Non-Volatile CPU(NVCPU) for Ultra Low-Power and High-Speed Computing

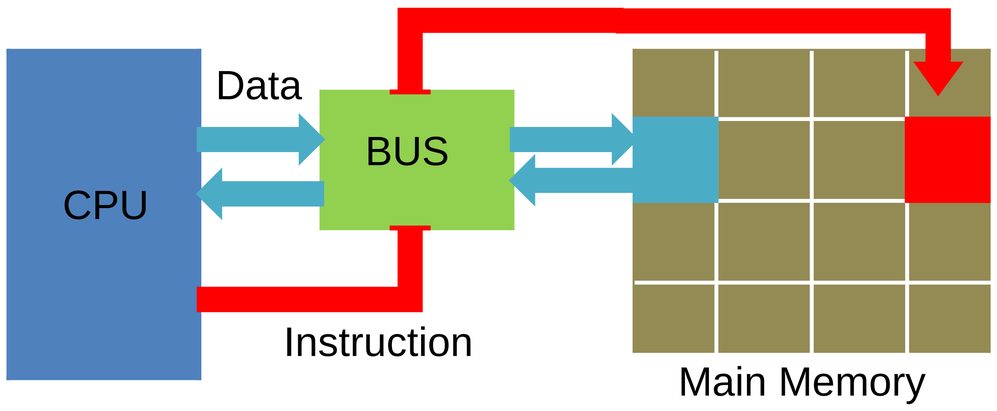
Weisheng ZHAO[1], Jacques-Olivier Klein[1], Damien Querlioz[1], Manuel Bibes[2], Julie Grollier[2], Gilles Sassatelli[3], Lionel Torres[3], Claude Chappert[1]

1. IEF, UMR 8622, Univ. Paris-Sud 11/CNRS, Orsay, France
2. UMPHY, UMR 137, CNRS/Thales, Palaiseau, France
3. LIRMM, UMR 5506, Univ. Montpellier/CNRS, Montpellier, France

Contact: weisheng.zhao@u-psud.fr

State of the art: Von-Neumann Architecture suffers from high power (both static and dynamic) and limited operating frequency

Von-Neumann Architecture



1. *Mainstream computing model* in current computing systems
2. Separation of CPU (>GHz) and computing memory (>100MHz)
3. A number of cache memories (SRAM) to faster the data and instruction access speed
4. Main memory is based on DRAM, which is volatile and with limited capacity

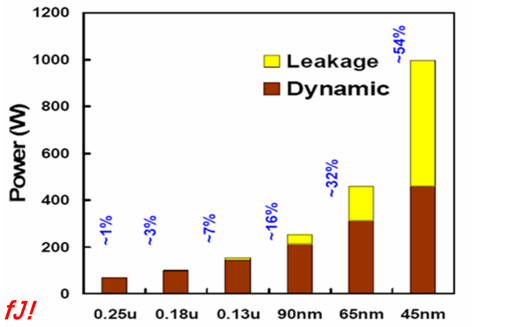
Bottlenecks:

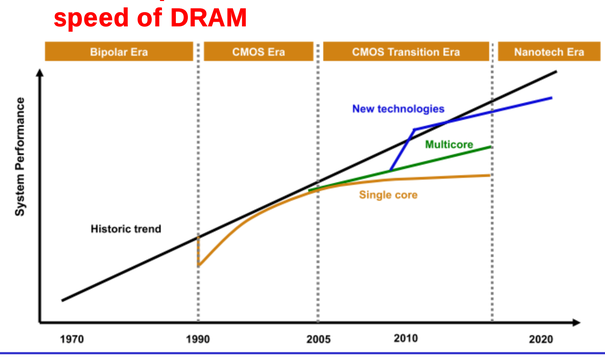
1. High static power due to the cache memories based on SRAM
2. High dynamic power due to the high capacity bus traffic

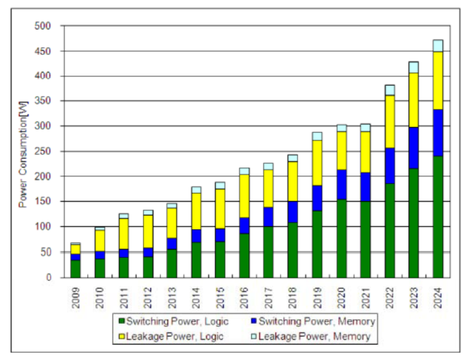
Change one bit in a transistor (22nm) costs *1 fJ!*

Access one data costs (22nm) *1 PJ/mm!*

3. Limited performance due to the low access speed of DRAM

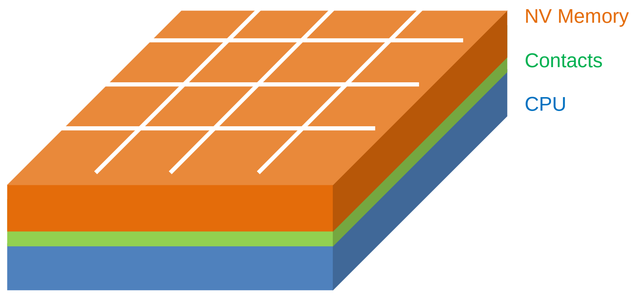




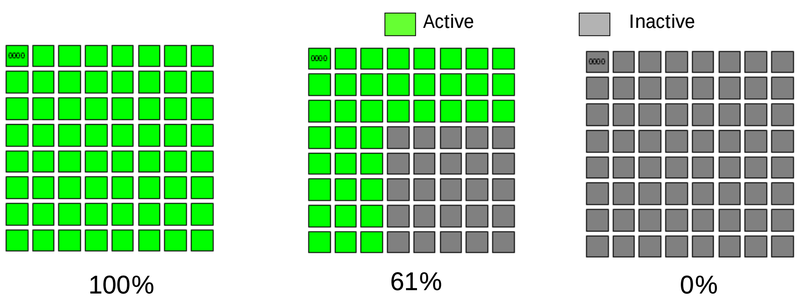


Non-Volatile CPU based on Non-volatile memory (NVM) promises ultra low power and fast computing speed

NV-CPU Architecture



1. NVM is embedded vertically above the CMOS based CPU
2. NVM is integrated locally at the computing operators
   1. Nearly zero standby power
   2. Low dynamic power dissipated by the data access
   3. Faster speed to process the data and restart the chips
   4. Low cost thanks to the delete of SRAM+ DRAM

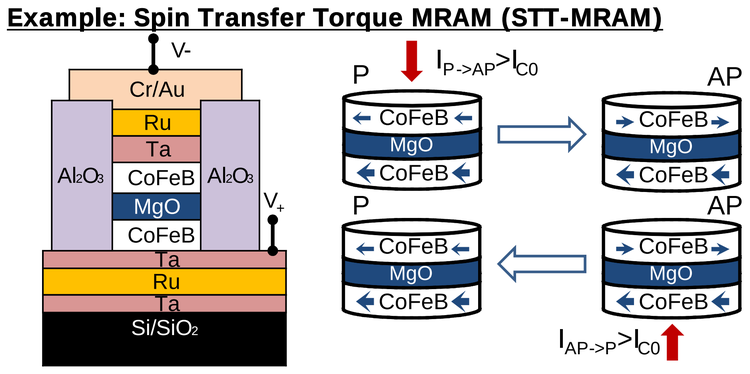


Current Challenges and research priorities:

1. High access speed (<1ns) and high density (>Gbits) NVM ?
2. Computing operators with NVM ?
3. System level optimization between NVM and CPU?

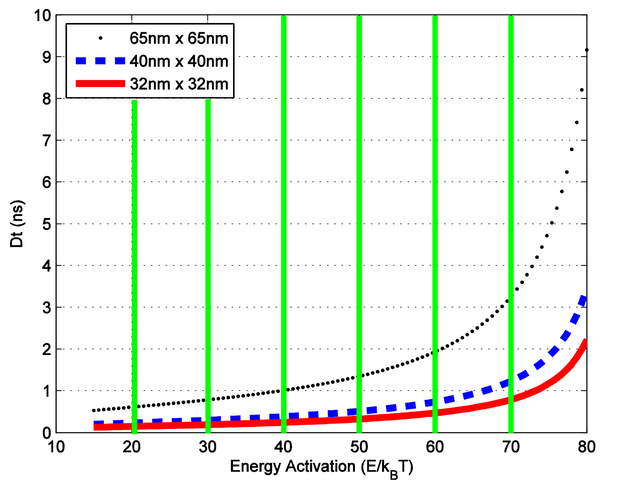
Potential NVM for NV-CPU: Nanospintronics, Memristor, Graphene etc.

Example: Spin Transfer Torque MRAM (STT-MRAM)



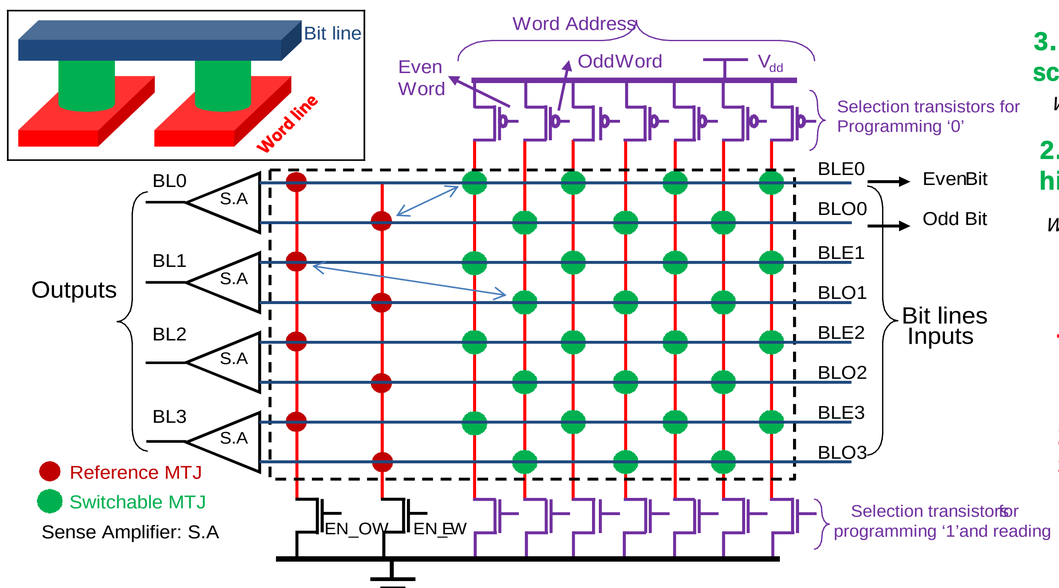
1. Vertical structure and easy 3D integration above CMOS

C. Chappert et al., Nature Materials, 6, .813-823, 2007



3. Faster speed to access following the scaling down of feature size (<1ns)

WS. Zhao et al., Procs .of IEEE VLSI-SOC, China, 2011



2. Cross Point architecture allows high density (>G bits)

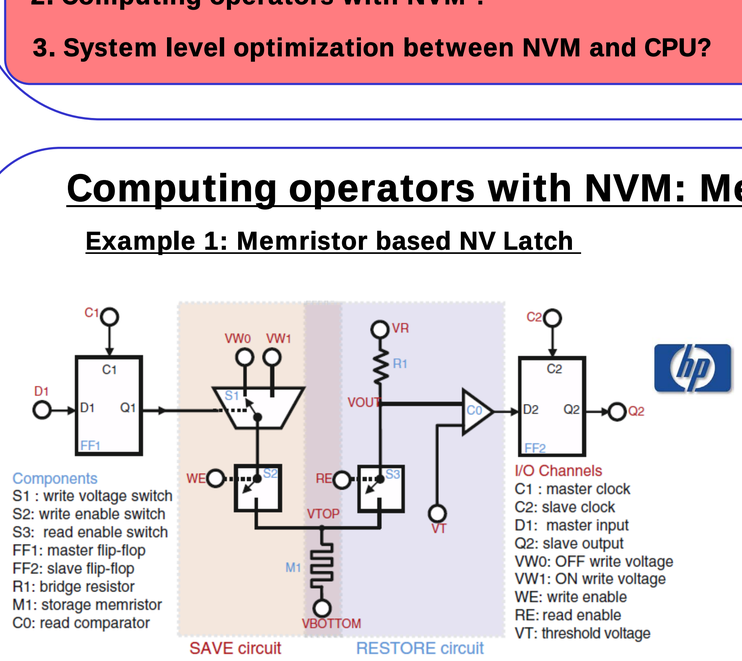
WS. Zhao et al., French Patent 1152472, 2011

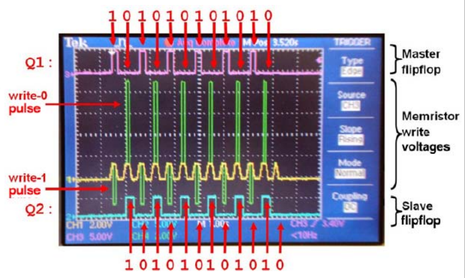
Technology challenges:

1. Low R_{on}/R_{off} (<600%)
2. Relatively high switching power
3. Immature manufacturing processes

Computing operators with NVM: Memristor NV Latch, Magnetic Flip-Flop, Magnetic NV Full Adder

Example 1: Memristor based NV Latch





NVM is integrated directly
High density
Zero standby power

W. Robinett et al., Nanotechnology, 21, 235203, 2010

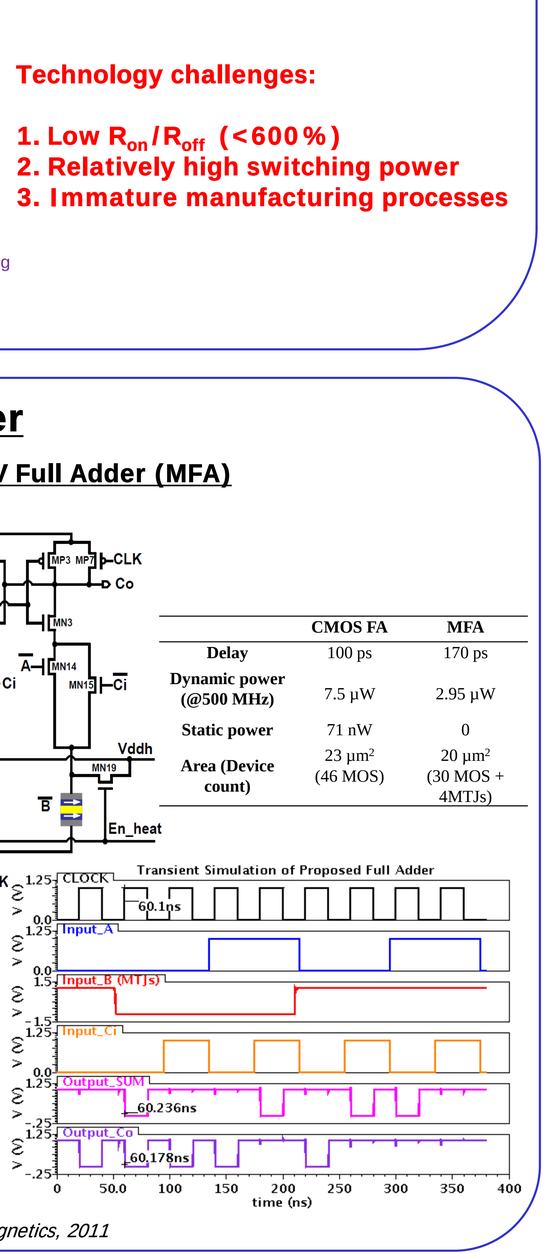
Example 2: Magnetic Flip-Flop (MFF)

NVM is integrated directly
High speed (3.3GHz)
Zero standby power

N. Sakimura et al., JSSCC, 2010

Example 3: Magnetic NV Full Adder (MFA)

	CMOS FA	MFA
Delay	100 ps	170 ps
Dynamic power (@500 MHz)	7.5 µW	2.95 µW
Static power	71 nW	0
Area (Device count)	23 µm² (46 MOS)	20 µm² (30 MOS + 4MTJs)



NVM is integrated directly
High computing speed
Zero standby power
Low dynamic power

Y. Gang et al., IEEE Transaction on Magnetics, 2011