Area, Power, and Latency Considerations of STT-MRAM to Substitute for Main Memory

Youngbin Jin, Mustafa Shihab, and Myoungsoo Jung

Computer Architecture and Memory Systems Laboratory

Department of Electrical Engineering

The University of Texas at Dallas

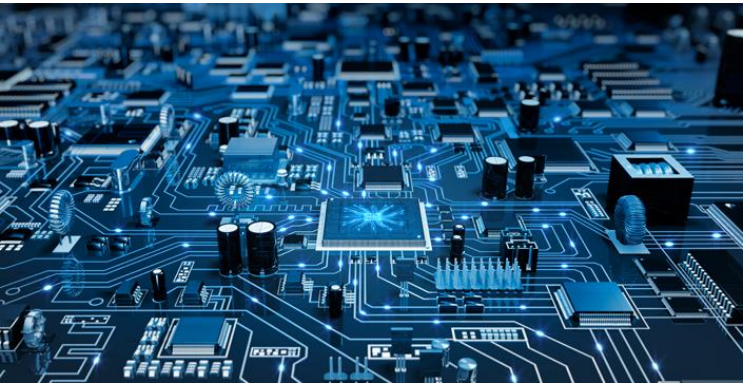
Takeaways

- ***Challenge:*** **Memory Scaling** and **Leakage and Refresh Power** consumption issues are constraining DRAM's long-term future as main memory
- ***Probable Solution:*** Replacing DRAM with new non-volatile memories such as STT-MRAM
- ***Underlying Issues:***
  - **Cell area of STT-MRAM** is nearly four times that of DRAM, restraining it from achieving cell density required by main memory
  - **Write latency and power** is too large and inhibit STT-MRAM's implementation as main memory
- ***Our Contribution:***
  - **Classified the critical device parameters**, including thermal stability factor, that can make STT-MRAM better suited for main memory
  - **Introduced an optimized STT-MRAM** that can offer DRAM-like density, lower power consumption, and shorter latency
- ***Results:*** Our early-stage optimized STT-MRAM can offer shorter latency and lower power consumption than a baseline DRAM by **18.4%** and **66.2%**, respectively

Overview

- Motivation
- MTJ, STT-MRAM Read and Write
- Critical Parameters
- Memory Level Optimizations
- Conclusion

DRAM: Importance and Challenges



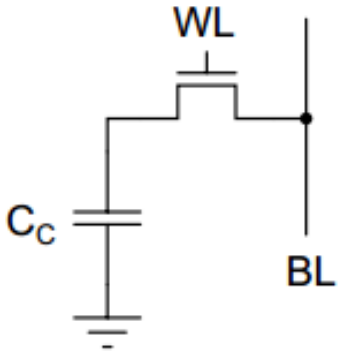
Embedded Systems

High Performance Computing

For around 3 decades DRAM has been THE main memory of choice

However….

- Cell leakage current is making scaling down for larger memory a critical bottleneck
- Refresh power for large memory is becoming almost as high as the processing power



Probable Solutions

Short-Term: Low-leakage, energy-efficient DRAMs

Long-Term: Replacing DRAM with newer and more efficient memory technology

STT-MRAM is a popular *Heir Apparent*

STT-MRAM is great:

- ✓Highly Scalable
- ✓Zero Leakage
- ✓High Endurance

But, nothing comes for free….

- Cell Area: $30F^2$ STT-MRAM cell is **4x larger** than DRAM (4 to 8 F^2)
- Write power and latency is **too high** as well

What do we do?

It is not a dead-end, rather an opportunity for exploration

We classify and tweak the Thermal Stability Factor and other critical parameters of STT-MRAM/MTJ to find Pareto optimal points

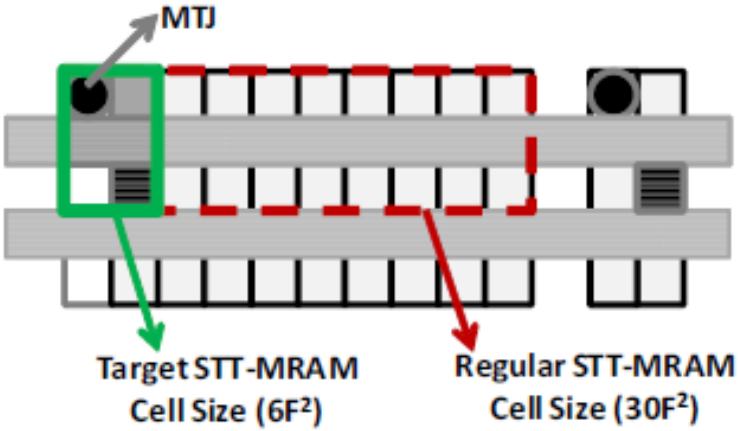
Optimized STT-MRAM can offer:

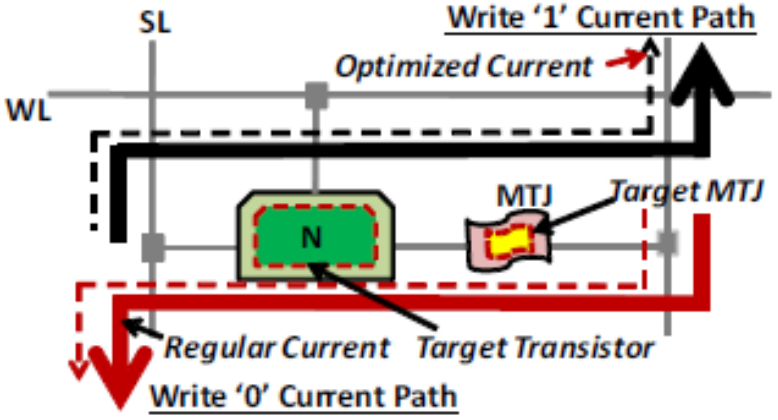
i) Higher density

ii) Lower power consumption on writes

iii) Shorter write latency

All comparable to modern DRAM technology

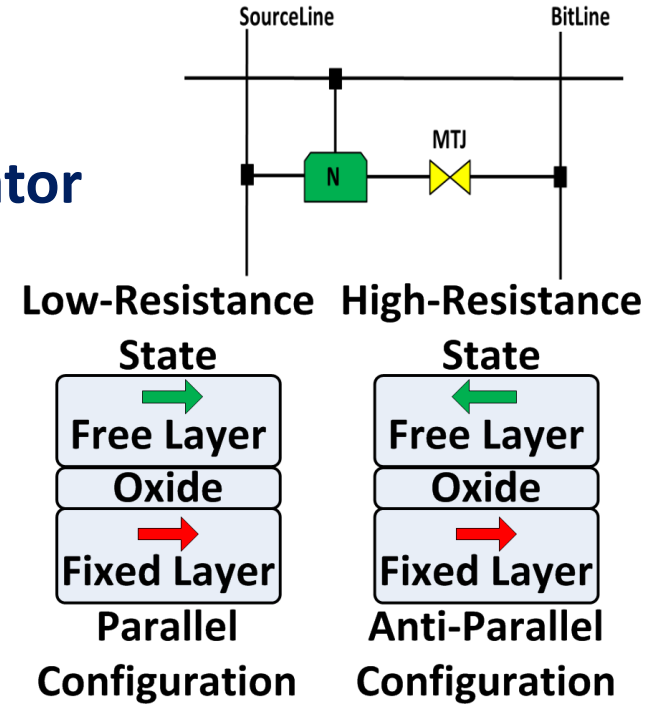




MTJ, STT-MRAM Read and Write

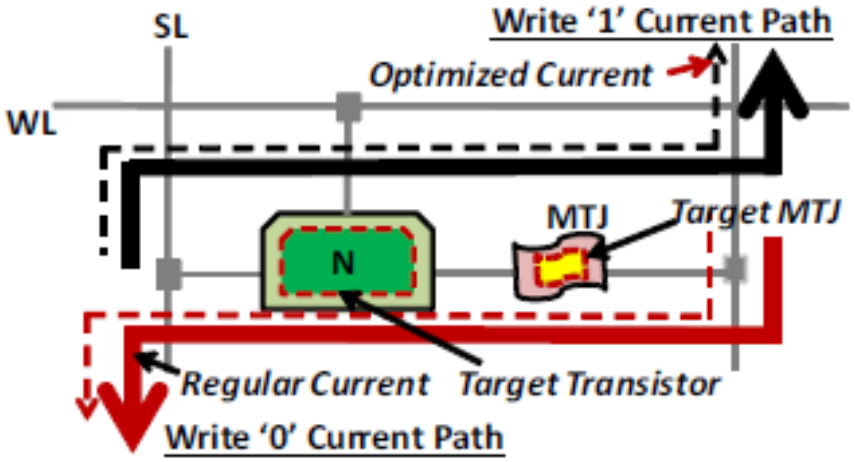
Magnetic Tunnel Junction (MTJ):

- Two ferromagnets separated by a thin insulator
- Stores data in terms of resistance level



Read Operation:

- Just sensing of the cell resistance
- Non-destructive read involves no write-back
- Only takes a couple of nanoseconds



Write Operation:

- Use polarized electrons to torque the magnetic state
- This process makes STTMRAM cell so large

If the transistor is not large enough to drive enough current then write failure occurs

Critical Parameters

Thermal Stability Factor (Δ):

- The stability is intrinsically linked to the magnitude of the energy barrier between the two orientations
- During a write process, energy barrier must be overcome

The higher the thermal stability factor, the higher the current (energy) or longer time it requires to write

$$\Delta = \frac{E_b}{k_B T} = \frac{H_K M_S V}{2 k_B T}$$

where, E_b = Energy Barrier, k_B= constant, T = Temperature,
H_K = Anisotropy Field Term, M_S = Saturation Magnetization, and
V = Volume of the MTJ = Area of the MTJ * Thickness of the MTJ = A. t_h

Critical Parameters

Critical Current (IC):

- **Determines the amount of current access transistor has to drive**

If the critical current can be reduced, access transistor size can be reduce …. we can achieve smaller cell area for STT-MRAM

$$I_{c0} = \left(\frac{4ek_BT}{h}\right) \cdot \frac{\alpha}{\eta} \cdot \Delta \cdot \left(1 + \frac{4\pi M_e ff}{2H_K}\right)$$

Where, e = electron charge, h = Plank's Constant, α = damping constant, η = STT-MRAM efficiency parameter, $4\pi M_{eff}$ = Effective demagnetization field, and Δ = Thermal Stability Factor

Critical Parameters

Retention Time:

- The expected time until a random bit-flip occurs in the cell
- Is determined by the thermal stability of the MTJ

High stability indicates the cell will have longer retention time, at the cost of higher write power and latency

$$\tau = \tau_0 \cdot exp(\frac{E_b}{k_B T})$$

where, τ_0 = operating frequency

Memory Level Optimization

DRAM Baseline:

Process technology: 45 nm

Cell Size:

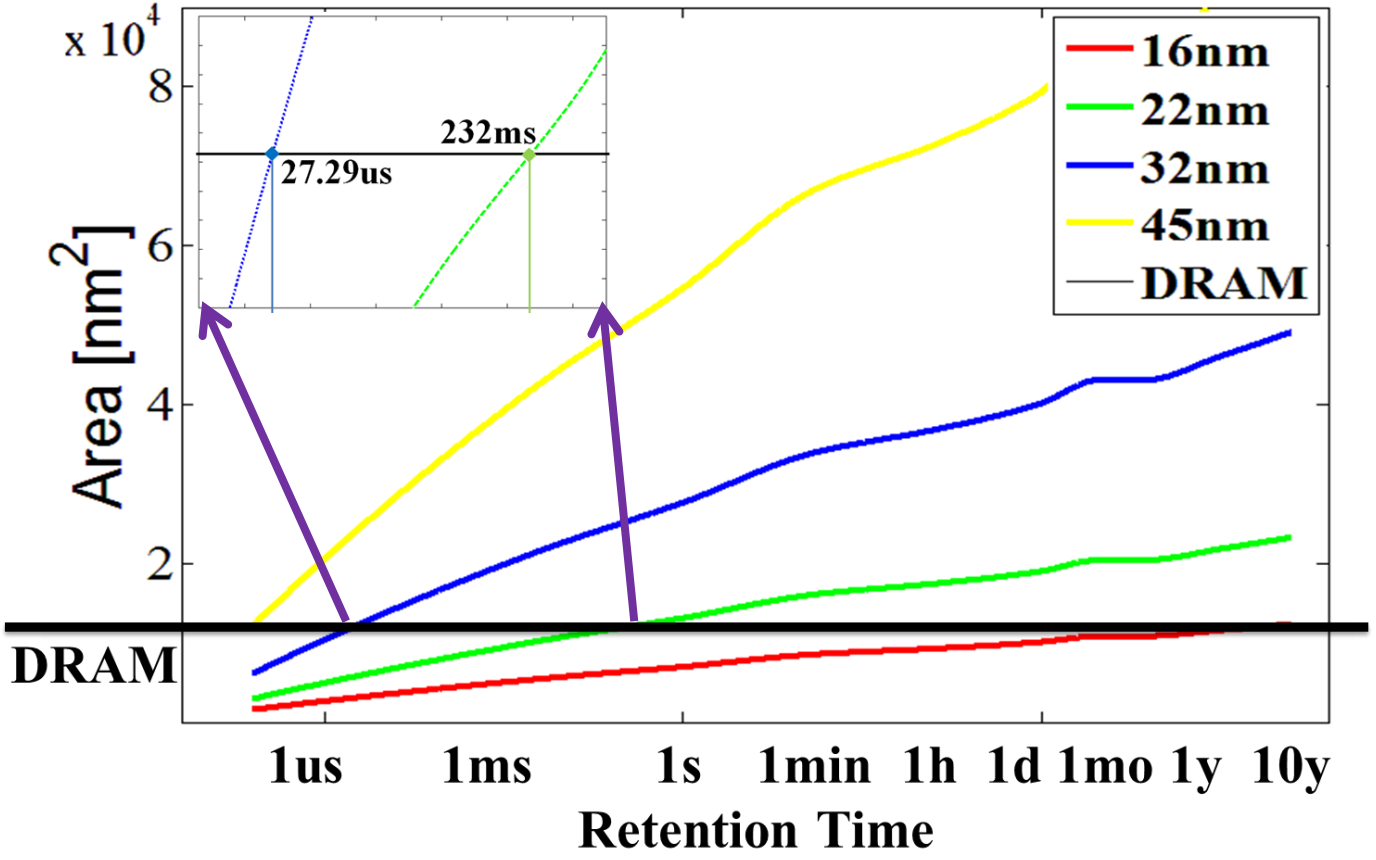
$6F^2$ ⇒ 12150 nm^2 (for 45 nm technology)

Write power: 88 µW*

*taken from [15]

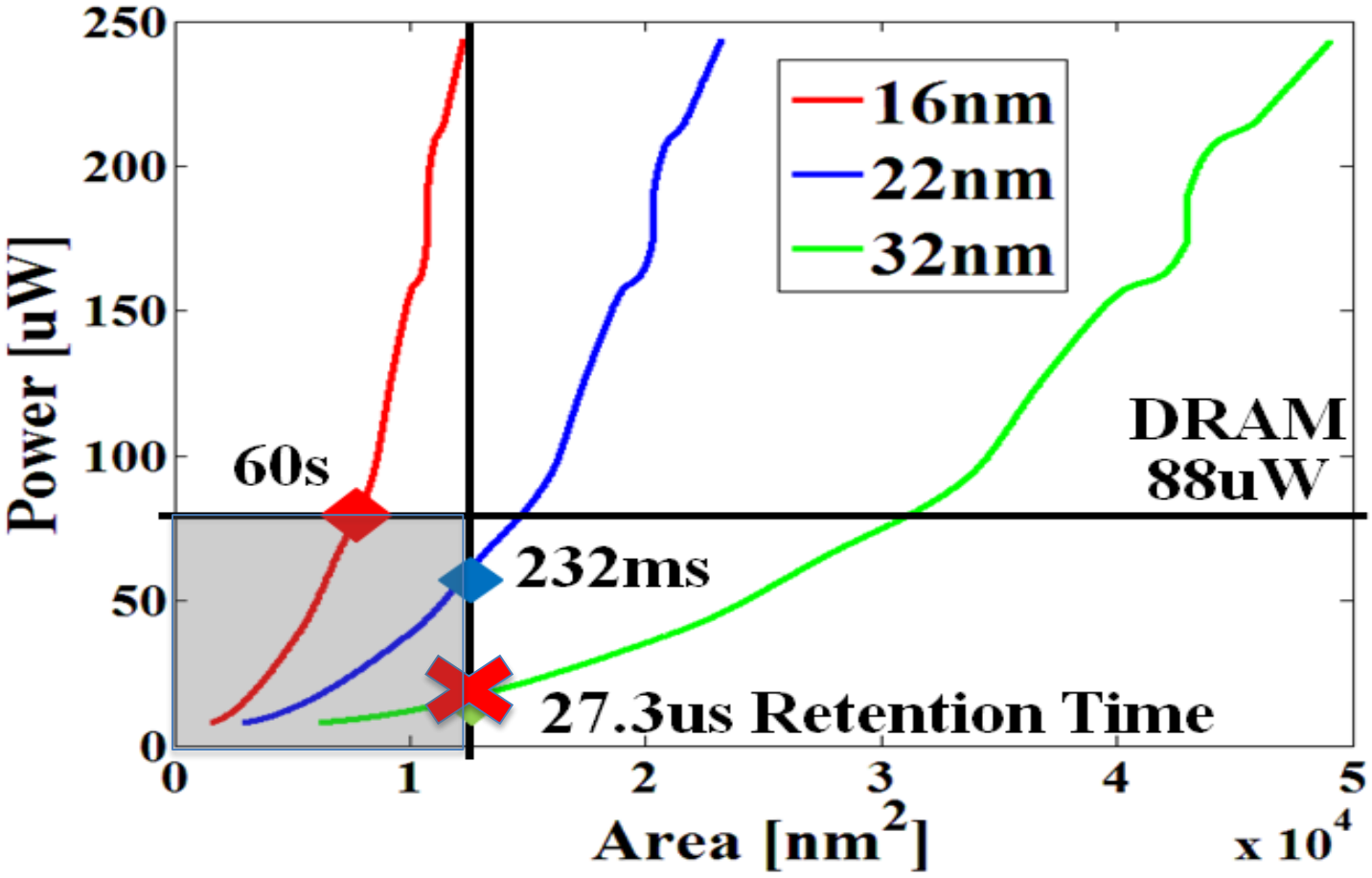
Memory Level Optimization

Area Optimization:



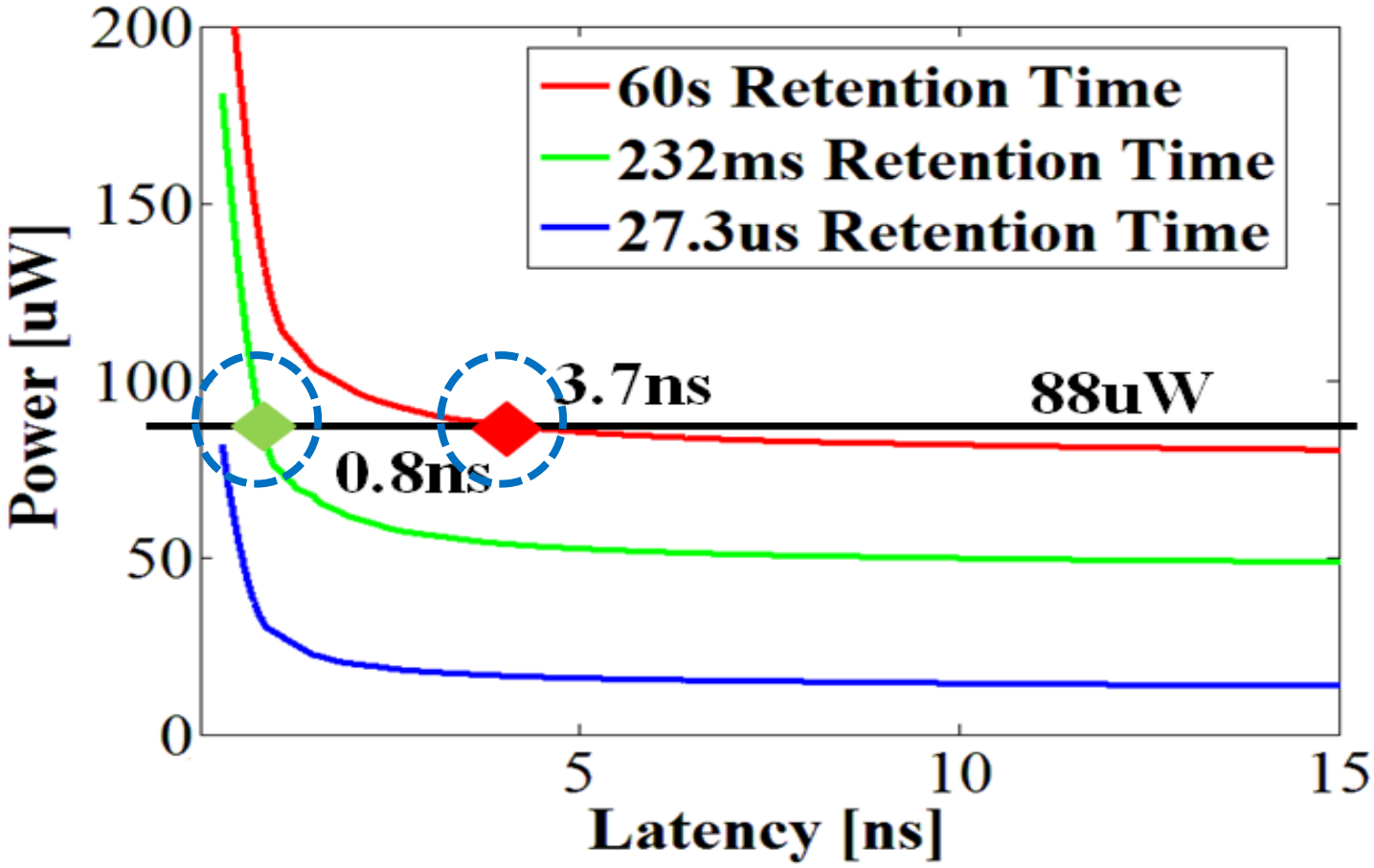
Memory Level Optimization

Power Optimization:

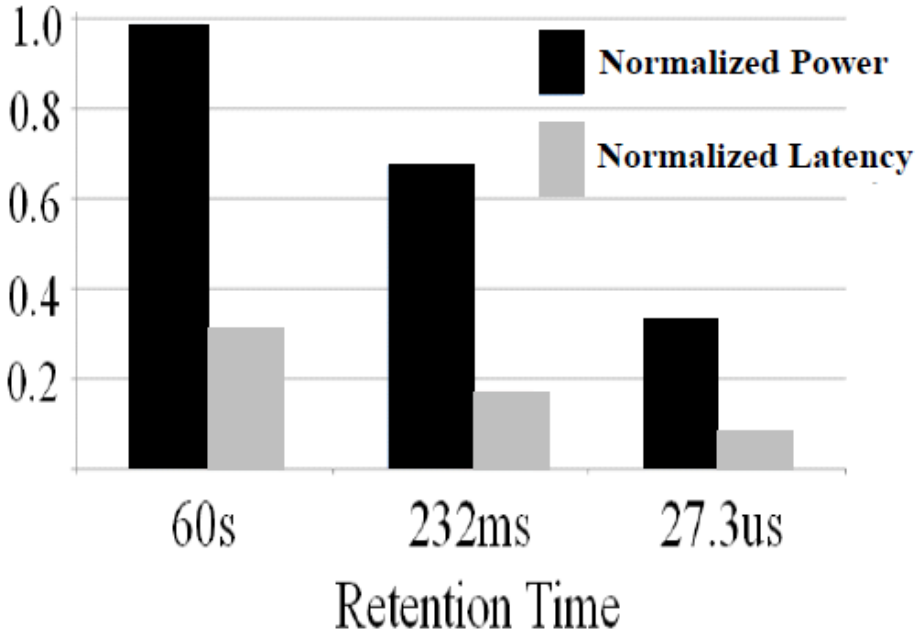


Memory Level Optimization

Write Latency Optimization:



Summary



Our early-stage optimized STT-MRAM:

- **Reduced power by 98.64%, 67.48%, and 32.48%**
- **Reduced normalized latency to 0.31, 0.17, and 0.074, compared to baseline DRAM**

Potential Implementations:

Process Technology	Thermal Stability Factor	Area (= 45nm DRAM)	Write Power (µW)	Write Latency (ns)	Retention Time
32 nm	27.26	**12150 nm²**	86.81	4.20	60 s
22 nm	19.26	**12150 nm²**	59.39	2.30	232 ms
16 nm	10.21	**12150 nm²**	28.59	1.00	27.3 µs

Related Prior Works

Zhenyu Sun et al. proposed both L1 cache and lower level cache designs using multi-retention level STT-MRAM cache, which can significantly reduce the total energy, while improving write for both level 2 and level 3 caches[16].

Clinton W. Smullen et al. claimed that retention-relaxed STT-MRAM can replace SRAM in processor caches by reducing the high dynamic energy and slow write latencies[12].

Emre Kultursay et al. explored the possibility of using STT-MRAM technology to completely replace DRAM in main memory by using partial write and row buffer write bypass[17].

Thank you

Questions?

References

[1] Khvalkovskiy, A. V., et al. *Basic principles of STT-MRAM cell operation in memory arrays*. Journal of Physics, 2013.

[2] Zhao, W. S., et al. *Failure and reliability analysis of STT-MRAM*. Microelectronics Reliability, 2012.

[3] Li, Jing, et al. *Design paradigm for robust spin-torque transfer magnetic RAM (STT MRAM) from circuit/architecture perspective*. In VLSI Systems, 2010.

[4] Sun, Jonathan Z., et al. *Spin angular momentum transfer in a current-perpendicular spin-valve nanomagnet*. Integrated Optoelectronic Devices, 2004.

[5] Dorrance, RichardWilliam. *Modeling and Design of STT-MRAMs*. PhD dissertation, 2011.

[6] Rizzo, N. D., et al. *Thermally activated magnetization reversal in submicron magnetic tunnel junctions for magnetoresistive random access memory*. Applied physics letters, 2002.

[7] Sun, J. Z. *Spin-current interaction with a monodomain magnetic body: A model study*. Physical Review, 2000.

[8] Raychowdhury, Arijit, et al. *Design space and scalability exploration of 1T-1STT MTJ memory arrays in the presence of variability and disturbances*. In IEDM, 2009.

[9] Huai, Yiming. *Spin-transfer torque MRAM (STT-MRAM): Challenges and prospects*. AAPPS Bulletin, 2008.

[10] Li, Hai, and Yiran Chen. *Nonvolatile Memory Design: Magnetic, Resistive, and Phase Change*. CRC Press, 2011.

References

[11] Lee, Benjamin C., et al. *Architecting phase change memory as a scalable dram alternative*. ACM SIGARCH Computer Architecture News, 2009.

[12] Smullen, Clinton W., et al. *Relaxing non-volatility for fast and energy-efficient STT-RAM caches*. In HPCA, 2011.

[13] Fong, Xuanyao, et al. *Failure Mitigation Techniques for 1T-1MTJ Spin-Transfer Torque MRAM Bit-cells*. 1-1.

[14] Diao, Zhitao, et al. *Spin-transfer torque switching in magnetic tunnel junctions and spin-transfer torque random access memory*. Journal of Physics, 2007.

[16] Sun, Zhenyu, et al. *Multi retention level STT-RAM cache designs with a dynamic refresh scheme*. In International Symposium on Microarchitecture, 2011.

[17] Kultursay, Emre, et al. *Evaluating STT-RAM as an energyefficient main memory alternative*. In ISPASS, 2013.

[18] Bergman, Keren, et al. *Exascale computing study: Technology challenges in achieving exascale systems.* DARPA IPTO, Tech. Rep 15, 2008.