Future Nanotechnology Process for Neuromorphic Systems

(NE 1)

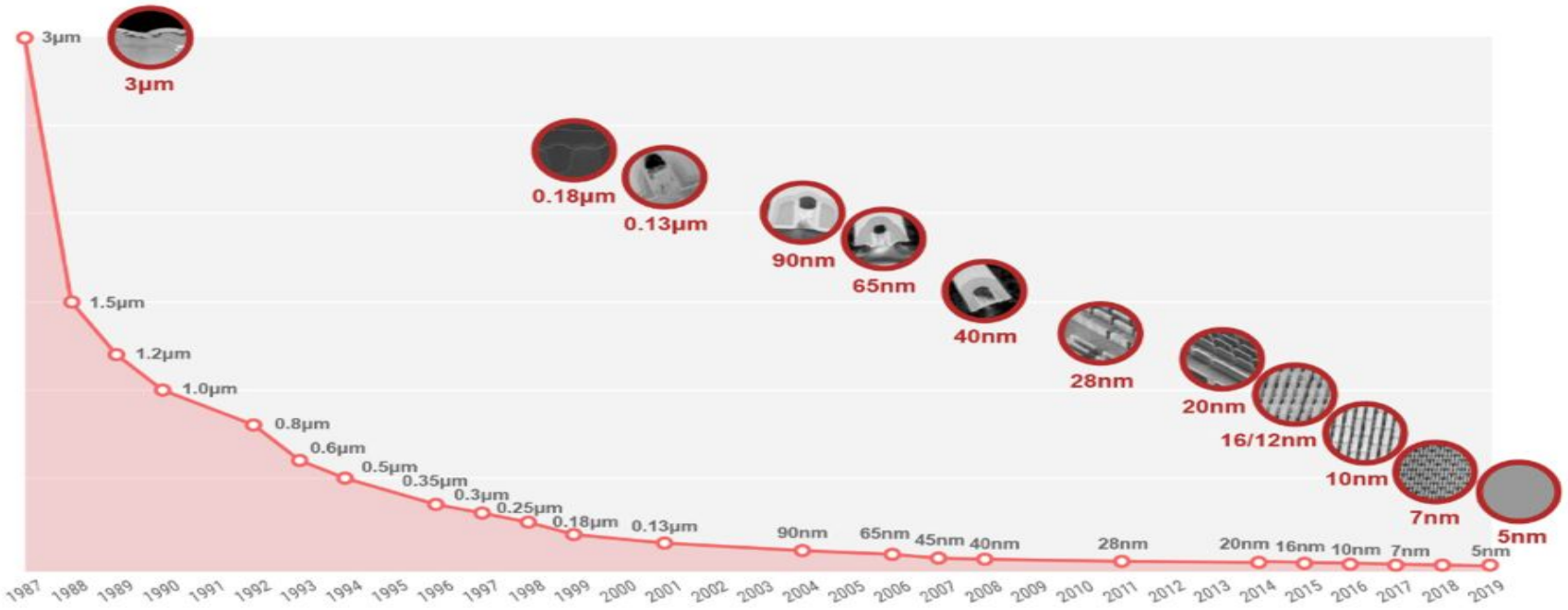
Shih-Chii Liu (edited by T Delbruck)

- Shrinking of the MOSFET
  - Overcoming short channel effects
  - Using SOI (Silicon on Insulator)
  - Moving to FINFET to Nanosheet FETs
- New non-volatile memory (NVM) technology
  - Going past DRAM and SRAM

Future Nanotechnology Process for Neuromorphic Systems

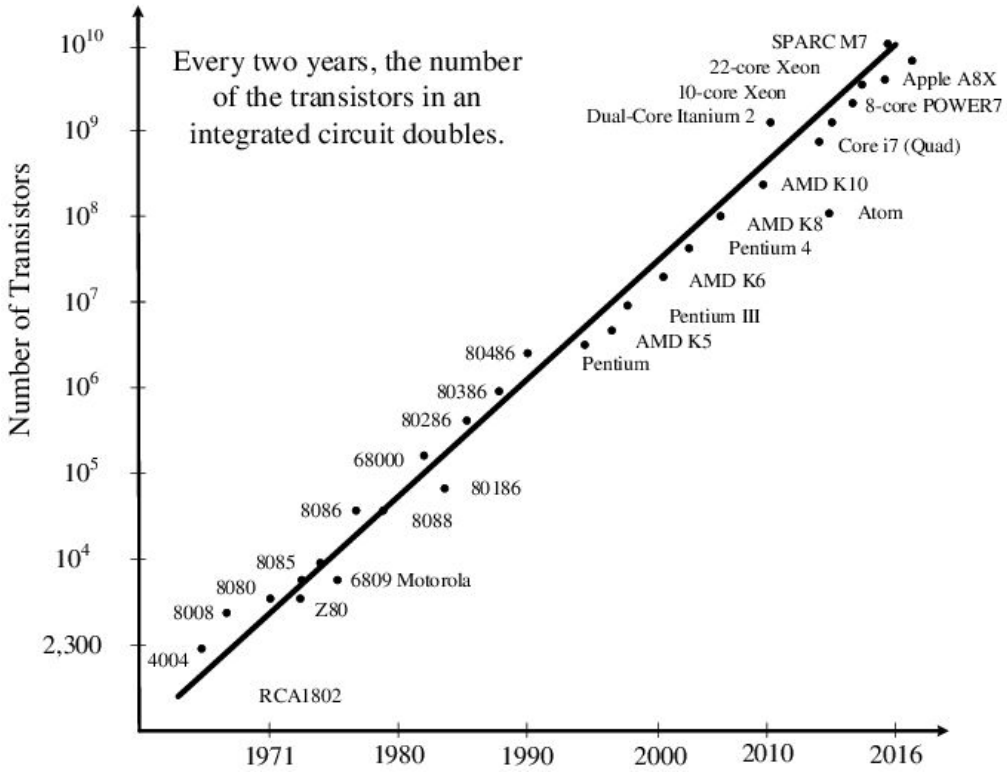
- Shrinking of the MOSFET
  - Overcoming short channel effects
  - Using SOI (Silicon on Insulator)
  - Moving to FINFET to Nanosheet FETs
- New non-volative memory (NVM) technology
  - Past DRAM and SRAM

Silicon process technology

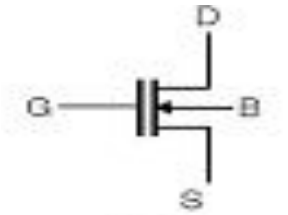


International Technology Roadmap for Semiconductors

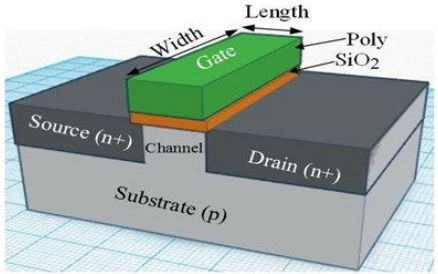
Moore's law

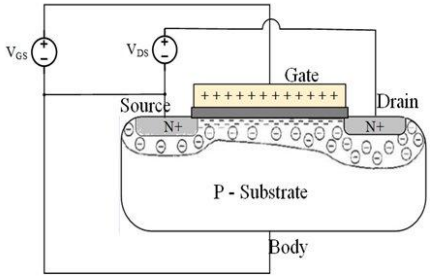


1965, Gordon Moore predicted the doubling of transistors every 24 months



(a)
NMOS transistor as four terminal device



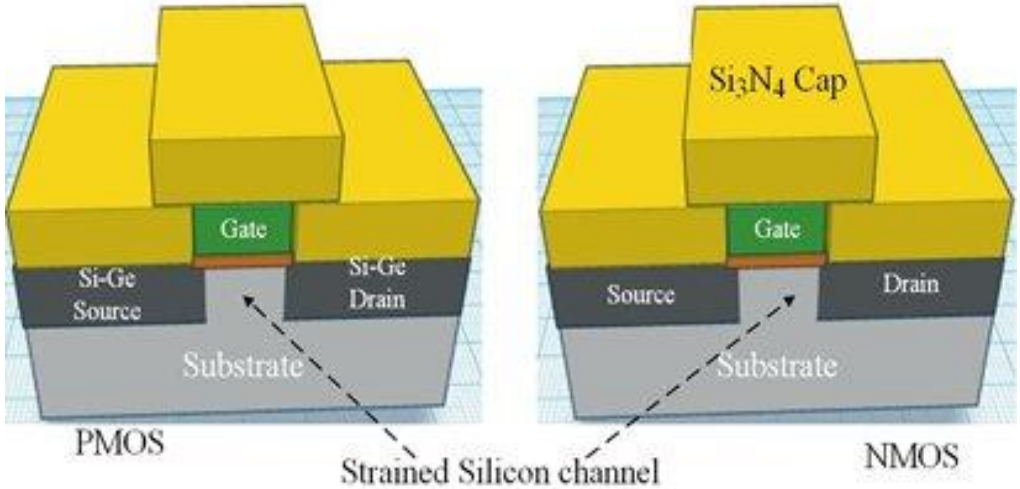


Undesirable effects in short channel transistors

1) Carrier Velocity Saturation
2) Drain Induced Barrier Lowering
3) Punch through

https://www.design-reuse.com/articles/41330/cmos-soi-finfet-technology-review-paper.html

Process improvements



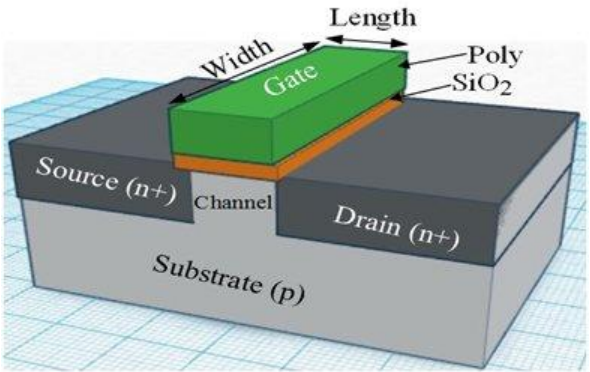
To control short channel effect

1) Use thinner gate oxide – gate can control channel better and reduce the depletion region underneath
2) Strained silicon technology to improve on mobility of carriers in the channel

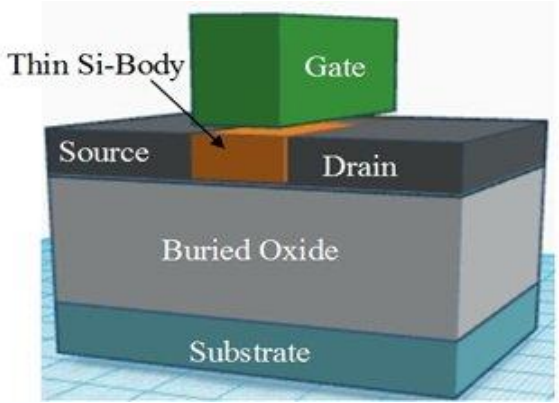
3) Reduce gate leakage by using high-K dielectric material (HfO2). Introduced by Intel in their 45nm process. Dielectric constant is 6 X that of silicon dioxide.

https://www.design-reuse.com/articles/41330/cmos-soi-finfet-technology-review-paper.html

New MOS structure: SOI



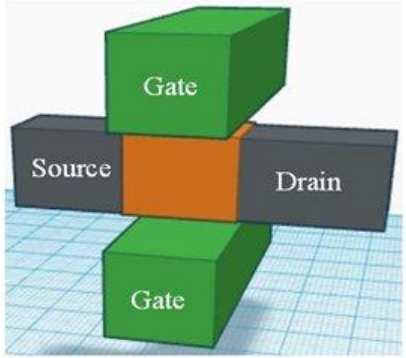
Bulk silicon

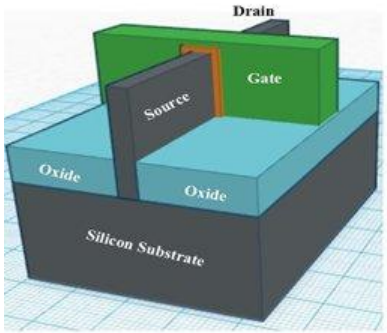


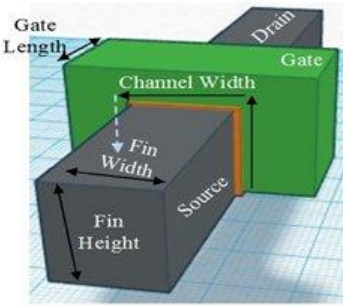
Silicon on insulator

https://www.design-reuse.com/articles/41330/cmos-soi-finfet-technology-review-paper.html

New MOS structure: FinFET



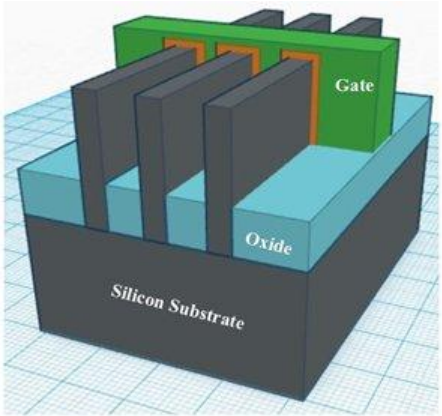




Chenming Hu and Team at Berkeley in 1999

https://www.design-reuse.com/articles/41330/cmos-soi-finfet-technology-review-paper.html

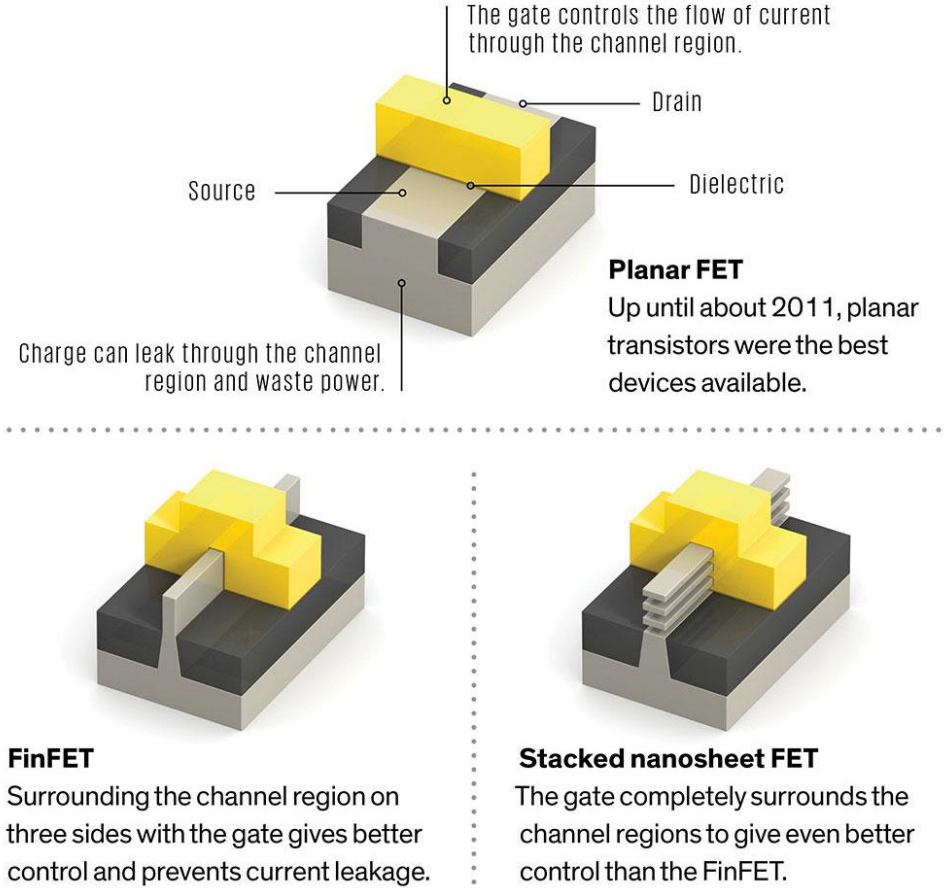
New MOS structure: FinFET



Multi-Fin FinFET
Intel introduce Trigate FETs in 2012
TSMC announced 16nm FinFET technology in 2014

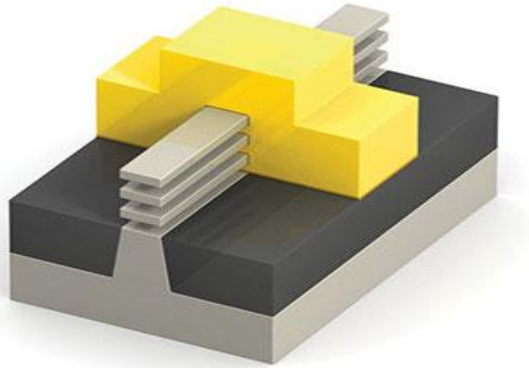
https://www.design-reuse.com/articles/41330/cmos-soi-finfet-technology-review-paper.html

Stacked **Nanosheet FET** for 3nm process



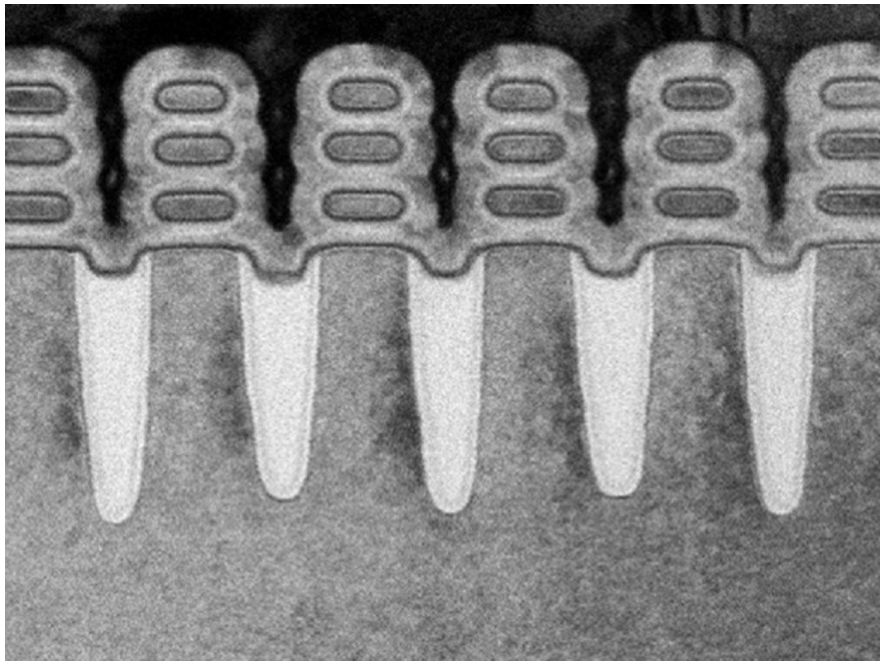
[Peide Ye, Thomas Ernst and Mukesh V. Khare. The Last Silicon Transistor. IEEE Spectrum, Sep 2019]

Stacked Nanosheet FET



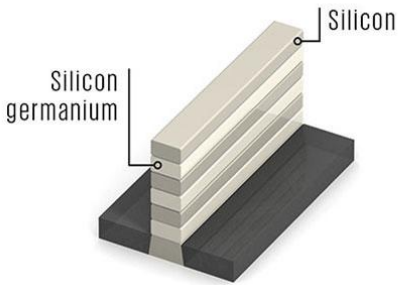
Stacked nanosheet FET

The gate completely surrounds the channel regions to give even better control than the FinFET.

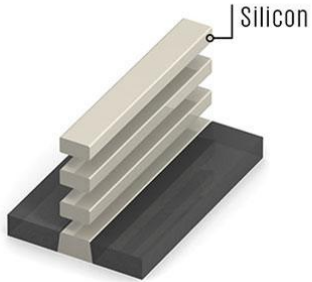


Peide Ye, Thomas Ernst and Mukesh V. Khare. The Last Silicon Transistor. IEEE Spectrum, Sep 2019

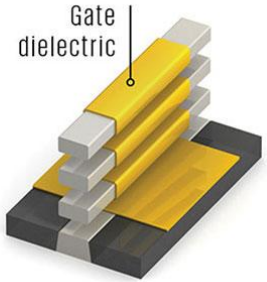
Stacked Nanosheet FET



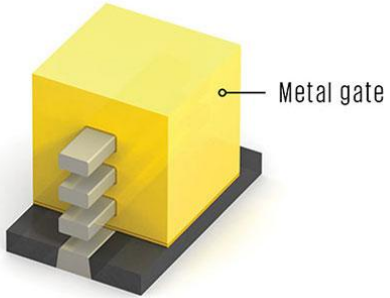
A superlattice of silicon and silicon germanium are grown atop the silicon substrate.



A chemical that etches away silicon germanium reveals the silicon channel regions.



Atomic layer deposition builds a thin layer of dielectric on the silicon channels, including on the underside.



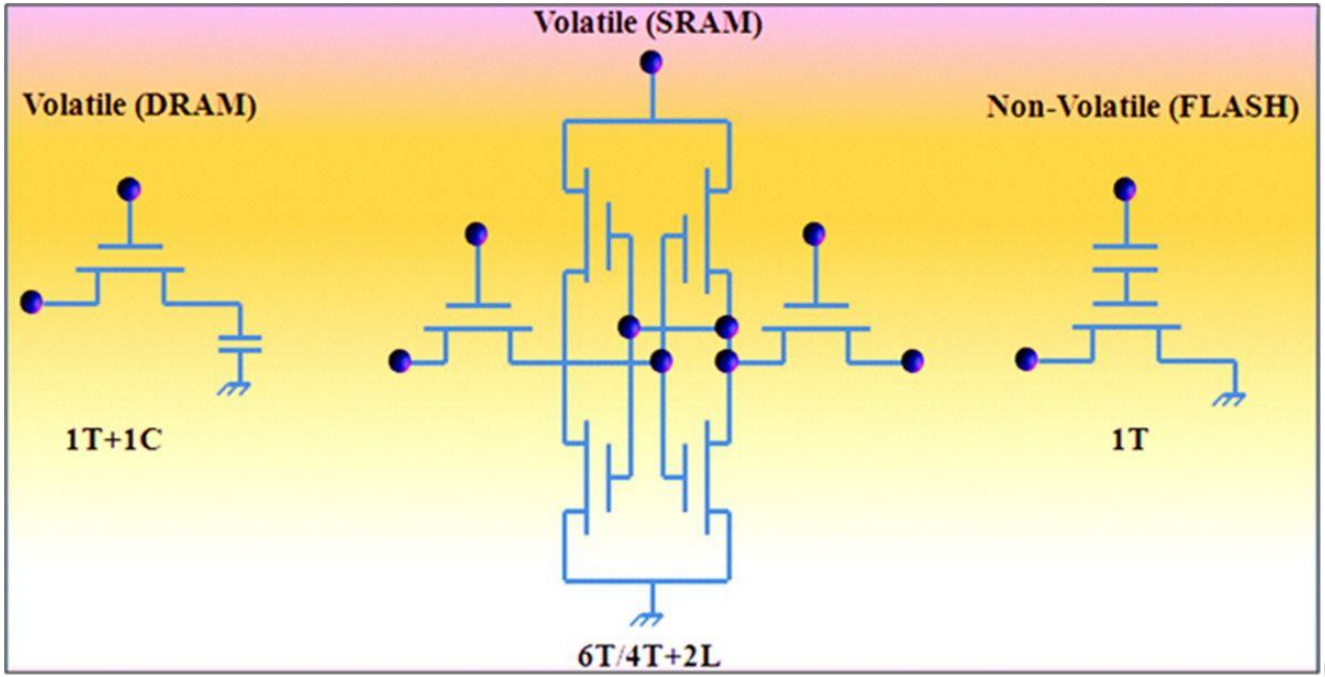
Atomic layer deposition builds the metal gate so that it completely surrounds the channel regions.

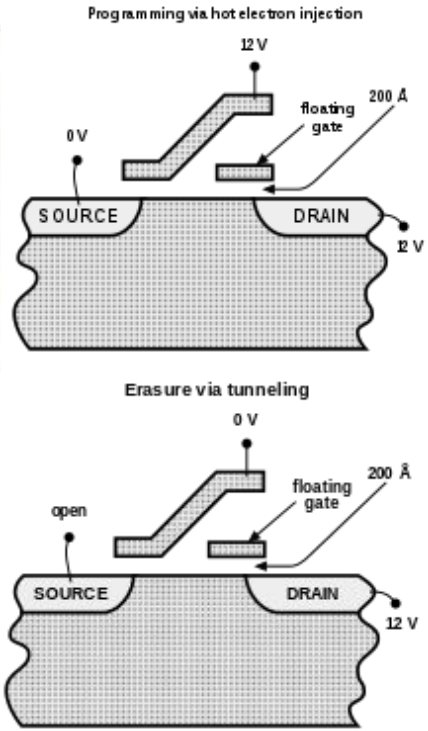
[Peide Ye, Thomas Ernst and Mukesh V. Khare. The Last Silicon Transistor. IEEE Spectrum, Sep 2019]

Future Nanotechnology Process for Neuromorphic Systems

- Shrinking of the MOSFET
  - Overcoming short channel effects
  - Using SOI (Silicon on Insulator)
  - Moving to FINFET to Nanosheet FETs
- New non-volative memory (NVM) technology
  - Past DRAM and SRAM

Types of Memory





Meena, J.S., Sze, S.M., Chand, U. et al. Overview of emerging nonvolatile memory technologies. Nanoscale Res Lett ***9,*** *526 (2014)*

Types of Memory

Table 1. Comparison of the different memory technologies.

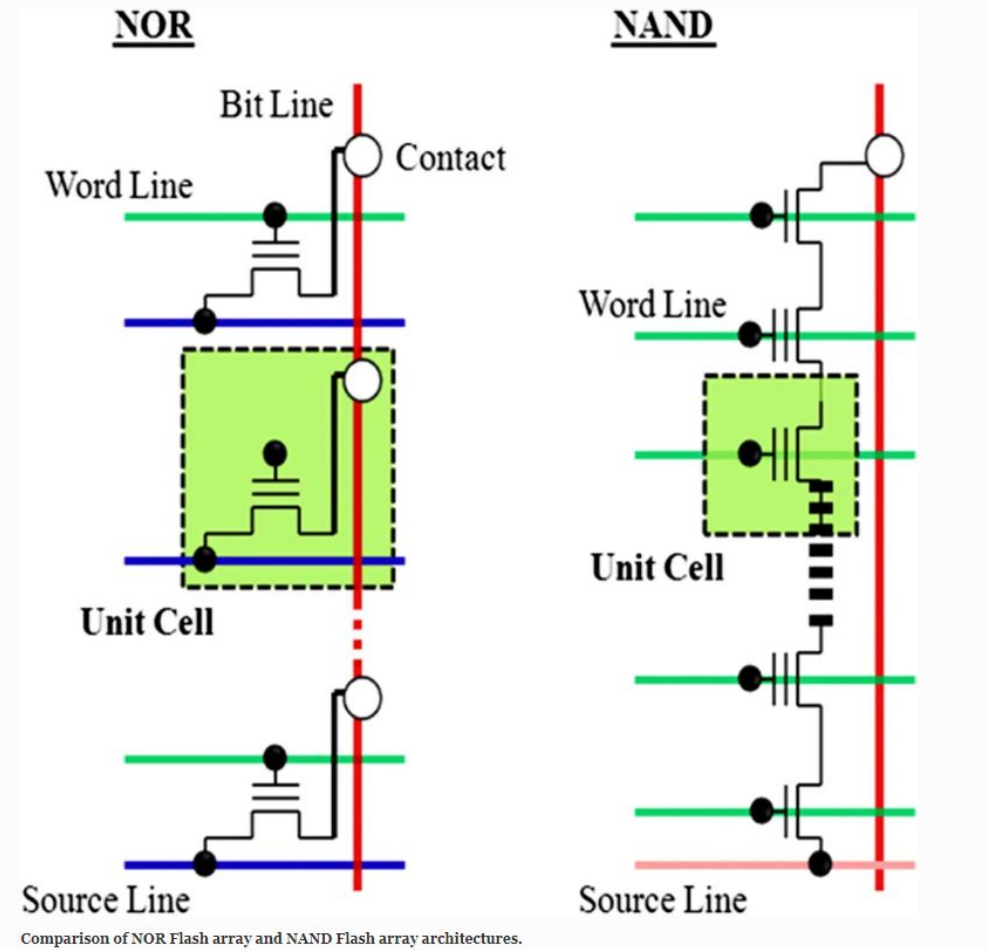
	SRAM	DRAM	NAND flash
Data retention	N	N	Y
Cell factor (F^2)	50–120	6–10	2–5
Read lat	1	30	50
Write latency (ns)	1	50	$>10^6$
Write numbers	10^{16}	10^{16}	10^5
Read/write power	Low	Low	High
Other power	Leakage	Refreshing	None

F=feature size

Flash memory invented by Masuoka in 1980 at Toshiba. Technology is an advance on EEPROM, Electrically Erasable Programmable Read Only Memory. Masuoka and colleagues presented NOR Flash in 1984, and NAND Flash in 1987.

Meena, Jagan Singh, Simon Min Sze, Umesh Chand, and Tseung-Yuen Tseng. 2014. "**Overview of Emerging Nonvolatile Memory Technologies**." *Nanoscale Research Letters* 9 (1): 526. https://doi.org/10.1186/1556-276X-9-526 .

Flash Memory



Comparison of NOR Flash array and NAND Flash array architectures.

Toshiba announced 3D VNAND Flash in 2007 and Samsung released first commercial chips in 2013.

*Meena, J.S., Sze, S.M., Chand, U. et al. Overview of emerging nonvolatile memory technologies. Nanoscale Res Lett **9**, 526 (2014)*

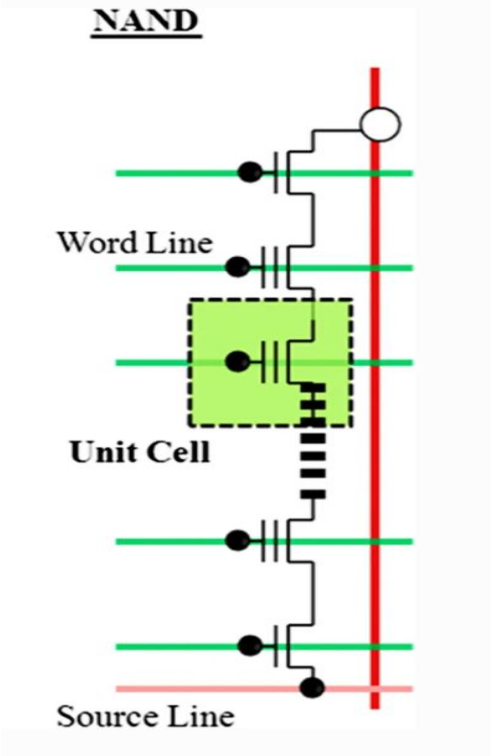
NOR/NAND Memory

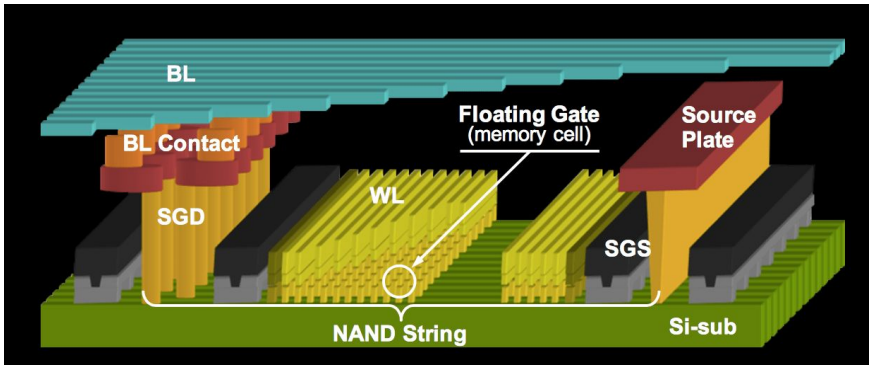
Features	NOR	NAND
Memory size	≤512 Mbit	1 to 8 Gbit
Sector size	Approximately 1 Mbit	Approximately 1 Mbit
Program time	9 µs/word	400 µs/page
Erase time	1 s/sector	1 ms/sector
Read access time	<80 ns	20 µs
Write parallelism	8 to 16 words	2 Kbyte
Output parallelism	Byte/word/dword	Byte/word
Read parallelism	8 to 16 words	2 Kbyte
Access method	Random	Sequential
Price	High	Very low
Reliability	Standard	Low

NOR Flash: Addressable through bit or word
NAND Flash: Addressable through word or page (4-8 KByte)

*Meena, J.S., Sze, S.M., Chand, U. et al. Overview of emerging nonvolatile memory technologies. Nanoscale Res Lett **9,** 526 (2014)*

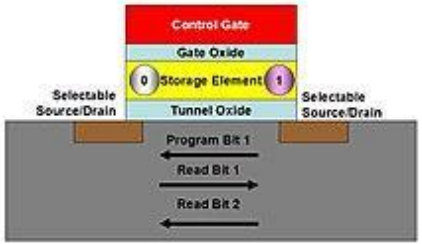
NAND Flash Technology

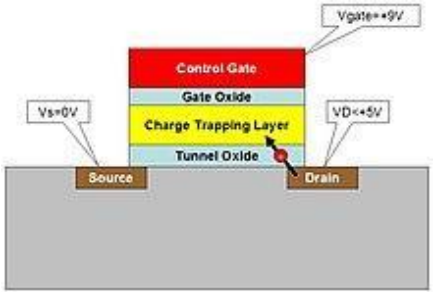


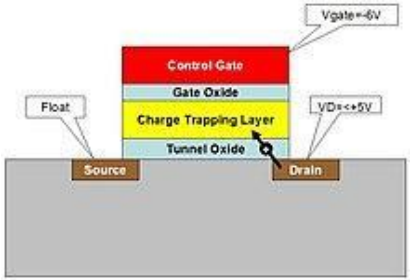


https://semiengineering.com/3d-nand-flash-wars-begin/ (Aug 2018)

3D VNAND Technology

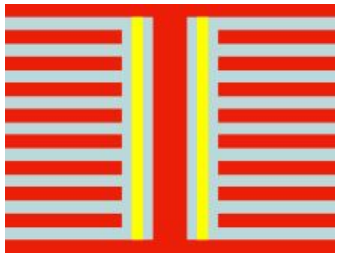






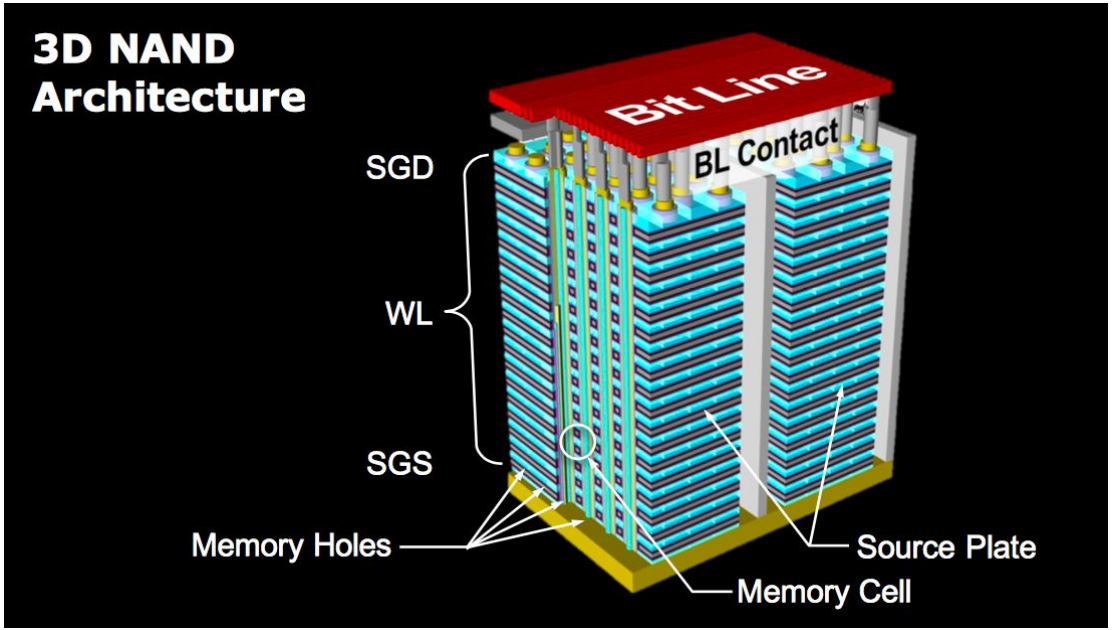
Wikipedia

Charge Trapping Flash instead of floating-gate technology. Charge is stored on Silicon Nitride film. Process is commercialized by AMD and Fujitsu in 2002.



Red is polysilicon: Yellow is Silicon Nitride storage element, Blue is Silicon dioxide insulator

3D VNAND Technology



256 layers thick
in 2020

Western Digital

https://semiengineering.com/3d-nand-flash-wars-begin/ (Aug 2018)

Types of Memory

Table 1. Comparison of the different memory technologies.

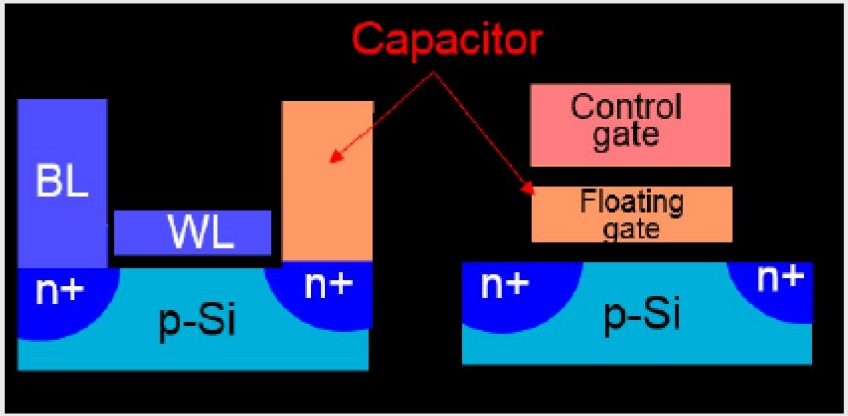
	SRAM	DRAM	NAND flash	PCM	STT-RAM	RRAM
Data retention	N	N	Y	Y	Y	Y
Cell factor (F^2)	50–120	6–10	2–5	6–12	4–20	<1
Read latency (ns)	1	30	50	20–50	2–20	<50
Write latency (ns)	1	50	$>10^6$	50–120	2–20	<100
Write numbers	10^{16}	10^{16}	10^5	10^{10}	10^{15}	10^{15}
Read/write power	Low	Low	High	High	Low	Low
Other power	Leakage	Refreshing	None	None	None	None

SRAM: Static random-access memory
STT-RAM: spin-transfer torque
PCM: phase-change memory
RRAM: resistive random-access memory

*Meena, J.S., Sze, S.M., Chand, U. et al. Overview of emerging nonvolatile memory technologies. Nanoscale Res Lett **9,** 526 (2014)*

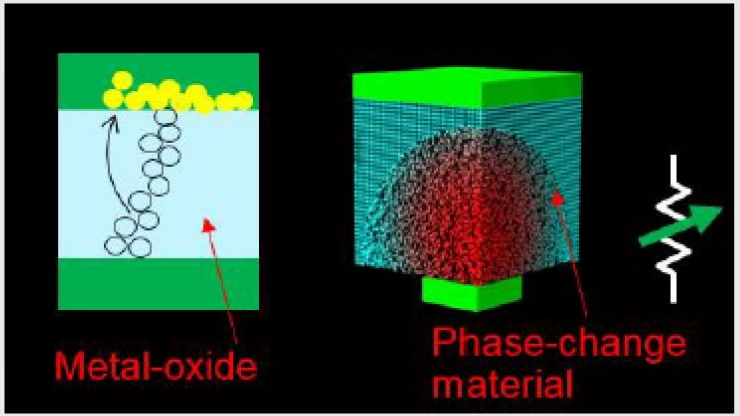
Resistive Memory Devices

"Charge on a capacitor"



Seshadri et al., ArXiv, 2016
Guo et al., IEDM, 2017
Gonugondla et al., ISSCC, 2018

"Alternate atomic arrangements"

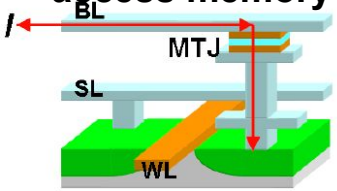


Wong, Salahuddin, Nature Nanotechnology, 2015
Chua, Appl. Phys. A., 2011
Waser, Aono, Nature Materials, 2007

- Difference in atomic arrangements induced by the application of electrical pulses and measured as a difference in electrical resistance
- **Resistive memory devices** or **memristive devices**
- Based on physical mechanisms such as **ionic drift** and **phase transition**

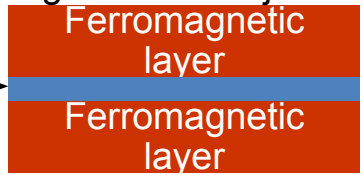
Memristive devices

Magneto-resistive random access memory (MRAM)

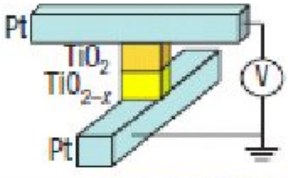


Magnetic tunnel junction

Tunneling insulating layer

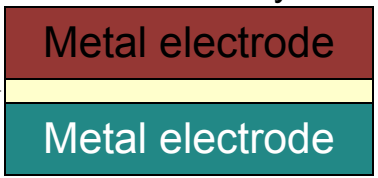


Redox resistive random access memory (RRAM)

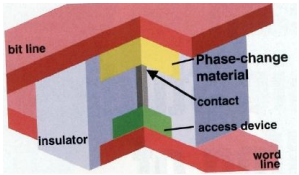


Redox memory cell

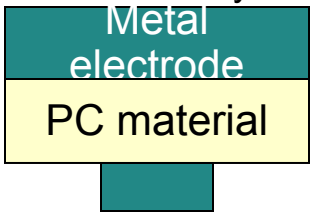
Transition metal oxide/ Solid electrolyte



Phase change memory (PCM)



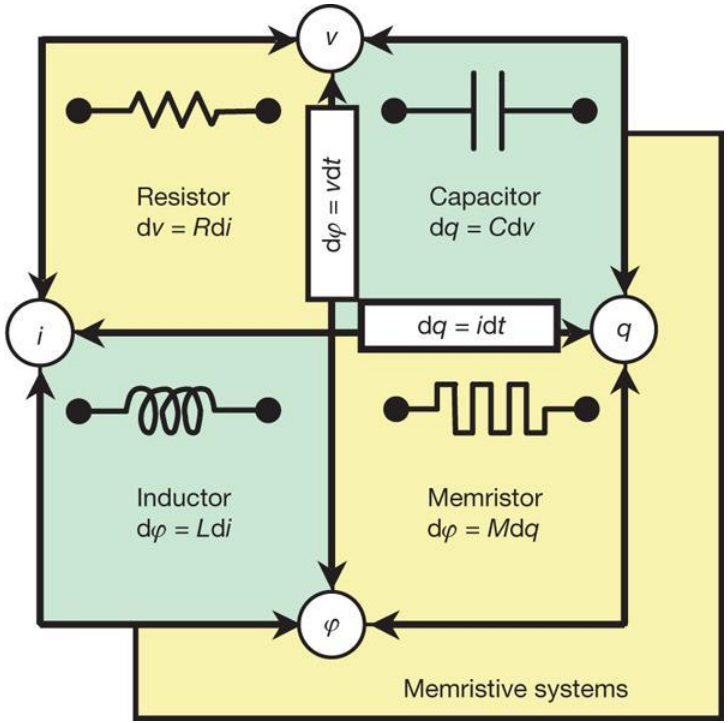
PCM memory cell



"All 2-terminal non-volatile memory devices based on resistance switching are memristors, regardless of the device material and physical operating mechanism" – Leon Chua

Chua, Appl. Phys. A, 2011

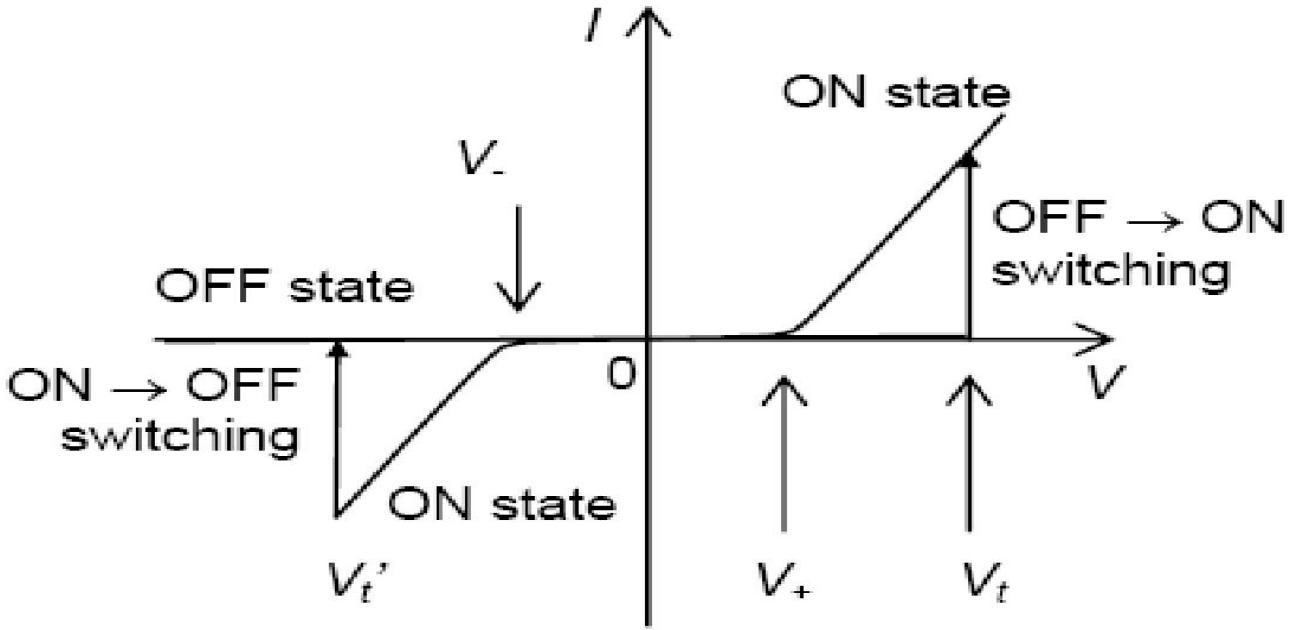
Two-terminal devices



4 basic electrical quantities: “v”, “i”, “q”, “f”.

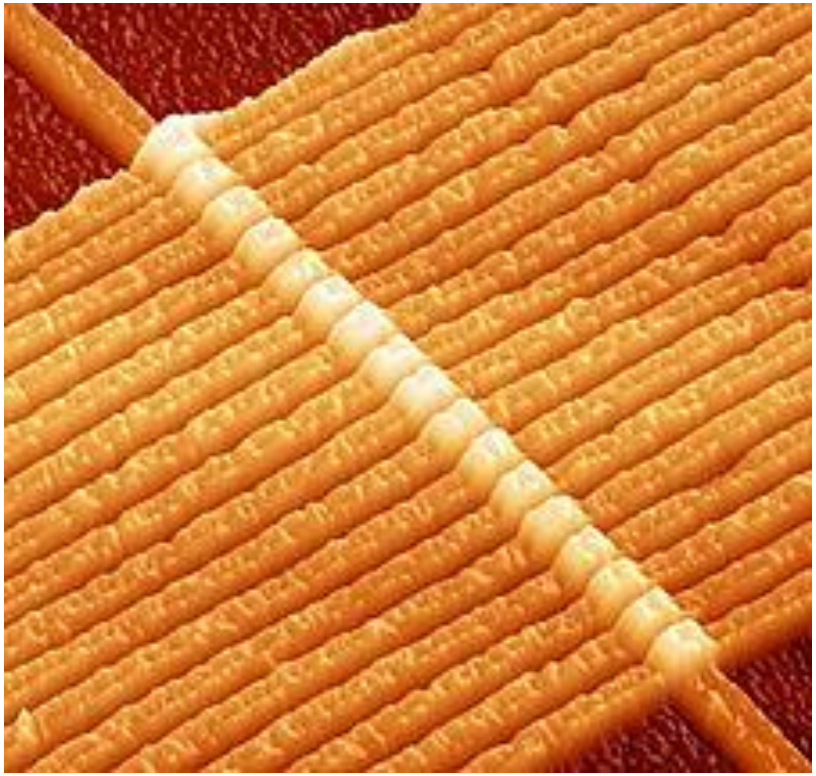
(Strukov et al, 2008)

I-V characteristic



1) Organics: Electrodes: Al, Interlayer: organic; 4-8V for switching,
2) Metal oxides: Electrodes: Cu/Pt, Interlayer: CuOx, TiO2, a-Si; 1-2 V for switching

Memristive wires (HP Labs)

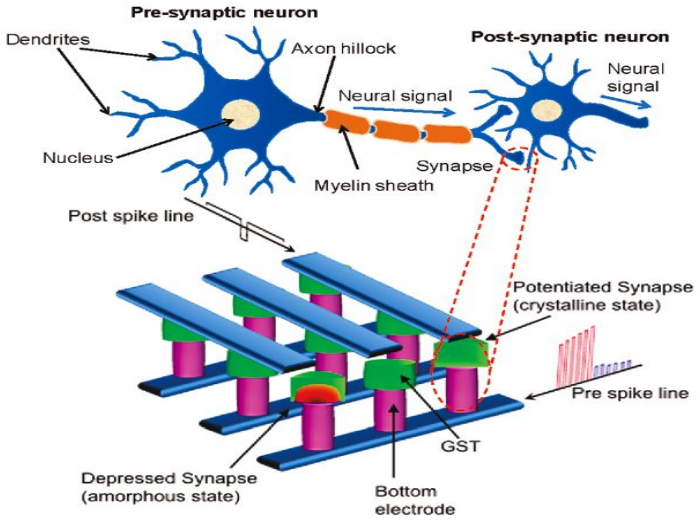


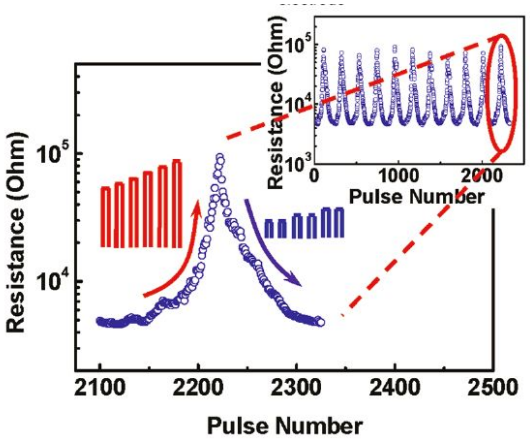
Problem: We need an “access” aka “selector” device to isolate individual crossings.

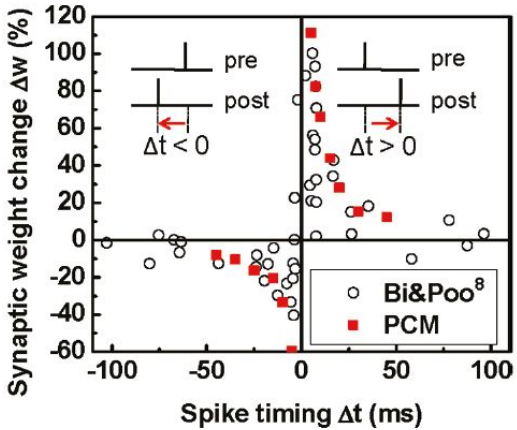
Solution: There now exist technologies for integrated nonlinear vertical access devices, e.g. in Intel’s 3D XPoint.

Picture of 17 memristor wires obtained using an Atomic Force Microscope, wires are 50nm or 150 atoms wide.

STDP rule using PCM crossbar







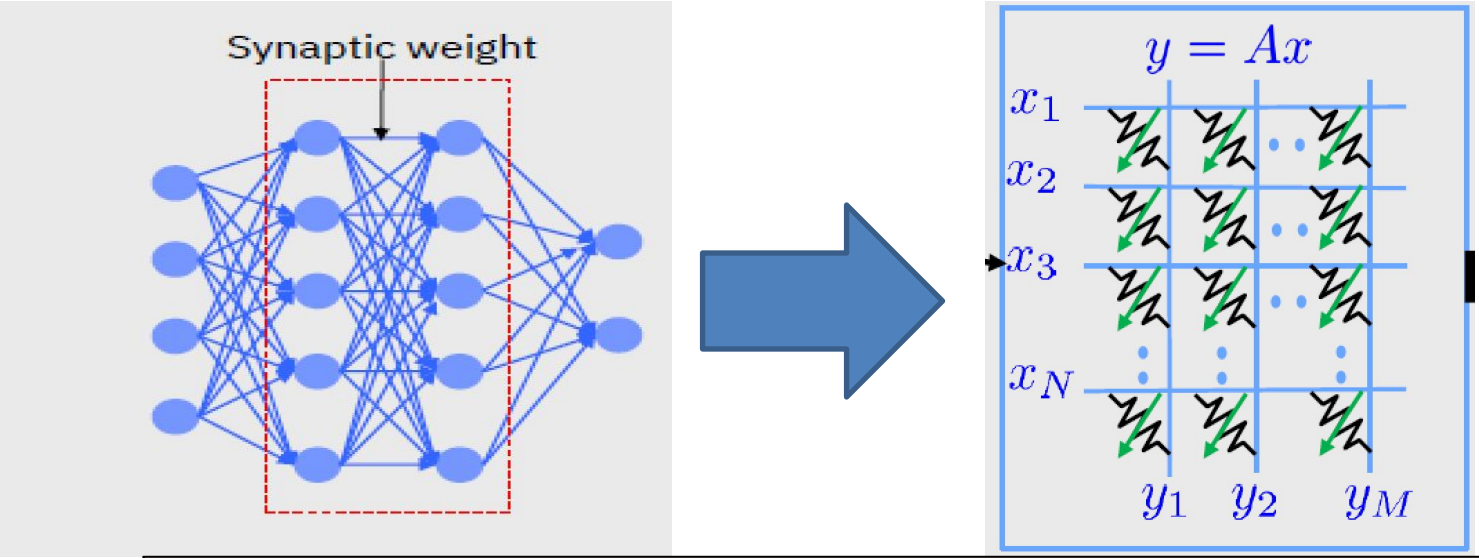
Problems:

1. Device variability is HUGE.
2. PCM state has very strong history and temperature dependence.

Kuzum et al., Nanoletters, 12, 2012

Providing local memory

for fully connected vector-matrix operations



Problems:

1. Device variability is HUGE.
2. Requires fully-differential operations to compensate some problems
3. Needs power-hungry DAC+ADCs for IO
4. Does not support shared-weight as in CNN

Abu Sebastian, IBM, 2019

Future Nanotechnology Process for Neuromorphic Systems

- Shrinking of the MOSFET
  - Overcoming short channel effects
  - Using SOI (Silicon on Insulator)
  - Moving to FINFET to Nanosheet FETs
- New non-volative memory (NVM) technology
  - Past DRAM and SRAM