SAND2012-2792C

ReRAM: How Advances in Nanotechnology Will Enable the Next Generation of Exascale Computers

April 6, 2012

Matt Marinella

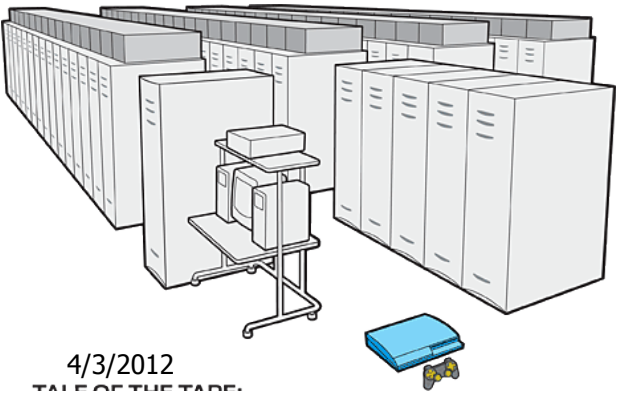
mmarine@sandia.gov

Outline

- Intro to High Performance Computing
- The Exascale Challenge
- Solutions Ready Today
- Opportunities for Emerging NVM
  - Storage & Hybrid Main Memory
  - Universal Memory
- Conclusions

Supercomputers

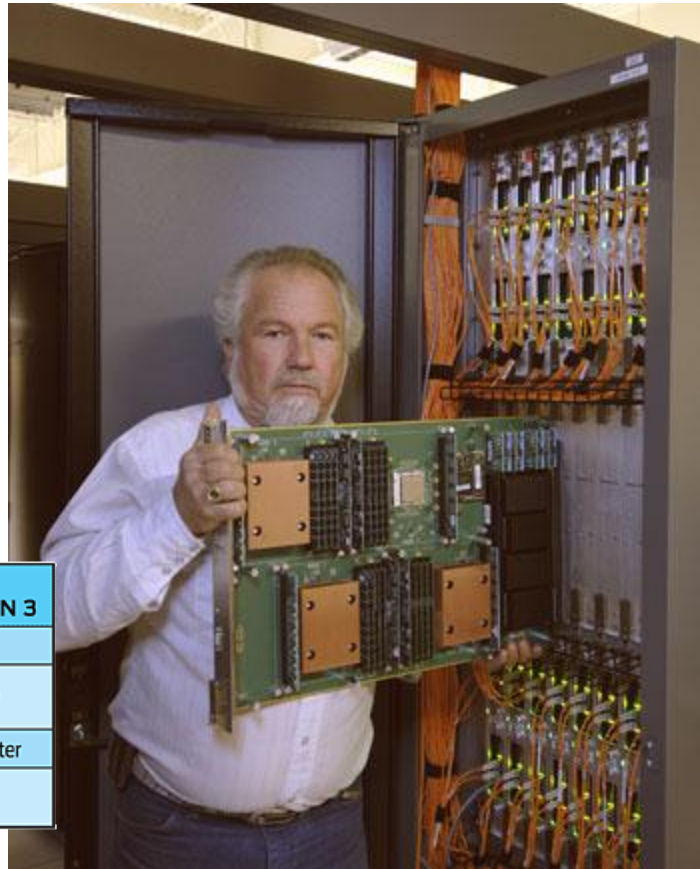
- Define the forefront of computing power
- Modern architectures are massively parallel systems
- Thousands of off the shelf processors
  - GPUs or CPUs
- Performance measured in FLOPS:
  - Floating point ops per second
- Benchmarked by LINPACK
  - Solve *n* x *n* system of linear eqns
- Out of date quickly



TALE OF THE TAPE: SUPERCOMPUTER VS. GAME CONSOLE

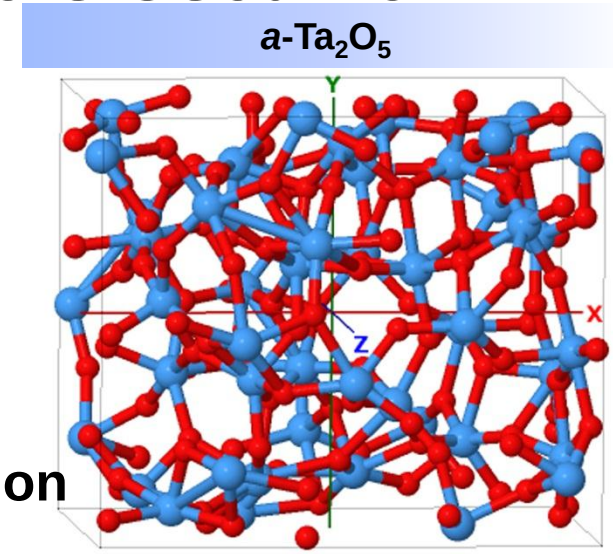
	SANDIA LAB'S ASCI RED	SONY PLAYSTATION 3
DATE OF ORIGIN	1997	2006
PEAK PERFORMANCE	1.8 teraflops	1.8 teraflops*
PHYSICAL SIZE	150 square meters	0.08 square meter
POWER CONSUMPTION	800 000 watts	<200 watts

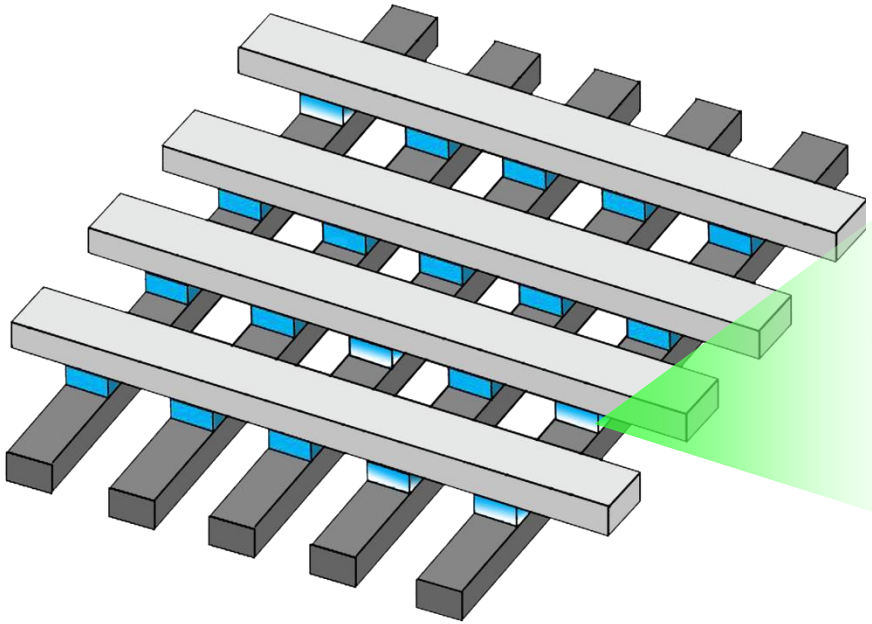
* For GPU; CPU adds another 0.2 teraflops

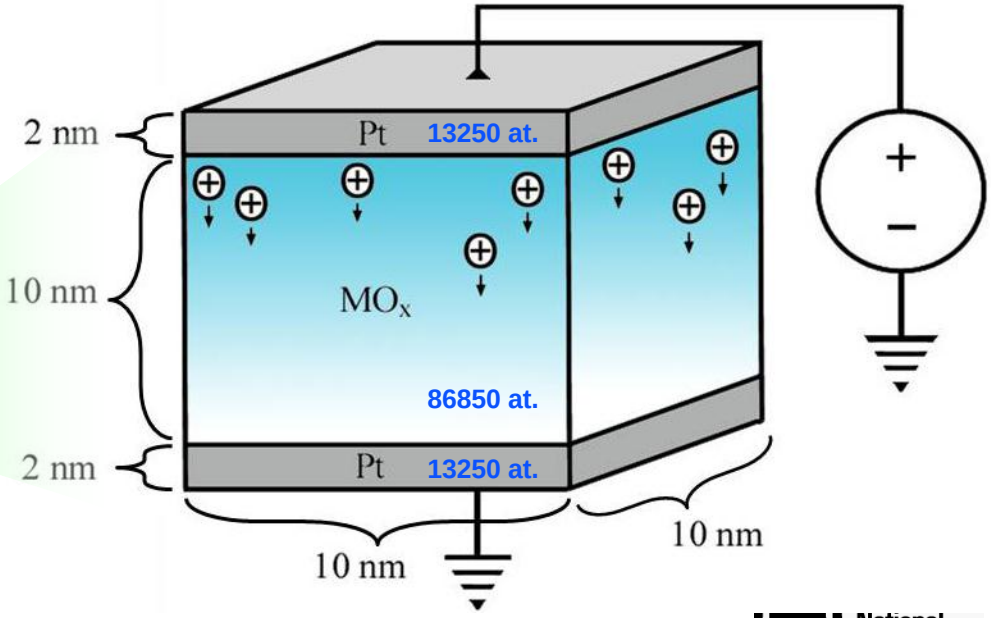


What are Supercomputers Used For?

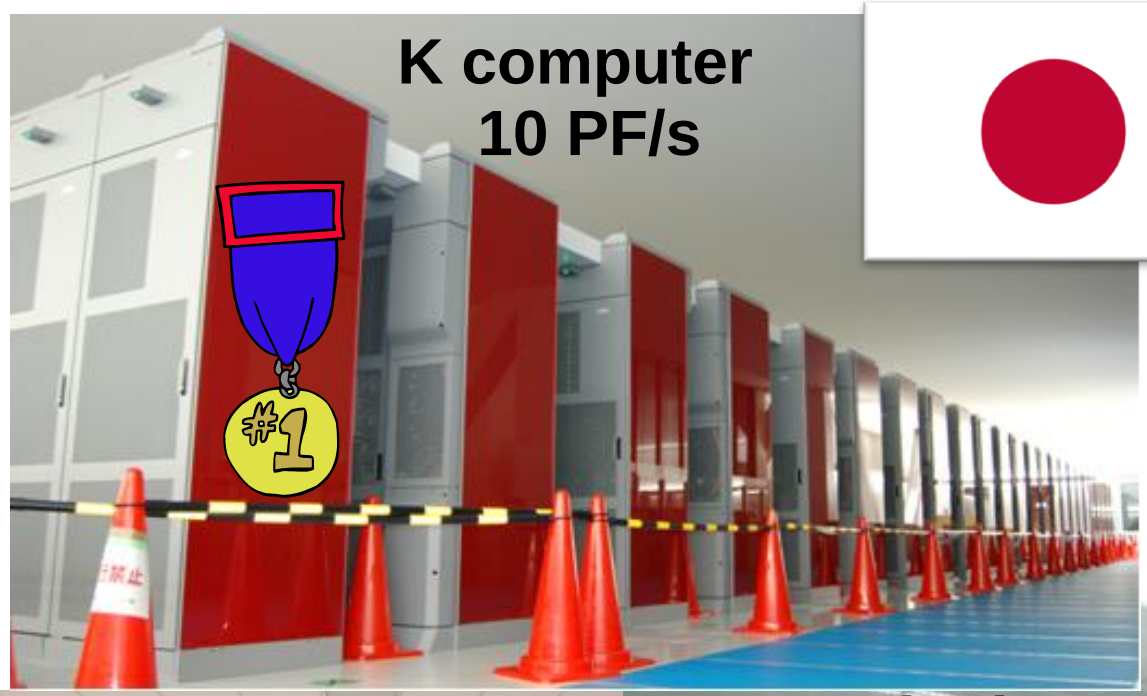
- Predicting weather
- Google searches
- Simulating nuclear explosions
- Quantum physics
- Molecular dynamics & DFT simulation

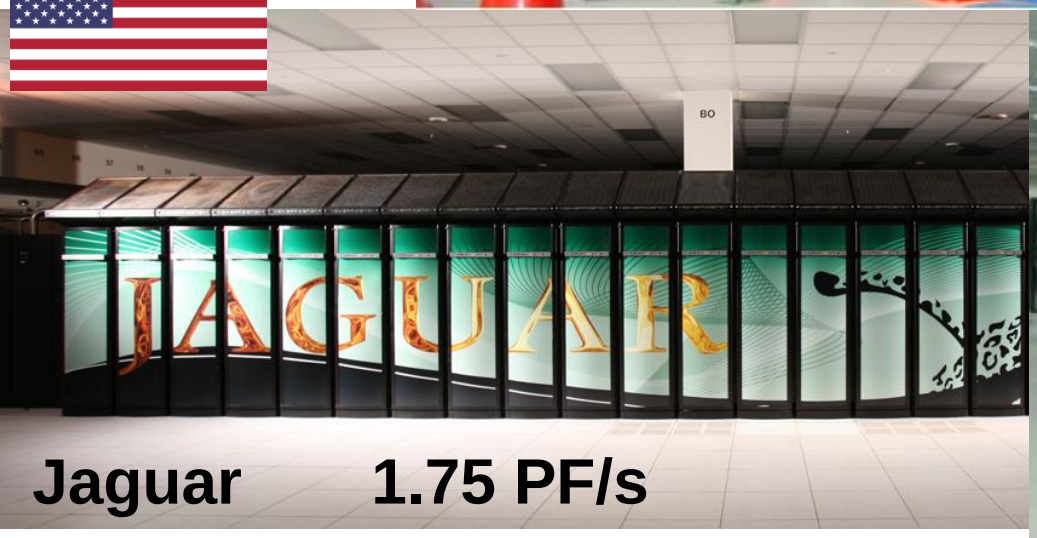






Supercomputing Olympics





The World's Best Computers

- **K computer (RIKEN, Japan)**
  - **Speed: 10.5 petaflops (Rmax)**
  - **Processors: 88k (SPARC64 2.0 GHz)**
  - **Memory: 1.4 PB**
- **Tianhe (China)**
  - **Speed: 2.5 petaflops (Rmax)**
  - **Processors: 186k (NVIDIA 2.93GHz GPU)**
  - **Memory: 229 TB**
- **Roadrunner (Oak Ridge National Lab, US)**
  - **Speed: 1.75 petaflops (Rmax)**
  - **Processors: 224k (Cray Opteron 6-cor, 2.6GHz)**
  - **Memory: 360 TB**
  - **Upgrade to GPUs this year – est. 20 petaflop**

Power

- K computer
  - Power: 13 MW
- Tianhe
  - Power: 4 MW
- Roadrunner
  - Power: 7 MW → enough to power 5000 homes
- Palo Verde Nuclear Generating Station
  - Power: 3 GW ⟵
- Typical Coal Fired Power Plant
  - Power: 500 MW
- 1 MW = $1,000,000/year power bill
- X pJ per operation = X MW per 10^{18} operations/sec (Exaflop)

Will Exascale need dedicated Nuclear Power Plant?

Outline

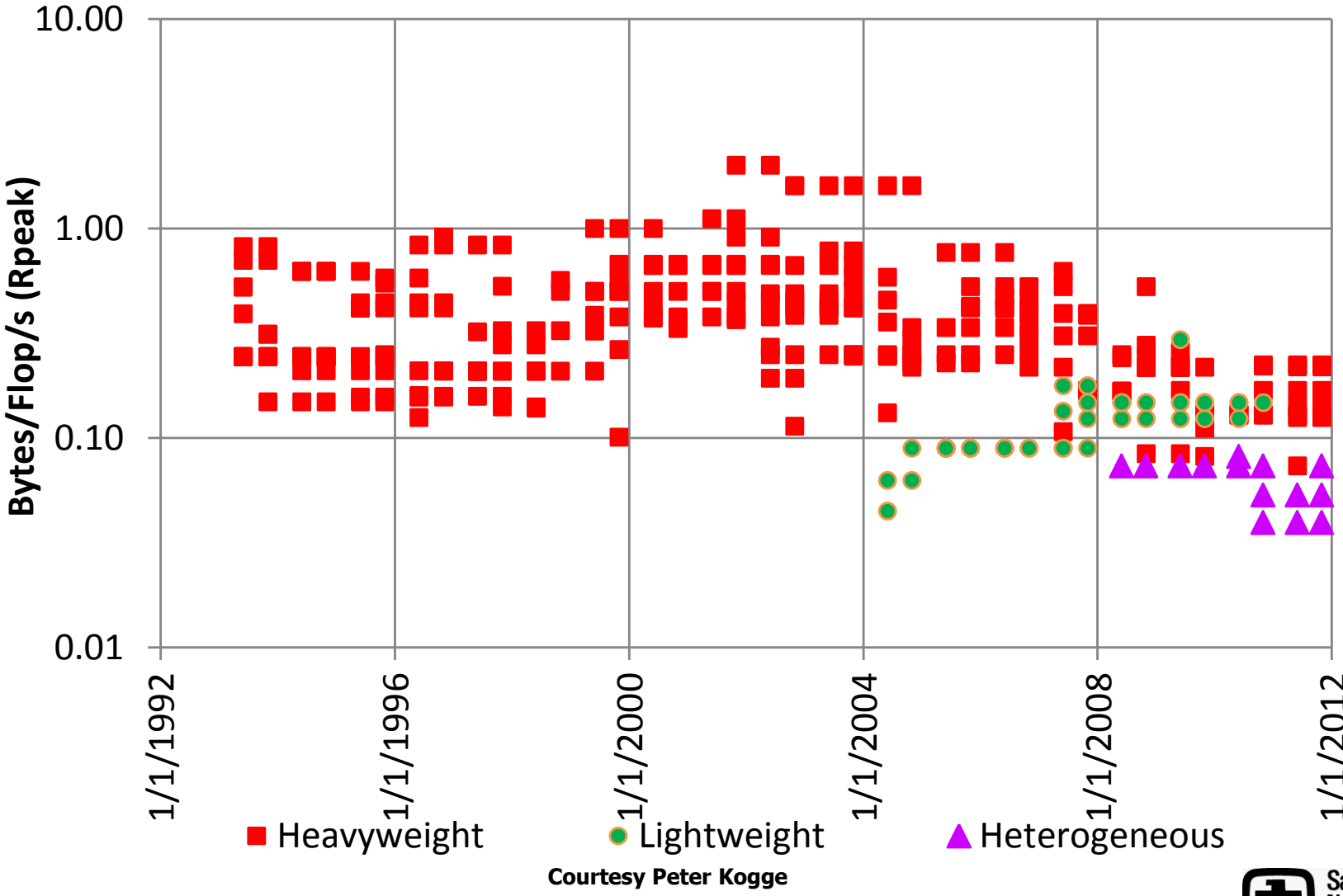
- Intro to High Performance Computing
- The Exascale Challenge
- Solutions Ready Today
- Opportunities for Emerging NVM
  - Storage & Hybrid Main Memory
  - Universal Memory
- Conclusions

Grand Challenges for Exascale

- The Exascale Report List Four Challenges:
  - Energy and Power
  - Memory and Storage
  - Concurrency and Locality
  - Resiliency
- US Dept. of Energy take action: Exascale Initiative

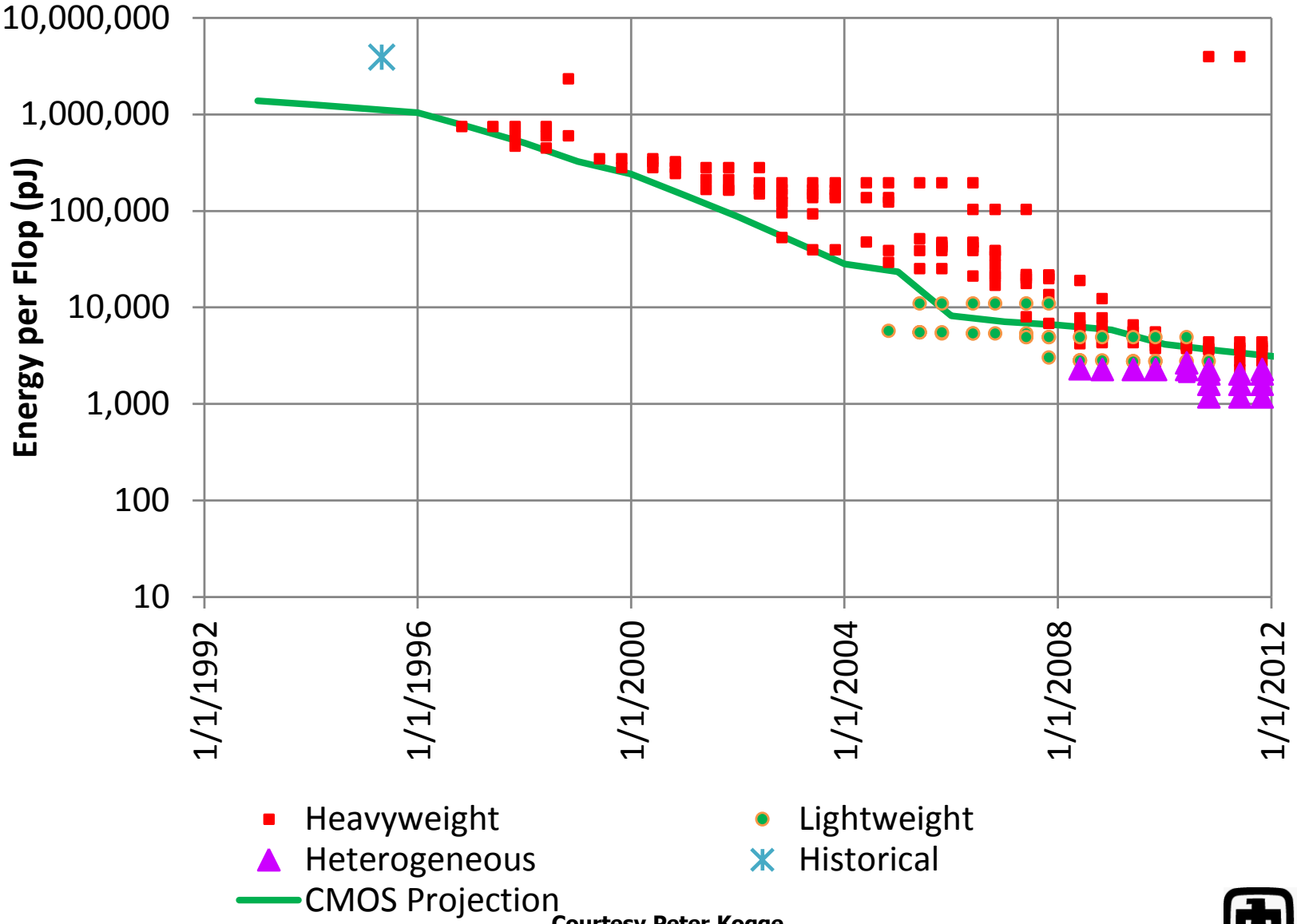
Major role for emerging nonvolatile memories in three of four challenges!

DRAM Byes per Flop



Courtesy Peter Kogge

Energy per Flop



Courtesy Peter Kogge
Matthew Marinella

Proposed Solutions

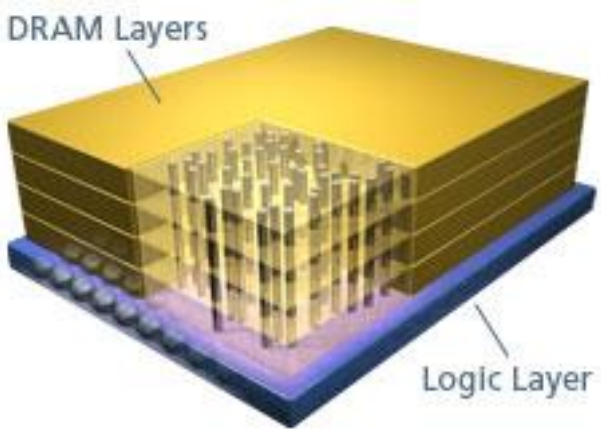
- Through silicon via (TSV) memory stacking
- Optical interconnects
- Near threshold logic
- Emerging nonvolatile memories

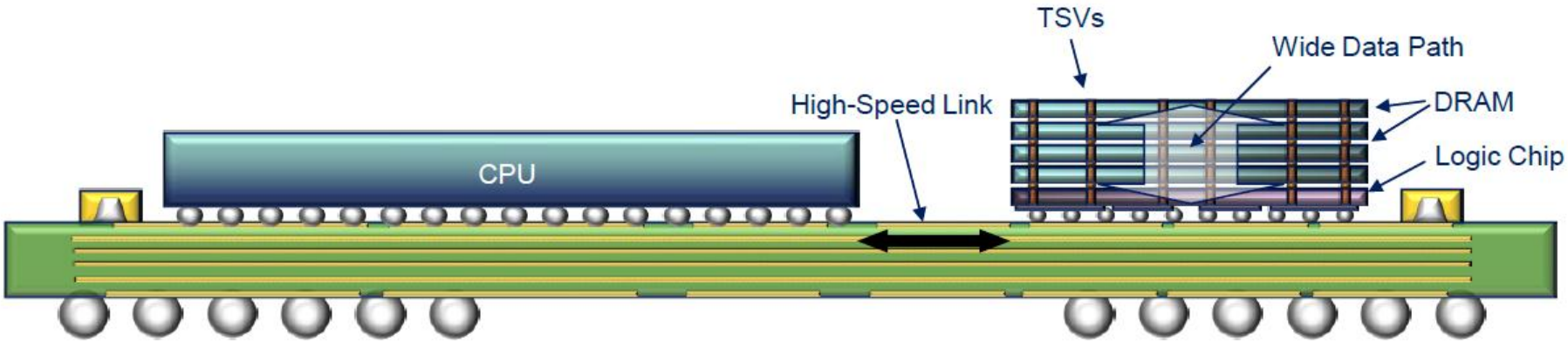
Outline

- Intro to High Performance Computing
- The Exascale Challenge
- Solutions Ready Today
- Opportunities for Emerging NVM
  - Storage & Hybrid Main Memory
  - Universal Memory
- Conclusions

Present Day Solution: 3D DRAM

- **Micron/Intel Hybrid Memory Cube**
- **DRAM die stacked on logic**
- **Connected via through-silicon-via**
- **Very clever combination of today's technologies**



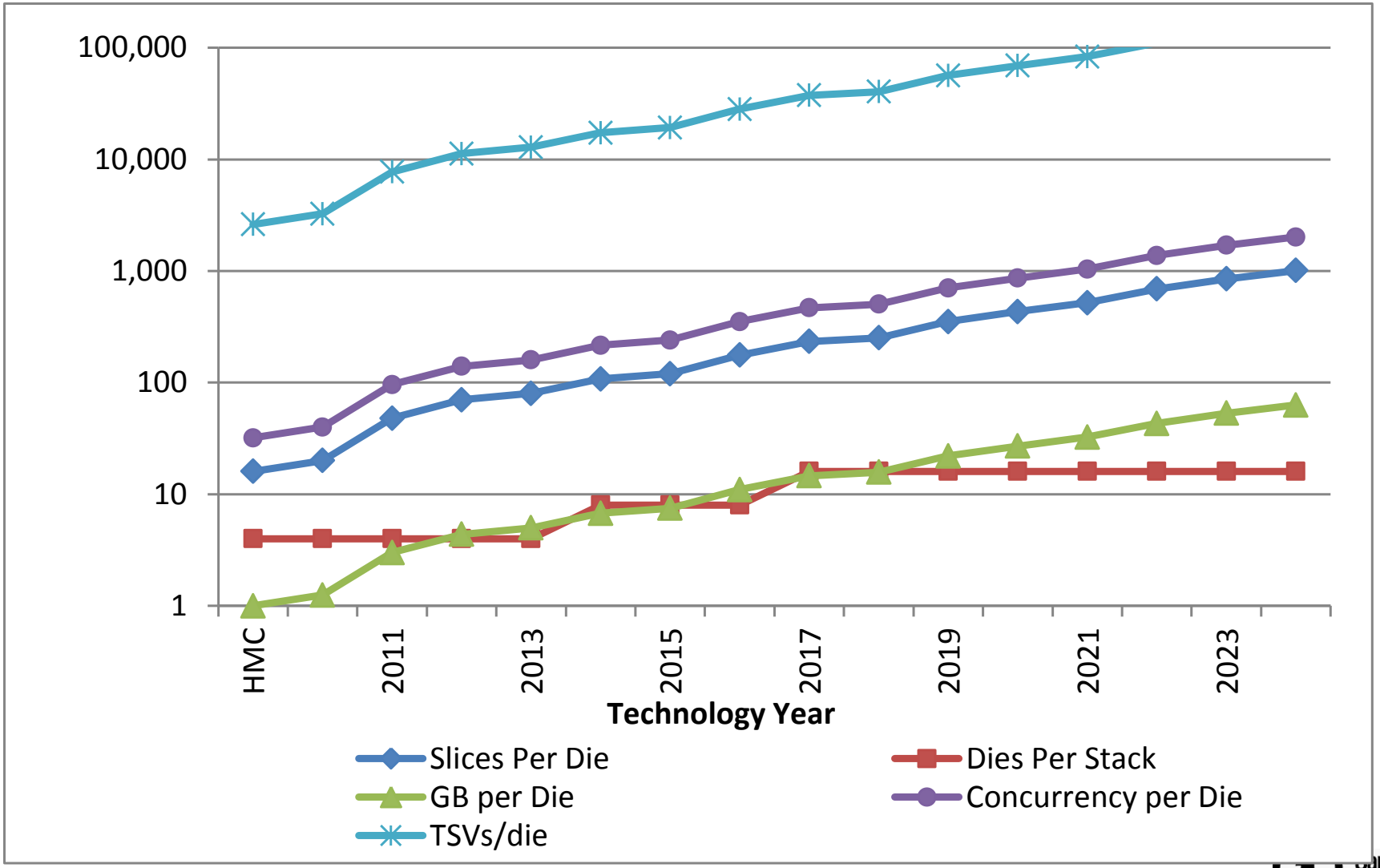


Hybrid Memory Cube Power Claims

- **Major power improvements**

Technology	VDD	IDD	BW GB/s	Power (W)	mW/GB/s	pj/bit	real pJ/bit
SDRAM PC133 1GB Module	3.3	1.50	1.06	4.96	4664.97	583.12	762
DDR-333 1GB Module	2.5	2.19	2.66	5.48	2057.06	257.13	245
DDRII-667 2GB Module	1.8	2.88	5.34	5.18	971.51	121.44	139
DDR3-1333 2GB Module	1.5	3.68	10.66	5.52	517.63	64.70	52
DDR4-2667 4GB Module	1.2	5.50	21.34	6.60	309.34	38.67	39
HMC, 4 DRAM w/ Logic	1.2	9.23	128.00	11.08	86.53	10.82	13.7

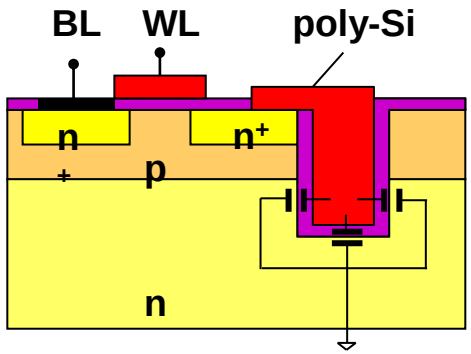
How Far Will TSV Stacking Take Us?

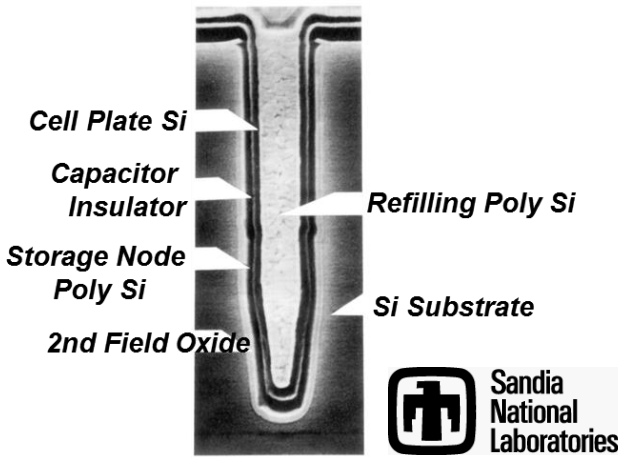


Courtesy Peter Kogge
Matthew Marinella

DRAM, TSV Limitations

- HMC-like technology should carry us to <100 pJ/operation
- DRAM Limits: From ITRS Roadmap PIDS Chapter
  - DRAMs struggling to maintain reasonable equivalent oxide thickness
  - Dielectric for cells 30nm to 20 nm still TBD
  - Is scaling possible below 20 nm?
  - Will always be volatile – not a Storage Class Memory
- Through Silicon Via
  - How many chips can you stack?
  - How many TSVs per chip?





Sandia National Laboratories

Outline

- Intro to High Performance Computing
- The Exascale Challenge
- Solutions Ready Today
- Opportunities for Emerging NVM
  - Storage & Hybrid Main Memory
  - Universal Memory
- Conclusions

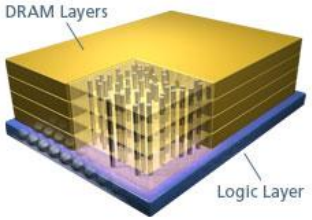
Timeline of Supercomputer Architectures

Standard Heavy/ Light node

10 Petaflop
1000 pJ/op

2012

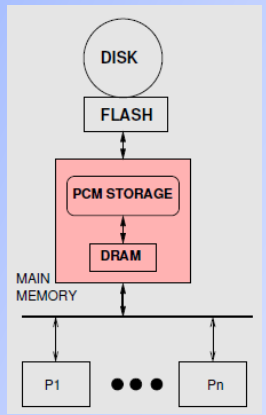
3D Stacked DRAM + chip-chip optical interconnect



100-1000 Petaflop
10-100 pJ/op

2015

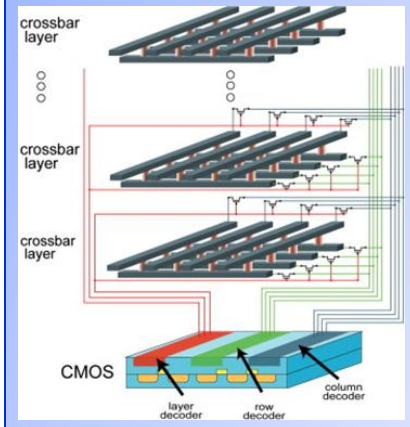
Emerging NVM for Storage, DRAM/NVM Hybrids



1-10 Exaflop
1-10 pJ/op

(Exact timeline TBD)

Storage Class Memory



10-100 Exaflop
0.1-1 pJ/op

2020

Neuromorphic, Crossbar Computers, RSFQ, Quantum…

Zetaflops+
<1 aJ/op

2030?
Timelines very fuzzy

Major opportunities for emerging NVM technologies

First Opportunity for NVM: Storage

- The most obvious goal for emerging NVM technologies
- Need to beat high end flash:

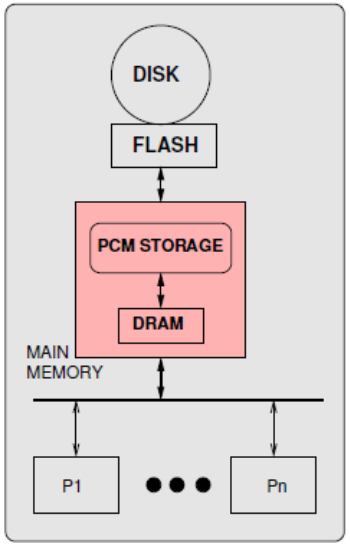
1. Voltage: < 15V
2. Endurance: > 10^4 W/E cycles
3. Scalability: < ~18 nm, 3D stackable
4. W/E time: < 100 µs
5. Retention: > 10 years

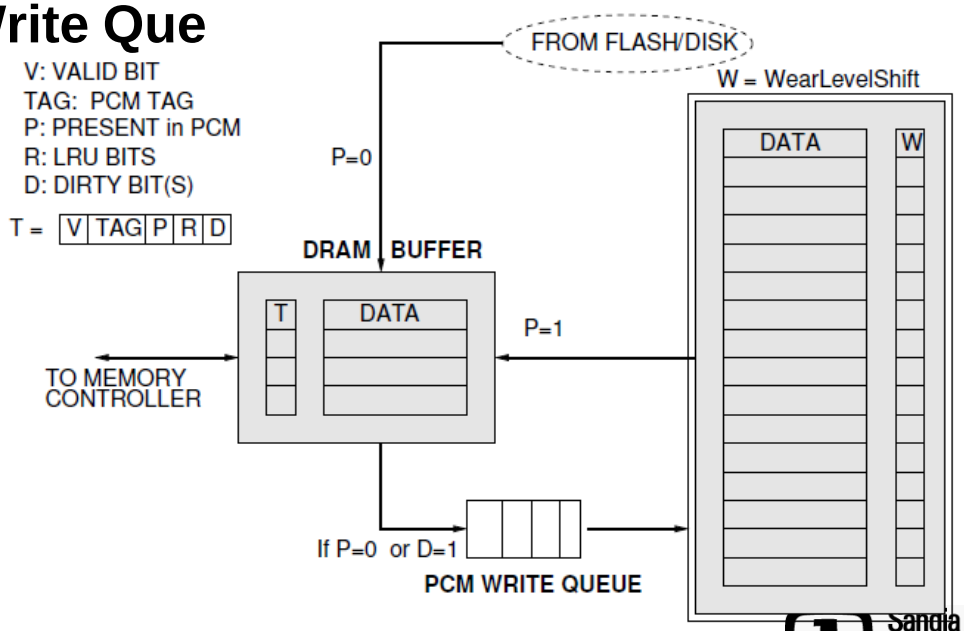
- All emerging technologies have proven these capabilities

Awaiting high capacity commercial parts!

Second Opportunity for NVM: DRAM/NVM Hybrid Main Memory

- Second interesting short term possibility for NVM
- Architecture would use emerging NMW with limitations
- DRAM buffer only needs 3% of main memory
- Lazy Write Organization (Quareshi, 2009)
  1. HDD→DRAM, allocate space in NVM
  2. If needed: DRAM→Write Que
  3. Write Que→NVM

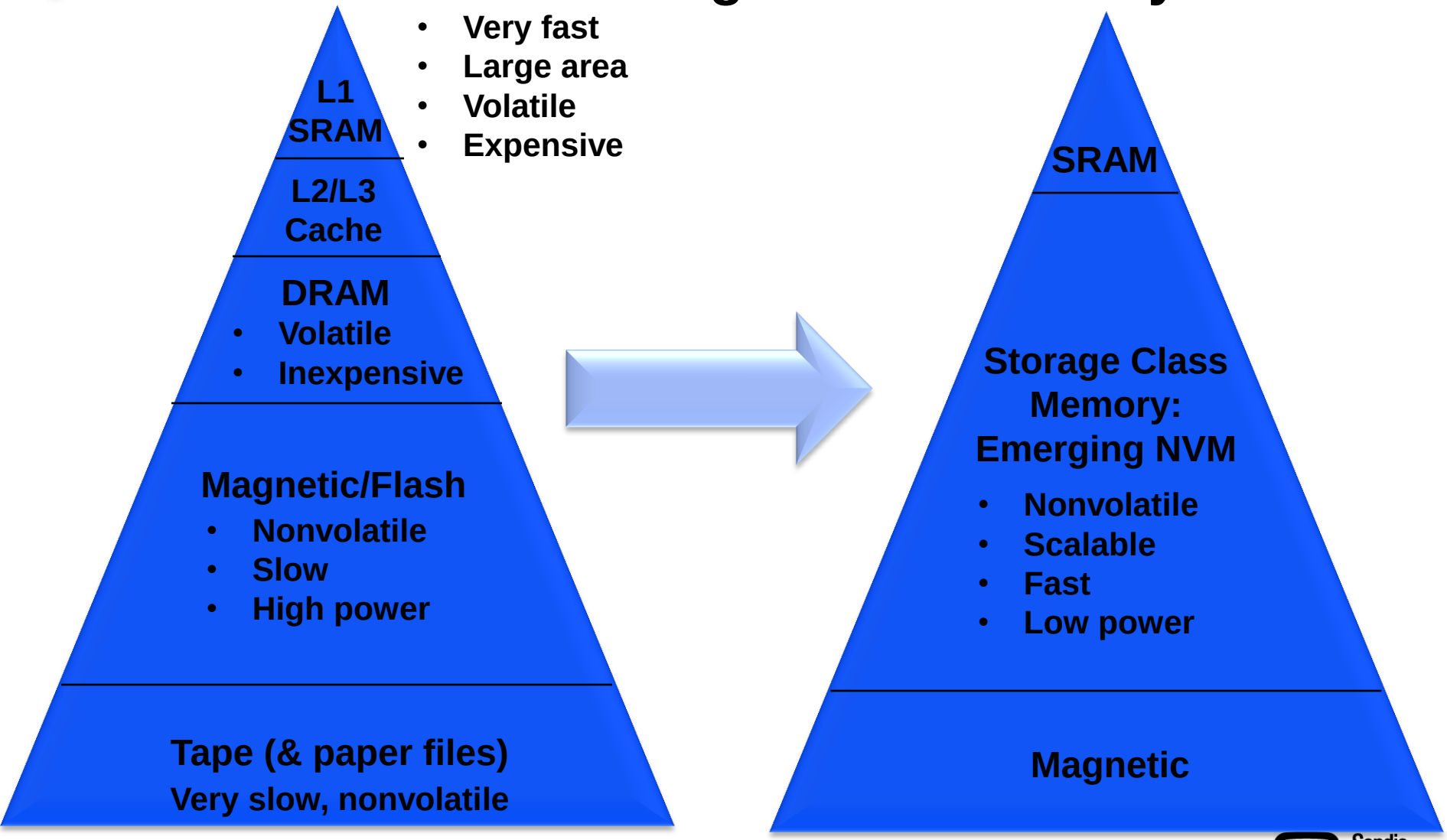




Resiliency

- Exascale computer will have a lot of hardware
- 10-100 petabytes main memory
  - 10-100 million DRAM chips
- 100's of exabytes storage
  - Millions of hard drives
- Failures are imminent! Could be a daily routine!
- Supercomputers must use checkpointing
- Traditional checkpointing at Exascale will not work
  - More time spent restoring than computing!
- Solution: Hardware checkpointing with NVM
  - Hybrid Main Memory
  - Storage Class/Universal Memory

Third Opportunity for NVM: Universal/Storage Class Memory



ITRS Requirements for SCM

Parameter	*Benchmark* [A]			*Target*	
	HDD [B]	NAND flash [C]	DRAM	Memory-type SCM	Storage-type SCM
Read/Write latency	3-5 ms	~100µs (block erase ~1 ms)	<100 ns	<100 ns	1-10µs
Endurance (cycles)	unlimited	10^4-10^5	unlimited	>10^9	>10^6
Retention	>10 years	~10 years	64 ms	>5 days	~10 years
ON power (W/GB)	~0.04	~0.01-0.04	0.4	<0.4	<0.04
Standby power	~20% ON power	<10% ON power	~25% ON power	<1% ON power	<1% ON power
Areal density	~ 10^{11} bit/cm^2	~ 10^{10} bit/cm^2	~ 10^9 bit/cm^2	>10^{10} bit/cm^2	>10^{10} bit/cm^2
Cost ($/GB)	0.1	2	10	<10	<3-4

Supercomputing SCM

Requirements* for SCM use in a Supercomputer

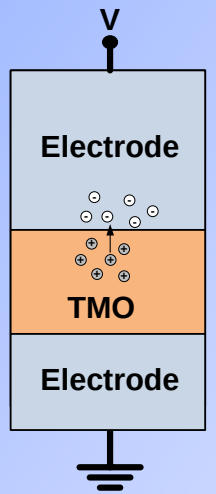
1. **Energy:** **< 1pJ per write/erase op**
2. **Endurance:** **> 10^{15} W/E cycles**
3. **Scalability:** **< 10 nm, 3D stackable, no select transistor**
4. **Read/Write:** **> 1 ns**
5. **Retention:** **> 10 years fully scaled at operation temp**
6. **Reliable Operation**

***Note: Requirements open to debate – especially retention**

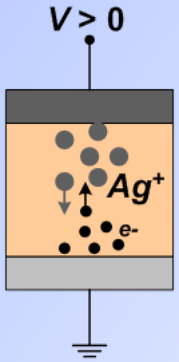
SCM Candidates

ReRAM

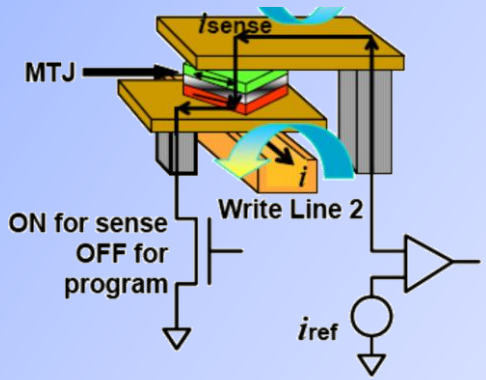
VCM



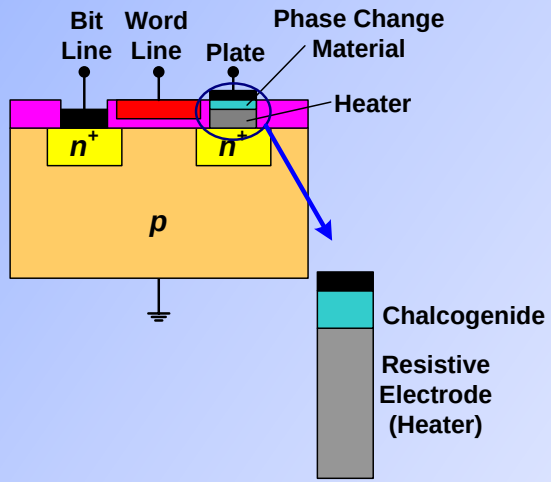
ECM

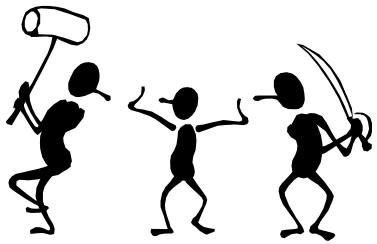


STT-MRAM



Phase Change





ECM, STT, PCM courtesy Dieter Schroder

Emerging Nonvolatile Memories

The infamous comparison chart

Biggest challenge for ReRAM: Catch-up

	DRAM	Flash (NOR-NAND)	ReRAM/Memoristor	STT-MRAM	PC-RAM
2012 Maturity	Production (30 nm)	Production (18 nm)	Development	Production (65 nm)	Production (45 nm)
Min device size (nm)	20	18	<10	16	<10
Density (F^2)	6	4	4	8-20	$4F^2$
Read Time (ns)	< 10	10^5	2	10	20
Write Time (ns)	< 10	10^6	2	13	50
Write Energy (pJ/bit)	0.005	100	<1	4	6
Endurance (W/E Cycles)	$>10^{16}$	10^4	10^{12}	10^{12}	$>10^9$
Retention	64 ms	> 10 y	> 10 y	weeks	> 10 y
BE Layers	FE	FE	4	10-12	4
Process complexity	High/FE	High/FE	Low/BE	High/BE	Low/BE

Biggest challenge for STT-MRAM: Retention/Scaling/Temperature

Biggest challenge for PCM: High erase current

A More Subjective Survey

	Prototypical (Table ERD3)			*Emerging (Table ERD5)*					
Parameter	**FeRAM**	**STT-MRAM**	**PCRAM**	**Emerging ferroelectric memory**	**Nanomechanical memory**	**Redox memory**	**Mott Memory**	**Macromolecular memory**	**Molecular Memory**
Scalability									
MLC									
3D integration									
Fabrication cost									
Endurance									

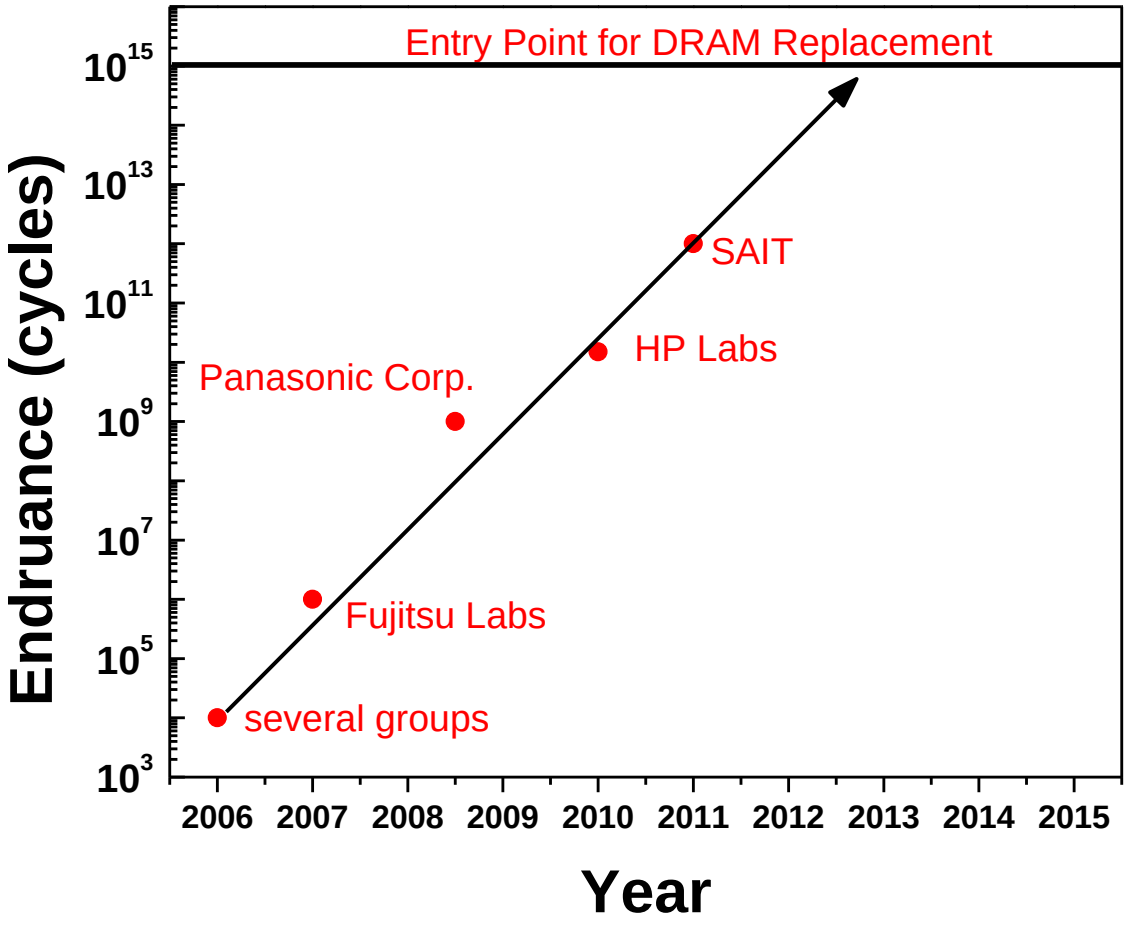
Scalability	F_{min} >45 nm
MLC	difficult
3D integration	difficult
Fabrication cost	high
Endurance	≤1E5 write cycles demonstrated

Scalability	F_{min}=10-45 nm
MLC	feasible
3D integration	feasible
Fabrication cost	medium
Endurance	≤1E10 write cycles demonstrated

Scalability	F_{min} <10 nm
MLC	solutions anticipated
3D integration	difficult
Fabrication cost	potentially low
Endurance	>1E10 write cycles demonstrated

ReRAM Endurance Improvements

Will this trend continue?



Courtesy J. Joshua Yang (HP Labs)

Key Challenges For ReRAM

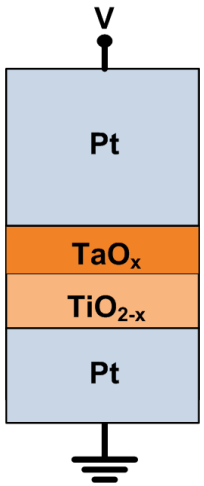
- ReRAM is the least mature of major contenders
  - Also (arguably) shows the greatest promise
- Endurance – continue improvements
  - Need > 10^{15} W/E cycles (10^{16})
- Scalable select device (no select transistor)
  - Still must eliminate sneak paths/parasitics
- Uniformity & reliability issues (can circuitry help?)
- Circuitry: read/write, wear leveling, error correction
- Prove 3D stacking

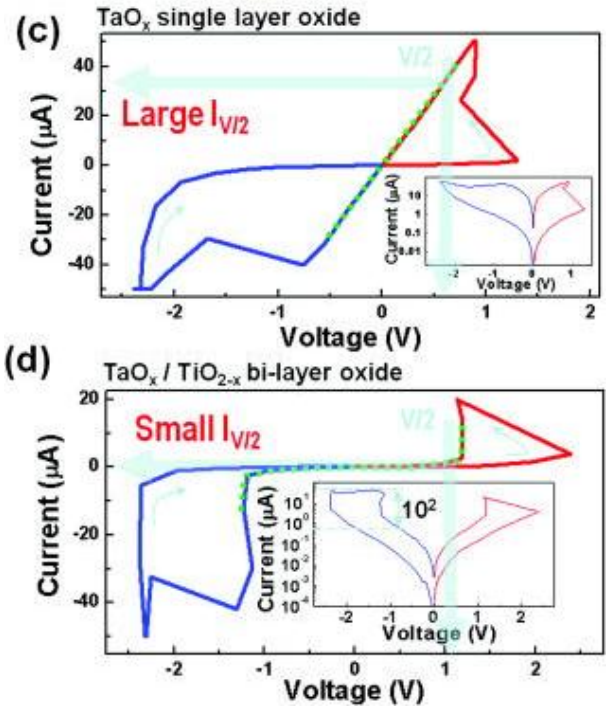
The Good News: Challenges all result from immaturity

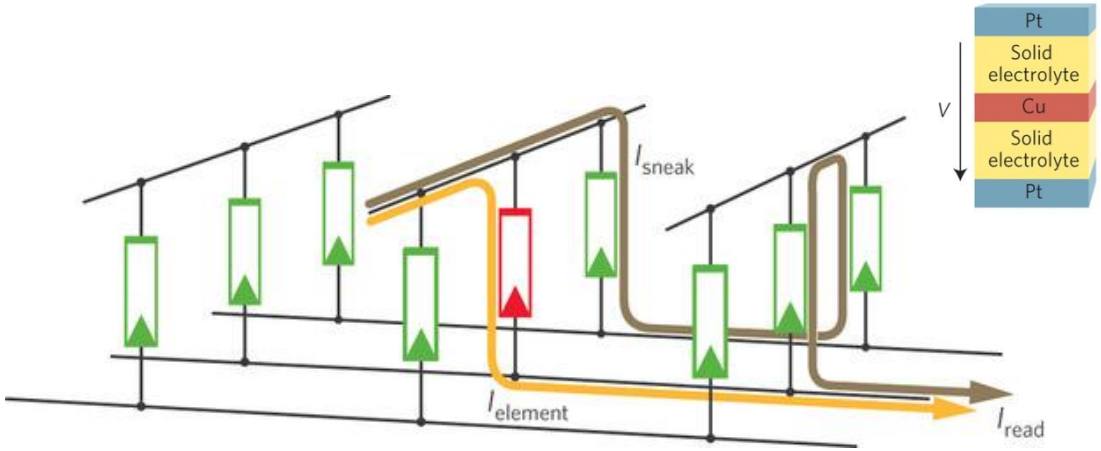
No fundamental physical showstoppers

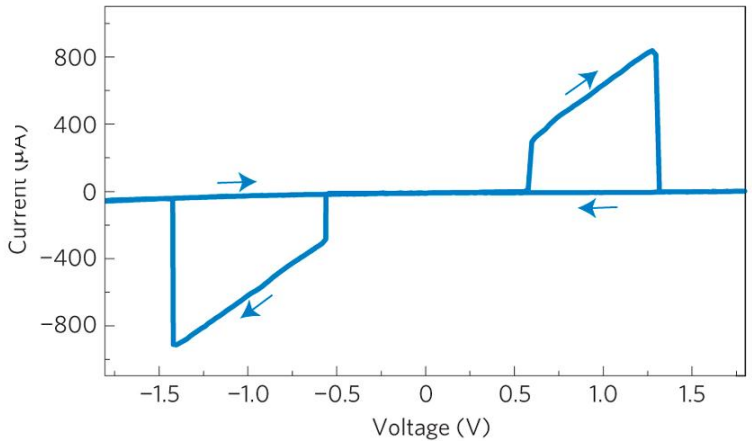
Select Device

- Major open issue with ReRAM
- I-V linearity governs array size
- Limits the array size
- DO NOT want a MOSFET
  - Kills scaling!
- Solutions:
  - Complementary Resistive Switch
  - Bilayer Nonlinearity



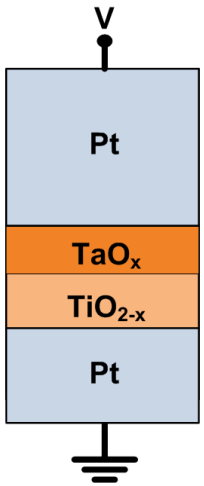


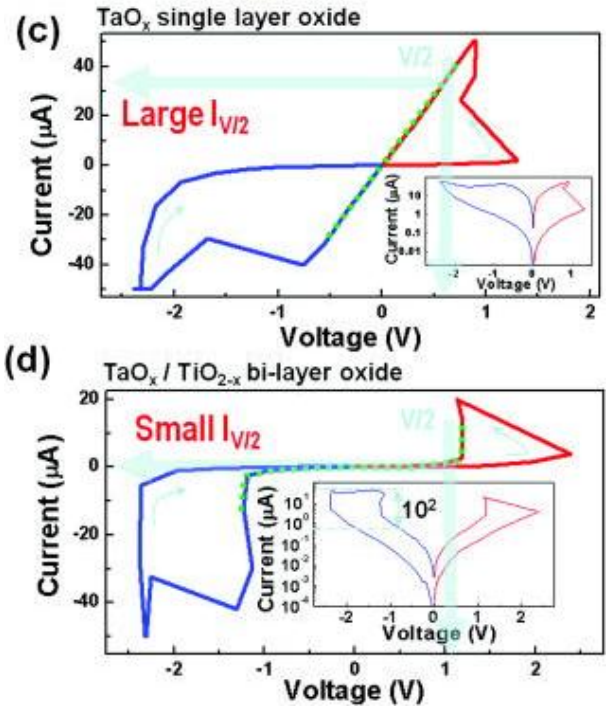


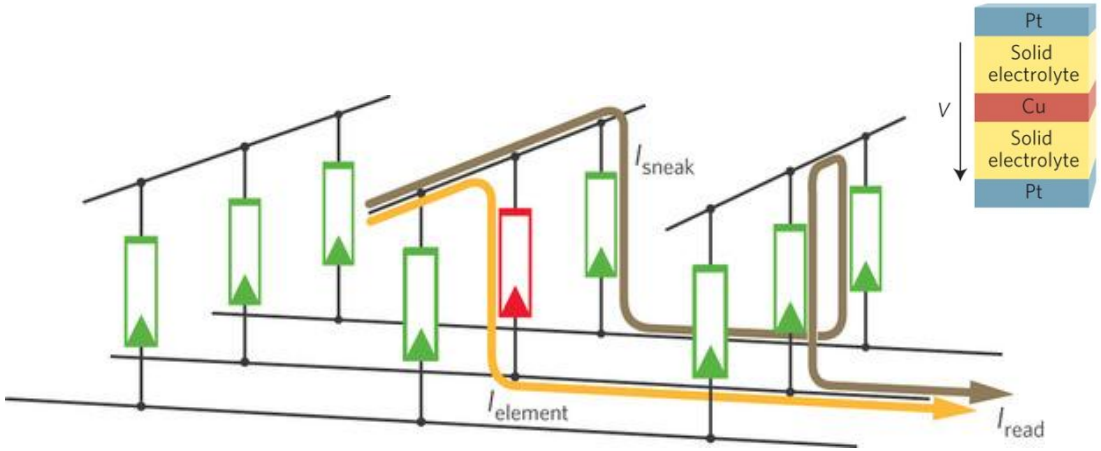


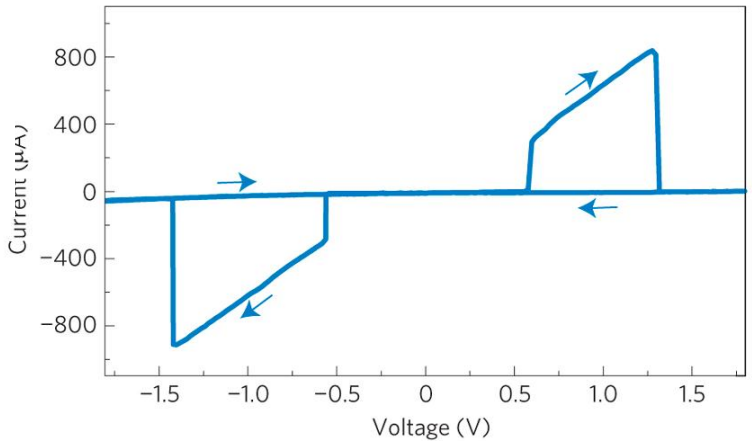
Array Considerations

- Major open issue with ReRAM
- I-V linearity governs array size
- Limits the array size
- DO NOT want a MOSFET
  - Kills scaling!
- Solutions:
  - Complementary Resistive Switch
  - Bilayer Nonlinearity



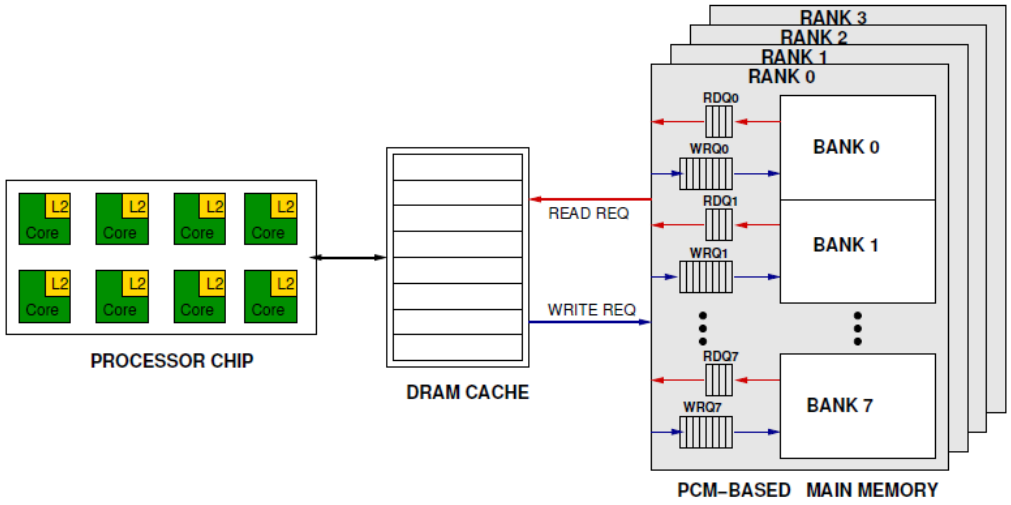


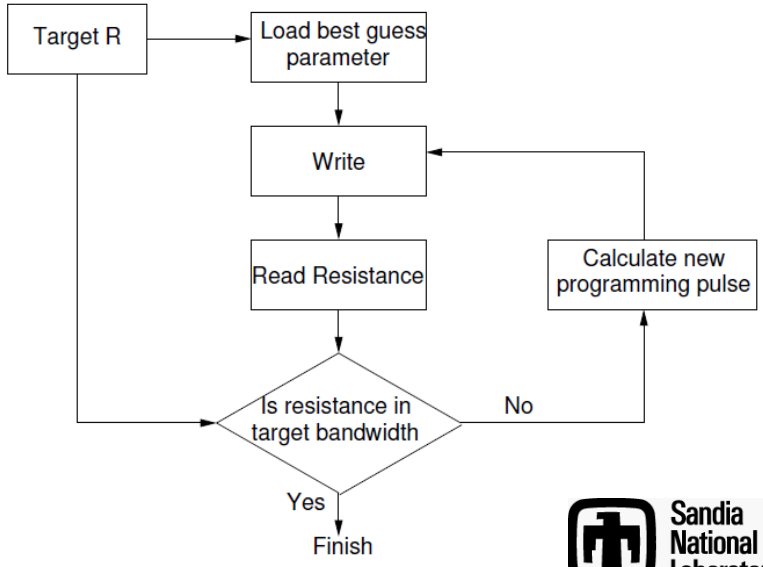




Array Architecture

- How do we architect ReRAM as a main memory array?
- What new issues will we face when converting from DRAM array → ReRAM
- Qureshi and colleagues have started this process for PCM
  - Example – PCM architecture and write scheme below
- Work needed for ReRAM (can learn from PCM techniques)





Summary

- **Exascale computing will be a tough road**
  - **Biggest challenge: memory**
- **Exascale will *need* a new memory solution**
  - **Whether this is at 0.1 or 10 exaflops is TBD**
- **Emerging nonvolatile memory devices**

References/Further Reading

- 2011 International Technology Roadmap for Semicondcutors, “Emerging Research Devices” and “Process, Integration, and Device Structures” Chapters. Available at itrs.net.
- Peter Kogge et al “Exascale Computing Study: Technology Challenges in Achieving Exascale Systems,”
- Qureshi et al “Phase Change Memory,” Morgan & Claypool, 2011.
- Qureshi et al “Improving Read Performance of Phase Change Memories via Write Cancellation and Write Pausing,” High Perf Computer Arch 2010.
- E. Linn et al, “Complementary resistive switches for passive nanocrossbar memories,” Nature Mater 9, 403-406, 2010.
- J.T. Pawlowski, “Hybrid Memory Cube (HMC),” HotChips 2011. Available at hotchips.org
- J. Joshua Yang et al, “Engineering nonlinearity into memristors for passive crossbar applications,” APL 100, 113501, 2012

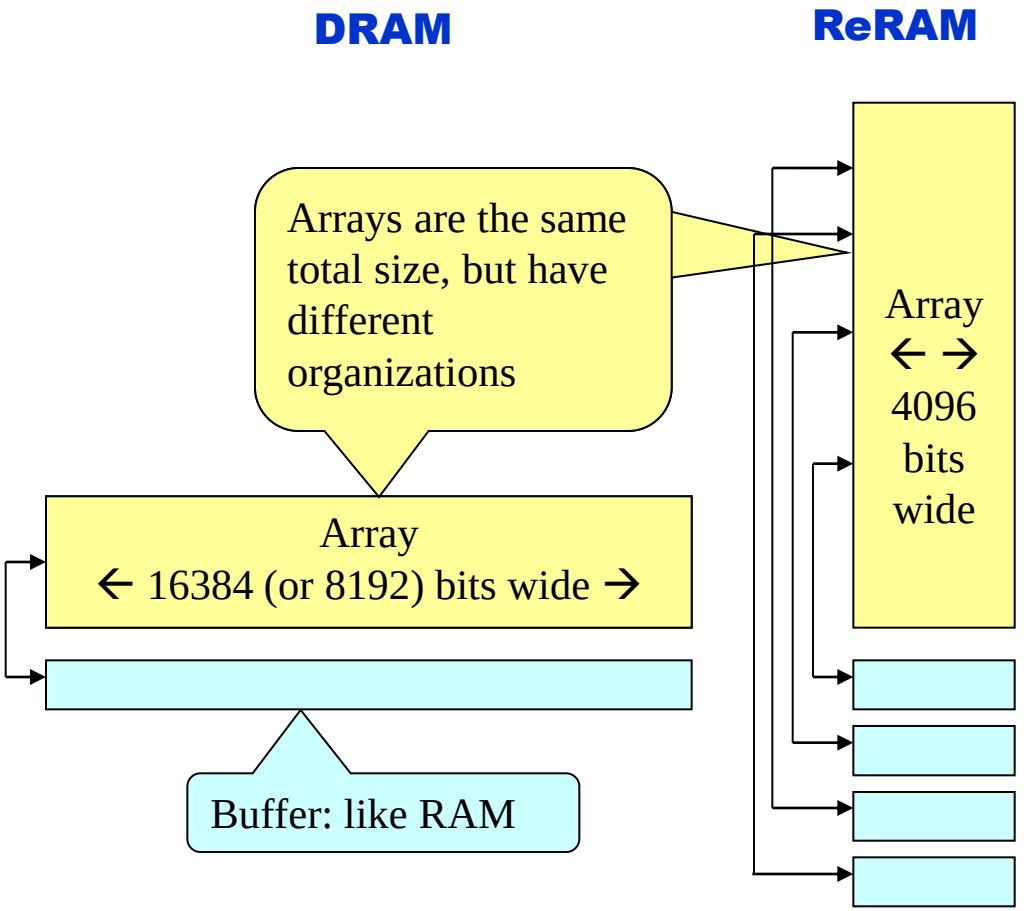
Extra Slides

Acknowledgements

- Useful discussions with Erik Debenedictis (SNL), Jianhua Yang (HPL), Rich Murphy (SNL), Greg Astfalk (HP), Stan Williams (HPL), Greg Snider (HPL), Alex Hsia (SNL) Jim Hutchby (SRC), Victor Zhirnov (SRC), Kim Denton-Hill (SNL)
- Peter Kogge & coauthors of DARPA Exascale Report
- ITRS colleagues esp. ERD, ERM, and PIDS workgroups
- ReRAM/Memristor Program at Sandia funded in part by Laboratory Directed Research and Development

Array Architecture

- Ideas for Phase Chance Memory apply to ReRAM
- Example: Lee et al
- Replace DRAM row buffer with 4 “cache-like” buffers of same cumulative length
  - Negligible reduction in reuse
  - Cuts bandwidth in/out of array by 4x
  - Also: Don’t write back if data has not changed (obvious but not an option for DRAM)
- Need continued work in this area

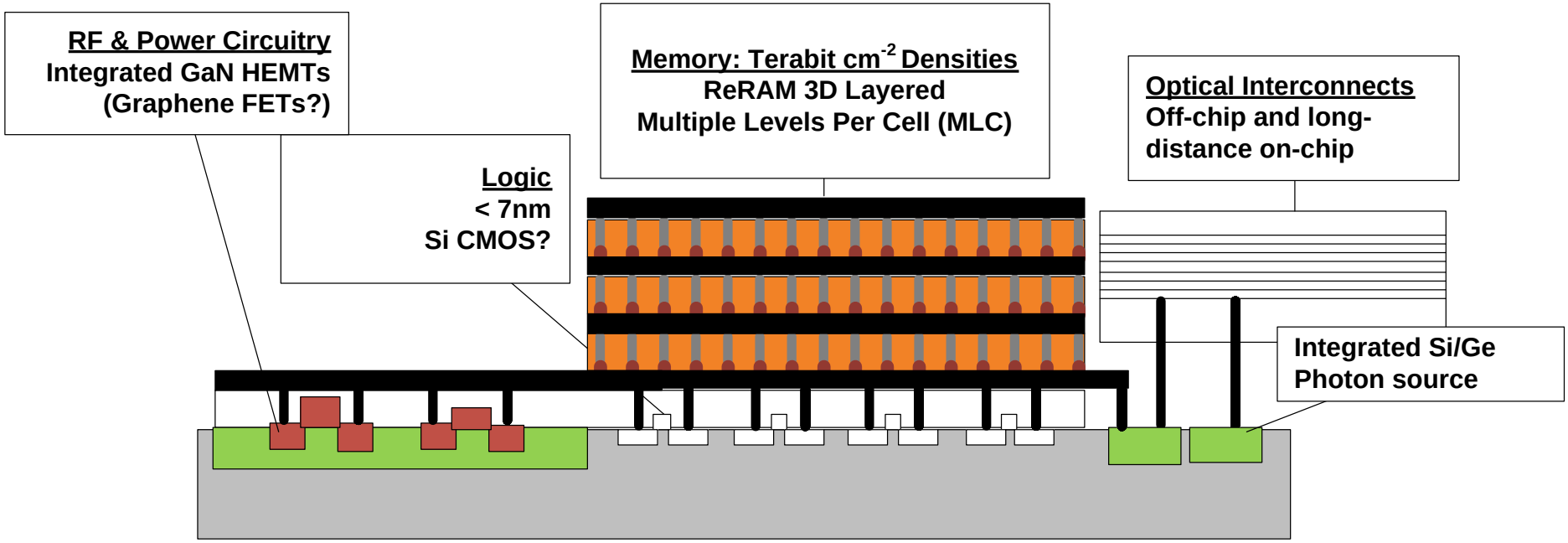


Race to Universal Memory

Another Possibility

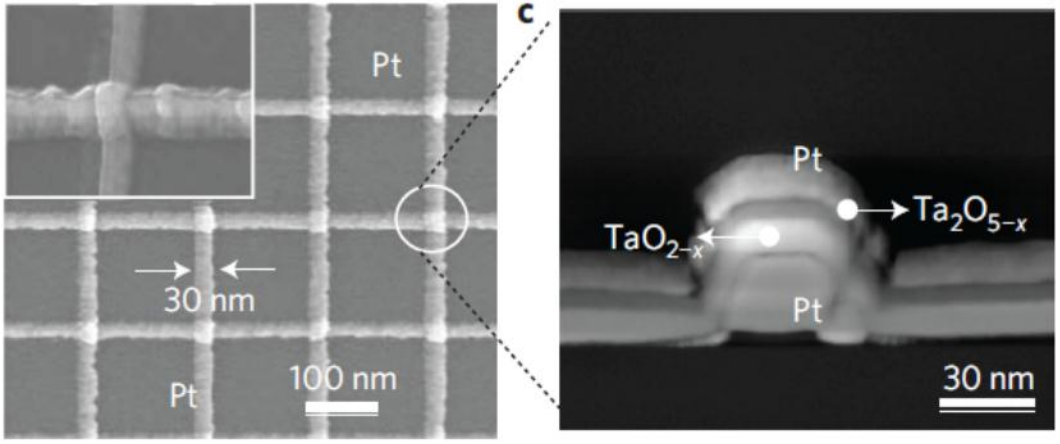
- STT-RAM has high speed and endurance
  - Retention/scaling issue
- ReRAM has high speed and retention
- Could these two coexist?

Chip of 2020

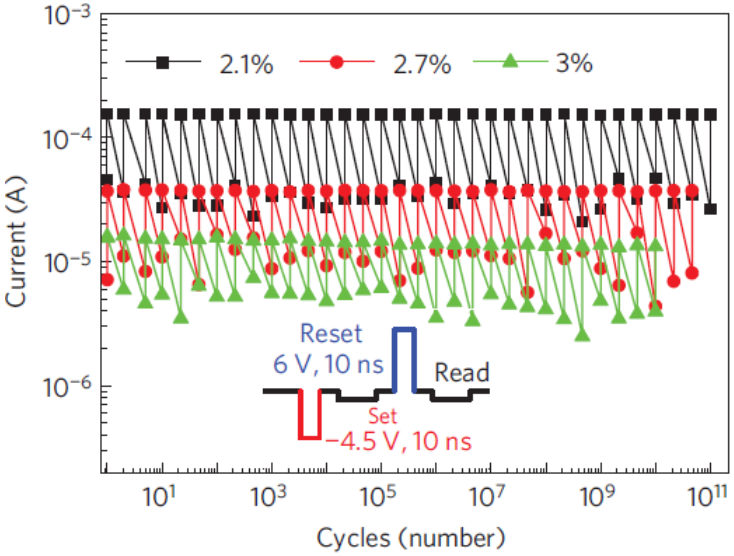

RF & Power Circuitry
Integrated GaN HEMTs
(Graphene FETs?)
Logic
< 7nm
Si CMOS?
Memory: Terabit cm-2 Densities
ReRAM 3D Layered
Multiple Levels Per Cell (MLC)
Optical Interconnects
Off-chip and long-distance on-chip
Integrated Si/Ge
Photon source

Current State of the Art

- **Records as of February 2012**
  - **Endurance 10^{12} (Samsung, TaO_x, shown below)**
  - **Scalability 10x10 nm^2 ($1F^2$), (IMEC HfO_x, right)**
  - **Switching time < 500 ps (HP Labs, TaO_x)**
  - **Retention >> 10 y (estimate by HP Labs),**
  - **Switching energy < 0.1 pJ/bit (HP Labs, TaO_x)**
- **State of the art is rapidly advancing**

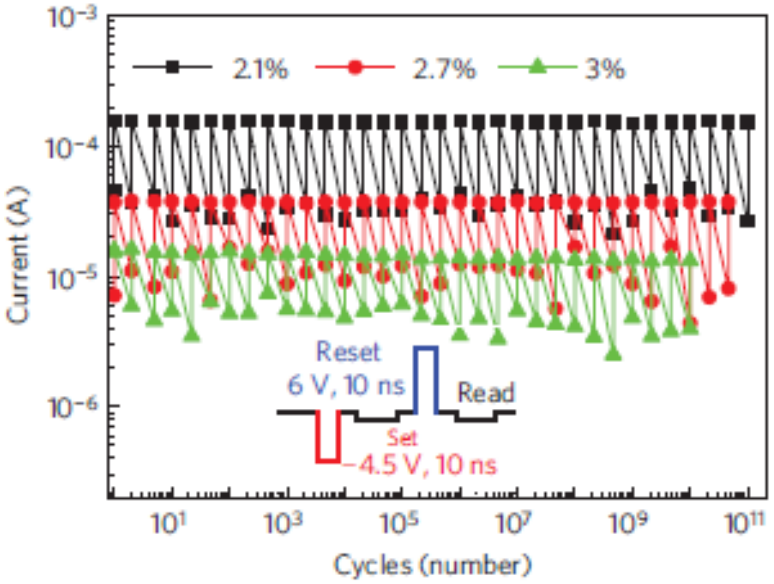


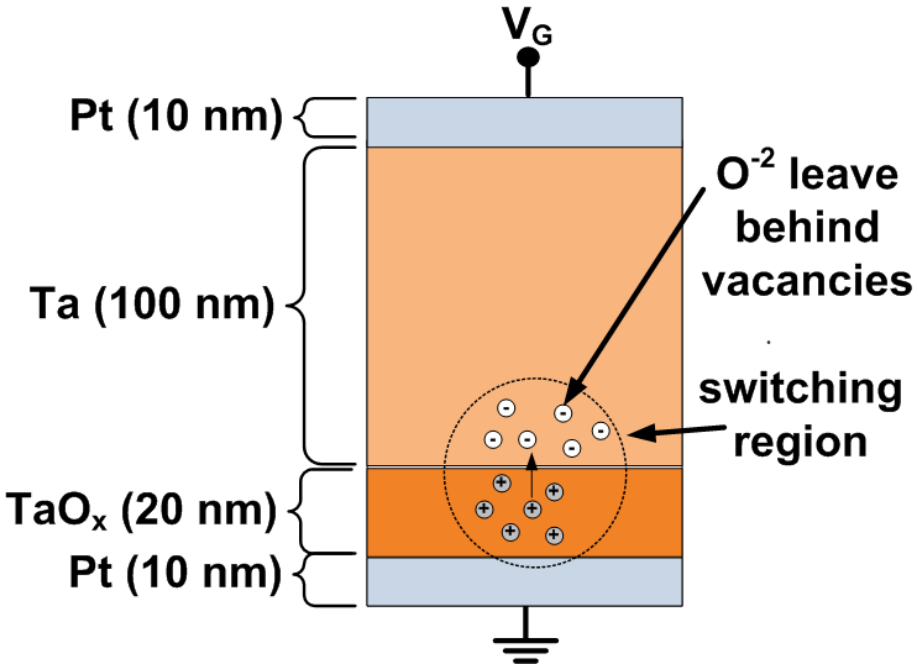
M.-J. Lee, *et al.*, *Nat Mater,* vol. 10, pp. 625-630, 2011.



TaO_x Memristor

- Recent endurance record: 10^{12} cycles! (Samsung)
- Very promising as a commercial memory
- Least understood material system
- HP Labs is a world leader in TaO_x memristor research

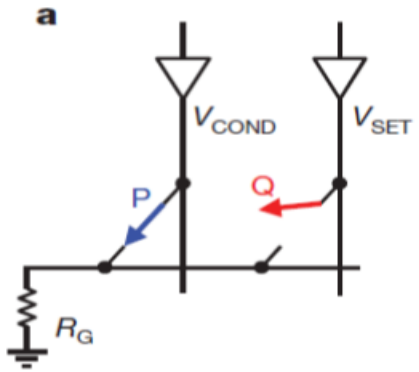




M.-J. Lee, *et al.*, *Nat Mater,* vol. 10, pp. 625-630, 2011.

Memristor Crossbar Computers

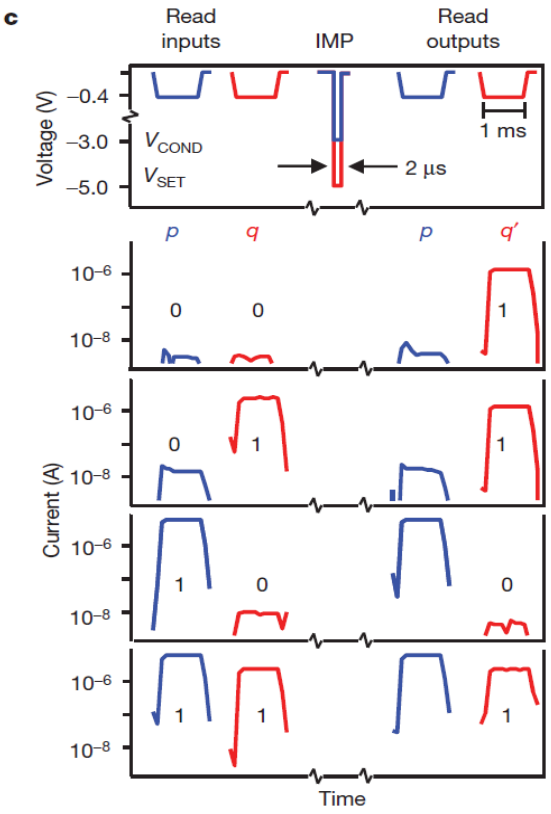
- **How?**
- **Possible Method: Material Implication Logic**

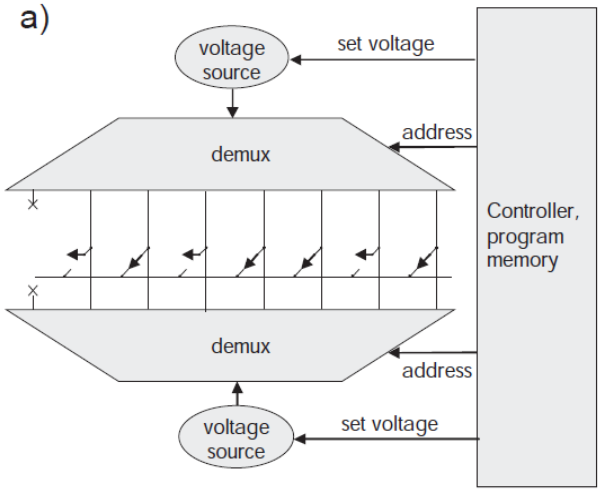


b

$q' \leftarrow p\text{IMP}q$

In p	In q	Out q'
0	0	1
0	1	1
1	0	0
1	1	1





Borghetti et al, Nature 464, 873-876, 2010