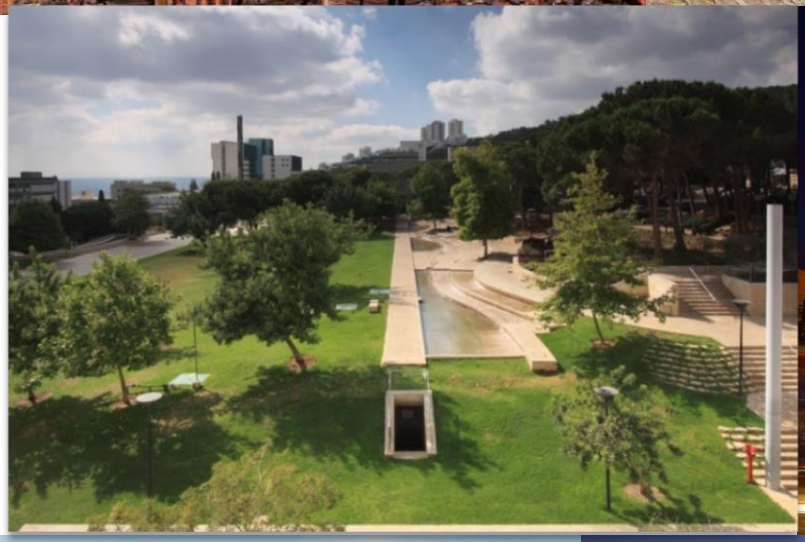
Real Processing in the Memory with Memristive Memory Processing Unit

Shahar Kvatinsky
Viterbi Faculty of Electrical Engineering
Technion – Israel Institute of Technology

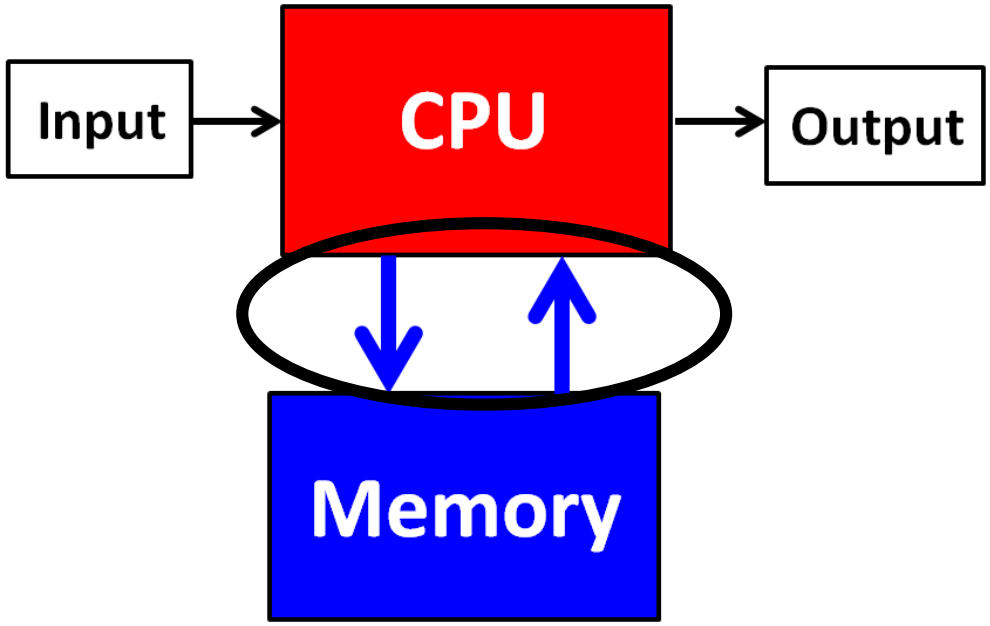
July 2019

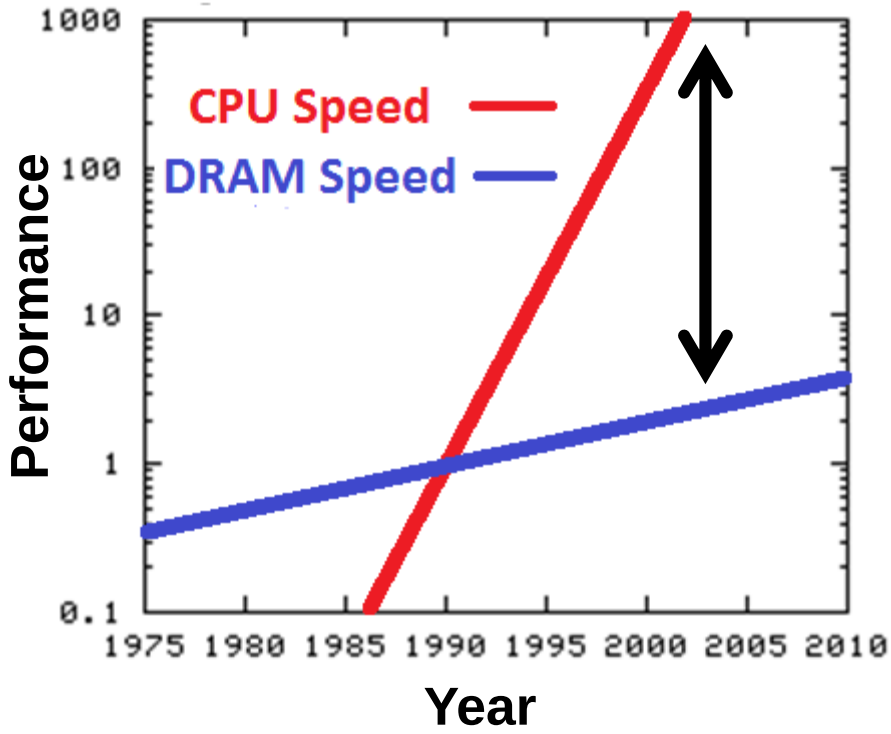


TECHNION
Israel Institute
of Technology

The External Memory Wall Problem

von Neumann (Architecture) Bottleneck





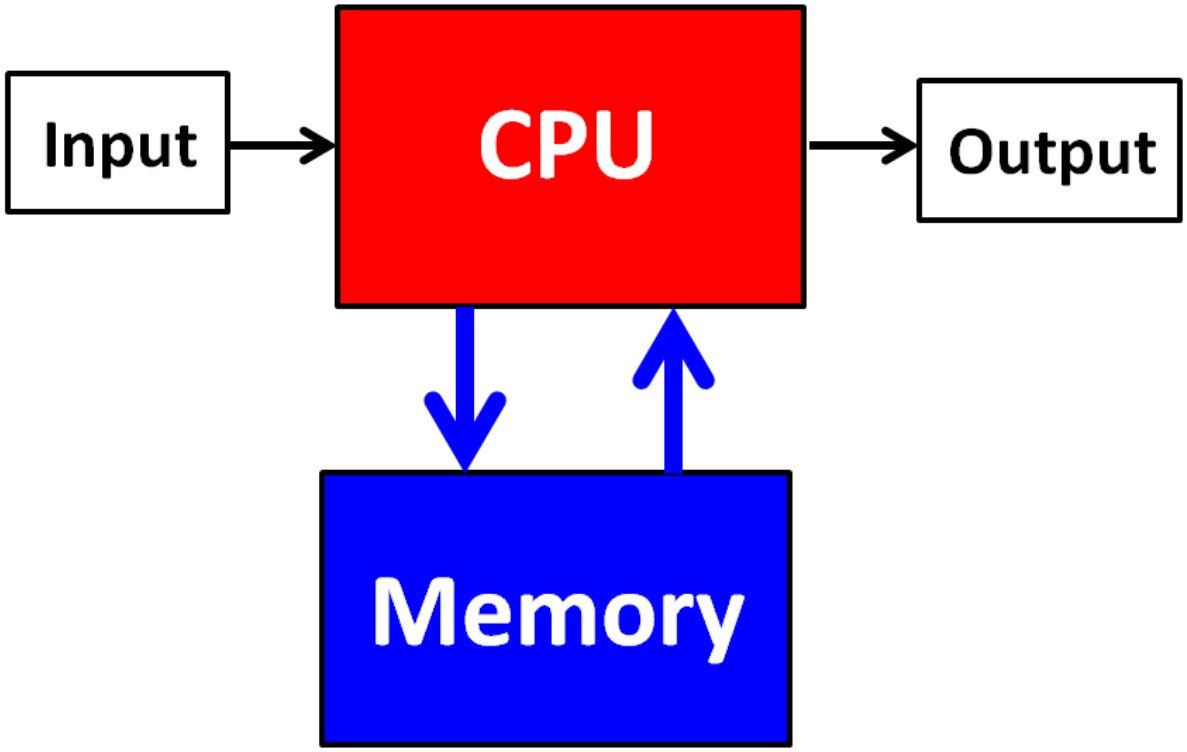
And a Huge Energy Bottleneck

Operation (16-bit operand)	**Energy/Op** (45 nm)	**Cost** (vs. Add)
Add operation	0.18 pJ	1X
Load from on-chip SRAM	11 pJ	61X
Send to off-chip DRAM	**640 pJ**	**3,556X**

>1000X more energy to go to memory

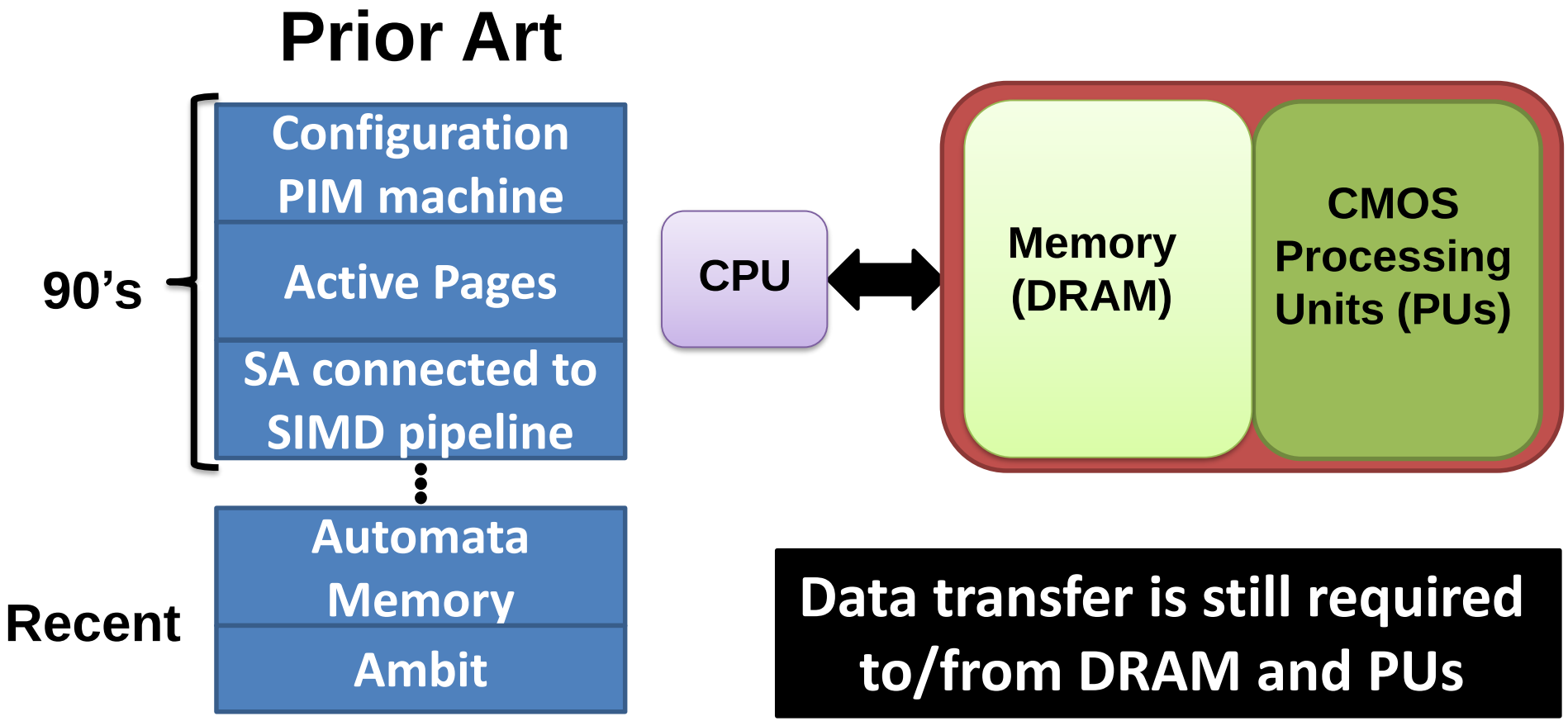
A. Pedram, S. Galal, S. Richardson, S. Kvatinsky, and M. Horowitz, "Dark Memory and Accelerator-Rich System Optimization in the Dark Silicon Era," IEEE Design & Test, April 2017

Moving Computation to Memory



Processing "In-Memory" (PIM)

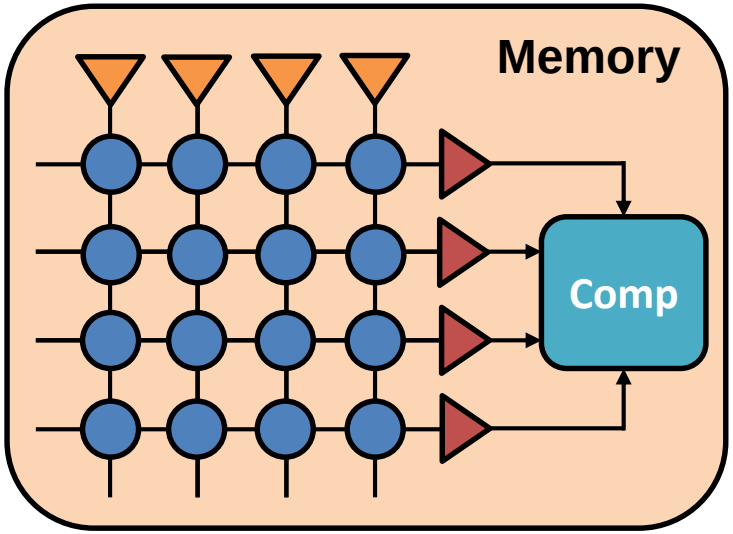
Reducing Data Movement



H. S. Stone, "A Logic-in-Memory Computer," *IEEE Transactions on Computers*, January 1970
M. Gokhale *et al.*, "Processing in memory: the Terasys massively parallel PIM array," *Computer*, 1995
M. Oskin *et al.*, "Active pages: A computation model for intelligent memory," *Comput. Archit. News*, 1998
D. Elliott *et al.*, "Computational ram: Implementing processors in memory," *IEEE Des. Test*, 1999
P. Dlugosch *et al.*, "An Efficient and Scalable Semiconductor Architecture for Parallel Automata Processing," *IEEE TPDS*, 2014
V. Seshadri *et al.*, "Ambit: In-Memory Accelerator for Bulk Bitwise Operations Using Commodity DRAM Technology," MICRO 2017

Alleviating the Data Transfer Problem

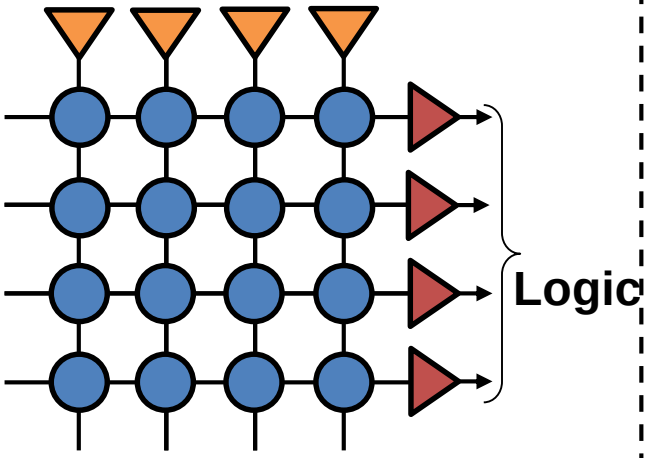
Computation using logic blocks



Dlugosch *et al.*, *IEEE TPDS*, 2014
Oskin *et al.*, *Comput. Archit. News*, 1998
Elliott *et al.*, *IEEE Des. Test*, 1999
Gokhale *et al.*, *Computer*, 1995

OUT-

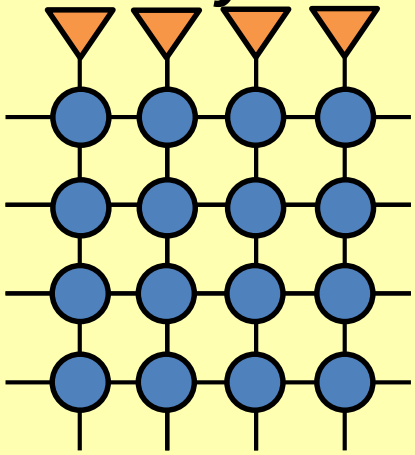
Computation using peripheral circuit



Li *et al.*, *DAC*, 2016
Seshadri *et al.*, MICRO 2017
Aga *et.al.* HPCA 2017
Eckert *et.al.* ISCA 2018

NEAR-

Computation ***using memory cells***



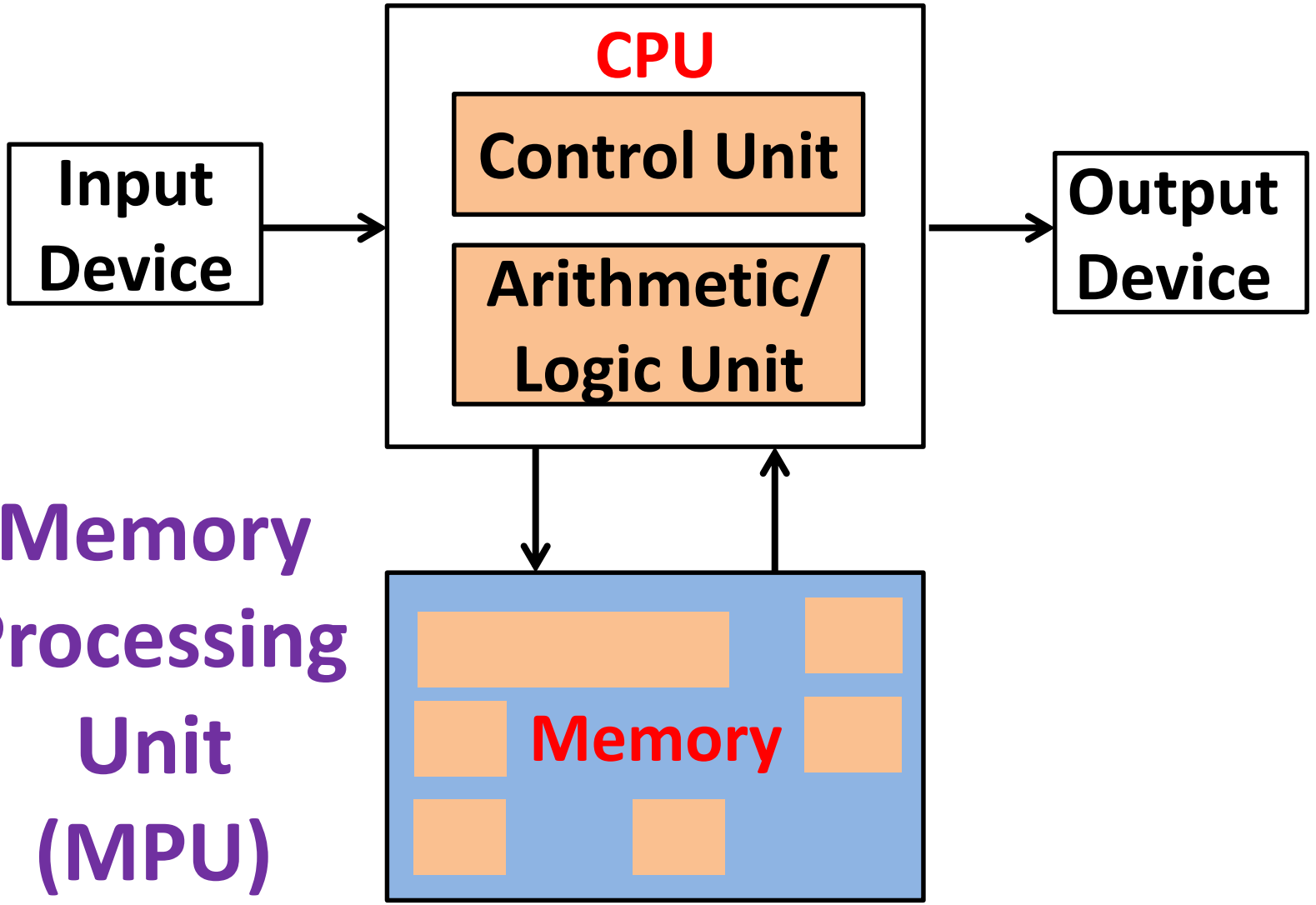
Kvatinsky *et al.*, *TCAS-II*, 2014
Talati *et al.*, TNANO 2016

IN-

J. Reuben, R. Ben Hur, N. Wald, N. Talati, A. Haj Ali, P.-E. Gaillardon, and S. Kvatinsky, "Memristive Logic: A Framework for Evaluation and Comparison", PATMOS 2017

Real Computing within the Memory

Beyond von Neumann Architecture

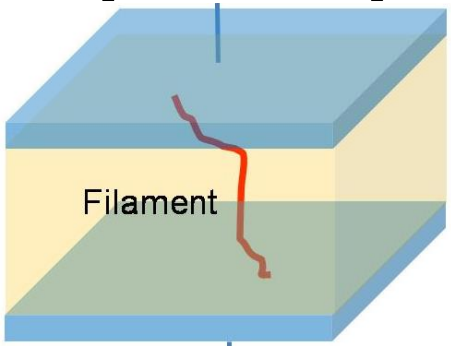


Memory Processing Unit (MPU)

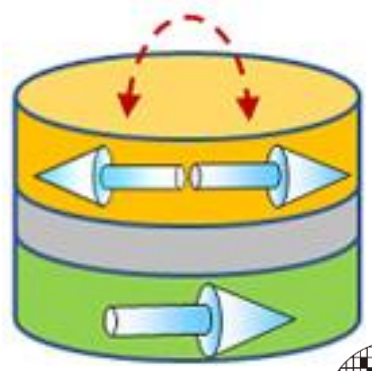
Memristors

Emerging Nonvolatile Memory Technologies

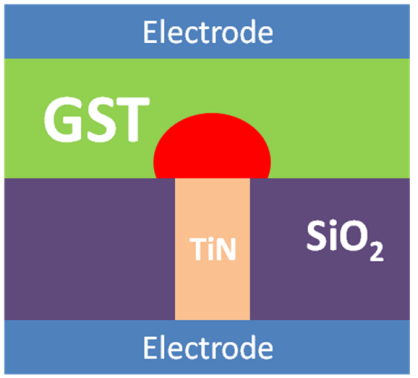
Resistive RAM (RRAM)



STT MRAM

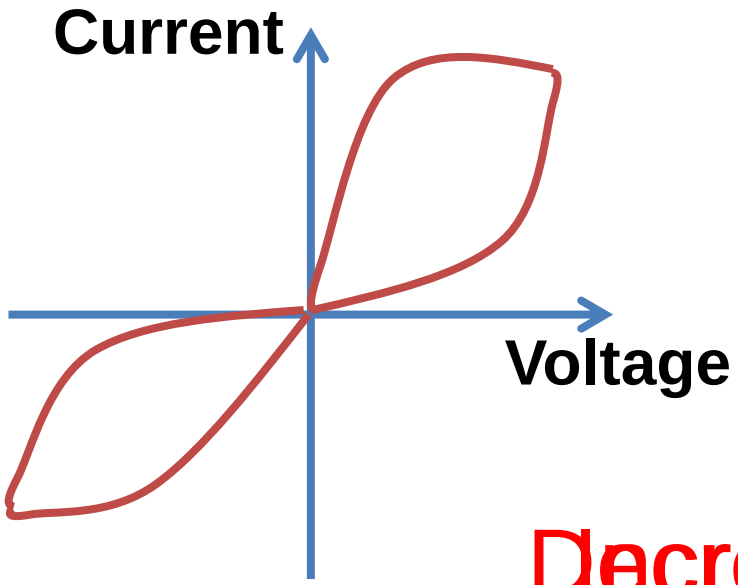


Phase Change Memory (PCM)



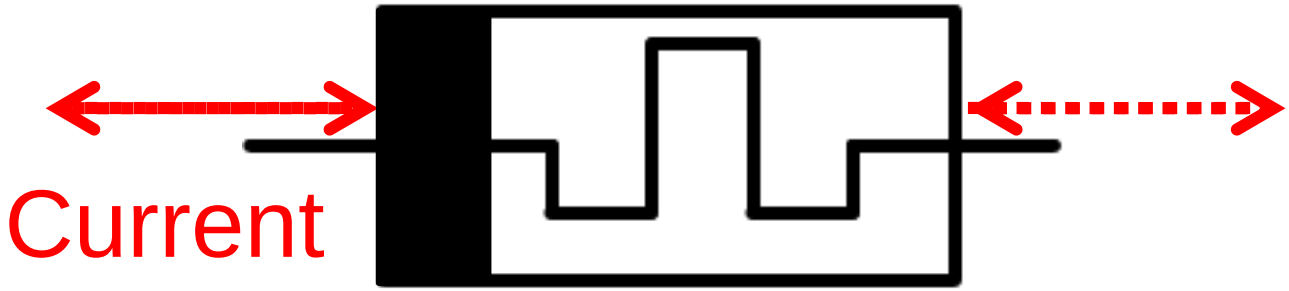
Memristor – Memory Resistor

Resistor with Varying Resistance



High resistive state
(R_{OFF}, HRS)

Decrease resistance



Memristor

Attractive for Memory!

CMOS compatible

Rad hard

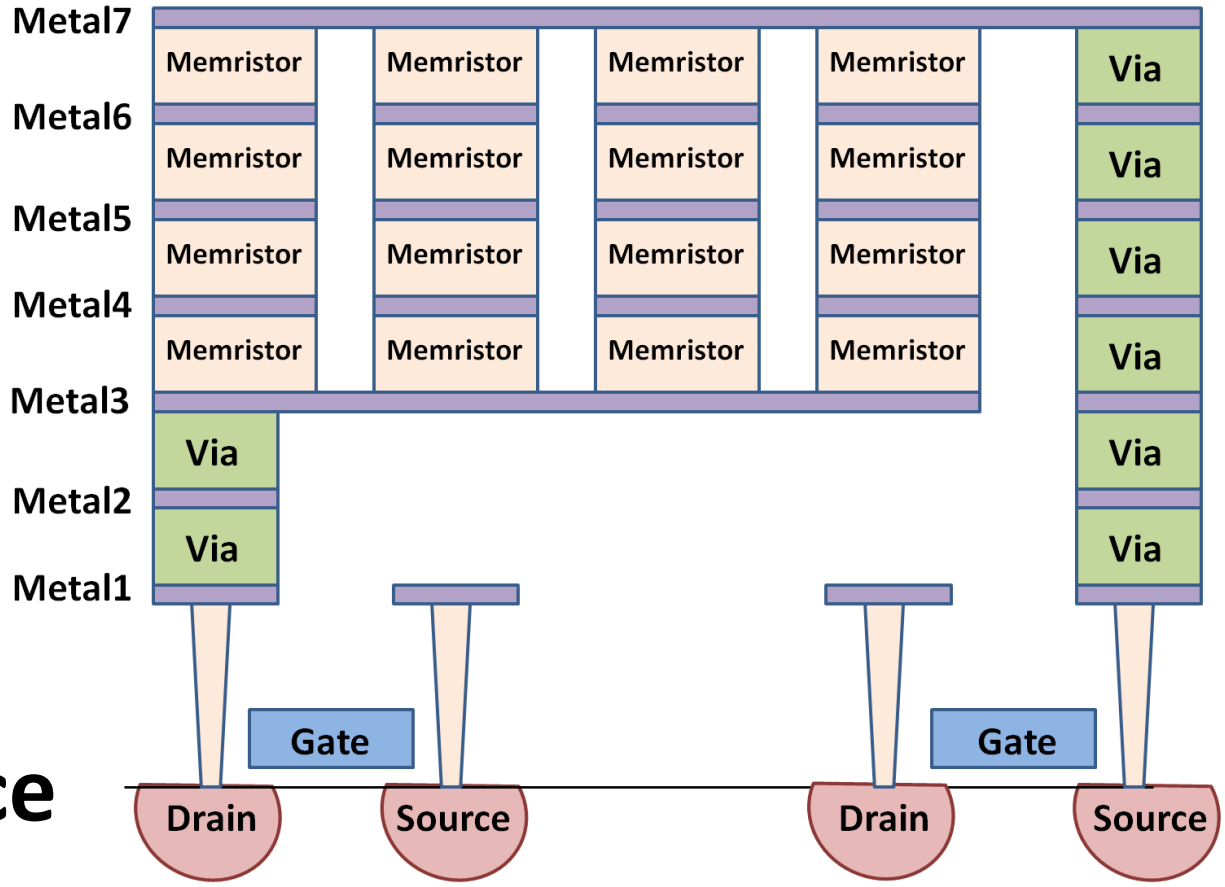
Dense

Nonvolatile

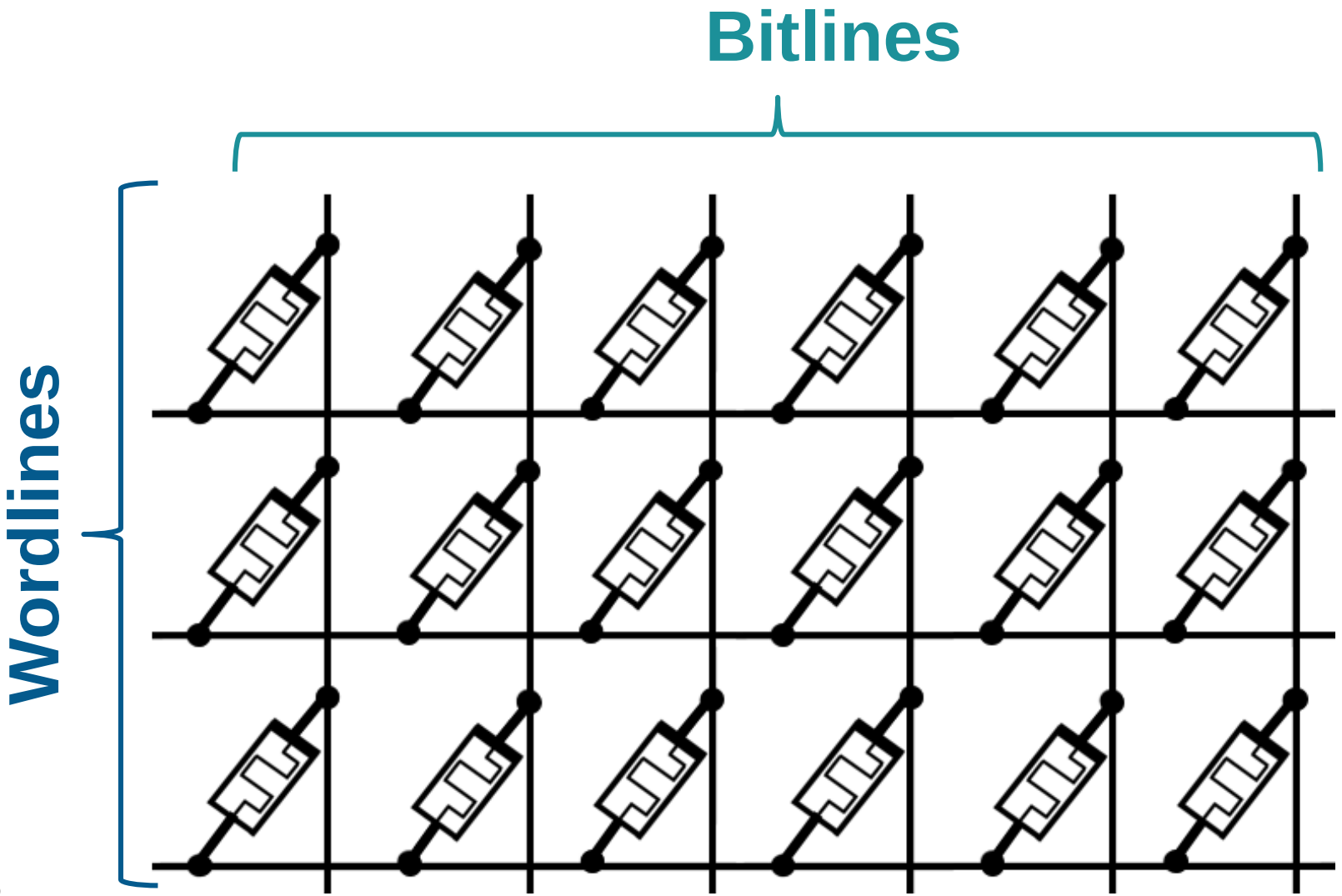
Fast

Low power

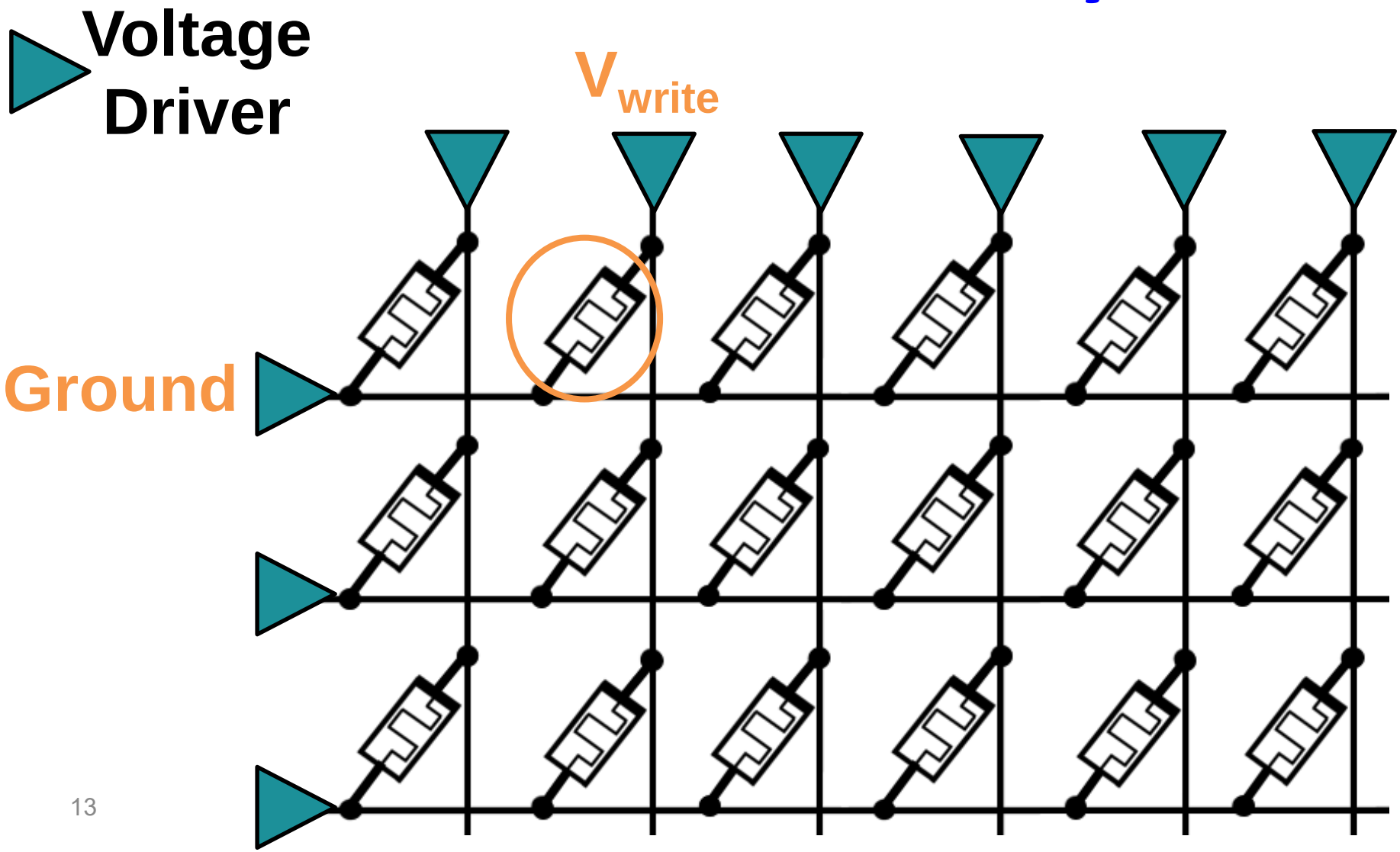
High endurance



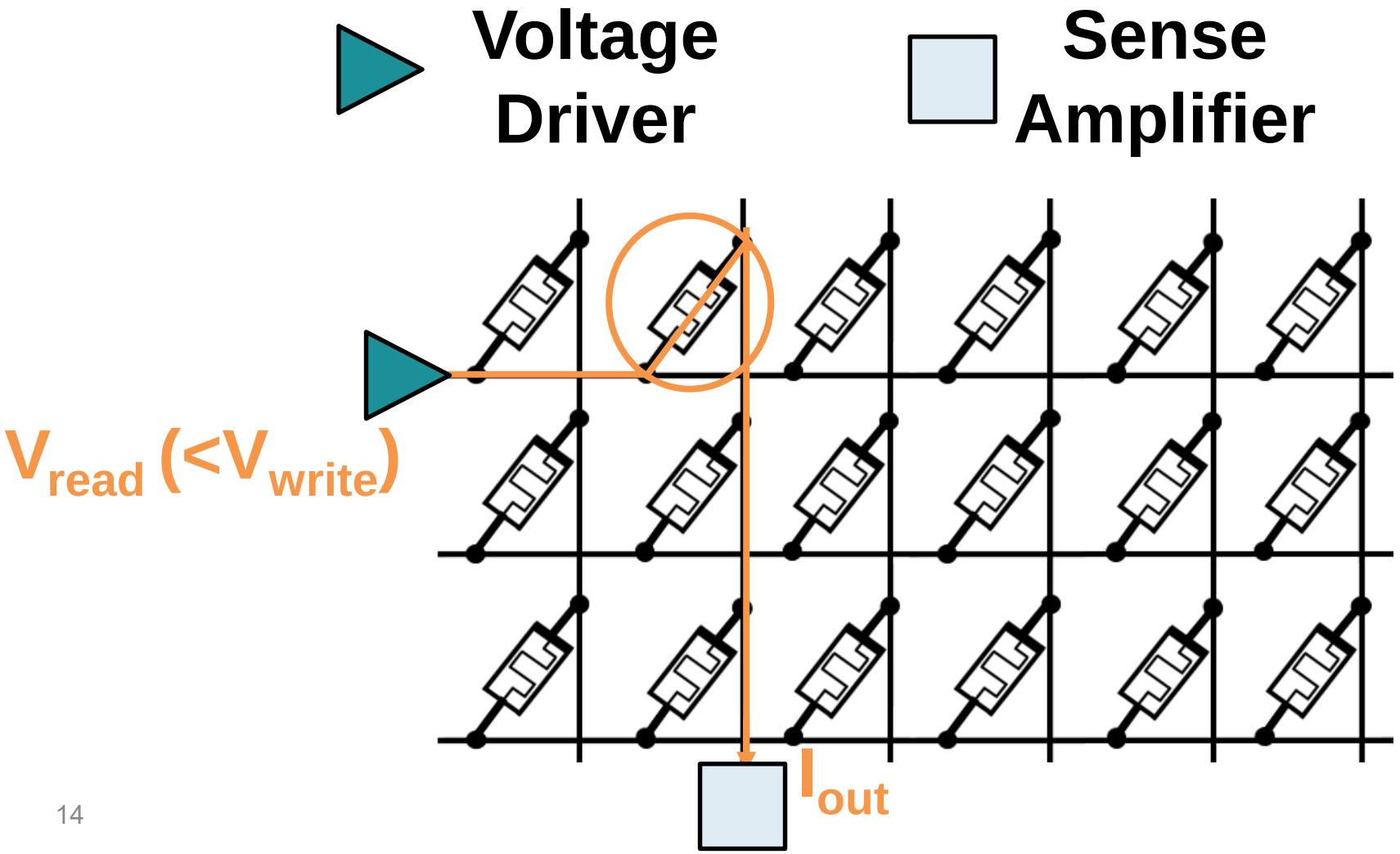
We Want to Compute within the Memristive Crossbar Memory



Write Operation in the Memristive Memory

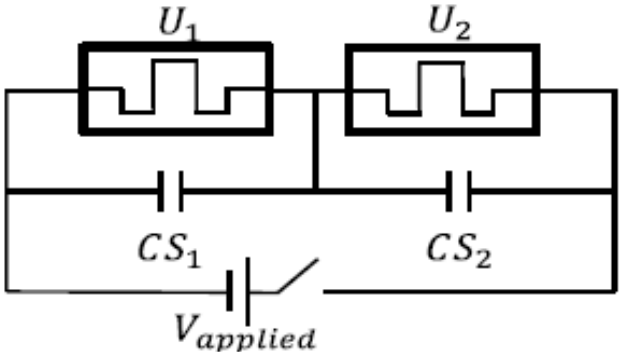


Read Operation in the Memristive Memory



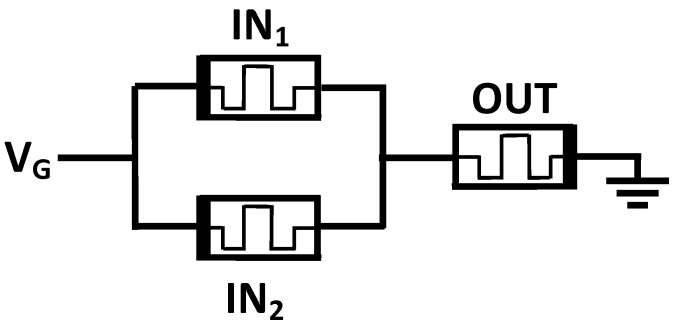
Processing In-Memory with Memristors

Logic Families



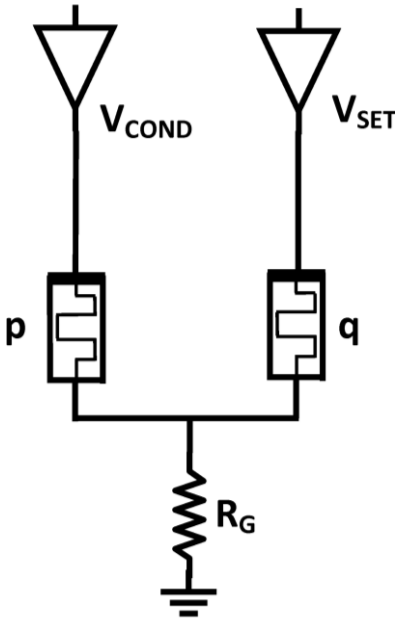
Unipolar logic

Amrani *et al.*, *VLSI-SoC* 2016



MAGIC

Kvatinsky *et al.*, TCAS-II 2014



IMPLY

Borghetti *et al.*, *Nature* 2010

J. Reuben, R. Ben Hur, N. Wald, N. Talati, A. Haj Ali, P.-E. Gaillardon, and S. Kvatinsky, "Memristive Logic: A Framework for Evaluation and Comparison", PATMOS 2017

MAGIC – Memristor Aided LoGIC

Example of MAGIC NOR

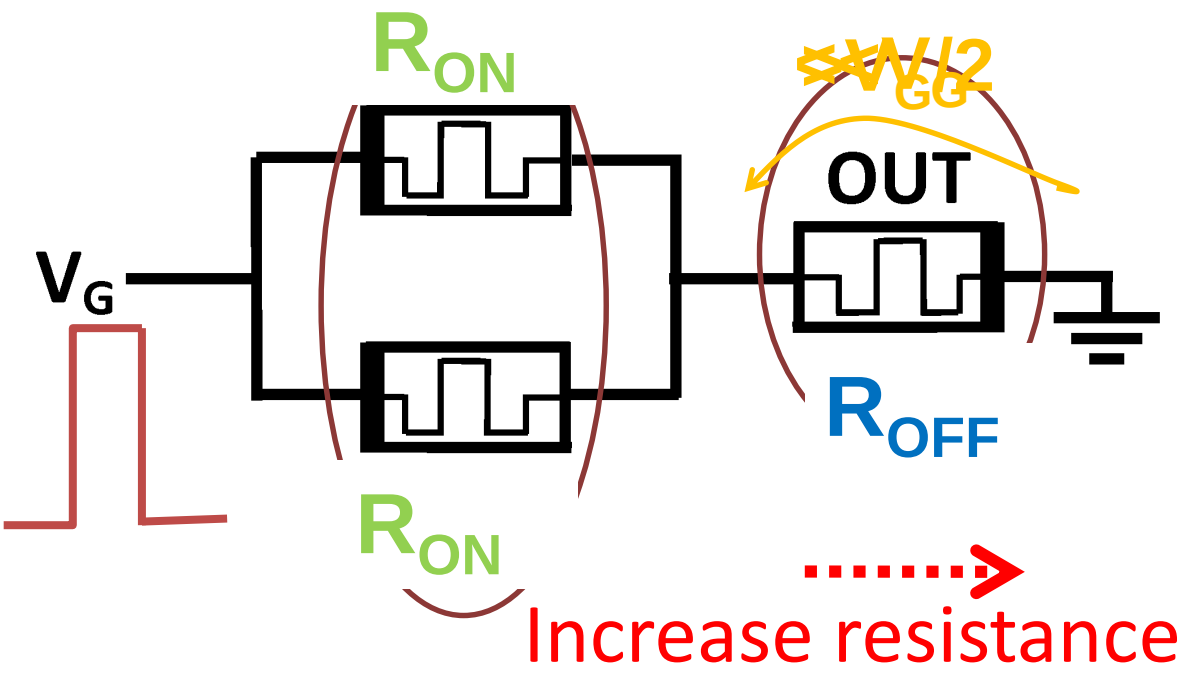
Initialize OUT to R_{ON}

$R_{OFF} >> R_{ON}$

R_{ON} =Logic '1'

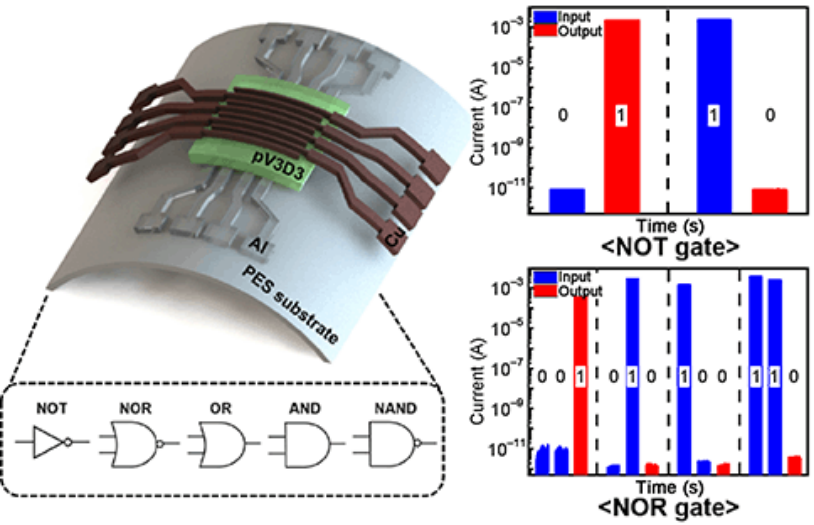
R_{OFF}=Logic '0'

IN_1	IN_2	*NOR*
0	0	1
0	1	0
1	0	0
1	1	0

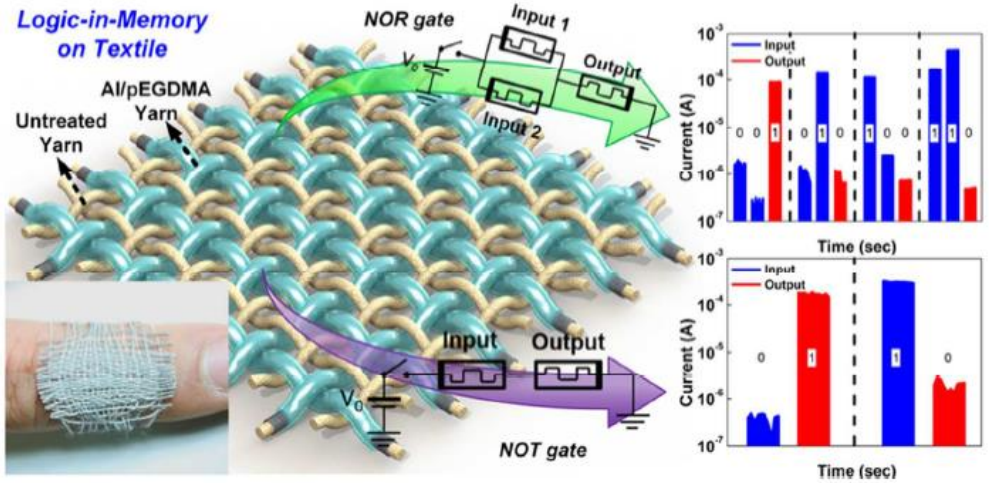


S. Kvatinsky, D. Belousov, S. Liman, G. Satat, N. Wald, E. G. Friedman, A. Kolodny, and U. C. Weiser, "MAGIC – Memristor Aided LoGIC," IEEE TCAS II, Nov. 2014

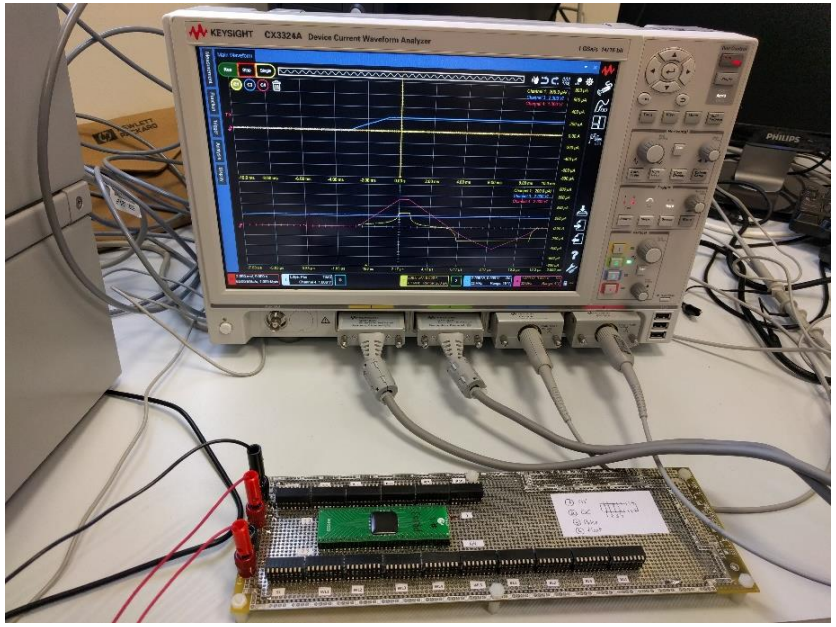
Real MAGIC



Jung *et al.*, *Nano Research*, July 2017



Bae *et al.*, *Nano Letters*, Oct. 2017



ASIC² winbond

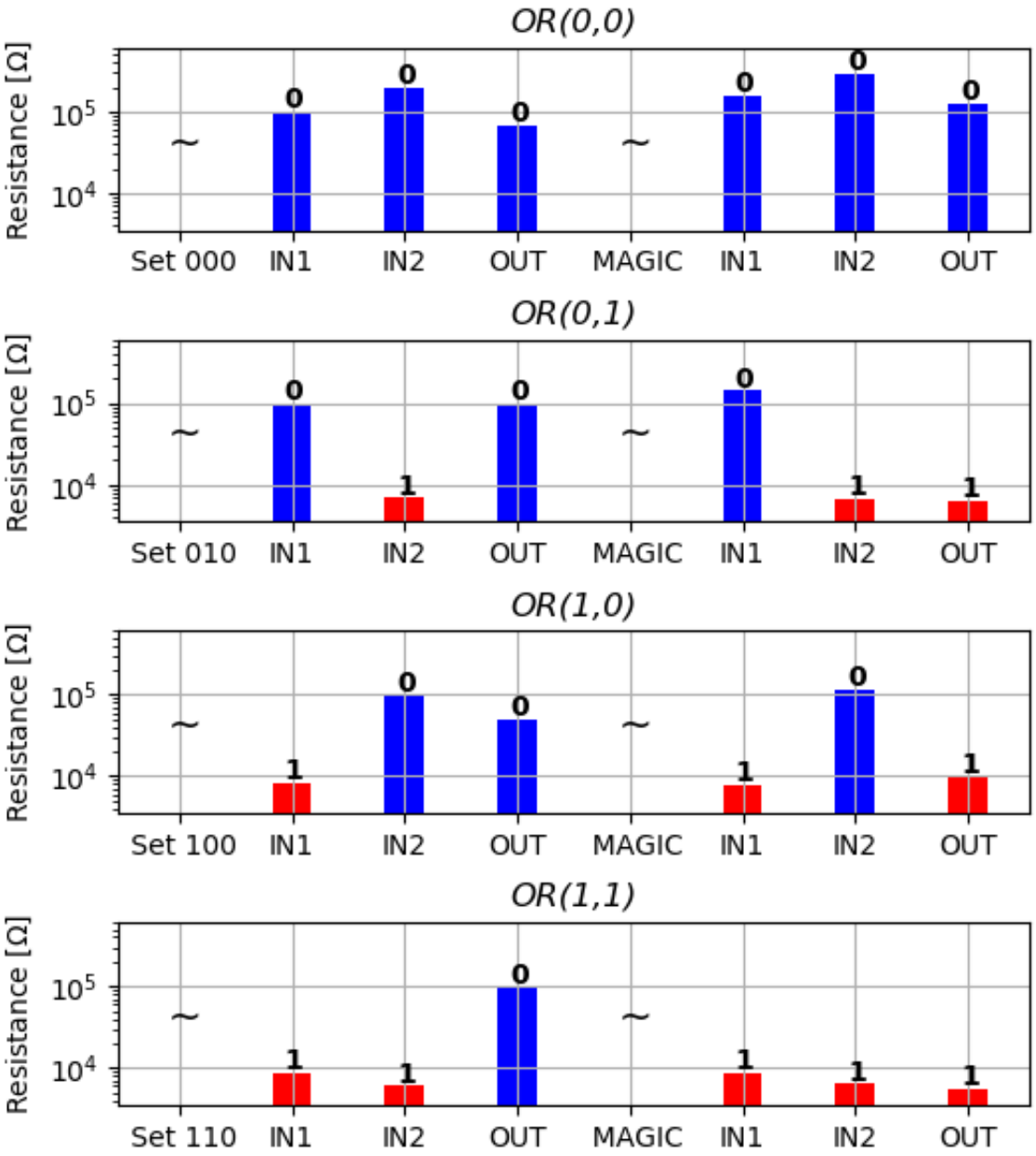
Our lab (HfOx based)

Lab Demonstration

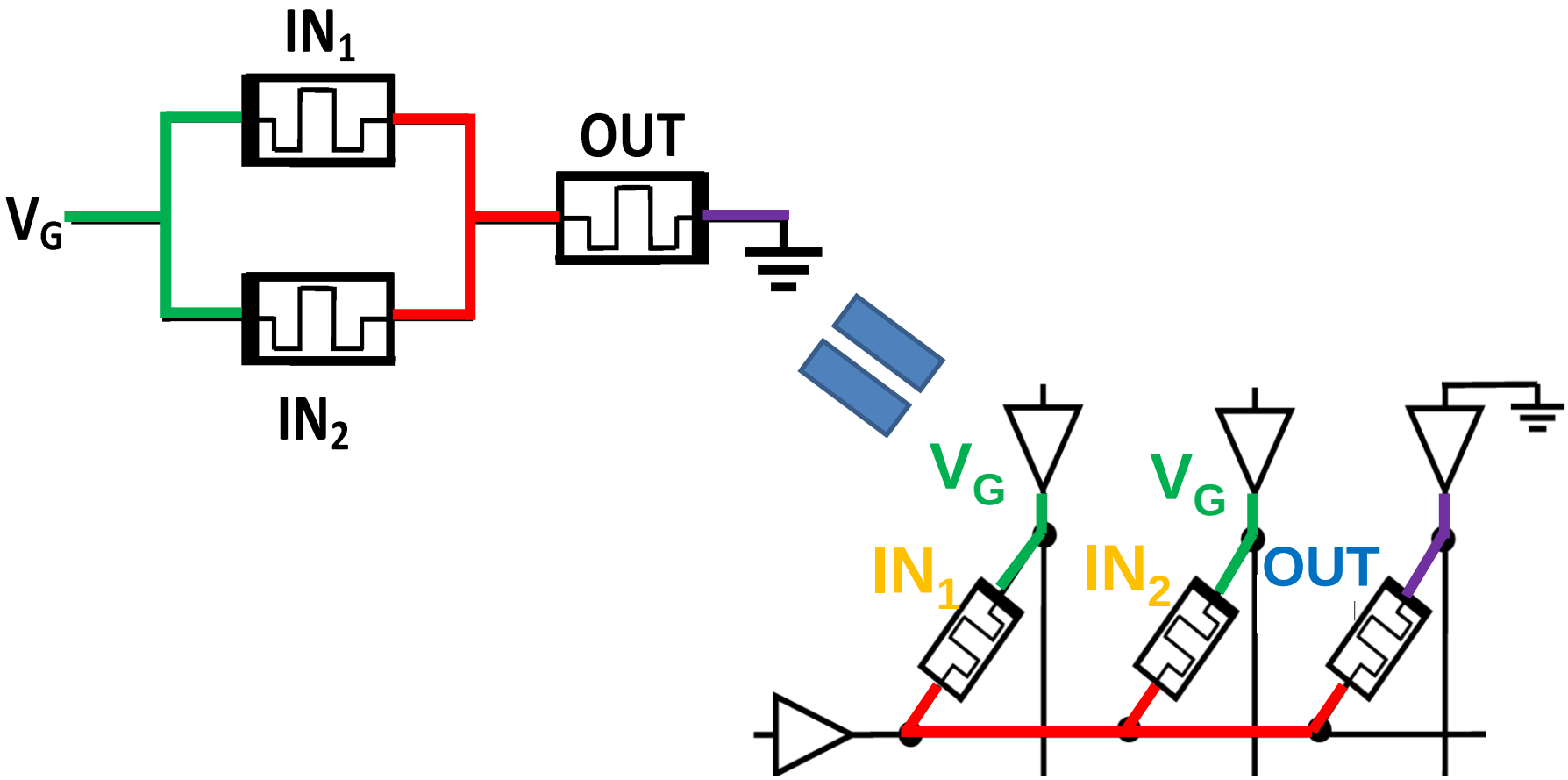
TaOx memristors

Fabricated by Vikas Rana (Julich)
Tested by Barak Hoffer (Technion)

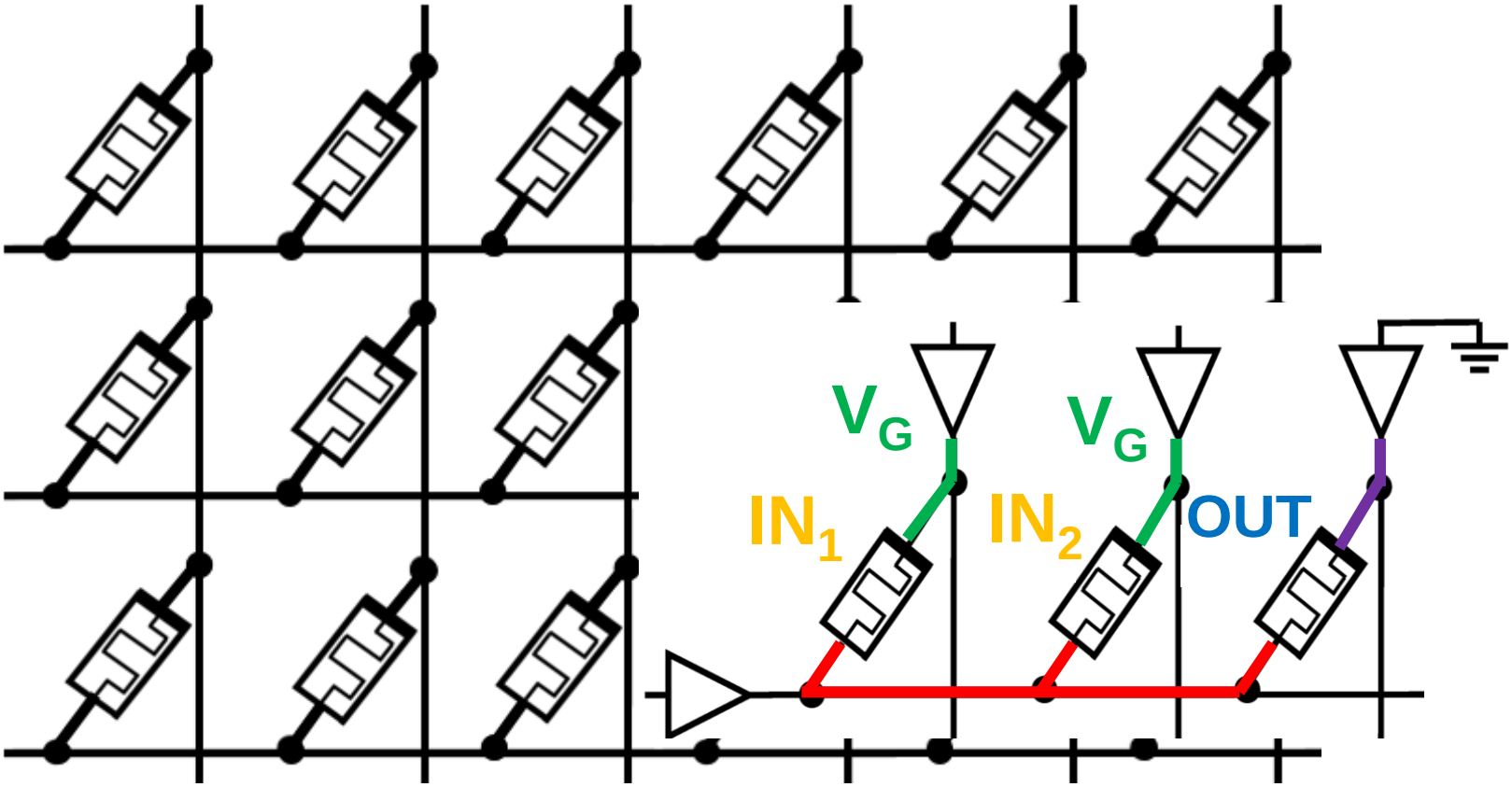
(Unpublished)



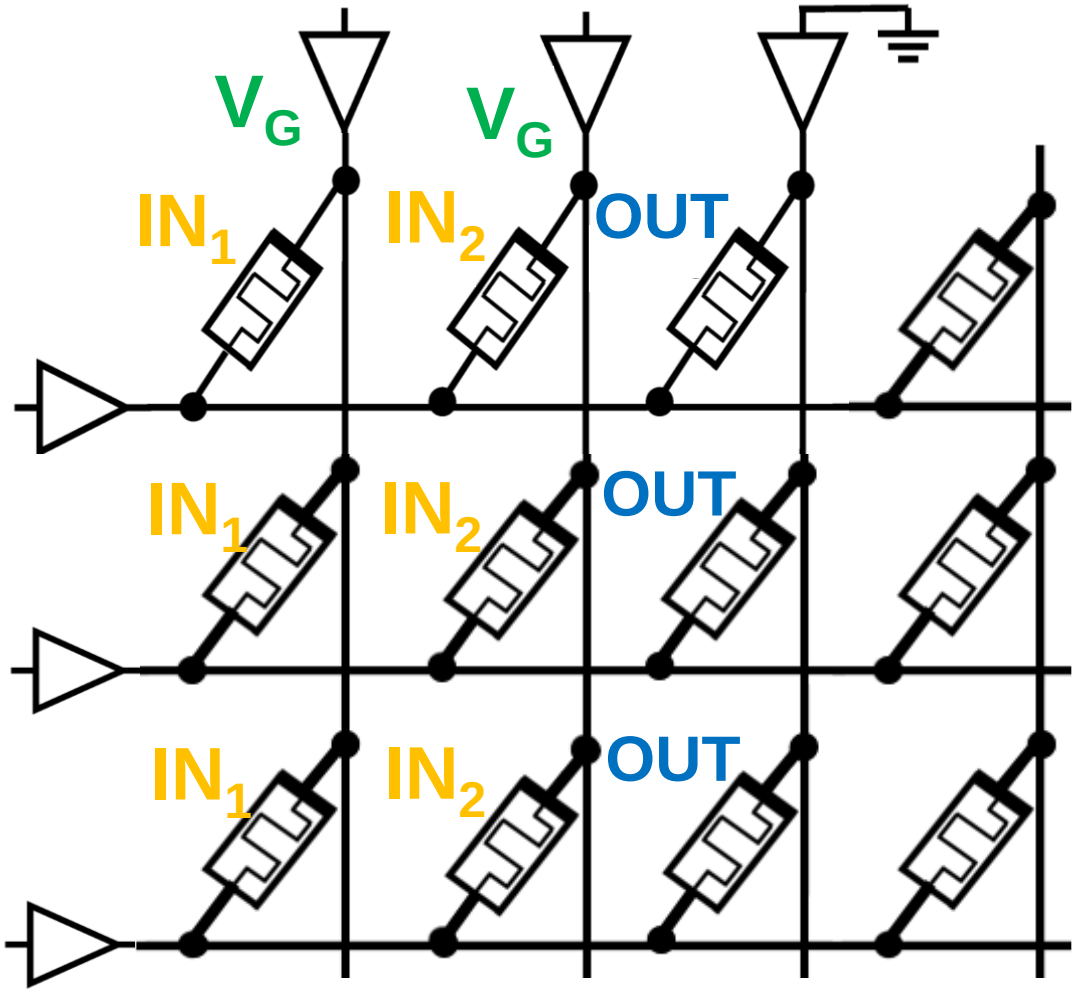
MAGIC NOR in a Crossbar



MAGIC NOR in a Crossbar



MAGIC NOR in a Memristive Memory



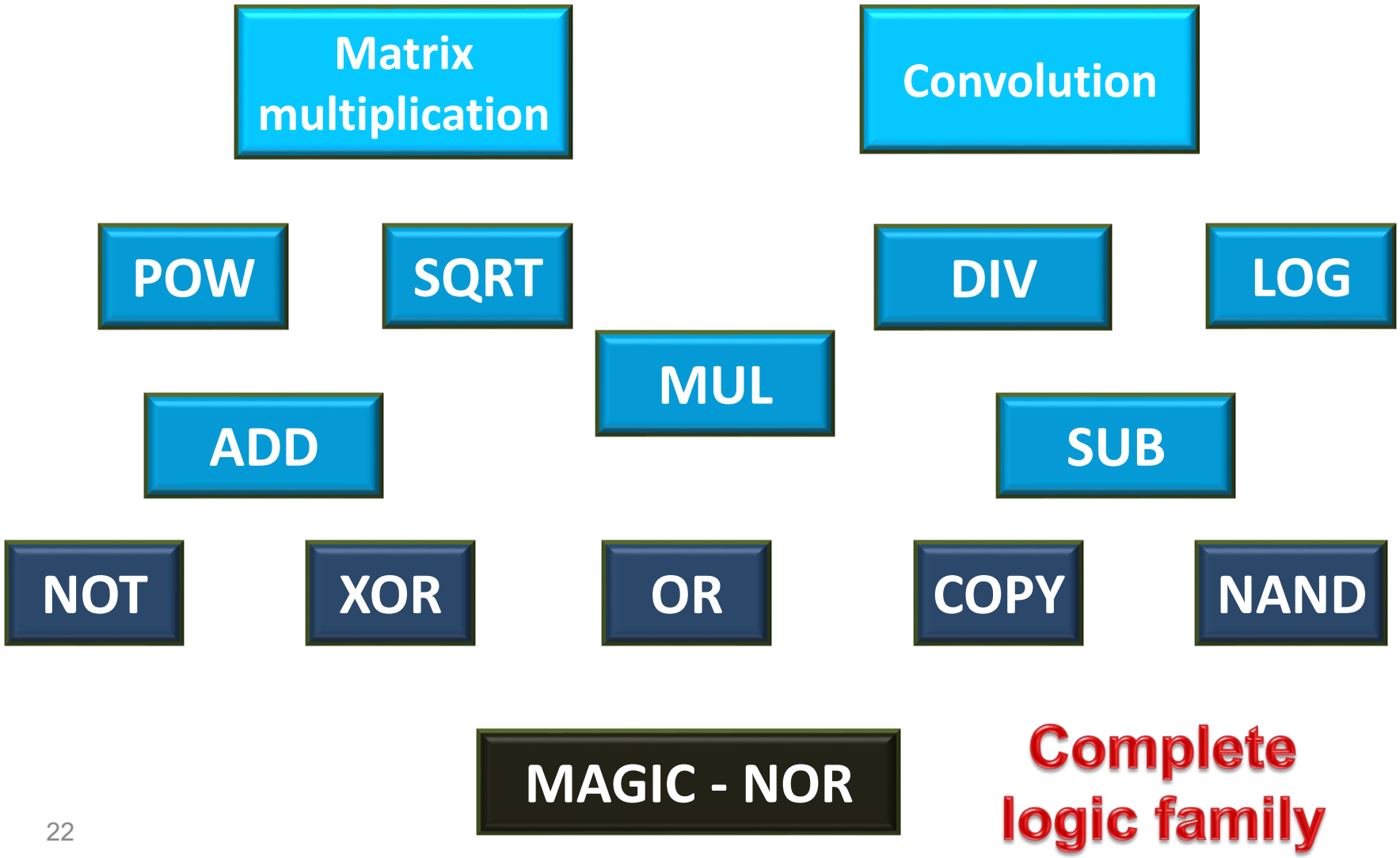
Parallelism

↓

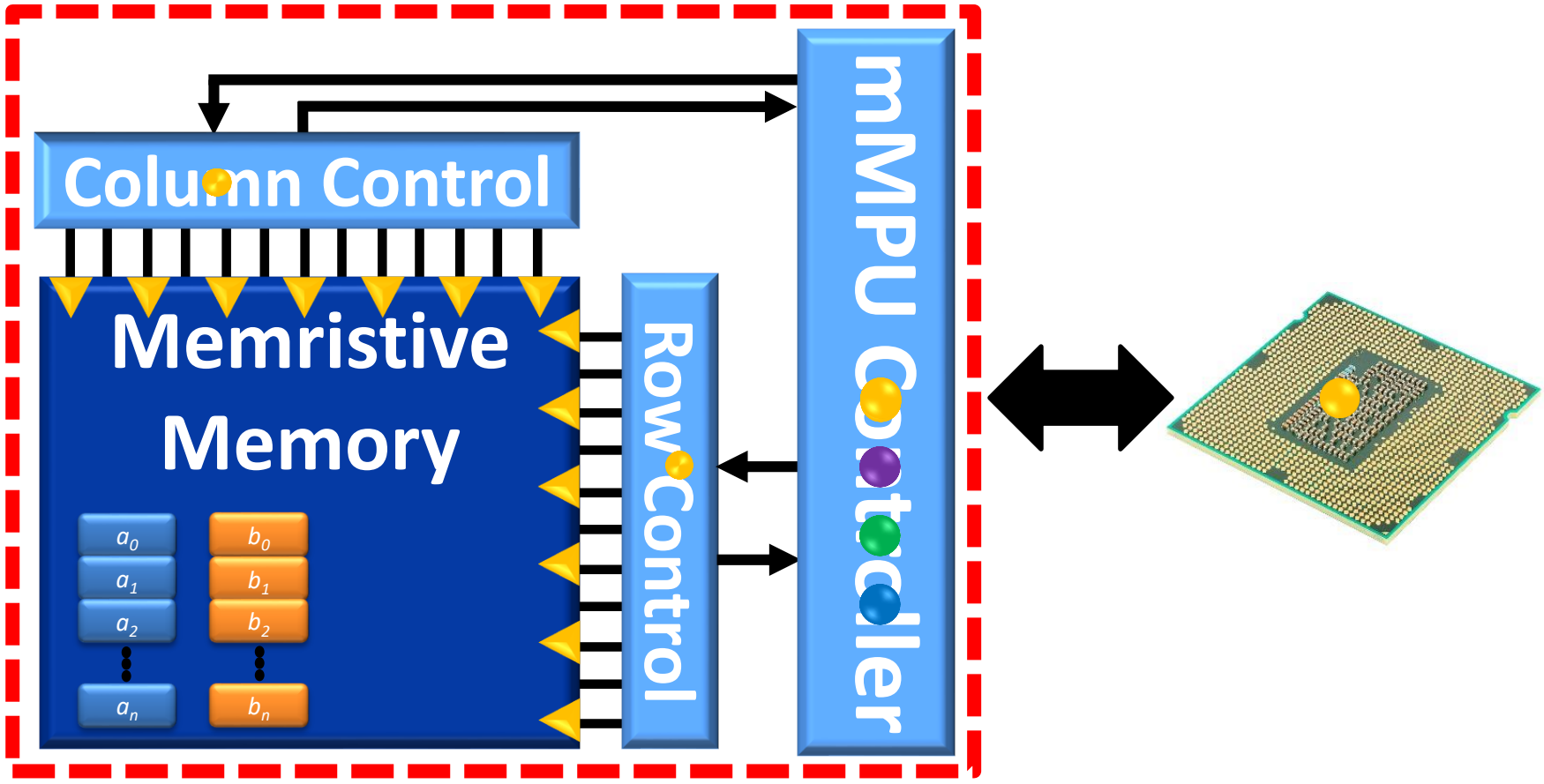
SIMD

N. Talati, S. Gupta, P. Mane, and S. Kvatinsky, "Logic Design within Memristive Memories Using MAGIC," IEEE Transactions on Nanotechnology, July 2016

Hierarchy of Logical Functions

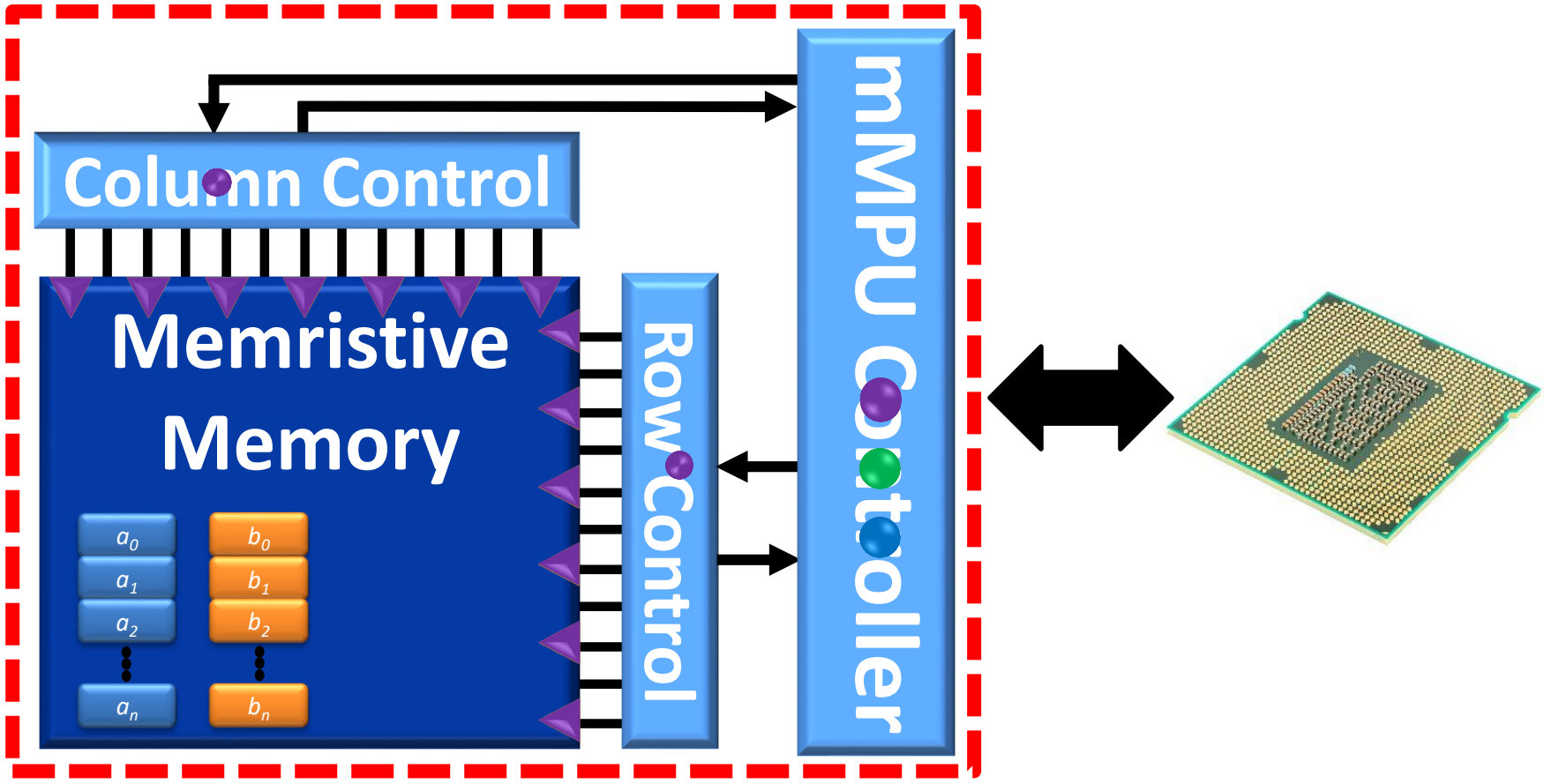


mMPU µArchitecture



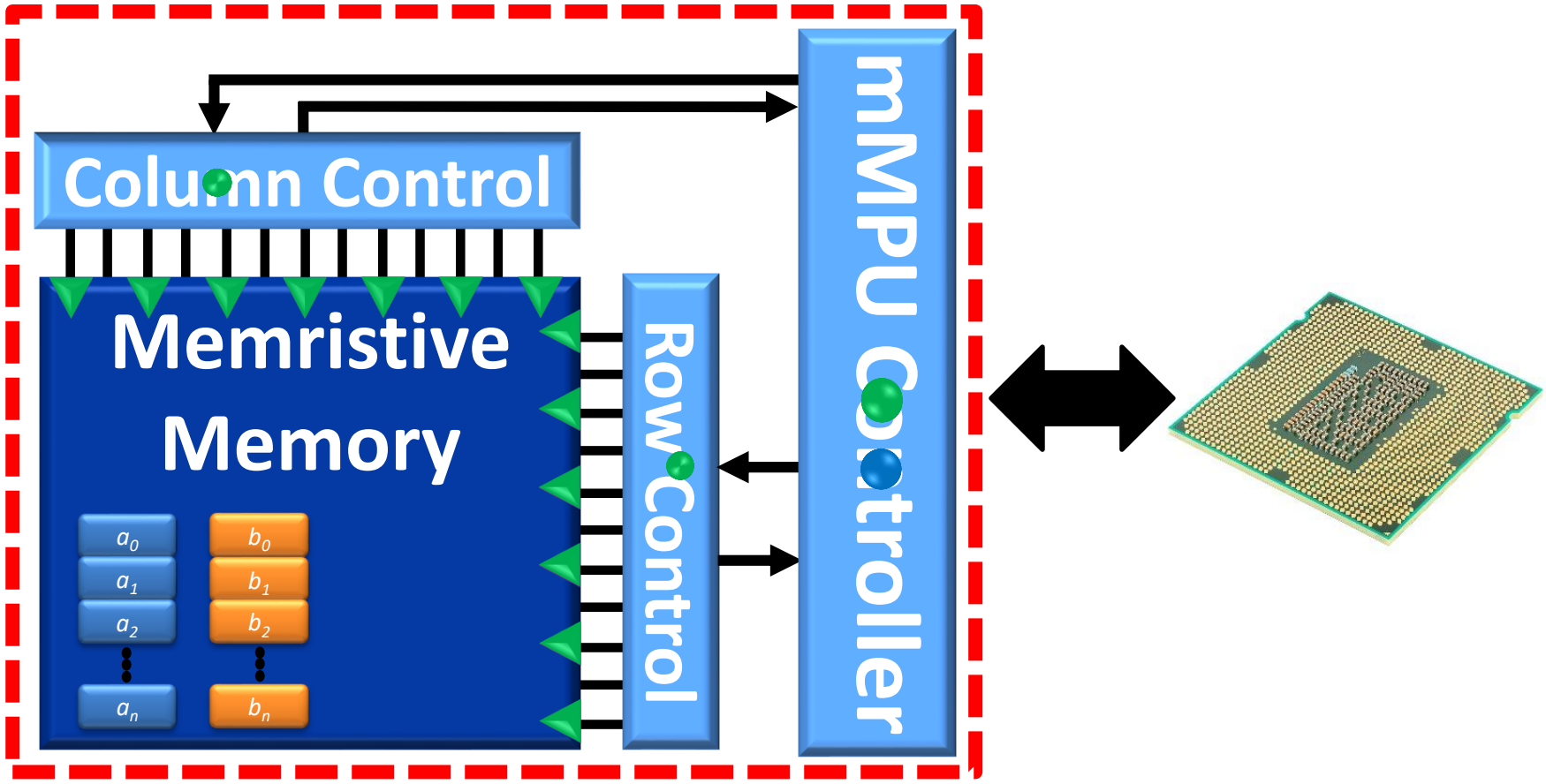
R. Ben-Hur and S. Kvatinsky, "Memory Processing Unit for In-Memory Processing," NANOARCH 2016

mMPU µArchitecture



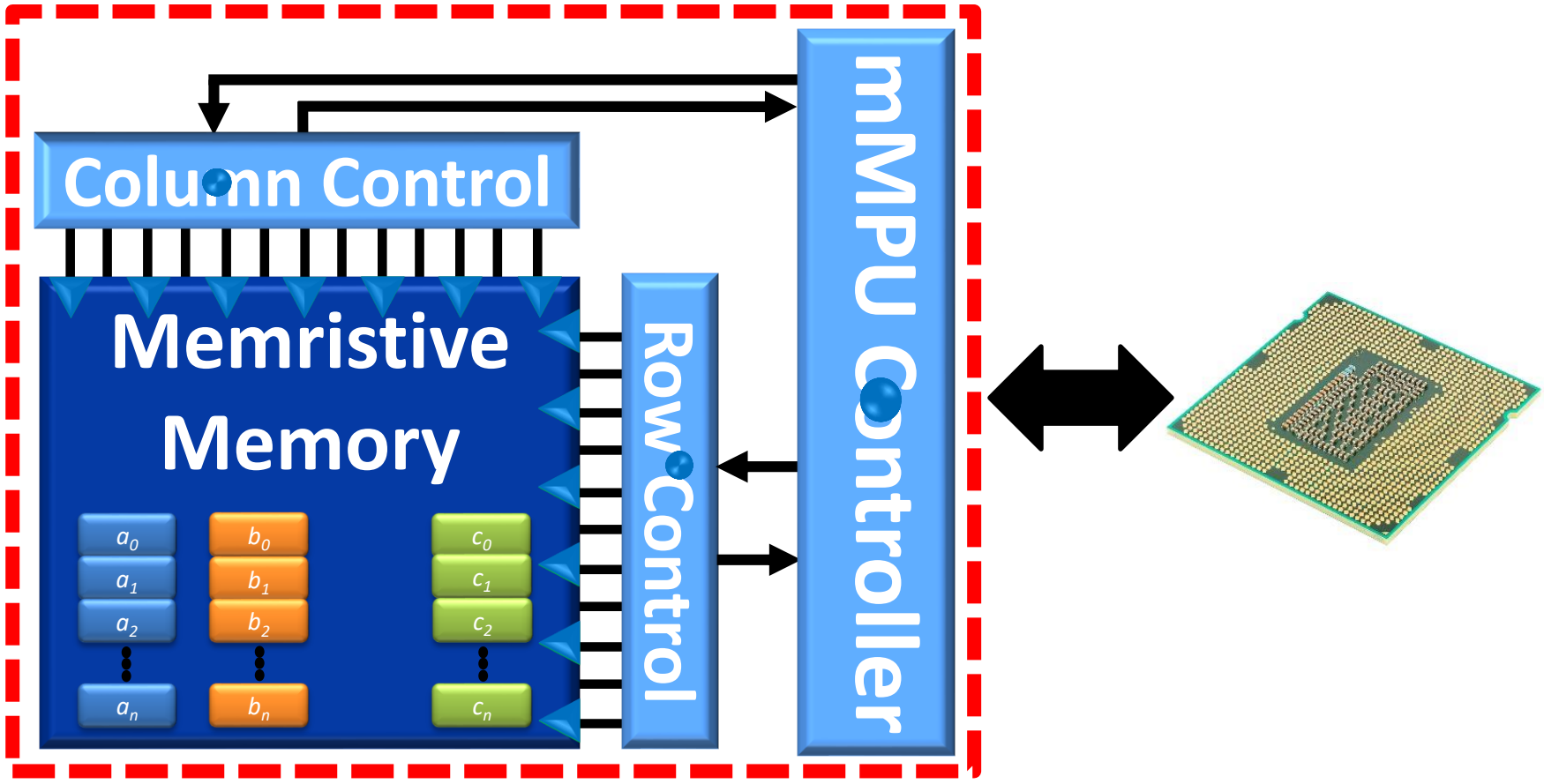
R. Ben-Hur and S. Kvatinsky, "Memory Processing Unit for In-Memory Processing," NANOARCH 2016

mMPU µArchitecture



R. Ben-Hur and S. Kvatinsky, "Memory Processing Unit for In-Memory Processing," NANOARCH 2016

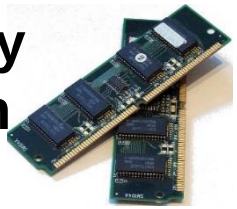
mMPU µArchitecture



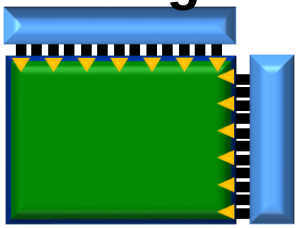
R. Ben-Hur and S. Kvatinsky, "Memory Processing Unit for In-Memory Processing," NANOARCH 2016

Issues Involved in mMPU System

Memory Design



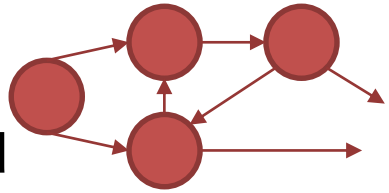
Periphery Design



Programming Model

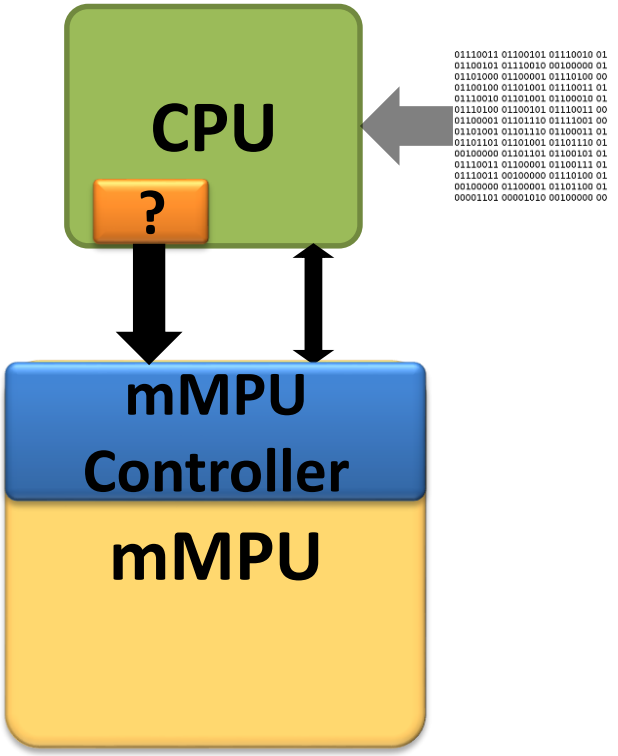
Software

mMPU Controller Design and Optimization



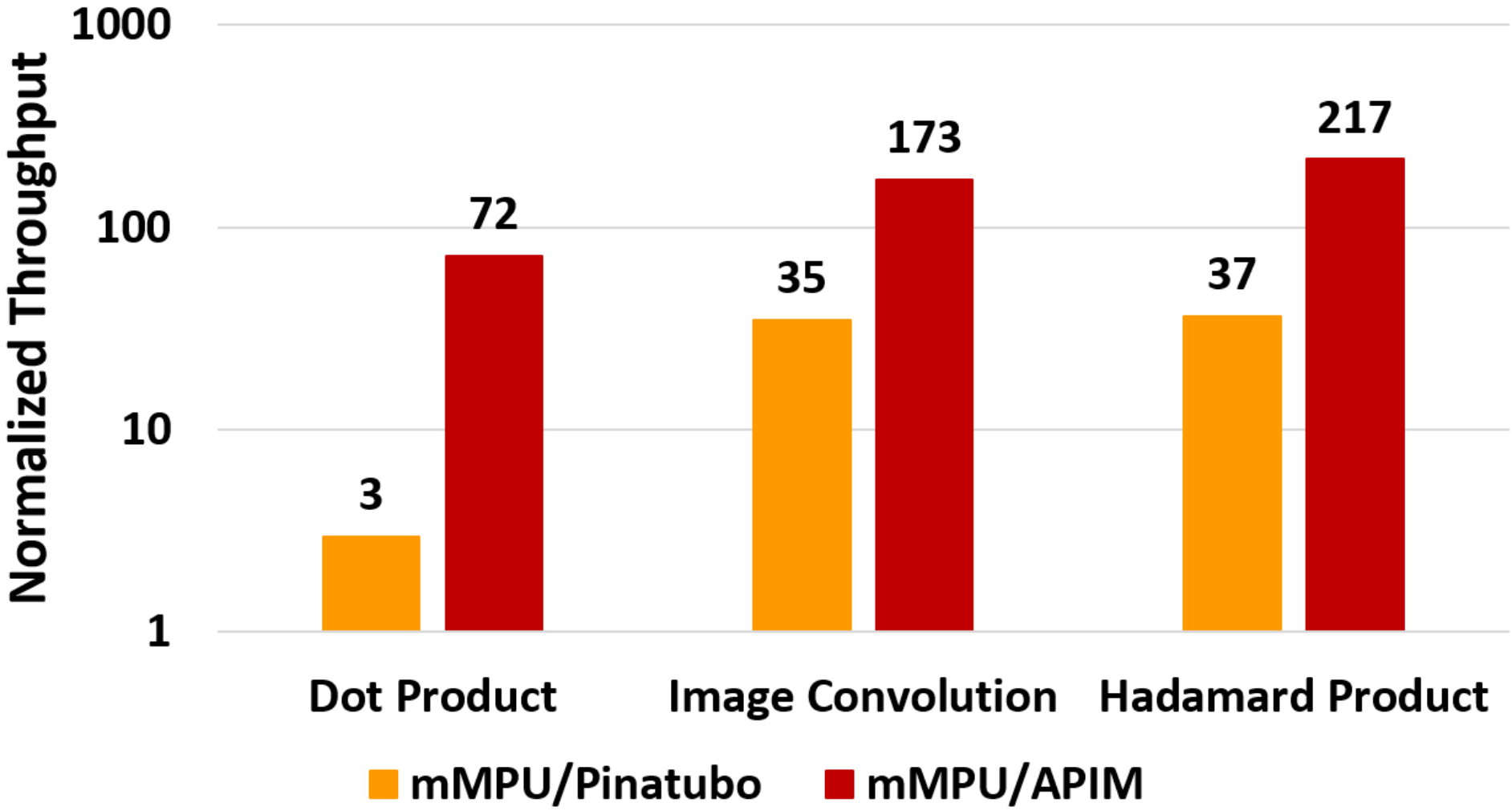
Applications

mMPU System



European Research Council

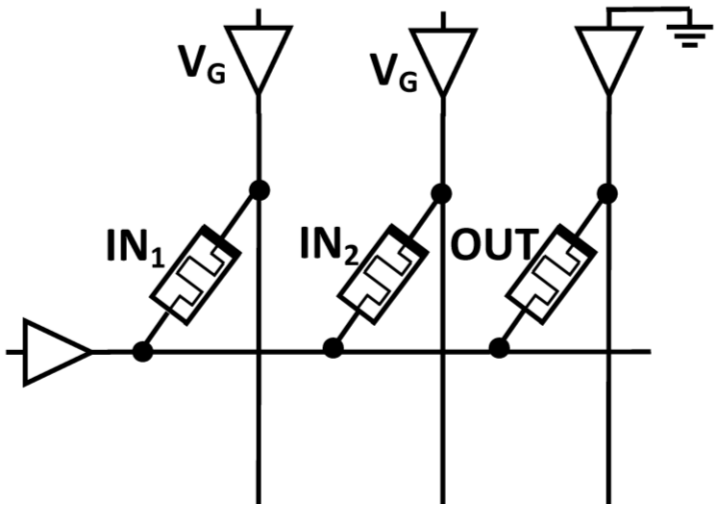
mMPU Performance Potential

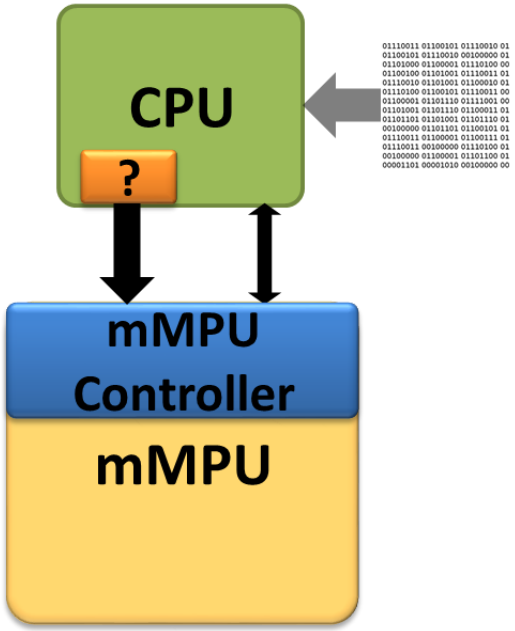


A. Haj-Ali *et al.*, "IMAGING - In-Memory AlGorithms for Image processING," IEEE TCAS I, December 2018
A. Haj-Ali *et al.*, "Not in Name Alone: a Memristive Memory Processing Unit for Real In-Memory Processing," IEEE Micro, October 2018
S. Li *et al.* , "Pinatubo: A Processing-in-Memory Architecture for Bulk Bitwise Operations in Emerging Non-volatile Memories," DAC 2016
M. Imani *et al.*, "Ultra-Efficient Processing In-Memory for Data Intensive Applications," DAC 2017

mMPU – Huge Potential

- Memristors enable non-von Neumann machines to overcome the memory wall
- mMPU – real processing in memory
  → new computing paradigm
- Our aim is to develop a working end-to-end mMPU system





Thanks!

ASIC² ARCHITECTURES SYSTEMS INTELLIGENT COMPUTING INTEGRATED CIRCUITS

Hiroshi Fujiwara Cyber Security Research Center

Prime Minister's Office National Cyber Bureau

HiPer

משרד המדע והטכנולוגיה

Israel Ministry of Science and Technology

ISRAEL SCIENCE FOUNDATION

RBNI

TCE technion computer engineering center

ICRI-CI Intel Collaborative Research Institute Computational Intelligence

Advanced Circuit Research Center ACRC

erc European Research Council

HUAWEI

רשות החדשנות Israel Innovation Authority

CISCO

BSF United States – Israel Binational Science Foundation