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

April 6, 2012

Matt Marinella
Sandia National Laboratories
mmarine@sandia.gov

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Outline

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

Supercomputers

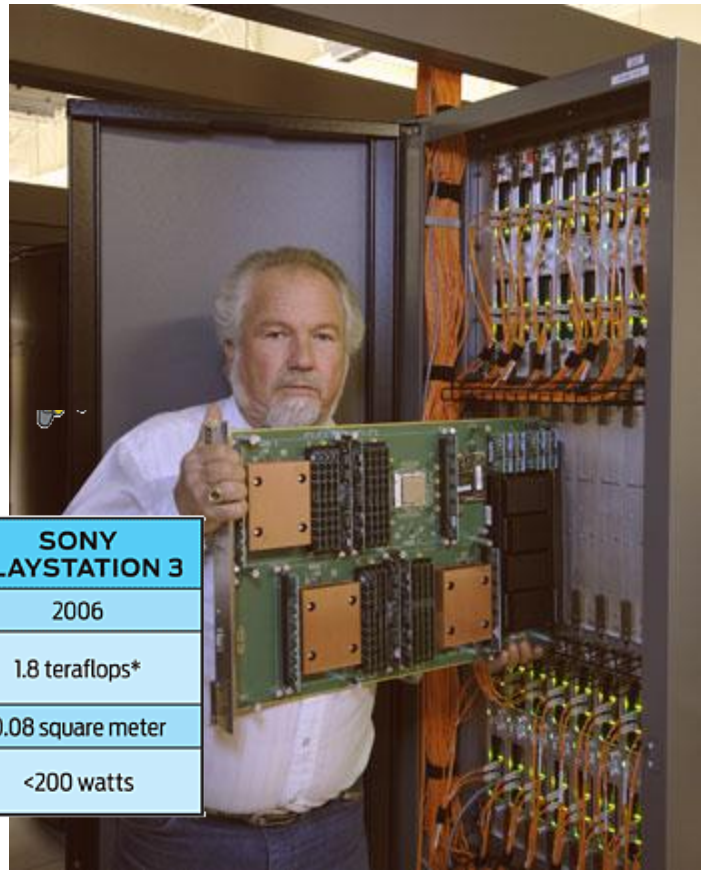
- Define the forefront of computing power
- Modern architectures are massively parallel systems
- Thousands of off the shelf processors & DRAM chips
  - GPUs or CPUs
- Performance measured in FLOPS:
  - Floating point ops per second
- Benchmarked by LINPACK
  - Solve *n* x *n* system of linear eqns
- Out of date quickly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

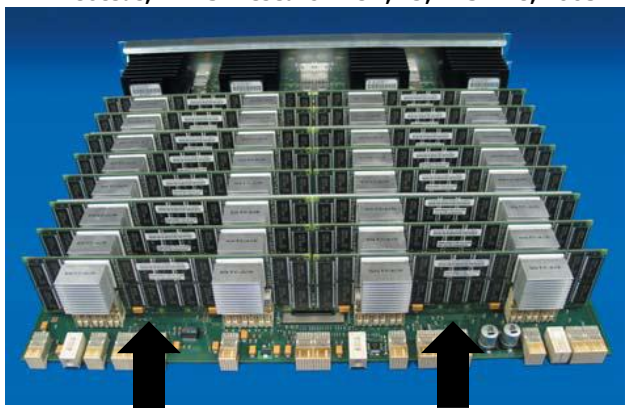
P. Kogge, IEEE Spectrum, 48, 48-54, 2011.



Modern Supercomputer Architecture

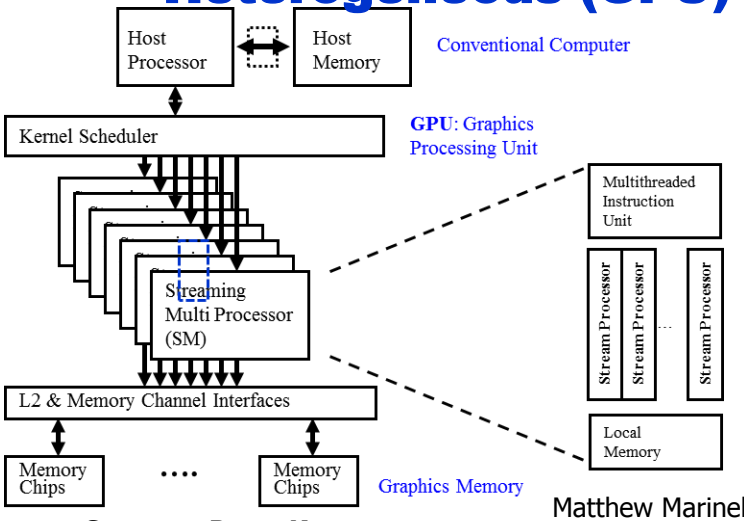
- As defined by Prof. Kogge:
- "Heavyweight" node
  - Commodity microprocessor
- "Lightweight" node
  - Custom microprocessors, low power
- "Heterogeneous" node
  - Uses GPUs + heavyweight master

P. Coteus, IBM J. Research Dev., 49, 213-248, 2005.



IBM Lightweight Nodes

Heterogeneous (GPU) Node & Architecture



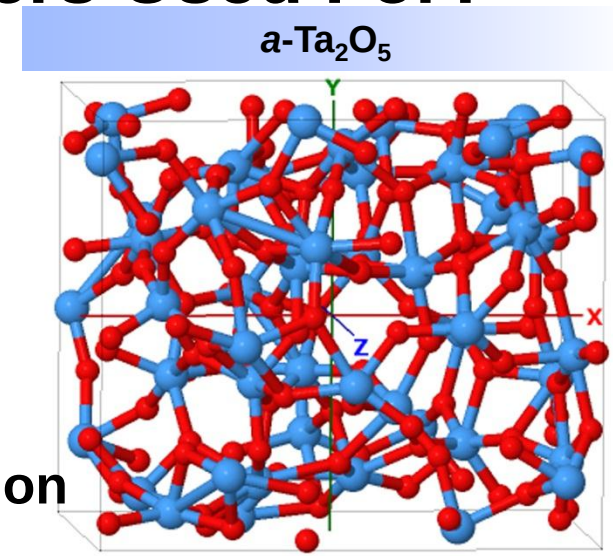
Courtesy Peter Kogge

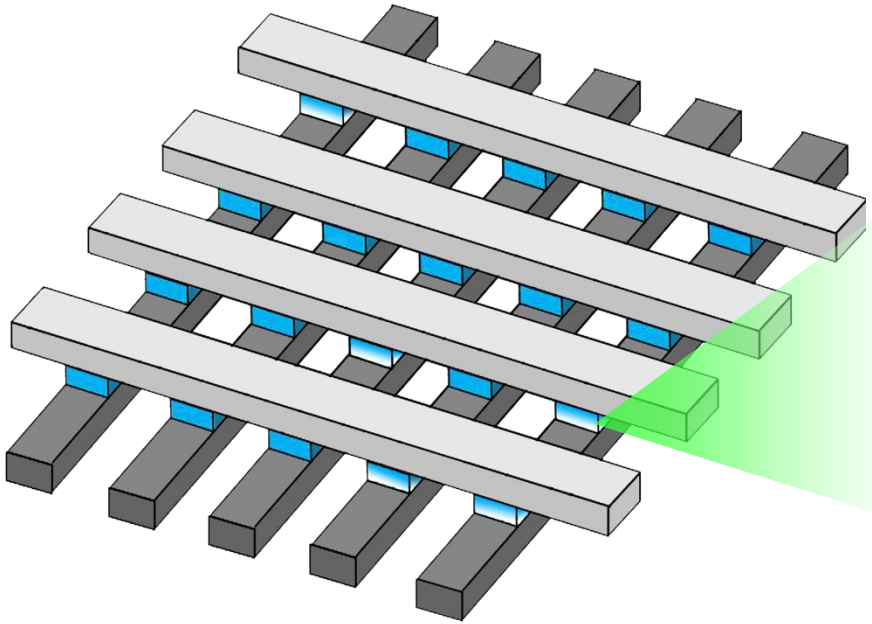
nvidia.com

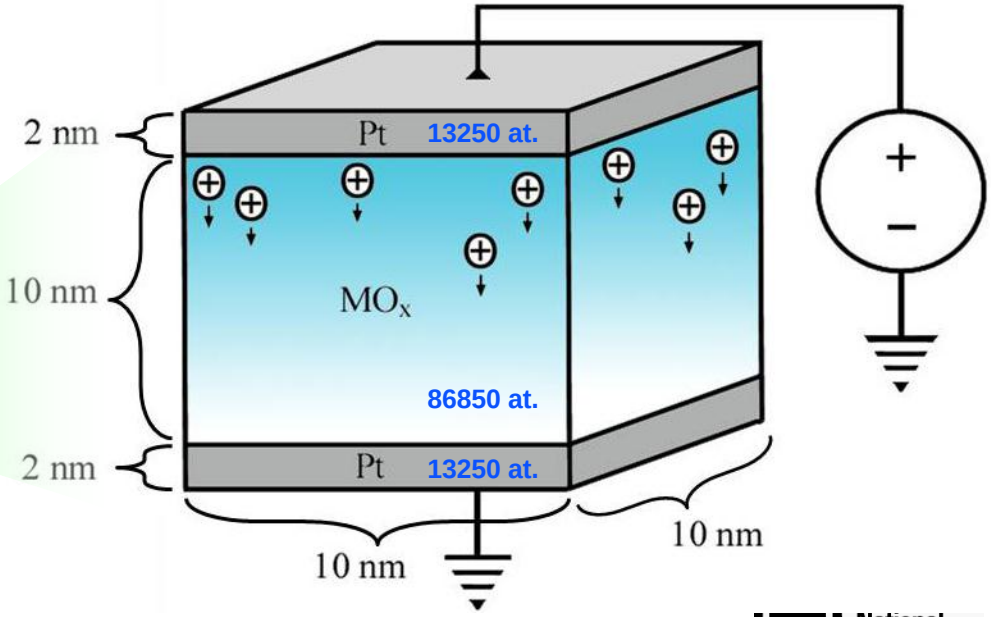
Matthew Marinella

What are Supercomputers Used For?

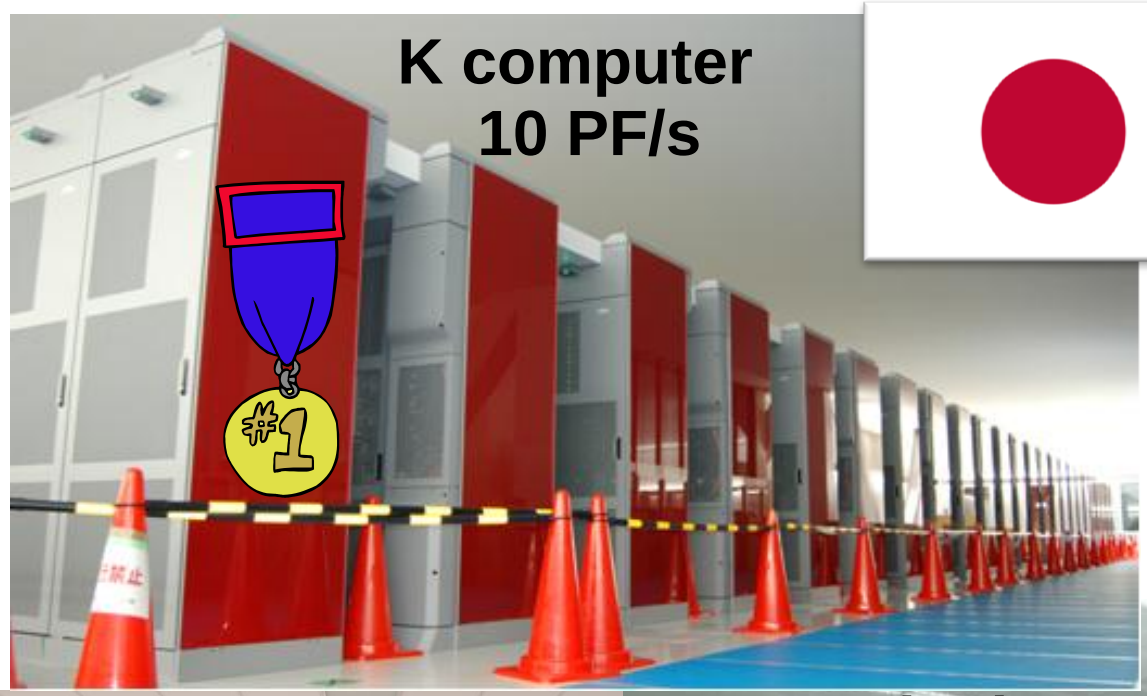
- Predicting the weather
- Google searches
- Simulating nuclear explosions
- Quantum physics
- Molecular dynamics & DFT simulation

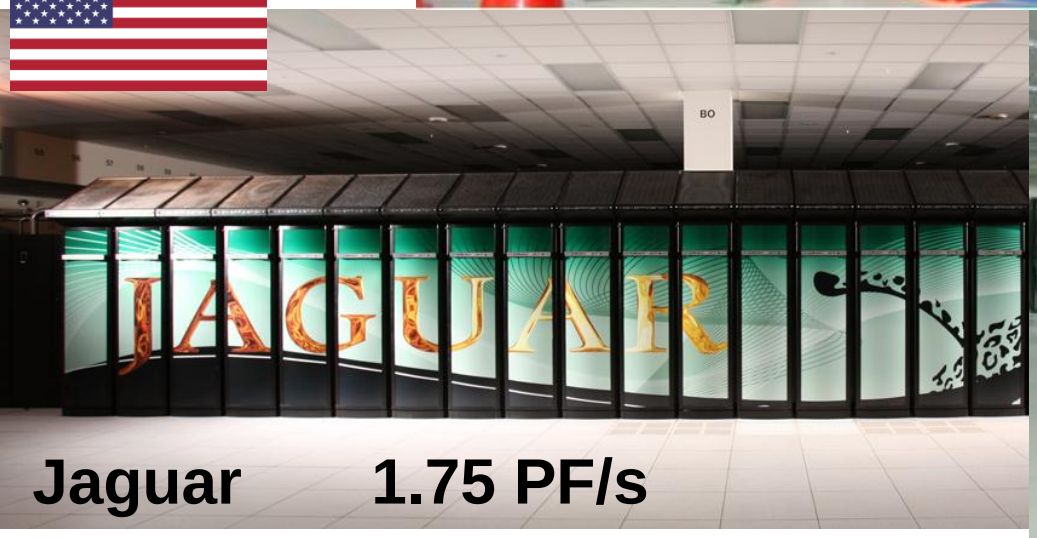






Supercomputing Olympics





The World's Best Computers

- **K computer (RIKEN, Japan)**
  - **Speed: 10.5 petaflops (Rmax)**
  - **Cores: 705k (SPARC64 2.0 GHz)**
  - **Memory: 1.4 PB**
- **Tianhe (China)**
  - **Speed: 2.5 petaflops (Rmax)**
  - **Cores: 186k (NVIDIA 2.93GHz GPU)**
  - **Memory: 229 TB**
- **Roadrunner (Oak Ridge National Lab, US)**
  - **Speed: 1.75 petaflops (Rmax)**
  - **Processors: 224k (Cray Opteron 6-core, 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

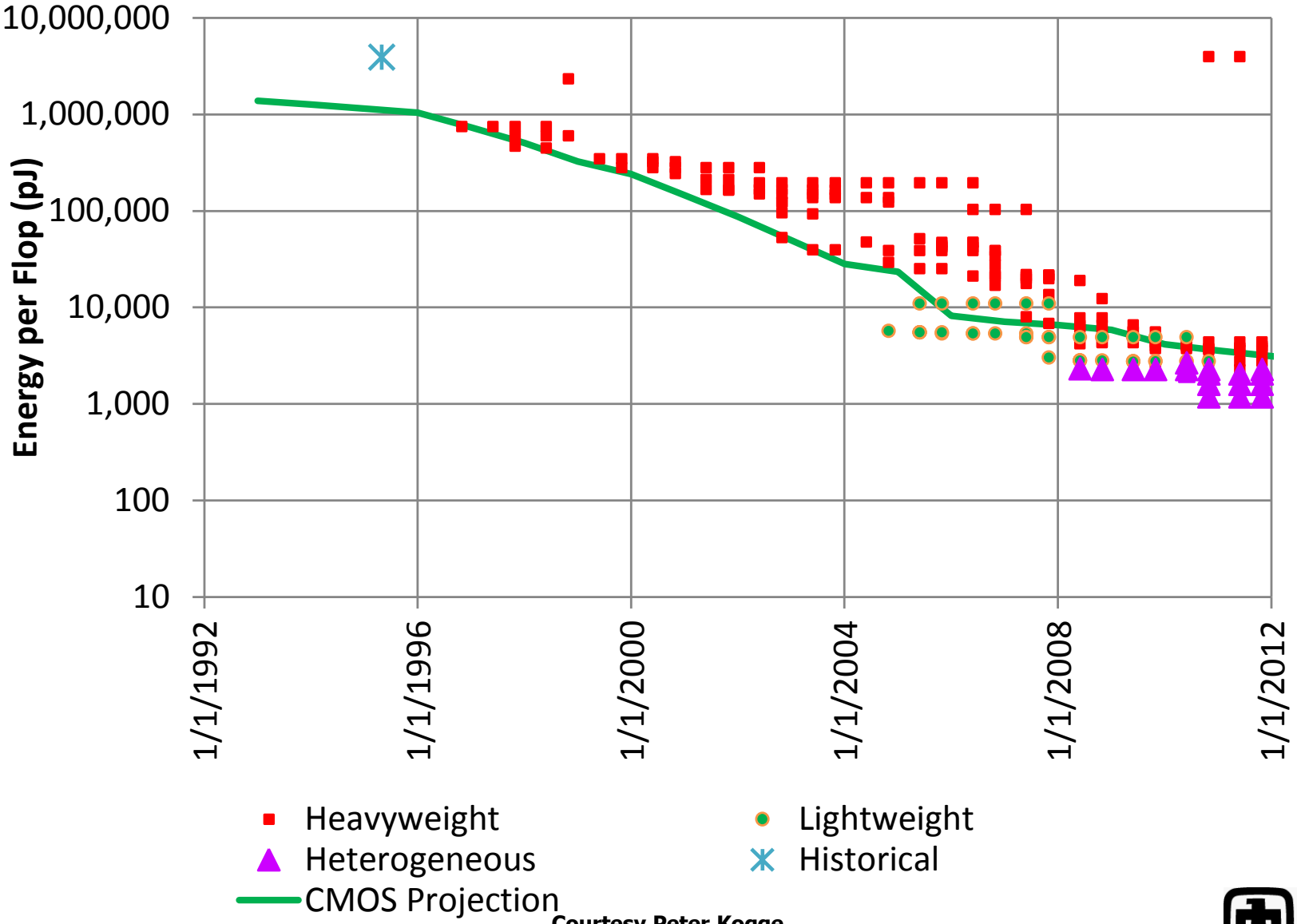
- Intro to High Performance Computing
- The Exascale Challenge
- Solutions Ready Today
- Opportunities for Emerging NVM
  - Storage & Hybrid Main Memory
  - Universal Memory
- Conclusions

Grand Challenges for Exascale

- DARPA Exascale Report Lists Four Main Challenges:
  - Energy and Power
  - Memory and Storage
  - Concurrency and Locality
  - Resiliency
- US Dept. of Energy takes action: Exascale Initiative

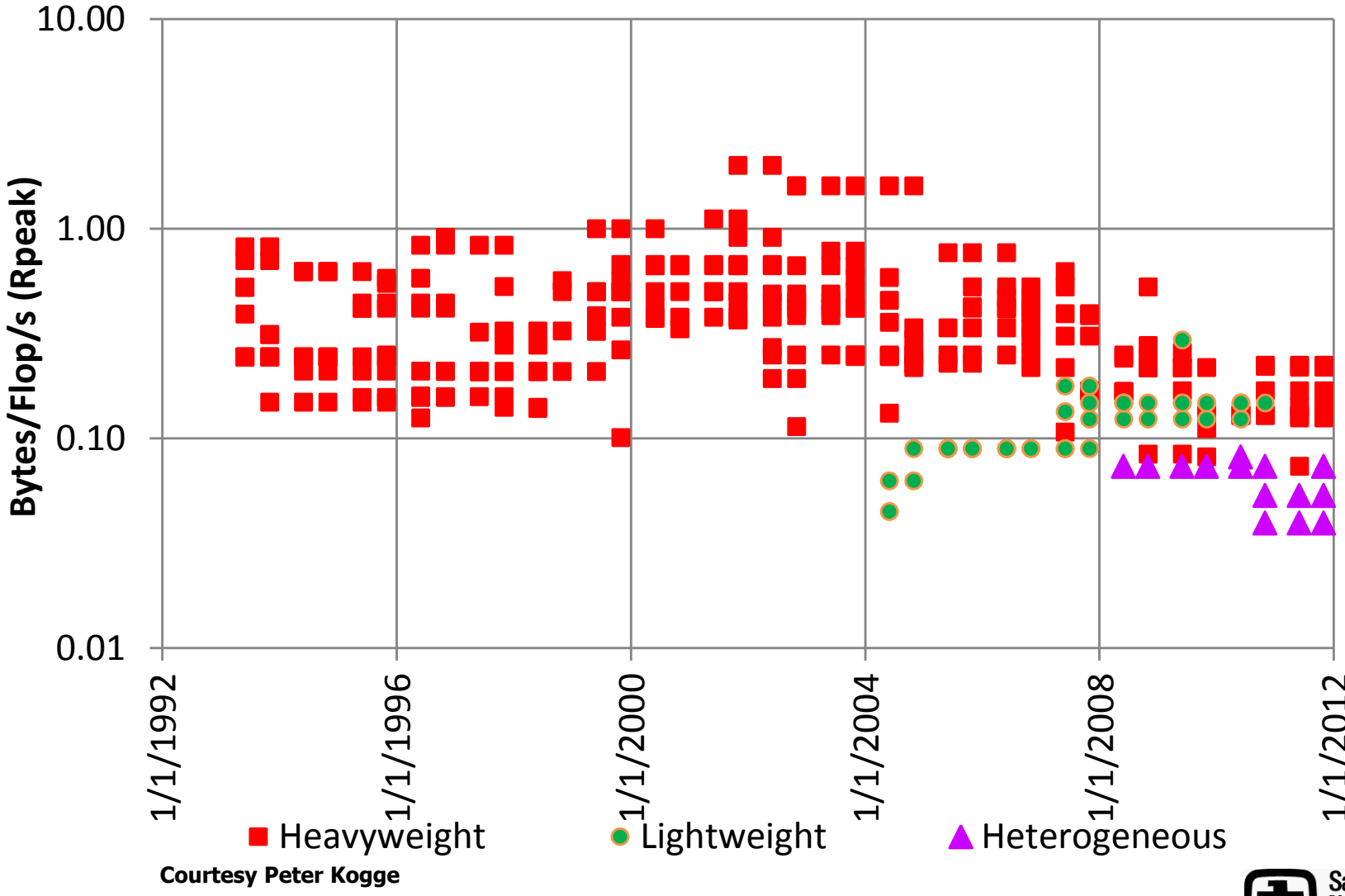
Major role for emerging nonvolatile memories in three of four challenges!

Energy per Flop



Courtesy Peter Kogge
Matthew Marinella

DRAM Bytes per Flop



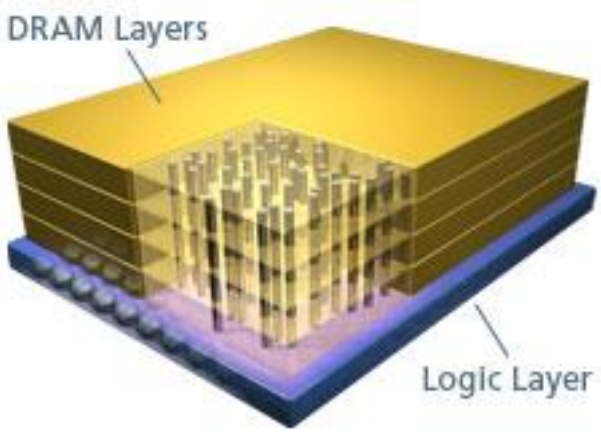
Courtesy Peter Kogge

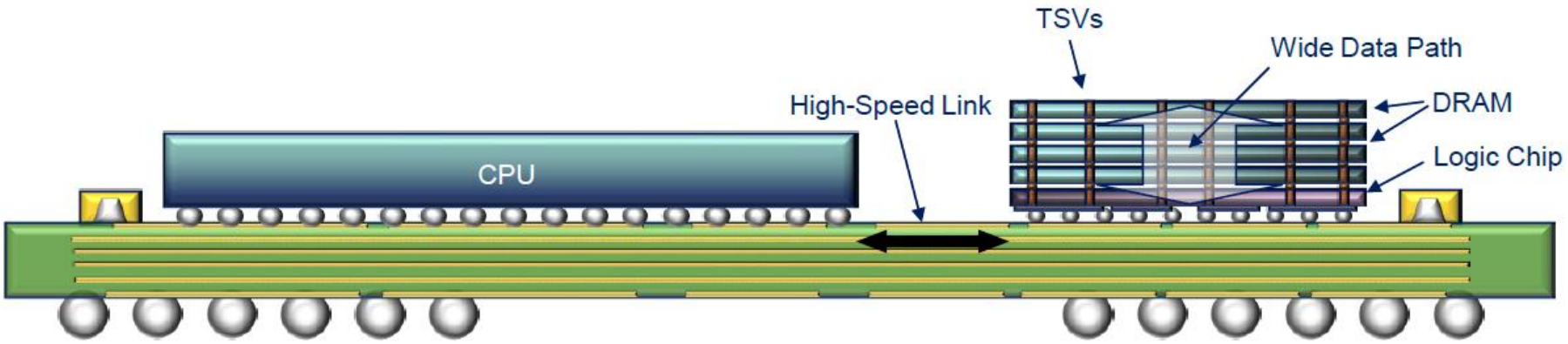
Outline

- 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
- Combine with on-chip optical interconnects





Micron, Hotchips 2011

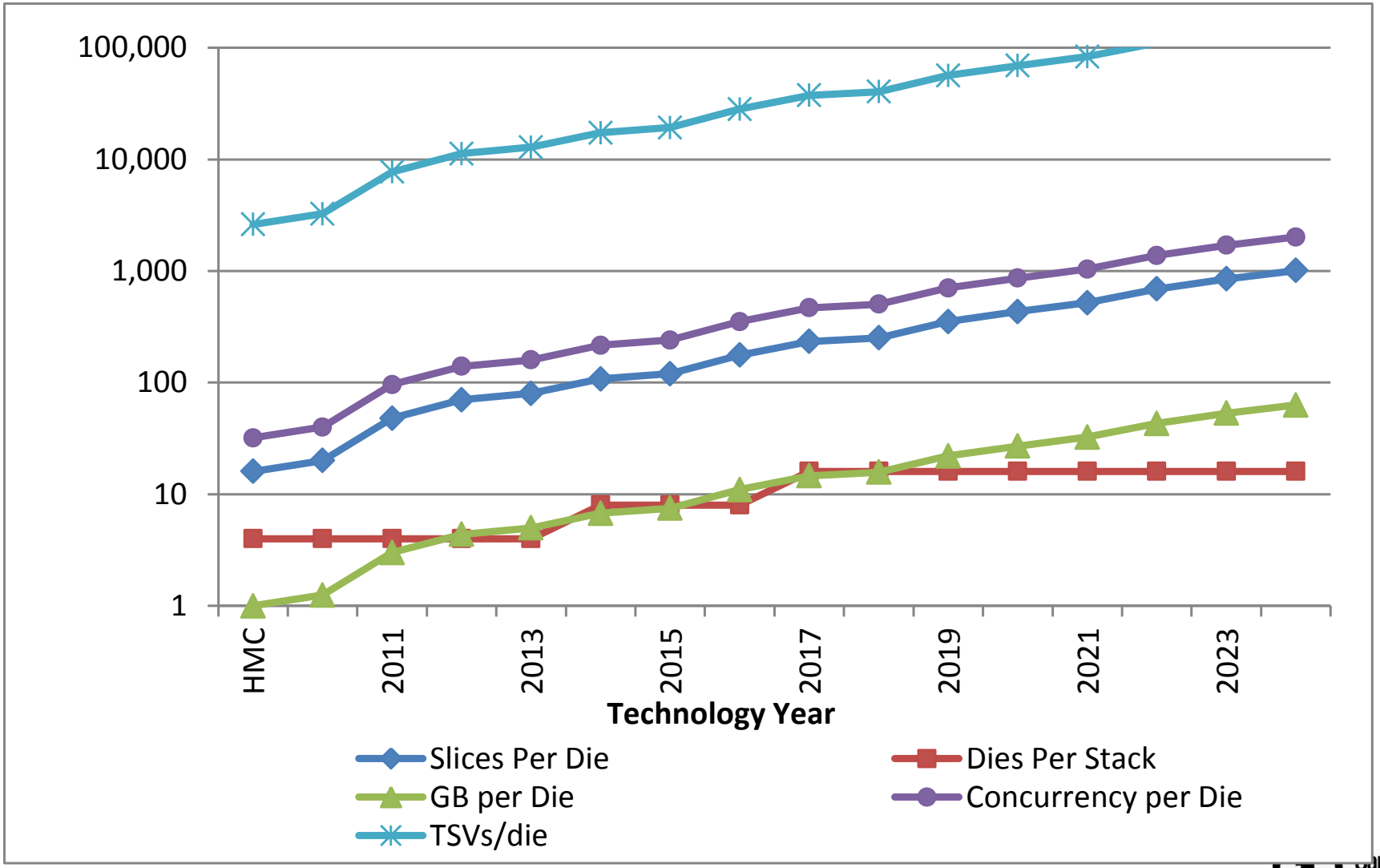
Hybrid Memory Cube Power Claims

- **Major power & bandwidth 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

Micron, Hotchips 2011

How Far Will TSV Stacking Take Us?

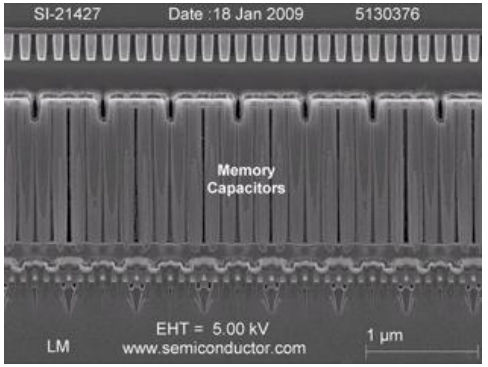


Courtesy Peter Kogge,
Matthew Marinella

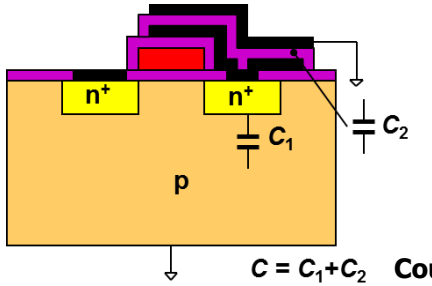
DRAM, TSV Limitations

- 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?

Micron Stacked DRAM



Stacked DRAM Cell



Courtesy Dieter Schroder, ASU

Outline

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

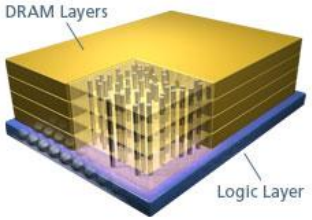
Timeline of Supercomputer Architectures

Standard Heavy/ Light node

10 Petaflop
1000 pJ/op

2012

3D Stacked DRAM + chip-chip optical interconnect

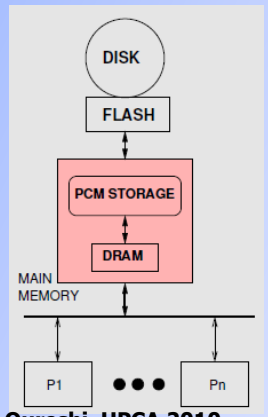


Micron Hotchips 2011

100-1000 Petaflop
10-100 pJ/op

2015

Emerging NVM for Storage, DRAM/NVM Hybrids

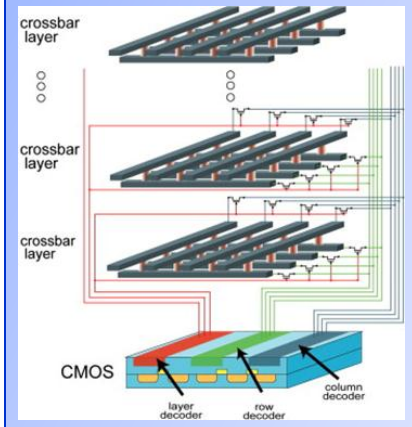


Qureshi, HPCA 2010

1-10 Exaflop
1-10 pJ/op

(Exact timeline TBD)

Storage Class Memory



Kügeler et al, SSE, 2009

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: File Storage

- First goal for emerging NVM technologies
- Need to beat high end flash:

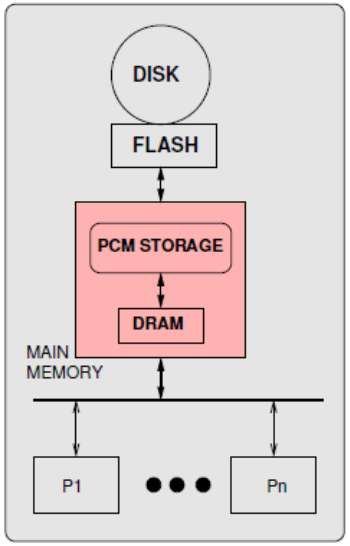
1. Voltage: < 15V
2. Endurance: > 10^4 W/E cycles
3. Scalability: < ~18 nm, 3D stackable
4. W/E time: < 100 µs
5. Retention: > 10 years

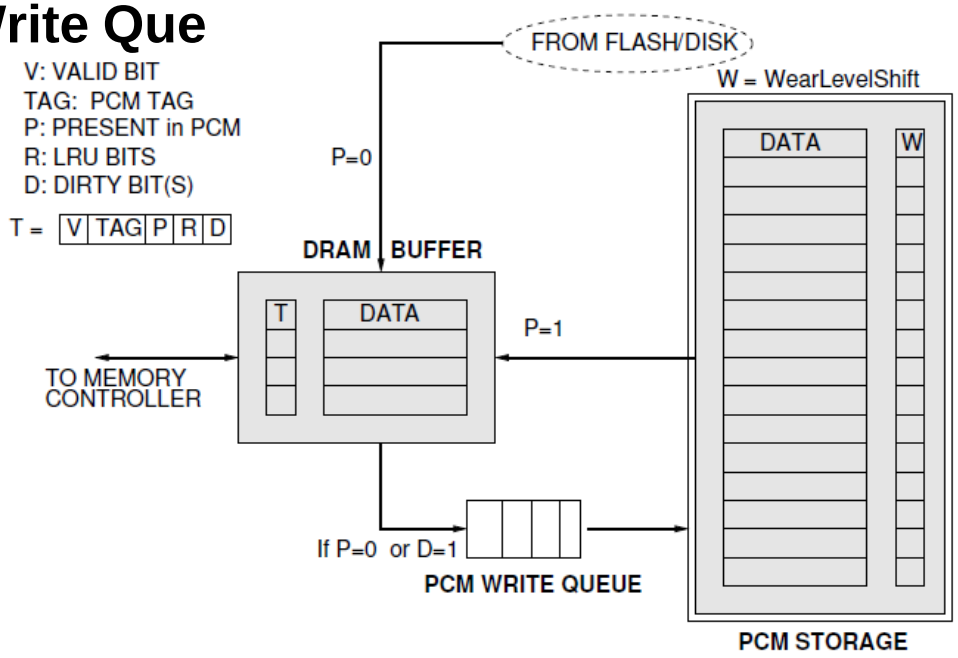
- All emerging technologies have proven these capabilities

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 (Qureshi, 2009)
  1. HDD→DRAM, allocate space in NVM
  2. If needed: DRAM→Write Que
  3. Write Que→NVM





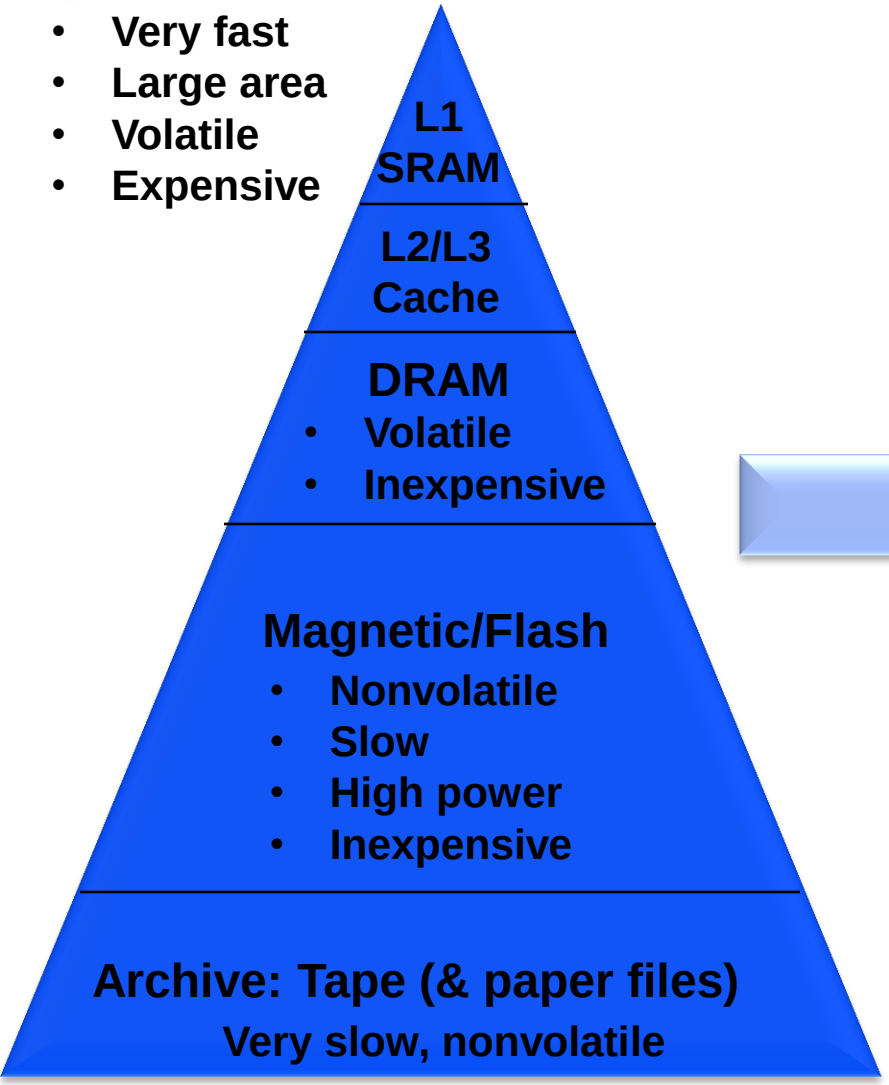
Qureshi, HPCA 2009

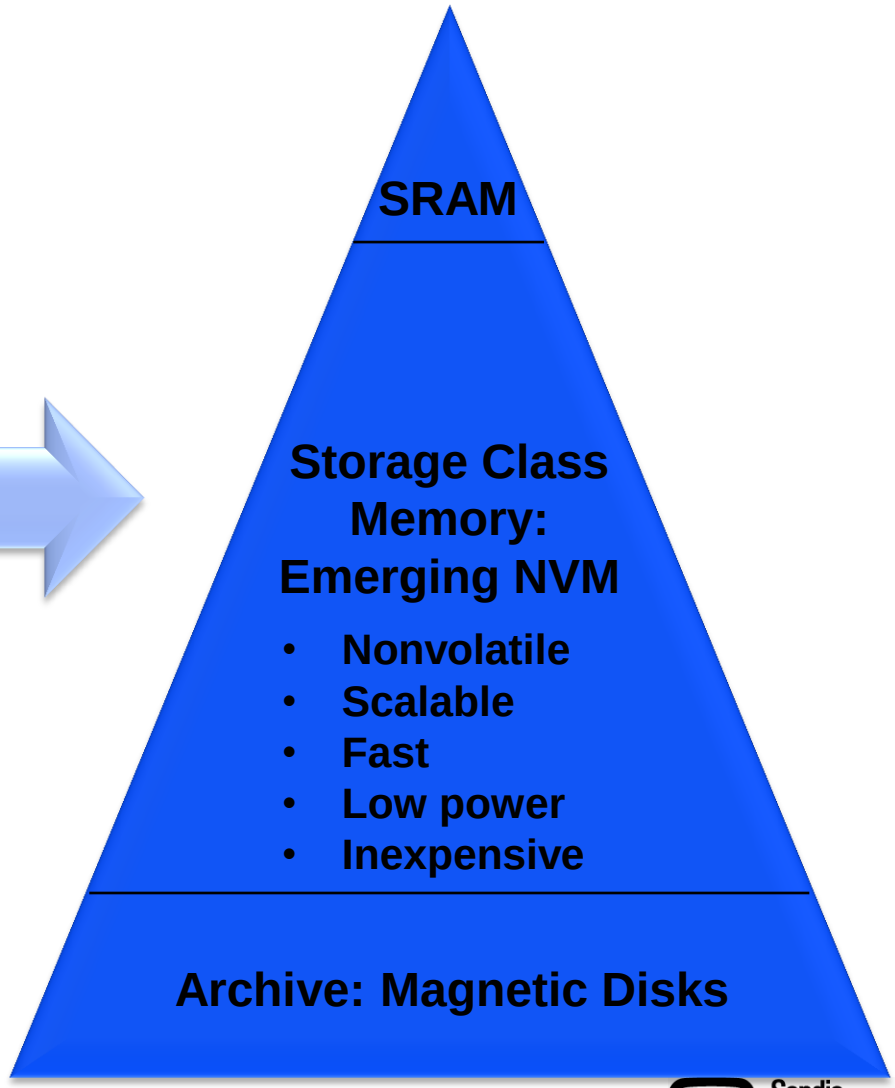
Resiliency

- Exascale computers 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

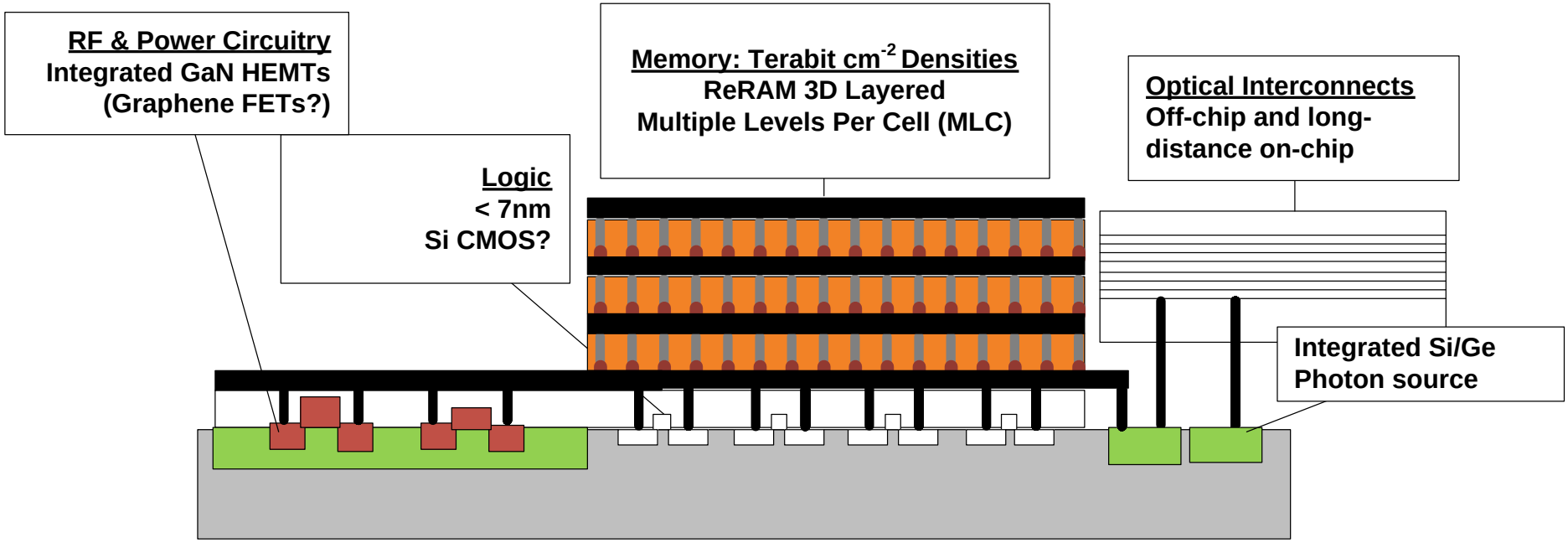
Universal/Storage Class Memory: A Game Changer

- Very fast
- Large area
- Volatile
- Expensive



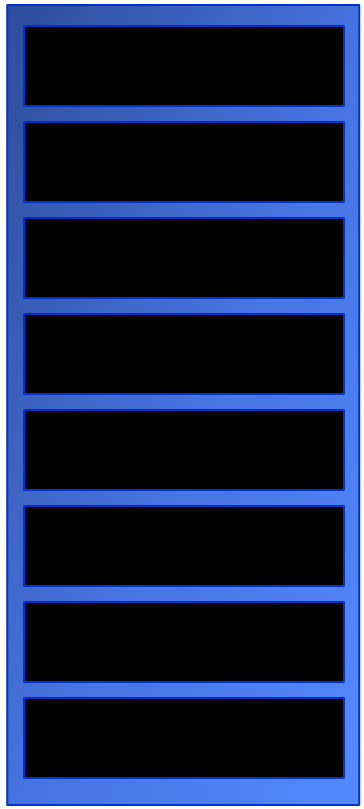


What will Universal Memory Look Like?



Universal Memory-Logic "Cube"

2012 Server Rack

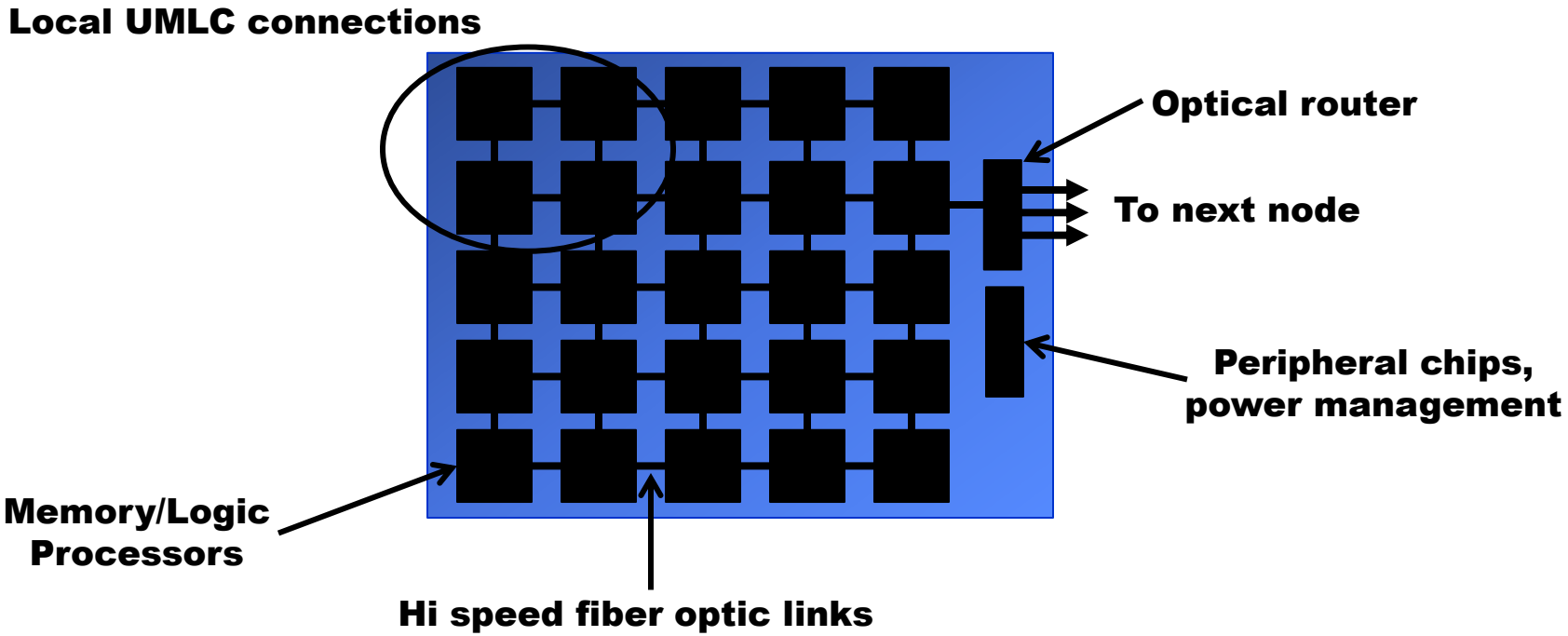


- Several hundred cores
- Hundreds of teraflops
- Main memory & storage
- Tens of terabytes
- Tens of Watts

Single Chip

Now We Are Ready for Exascale

- New exascale strawman: Universal Memory Logic Cube Node
- This will solve a lot of problems!



Many new architectural questions must be answered

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

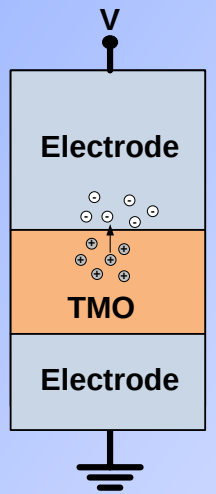
1. **Energy:** **< 1pJ per write/erase op**
2. **Endurance:** **> 10^{15} W/E cycles**
3. **Scalability:** **< 10 nm, 3D stackable, no select transistor**
4. **Read/Write:** **> 1 ns**
5. **Retention:** **> 10 years fully scaled at operation temp**
6. **Reliable Operation**

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

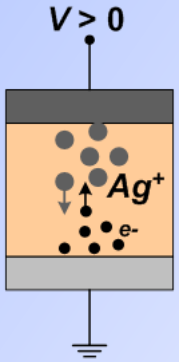
SCM Candidates

ReRAM

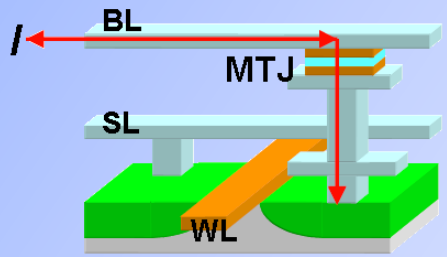
VCM



ECM

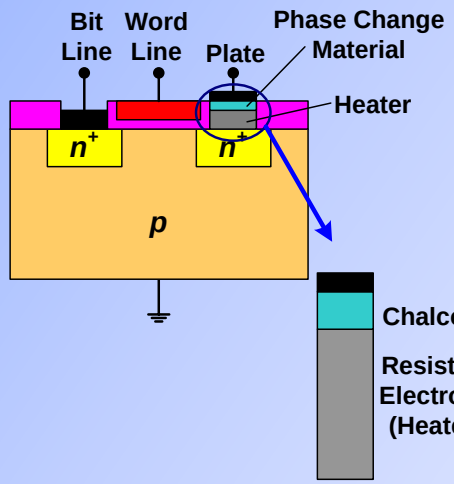


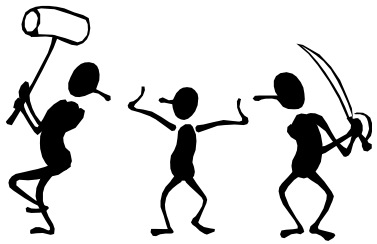
STT-MRAM



ITRS ERD 2010 Memory Rept.

Phase Change





ECM, 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									

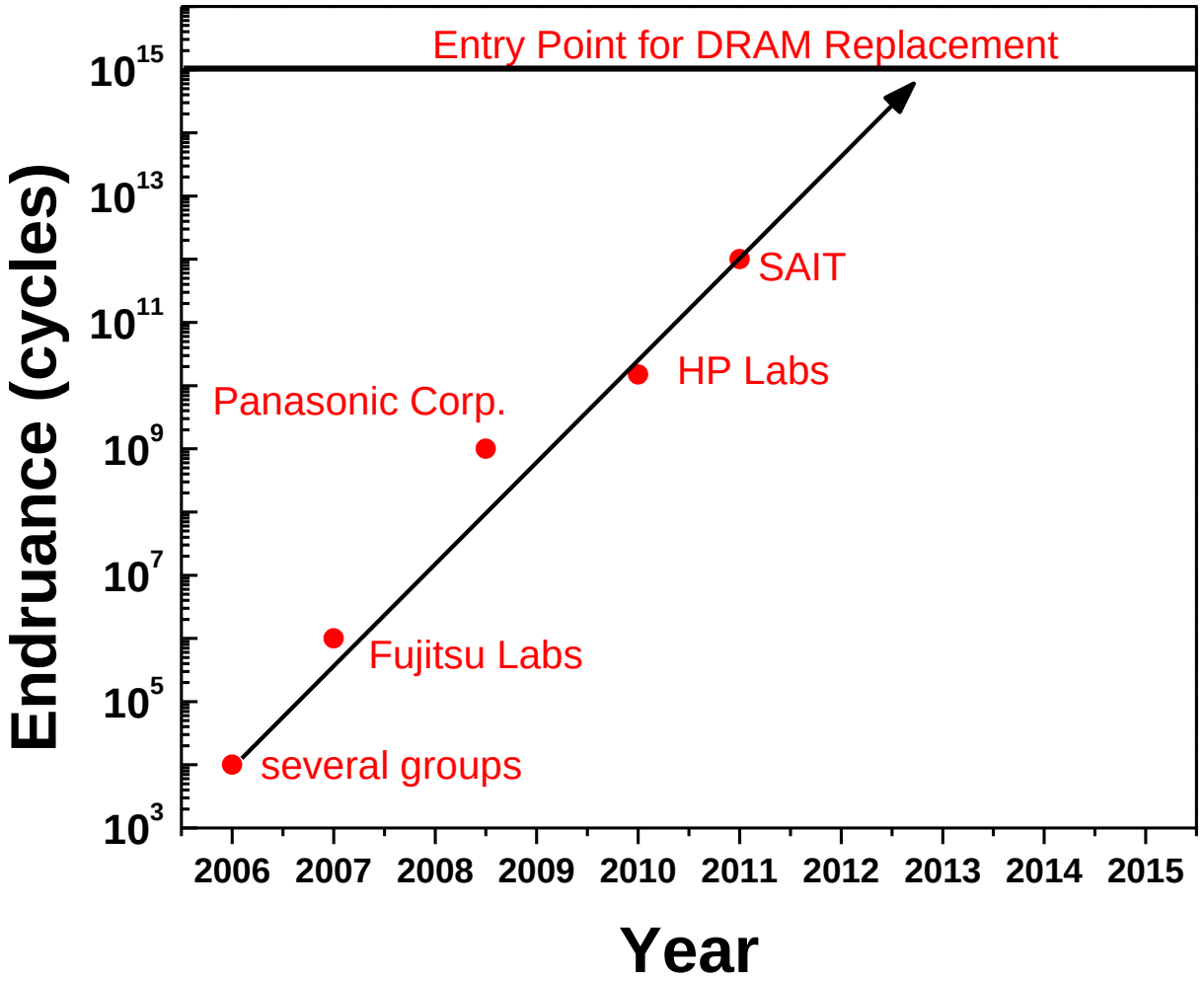
Scalability	F_{min} >45 nm
MLC	difficult
3D integration	difficult
Fabrication cost	high
Endurance	≤1E5 write cycles demonstrated

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)

ReRAM's Path to Universal Memory

- ReRAM is the least mature of major contenders
  - Also (arguably) shows the greatest promise
- Device/material level improvements
  - Need > 10^{15} W/E cycles (10^{16} desired)
  - Scalable select device (no select transistor)
  - Uniformity & reliability issues (can circuitry help?)
  - Still must eliminate sneak paths/parasitics
- Circuitry improvements
  - Read/write, wear leveling, error correction
- Architectural
  - Reorganize buffers, row/col for max efficiency
  - 3D addressing for stacked memory

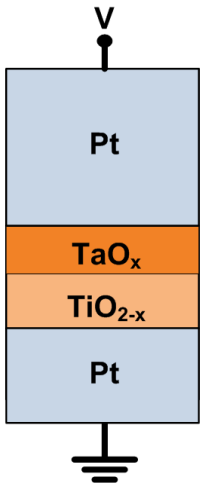
The Good News: Challenges all result from immaturity

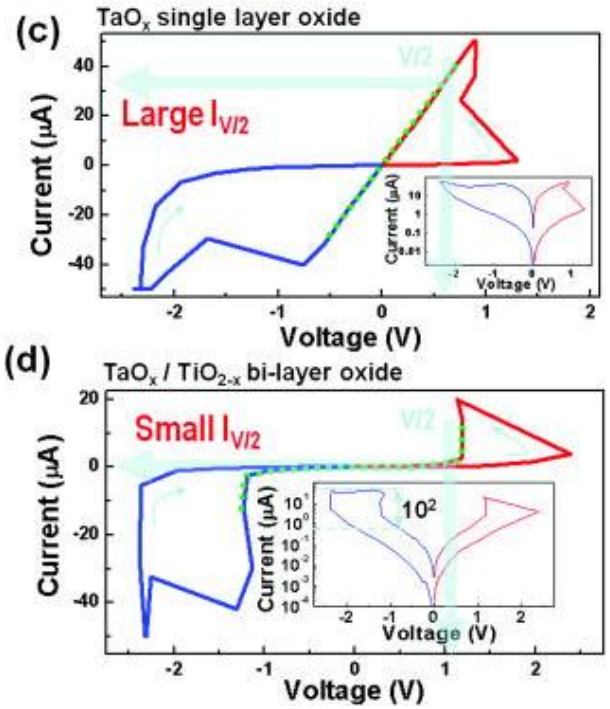
No fundamental physical showstoppers

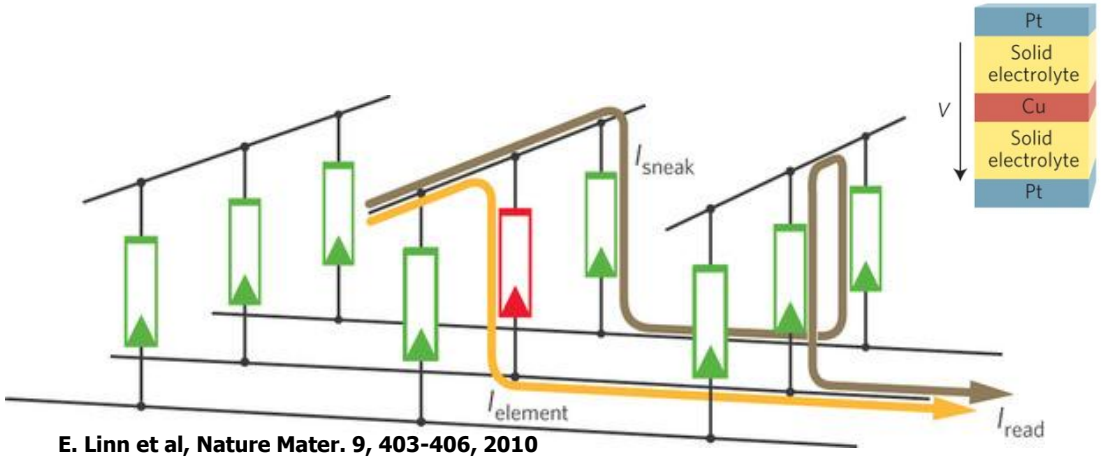
Select Device

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

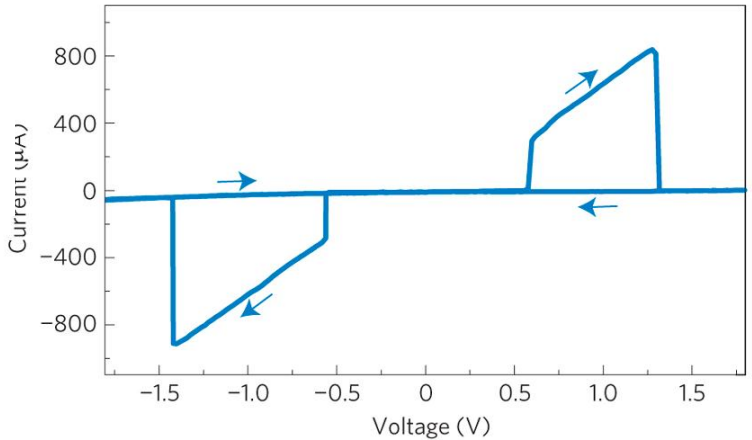
J. Yang et al, APL 100, 113501, 2012.





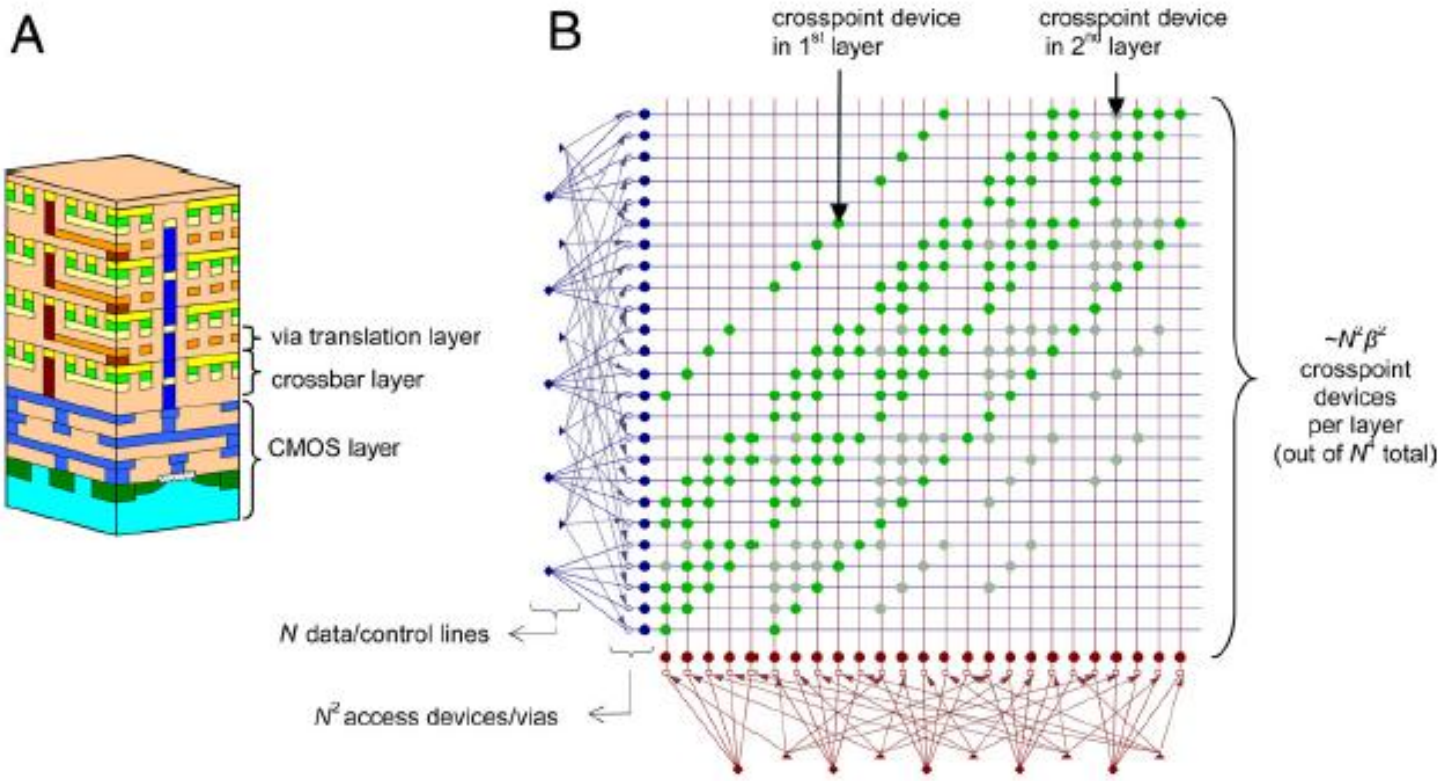


E. Linn et al, Nature Mater. 9, 403-406, 2010



3D Stack Addressing

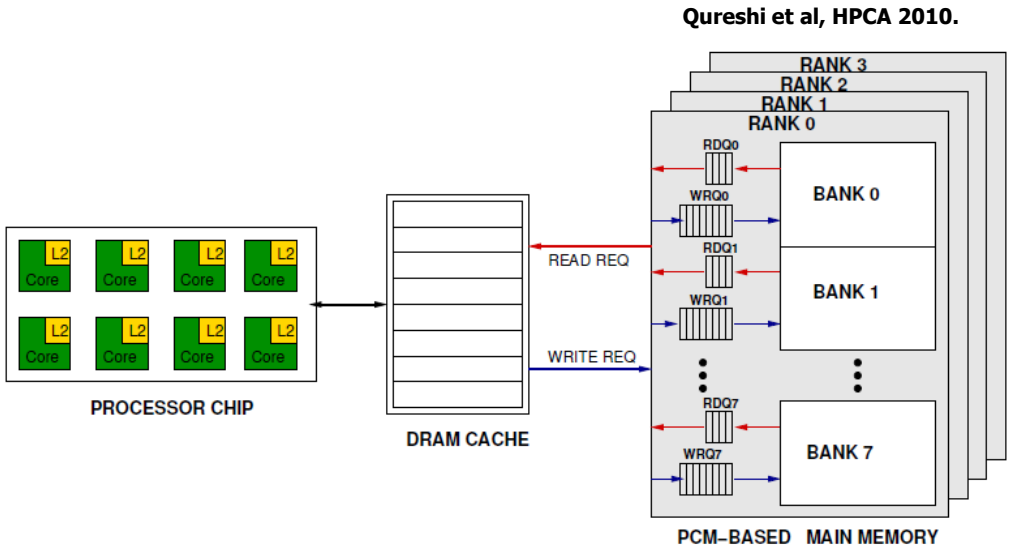
- **How do we control many layers with a CMOS base layer?**

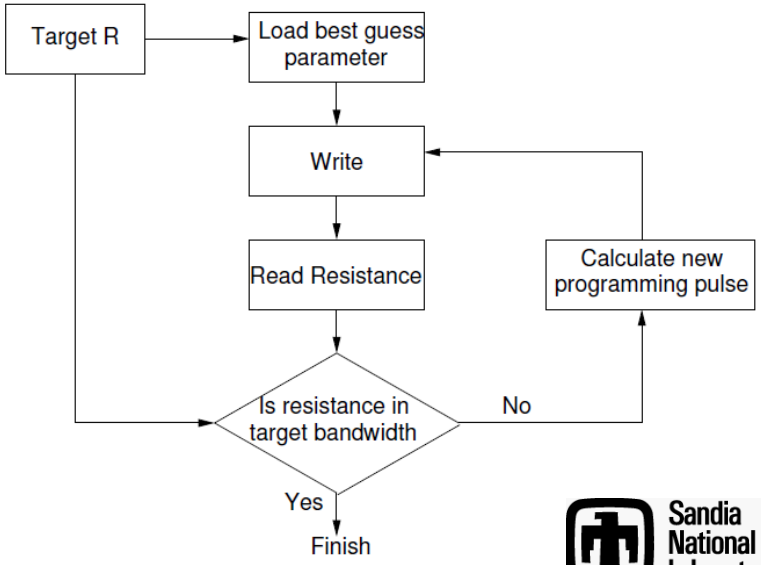


Strukov et al, PNAS, 2009

Array Architecture

- How do we architect ReRAM as a main memory array?
- What new issues will we face when converting from DRAM array → ReRAM
- This process has been started for PCM
  - Example – PCM architecture and write scheme below
- Do we need wear leveling?
- Work needed for ReRAM (can learn from PCM techniques)





Summary

- Exascale computing will be a tough road
  - Biggest challenges: Power and memory
- Exascale will *need* an advanced memory solution
  - Whether this is at 0.1 or 10 exaflops is TBD
- Current solution: Hybrid Memory Cube & optical interconnects
- Next generation: Hybrid DRAM & Advanced Memory
- Universal/Storage Class Memory is coming
  - This is a game changer for all scales of computing
  - Will be a major technological breakthrough of this decade
- Universal Memory based on ReRAM will enable a new generation of supercomputers

Acknowledgements

- Useful discussions with Erik Debenedictis (SNL), Jianhua Yang (HPL), Rich Murphy (SNL), Greg Astfalk (HP), Stan Williams (HPL), Greg Snider (HPL), Michael Kozicki (ASU), Dieter Schroder (ASU), 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

References/Further Reading

1. 2011 International Technology Roadmap for Semicondcutors, “Emerging Research Devices” and “Process, Integration, and Device Structures” Chapters. Available at itrs.net.
2. P. Kogge et al “Exascale Computing Study: Technology Challenges in Achieving Exascale Systems,” available online at: www.cse.nd.edu/Reports/2008/TR-2008-13.pdf.
3. P. Kogge, “The tops in flops,” IEEESpectrum, 48, 48-54, 2011.
4. P. Coteus, “Packaging the Blue Gene/L supercomputer,” IBM J. Research Dev., 49, 213-248, 2005.
5. J.T. Pawlowski, “Hybrid Memory Cube (HMC),” in HotChips 2011. Available at hotchips.org.
6. C Kügeler, et al, "High density 3D memory architecture based on the resistive switching effect." Solid-State Electronics 53, 1287-1292, 2009.
7. M.K. Qureshi et al, *Phase Change Memory*, Morgan & Claypool, 2011.
8. M.K. Qureshi et al, “Scalable High Performance Main Memory System Using
9. Phase-Change Memory Technology” in HCPA 2009.
10. E. Linn et al, “Complementary resistive switches for passive nanocrossbar memories,” Nature Mater. 9, 403-406, 2010.
11. J. Yang et al, “Engineering nonlinearity into memristors for passive crossbar applications,” APL 100, 113501, 2012.
12. Strukov et al, “Four-dimensional address topology for circuits with stacked multilayer crossbar arrays,” Proc NAS, 48, 20155–20158, 2009
13. M.K. Qureshi et al “Improving Read Performance of Phase Change Memories via Write Cancellation and Write Pausing,” in HPCA 2010.