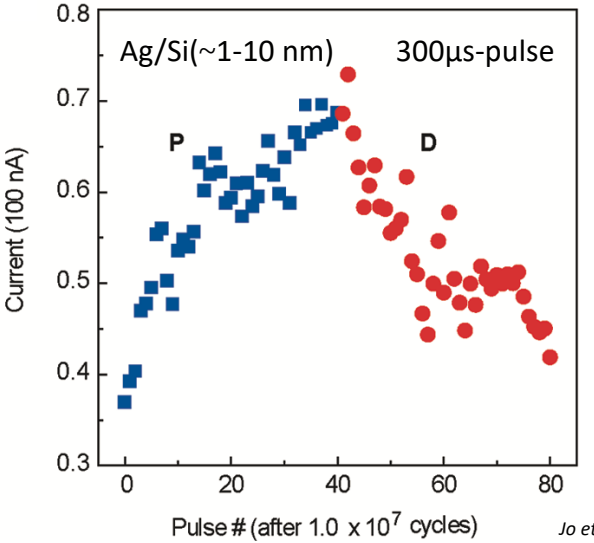


Bipolar switching

Micron/Sony 16Gb, 10µs prototype binary CBRAM

Yu (ed.), Neuro-inspired Computing Using Resistive Synaptic Devices, Springer (2017)
Burr et al. Adv. Phys. X 2, 89 (2017)

Analog synapse

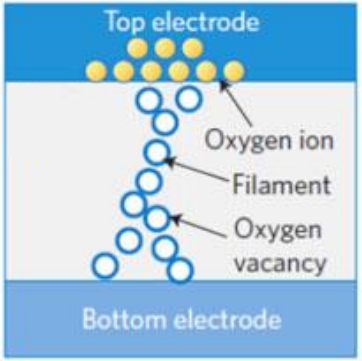


Jo et al. Nano Lett., 10, 1297 (2010)

Resistive RAM

EEPROM	Panasonic/Fujitsu RRAM **10ms & 8Mb**

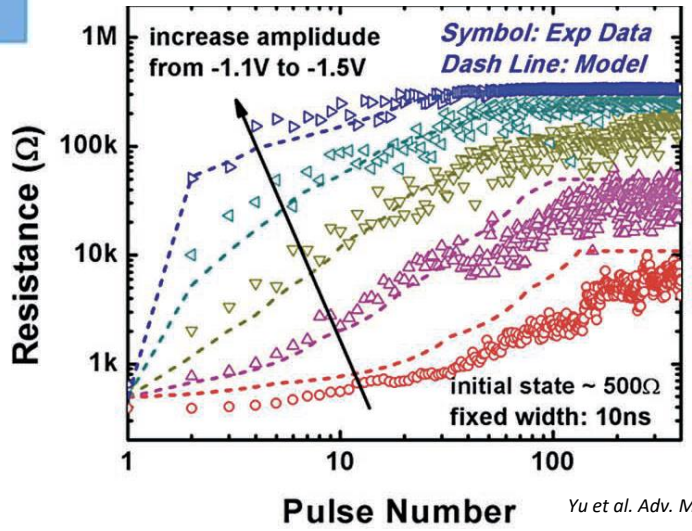
Filamentary



Bipolar switching
SanDisk/Toshiba 32Gb, 230µs, prototype binary RRAM

Yu (ed.), Neuro-inspired Computing Using Resistive Synaptic Devices, Springer (2017)
Burr et al. Adv. Phys. X 2, 89 (2017)

Analog synapse

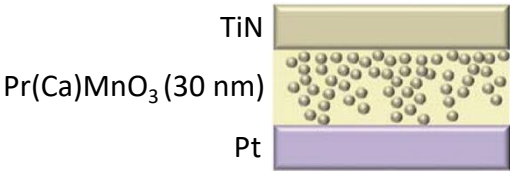


HfO_x (4 nm)

Yu et al. Adv. Mater. 25, 1774 (2013)

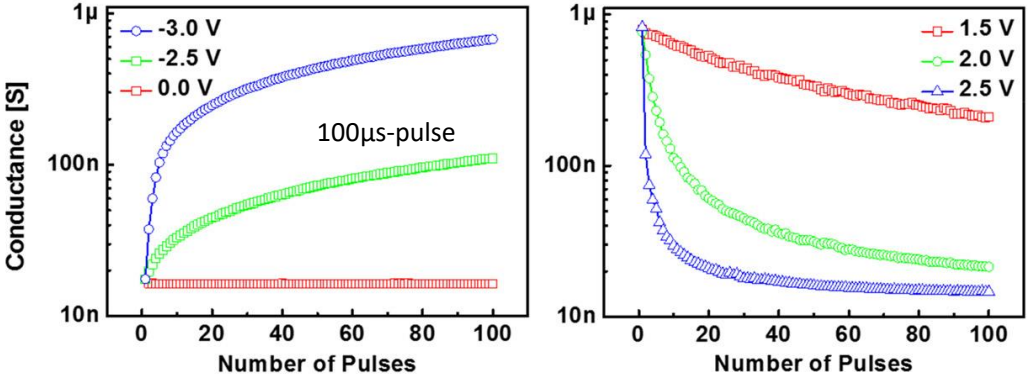
Resistive RAM

Non-filamentary



Bipolar switching

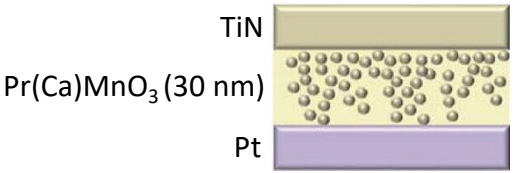
Analog synapse



Moon et al. Nanotechnology 25, 495204 (2014)

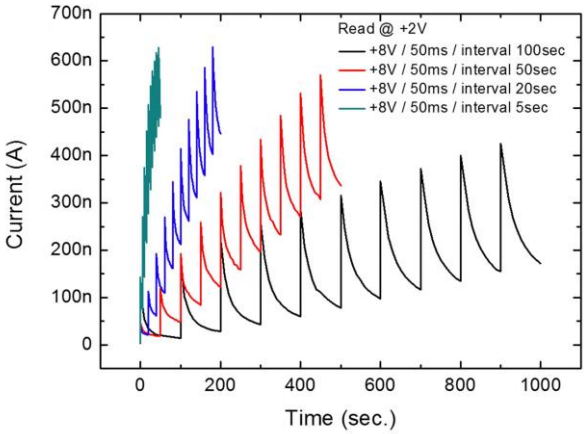
Resistive RAM

Non-filamentary

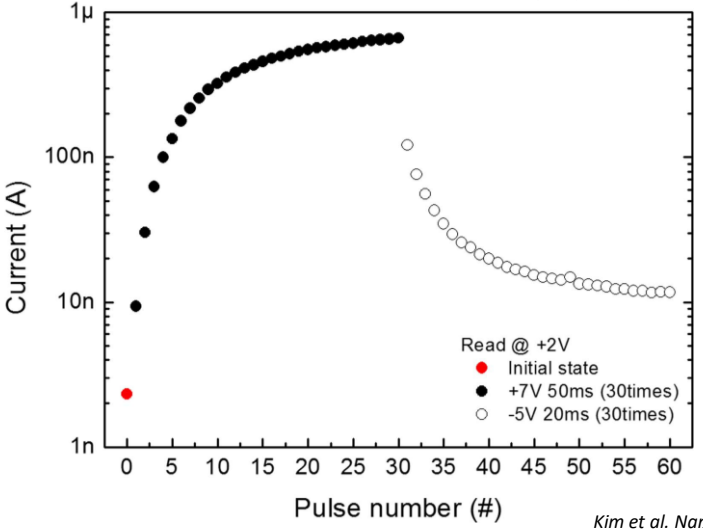


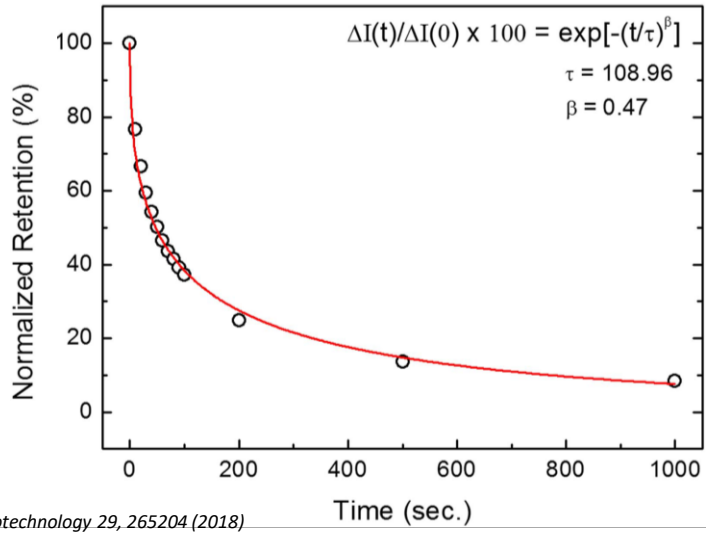
Bipolar switching

Analog leaky-sum-and-fire neuron



Analog synapse with short term memory

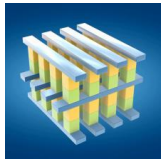


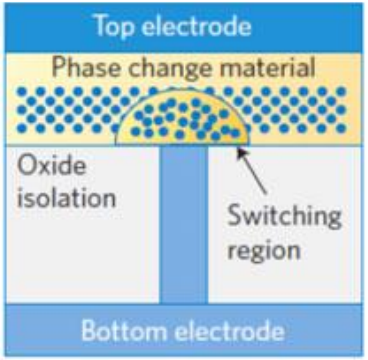


Kim et al. Nanotechnology 29, 265204 (2018)

Phase change RAM

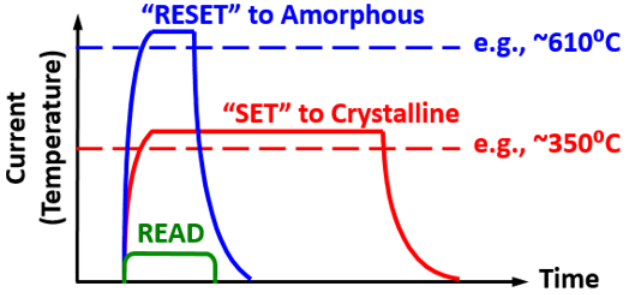
Intel/Micron
XPoint-PCRAM
1µs & 128Gb

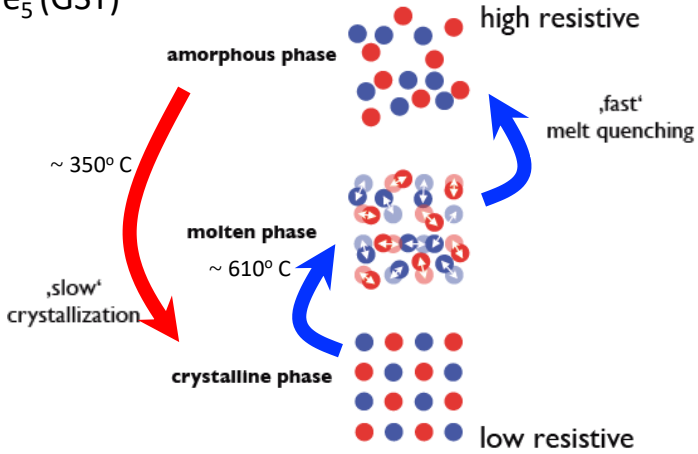




Unipolar switching

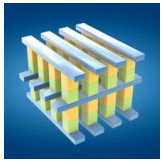
$Ge_2Sb_2Te_5$ (GST)

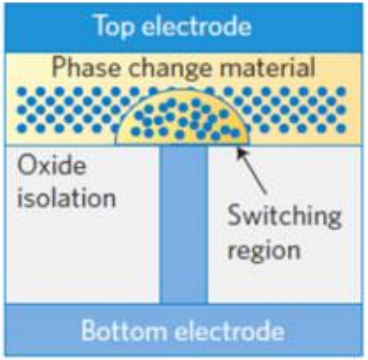




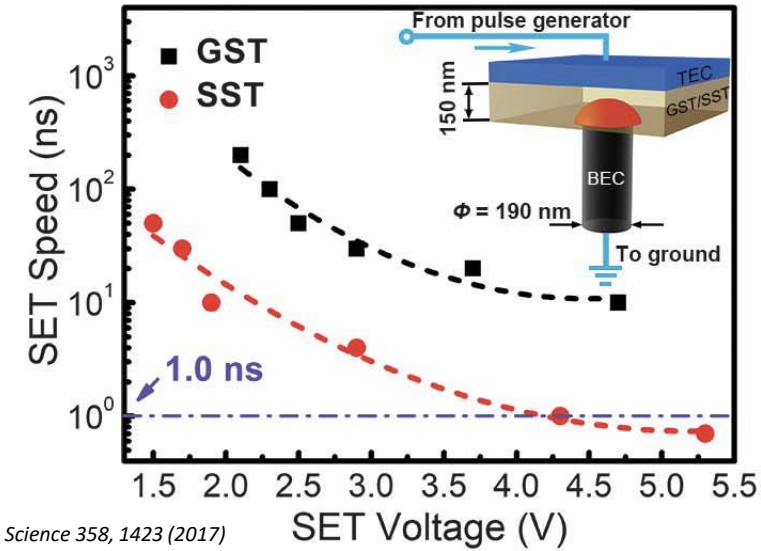
Phase change RAM

Intel/Micron
XPoint-PCRAM
1µs & 128Gb





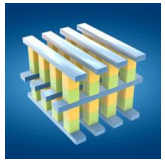
Unipolar switching

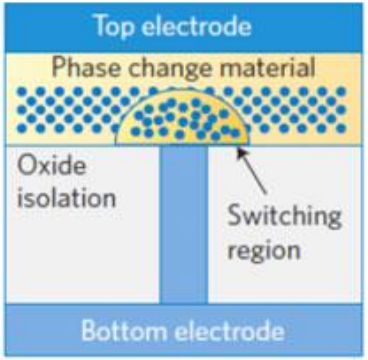


Rao et al., Science 358, 1423 (2017)

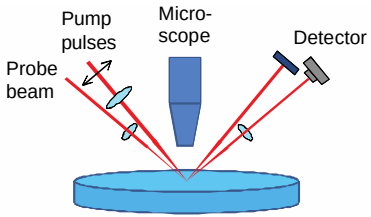
Phase change RAM

Intel/Micron
XPoint-PCRAM
1µs & 128Gb

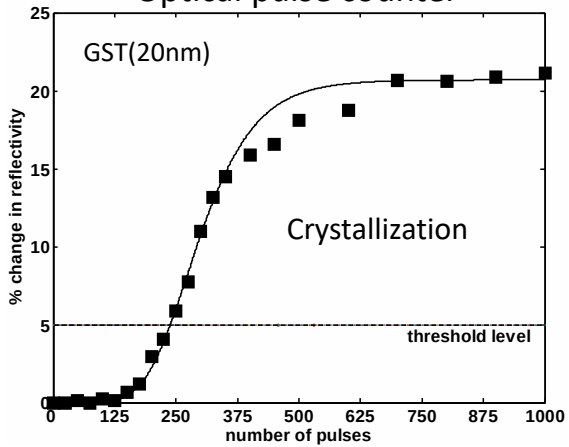




Unipolar switching



Optical pulse counter

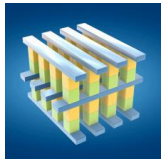


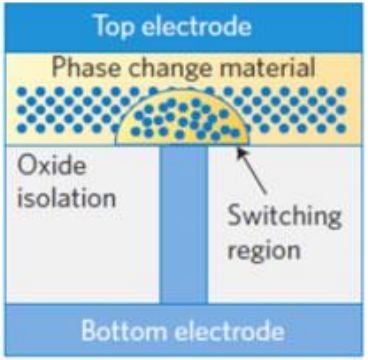
Amorphization
85 fs, 3.61 mJ/cm^2 pulse

Wright et al. Adv. Mater. 23, 3408 (2011)

Phase change RAM

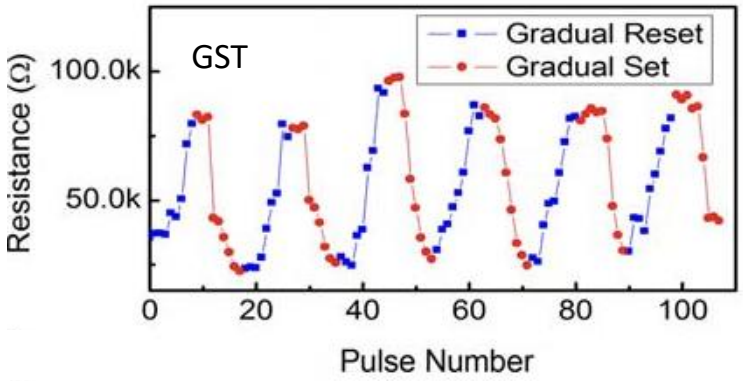
Intel/Micron
XPoint-PCRAM
1µs & 128Gb





Unipolar switching

Analog synapse



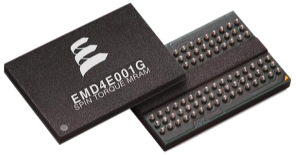
1.5–2 V, 40 ns
0.75–1.25 V, 500ns

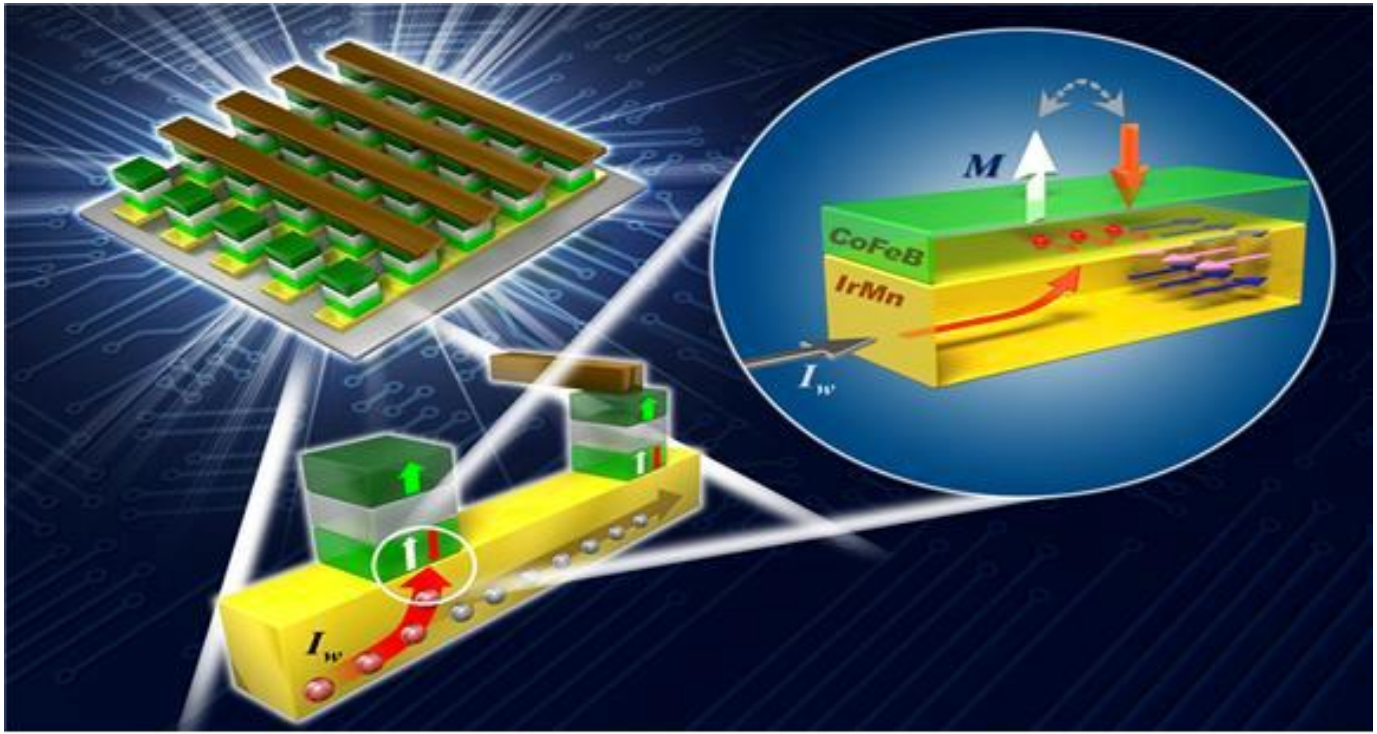
Zhong et al. Phys. Stat. Sol. RRL 9, 414 (2015)

Magnetic RAM

Bipolar switching

Everspin
MRAM
10ns & 1Gb

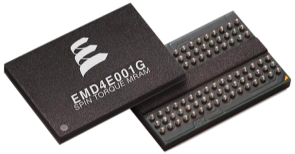


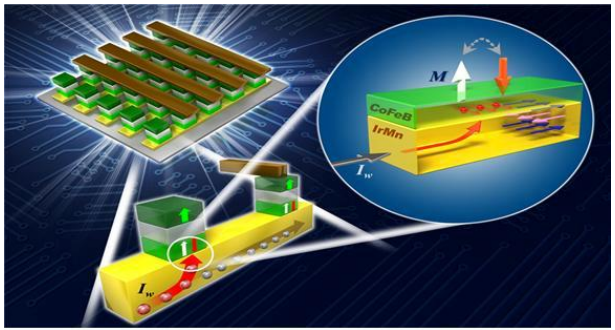


Magnetic RAM

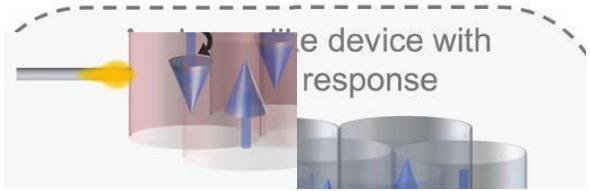
Bipolar switching

Everspin
MRAM
10ns & 1Gb

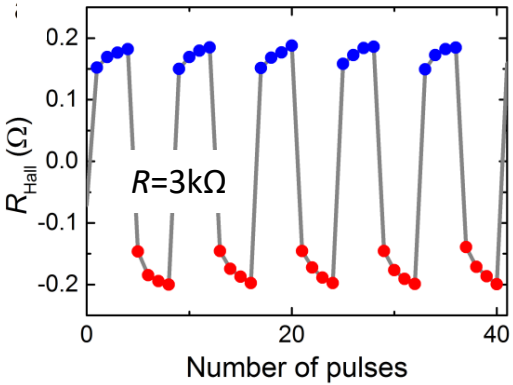




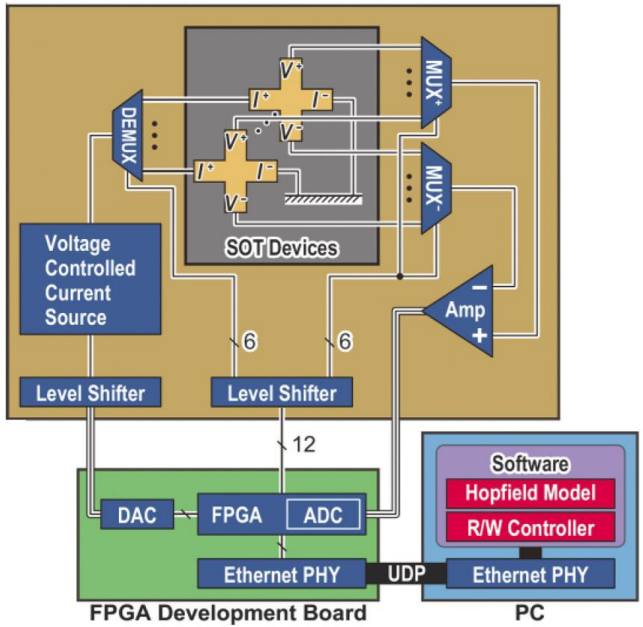
Ferromagnetic domains



Analog synapse



Artificial NN

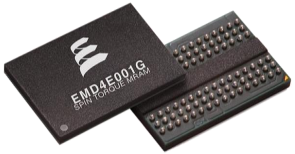


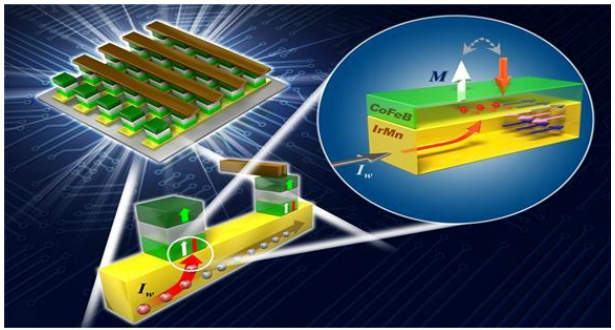
Borders et al. Appl. Phys. Exp. 10, 013007 (2017)

Magnetic RAM

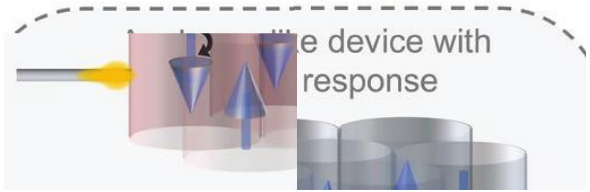
Bipolar switching

Everspin
MRAM
10ns & 1Gb

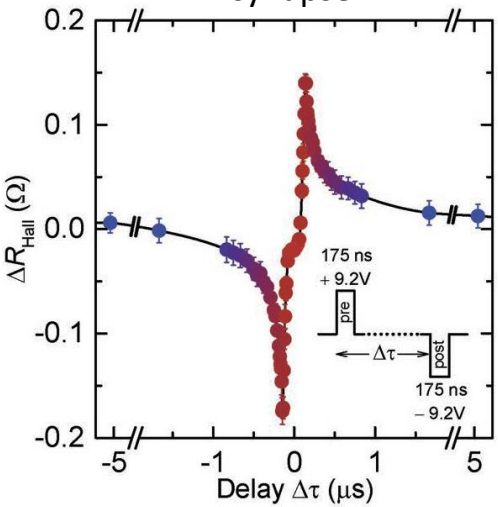




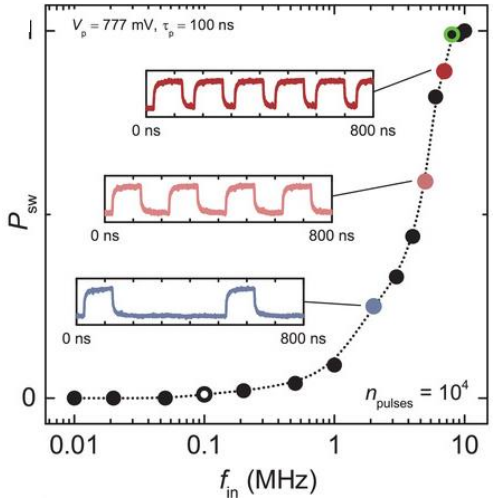
Ferromagnetic domains



Spiking time dependent plasticity synapse



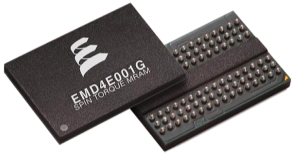
Spiking sum-and-fire neuron



Kurenkov et al. Adv. Mater. 31, 1900636 (2019)

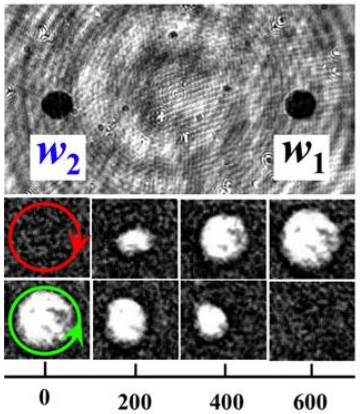
Magnetic RAM

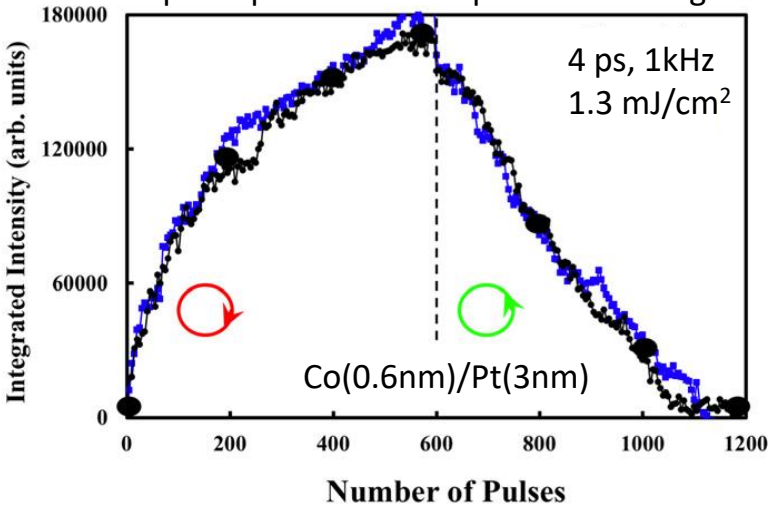
Everspin
MRAM
10ns & 1Gb



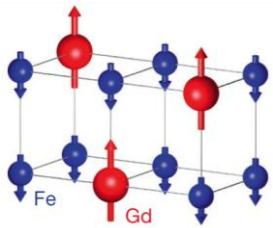
Bipolar switching

Analog synapse
optical polarization-dependent writing





Binary switching in ferrimagnets GdFeCo(20nm), YIG:Co(7.5 µm):
Single 100 fs pulse of ~1-10s mJ/cm^2 and reversal time ~10s ps



Ostler et al. Nat. Comm. 3, 666 (2012)
Stupakiewicz et al. Nature 542, 71 (2017)