# Improving Rad Performance of STT-MRAM based Main Memories through Smash Read and Flexible Read

Lei Jiang: Advanced Micro Devices, Inc
Wujie Wen: Forida International University
Danghui Wang: Northwestern Polytechnical University
Lide Duan: University of Texas at San Antonio

# Outline

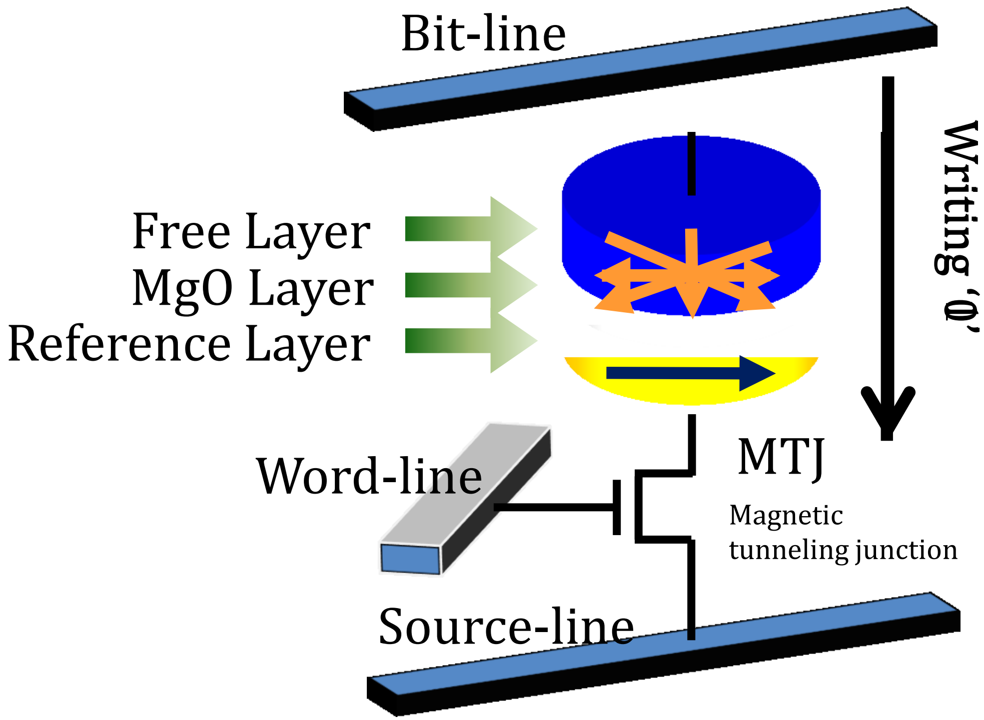
- Introduction
- Background and Motivation
- Proposed Methods
- Experimental
- Conclusion

# STT-MRAM Act as Main Memory

- DRAM
  - Power Hungry , due to periodic refreshes
  - Consumes 38.5% total energy in a smart-phone
  - Cell size: $6F^2$
- STT-MRAM
  - Non-volatility
  - Saves ~20% energy, compared to DRAM based main memory
  - Cell size: $8F^2$

# What is STT-RAM?

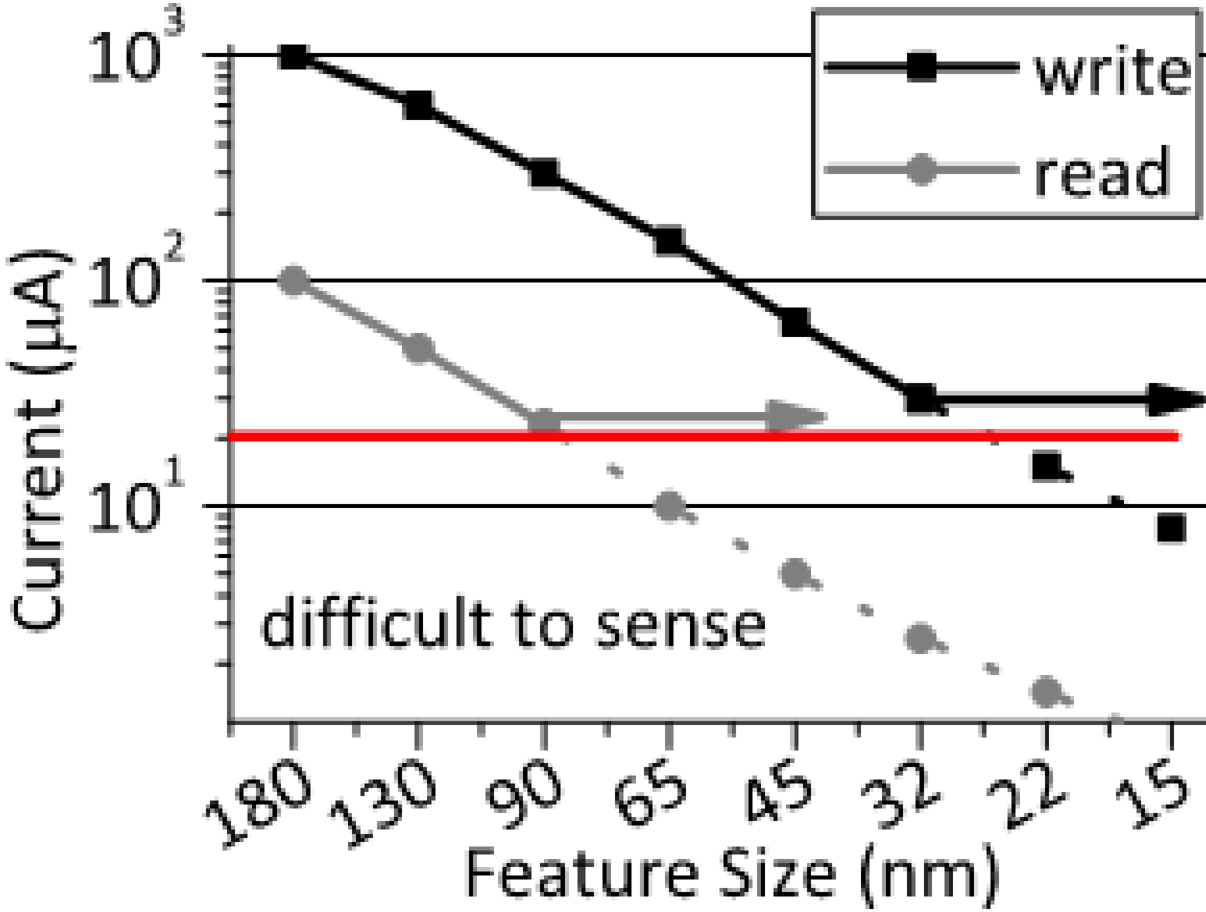
**Spin-Transfer Torque Random Access Memory**




1T-1MTJ STT-RAM Schematic

*A scalable technology*

# Write Current Scaling



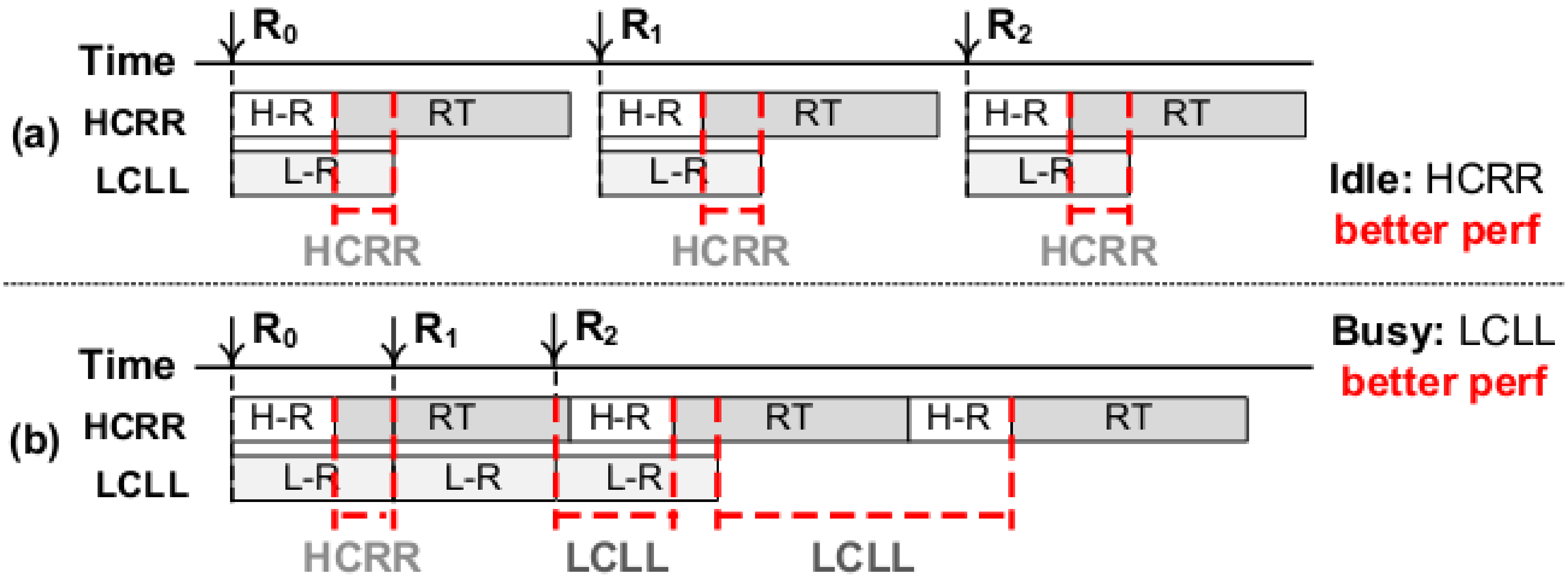

**Shrinking read current is challenging for small feature size**

**Read Disturb Errors!**

# How to overcome RDEs

- Destructive Read
  - High Current Restore Required (HCRR) read
  - Restore operation Increases the bank busy time, may block the following reads operations.
- Non-destructive Read
  - Low Current Long Latency (LCLL) Read
  - Prolong the latency , and has direct impact on the performance

# Motivation



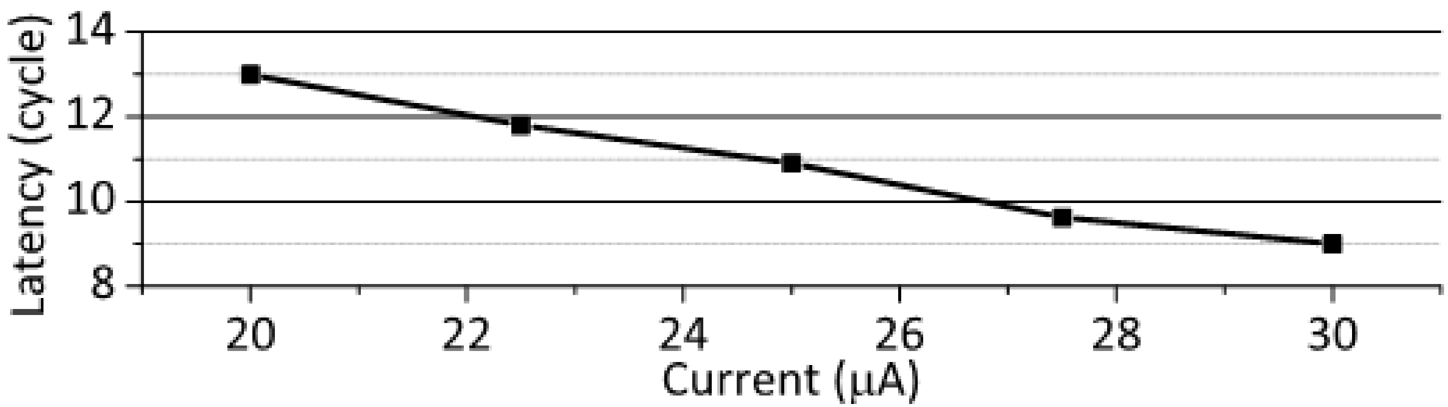

One single read scheme cannot always have the best performance for all applications.

An adaptive read method switching between two read schemes is a must.

# Smash Read

- The relationship between read current and read latency

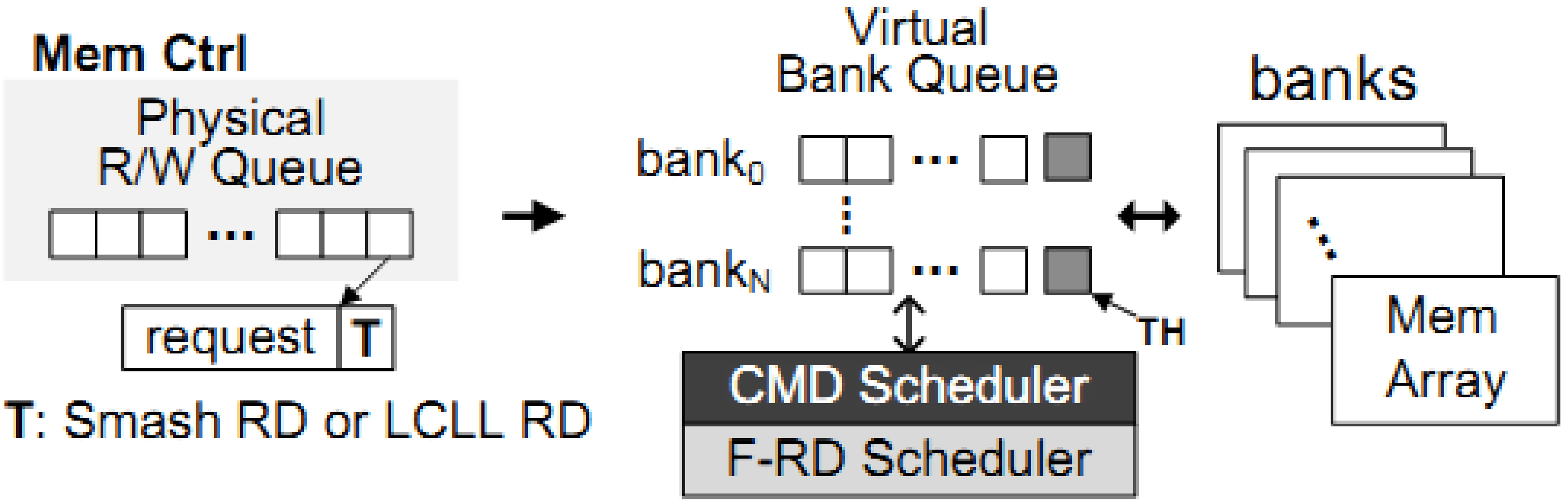
$$T_{RD} = \frac{C_{BL} \bullet \Delta V_{BL}}{I_{RD}}$$



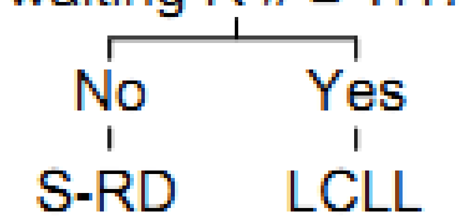

Read Latency VS Read Current

By boosting the read current from 20uA to 30uA, the read latency decreases from 13 cycles to 9 cycles.

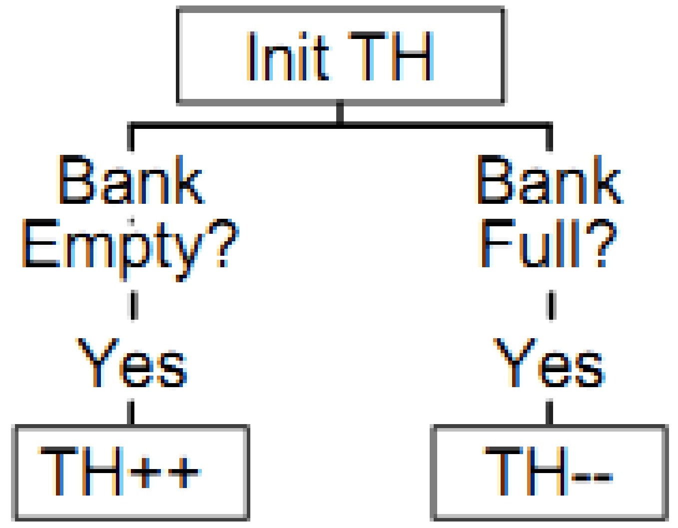
# Flexible Read



# Flexible Read Policy

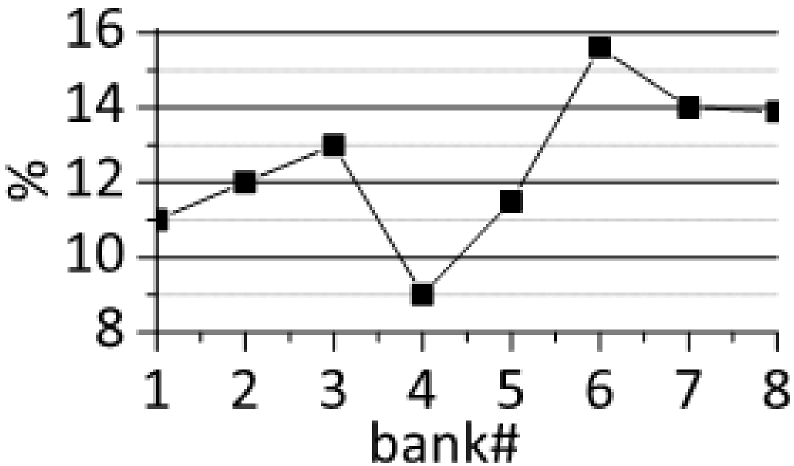



(a) Threshold.

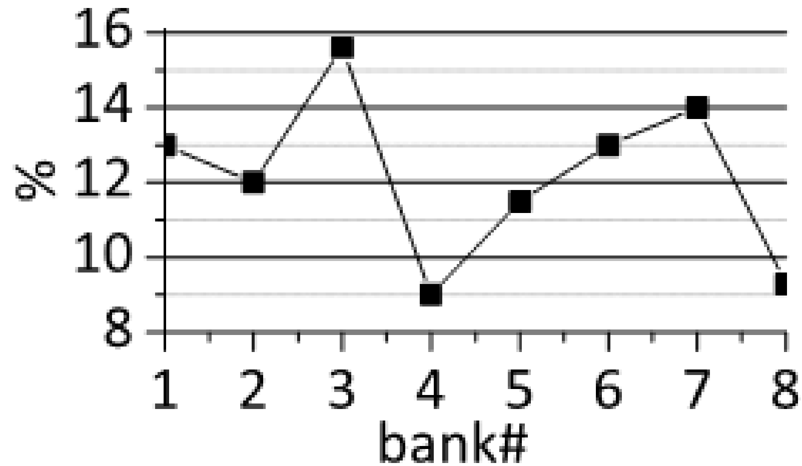



(b) Threshold Control.

# Read Distribution




(a) astar.




(b) sjeng.

Requests from CPU cores do not distribute evenly among all banks.

# Design Overhead

- Smash Read
  - Does not require any additional hardware
  - Increases read energy by ~50% over HCRR, evaluated by NVsim
  - F-RD tries to issue less S-RD
- Flexible Read
  - One bit in each read/write queue entry to indicate whether this entry stores a S-RD request or a LCLL Request. Totally 8 bytes.
  - One TH counter for every bank. Totally 3 bytes.
  - A R-RD scheduler is integrated into the CMD scheuler. ~0.8K gates.
  - Each scheuling operation increases 0.03pJ energy.

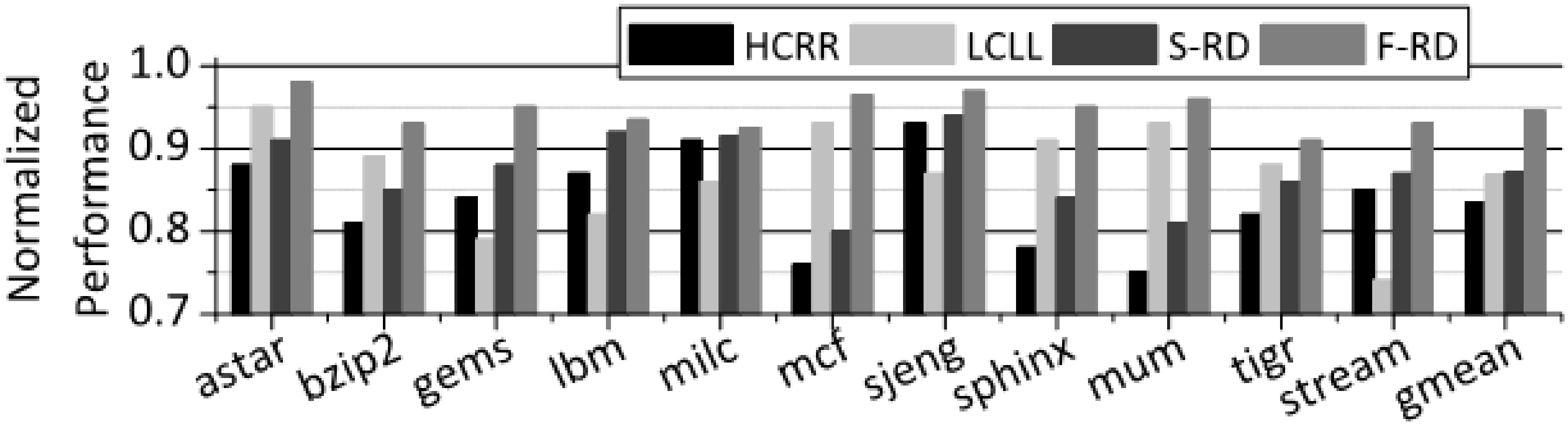
# Experimental Configuration

BASELINE CONFIGURATION

| | |
|---|---|
| CPU | 4 ARMv7 cores, 2GHz, out-of-order |
| SRAM L1 | private, I/D separate, 32KB/core, 64B line |
| SRAM L2 | shared, 8MB, 16-way LRU, 64B line, write back |
| Mem. Ctrl | on-chip, 64-entry R/W queues, close-page, FR-FCFS |
| LPDDR3 STT-MRAM based Main Memory | 1 channel, 1 rank-per-channel, 8 banks-per-rank. *Ideal*: the LPDDR3 timing and current values are configured as [32]. *HCRR*: *tRCD* 13, *tRC* 34, *IDD0(1.2V)* 61.44$mA$ *LCLL*: *tRCD* 16, *tRC* 21, *IDD0(1.2V)* 52.13$mA$ *S-RD*: *tRCD* 9, *tRC* 30, *IDD0(1.2V)* 62.78$mA$ |

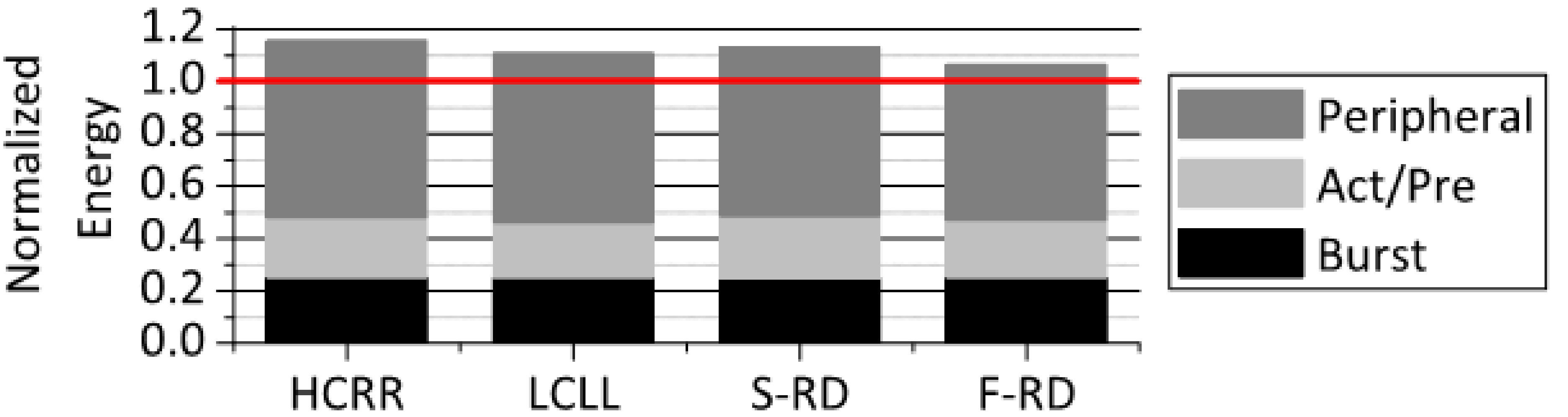
Workload:A subset of simulation benchmaiks from SPEC CPU2006, Bio-Bench and STREAM

# Evalution-Performance



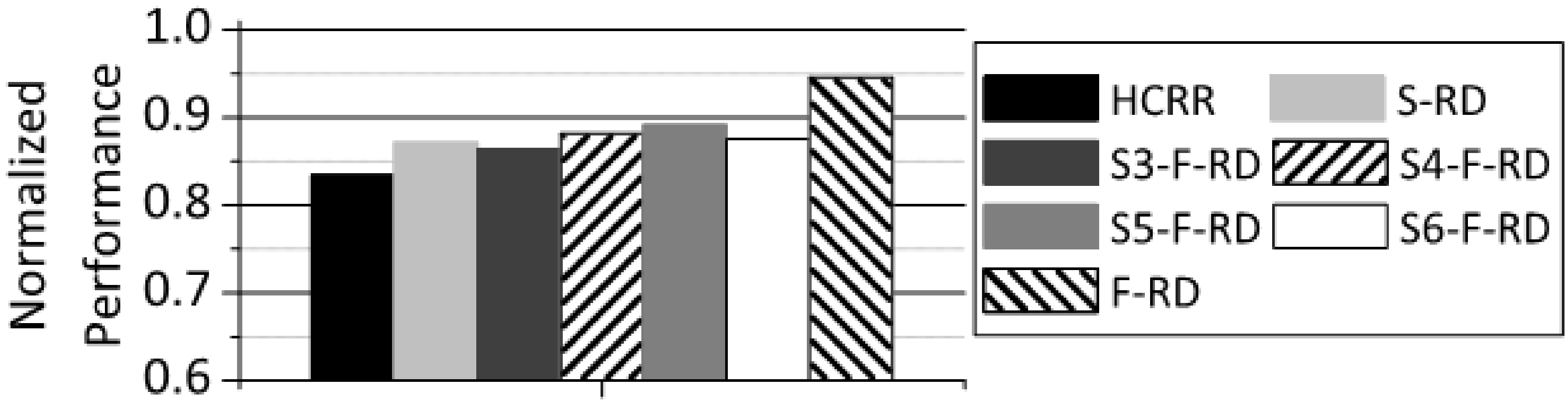

Our F-RD boosts system performance by 13.3% and 8.9% over HCRR and LCLL, respectively

# Evalution-Energy




HCRR adds additional 15.6% total main memory energy.
LCLL boosts main memory energy by 10.9%.
S-RD saves 3% main memory energy.
F-RD reduces main memory energy by 8% and 4% over HCRR and LCLL, respectively.

# Evalution- F-RD Threshold




*F-RD* improves main memory system performance by 6.1% over the best static scheme, i.e., *S5-F-RD*..

# Conclusion

- With fast write current scaling, the read disturbance has become an inevitable reliability issue for STT-MRAM.
- Neither HCRR reads nor LCLL reads can always achieve the best performance.
- We propose Smash Read to accelerate HCRR reads by issuing a larger read current.
- We further improve the STT-MRAM based main memory performance by Flexible Read.
- Experimental results show that Flexible Read gains the best performance in a LPDDR3 STT-MRAM based main memory system for a wide variety of applications.

# Thank you!