ROCKY: A Robust Hybrid On-chip Memory Kit For The Processors With STT-MRAM Cache Technology

Mahdi Talebi, Arash Salahvarzi, Amir Mahdi Hosseini Monazzah, Kevin Skadron, And Mahdi Fazeli

Arash Salahvarzi

RA @ CPS-Lab

IUST

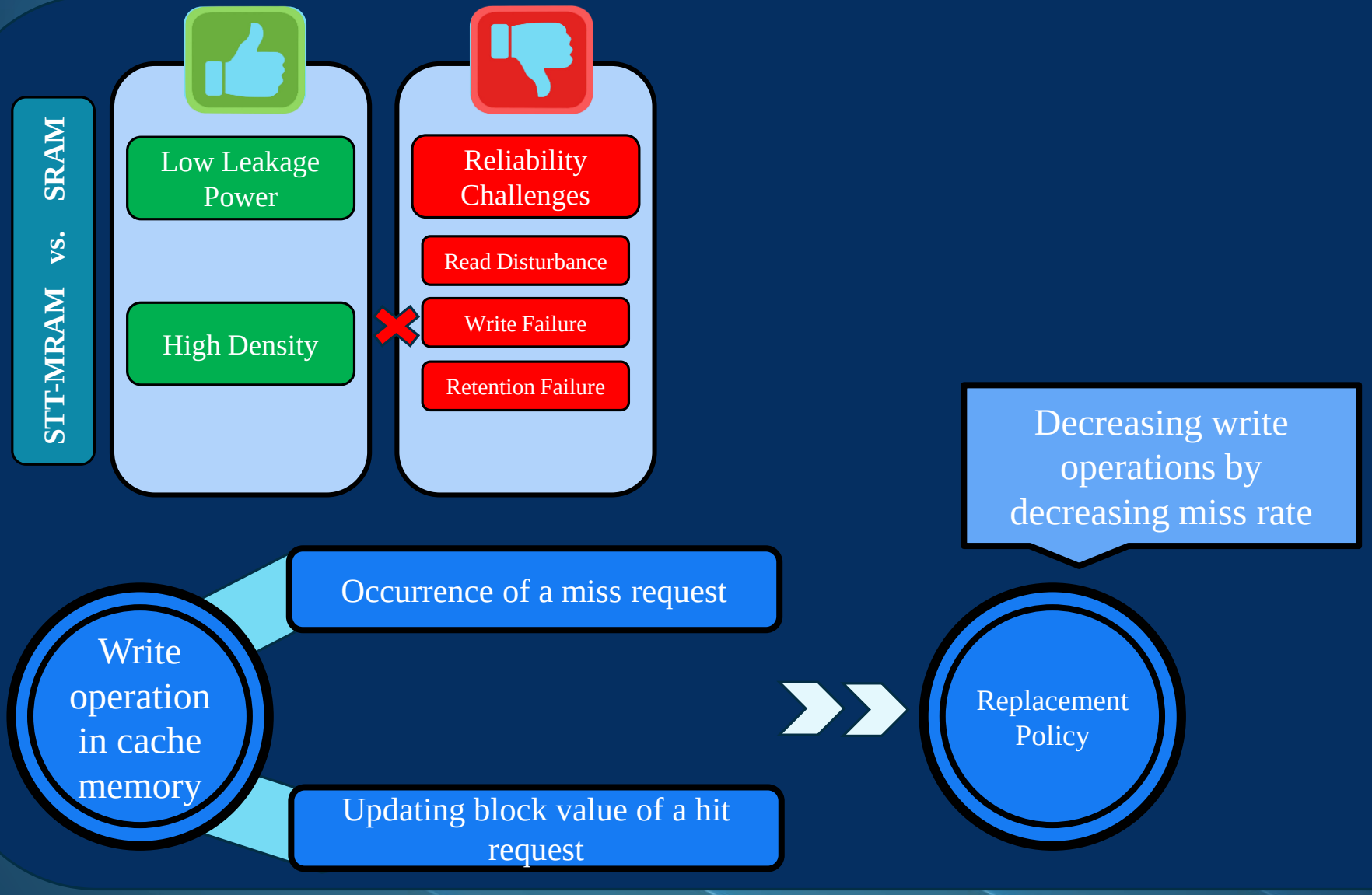
Arash_salahvarzi@alumni.iust.ac.ir

IEEE Transactions on Computers

Cyber-Physical Systems Laboratory

Iran University of Science and Technology

Outline



The Main Contributions

- Exploring the effects of different replacement policies on the reliability of STT-MRAM caches
- Proposing a set of replacement policies which are efficiently coupled with each other
- Design and evaluating the underlying cache memory hierarchy architecture to utilize the proposed replacement policies.

IEEE Transactions on Computers

Read Operation

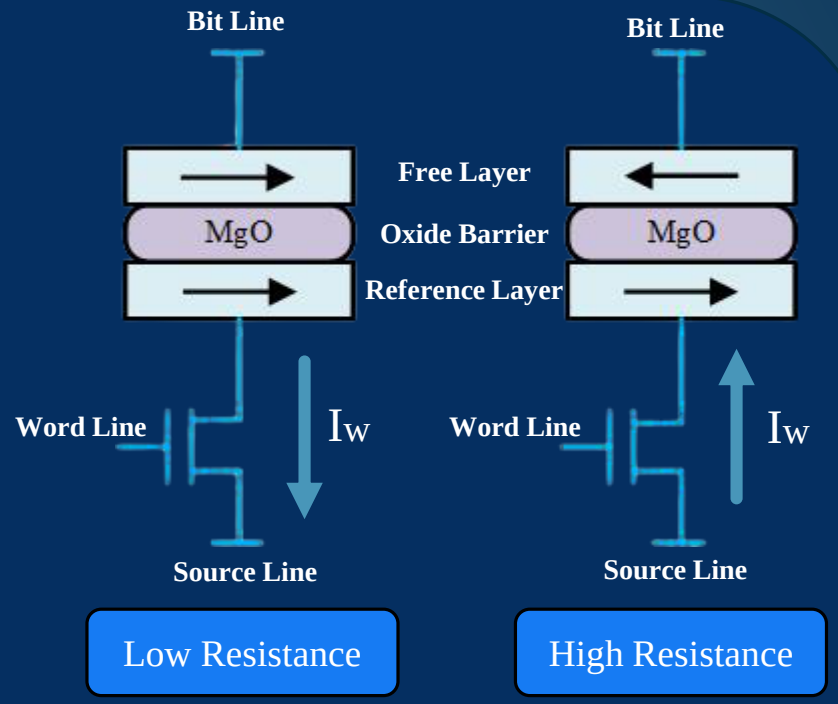
Step 1

Activation of WL

Applying a low sensing current through the cell

Step 2

Comparing the sensing to the reference current



Write Operation

0

Step1	Step2	Step3
Activation of WL	BL =High SL=Low	Injecting a current to MTJ, BL to SL

1

Step1	Step2	Step3
Activation of WL	BL =Low SL=High	Injecting a current to MTJ, SL to BL

Read Failure

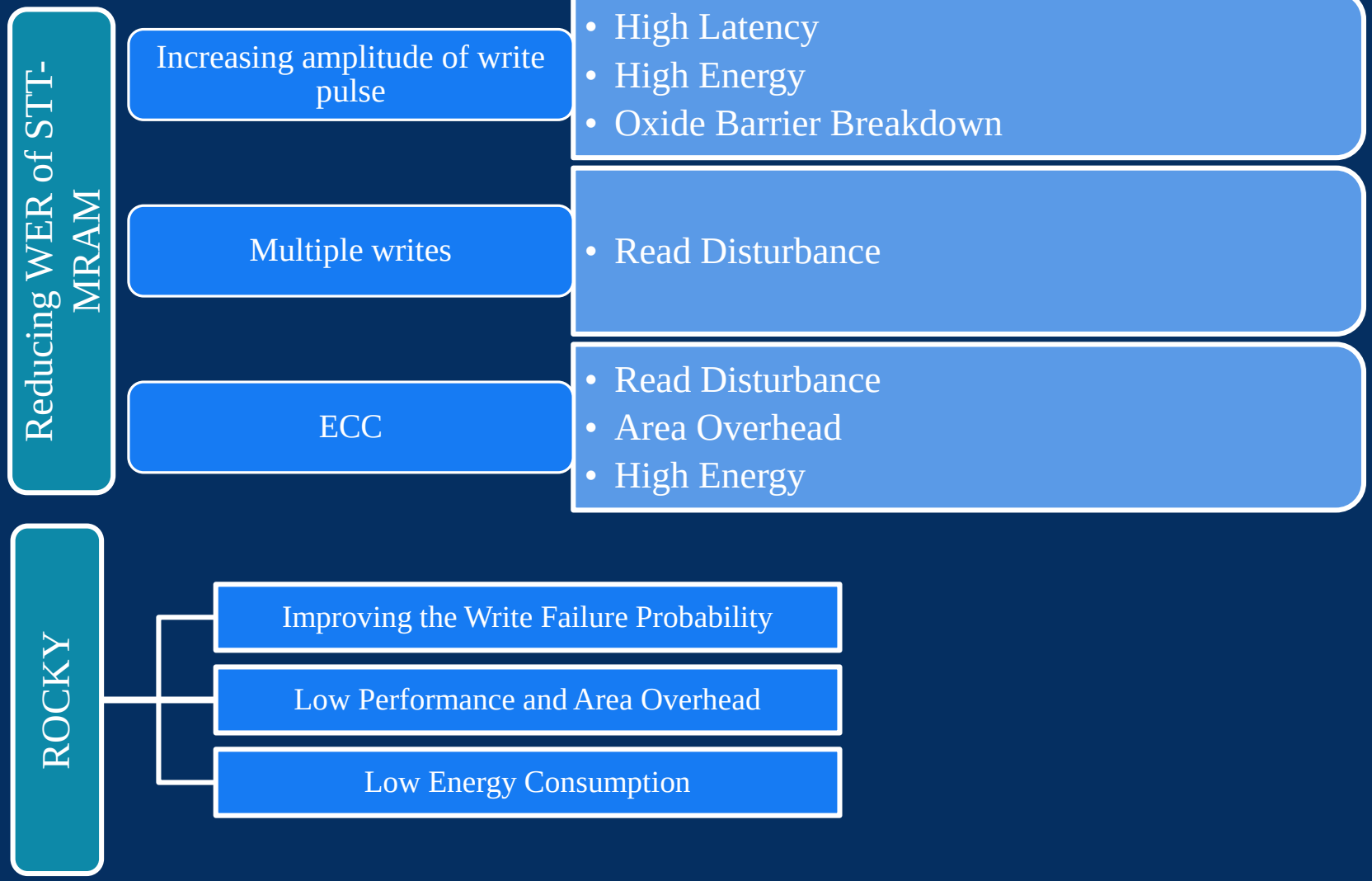
- Applied read current higher than the critical switching current
- Changing the content of the cell unwantedly
- $P_{RD}(t_r) = 1 - \exp(\frac{-t_r}{\tau} \times \exp\left(\frac{-\Delta(I_r - I_{C0})}{I_{C0}}\right))$

Retention Failure

- Changing the content of STT-MRAM after being idle for a time
- $P_{RF}(t) = 1 - \exp(\frac{-t}{\exp(\Delta)})$

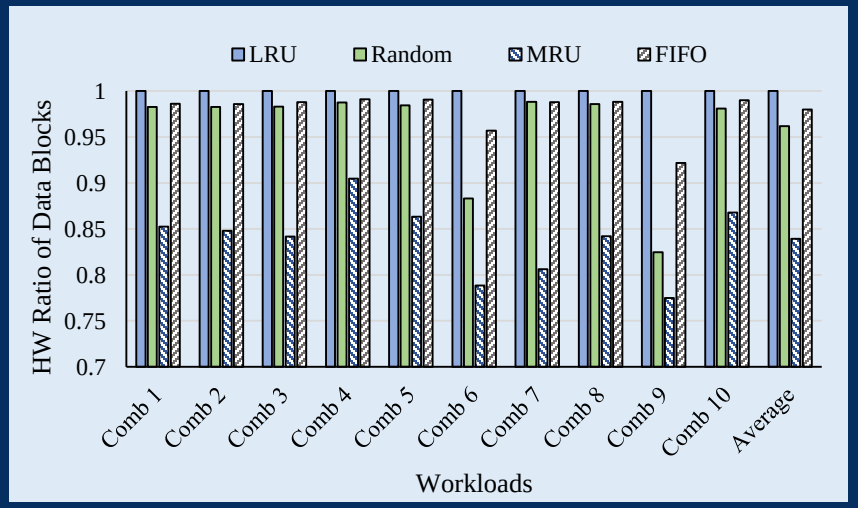
Write Failure

- A write current signal period shorter than the MTJ switching time
- $P_{WF}(t_w) = \exp(-t_w \times \frac{2\mu_B P(I_w - I_{C0})}{(c + \ln\left(\pi^2 \frac{\Delta}{4}\right)) \times (em(1+p^2))})$

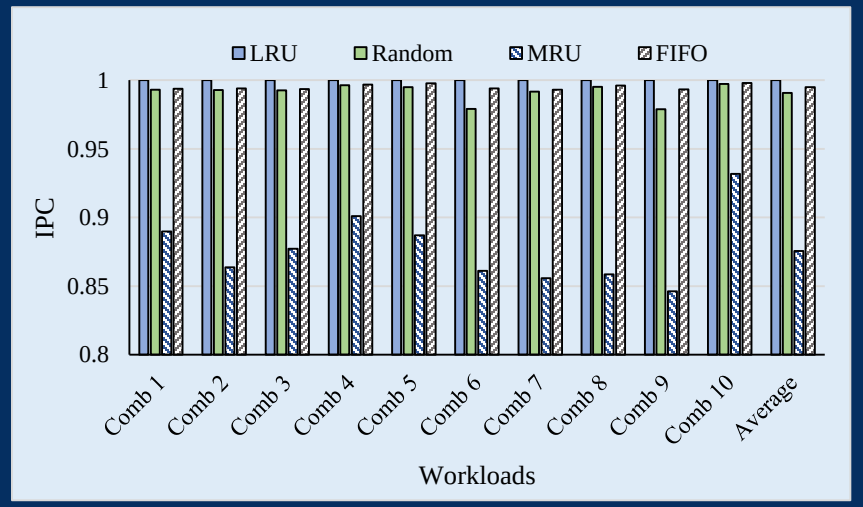


Higher switching time of 0→1 compared to 1→0

0→1 ∝ Hamming Weight

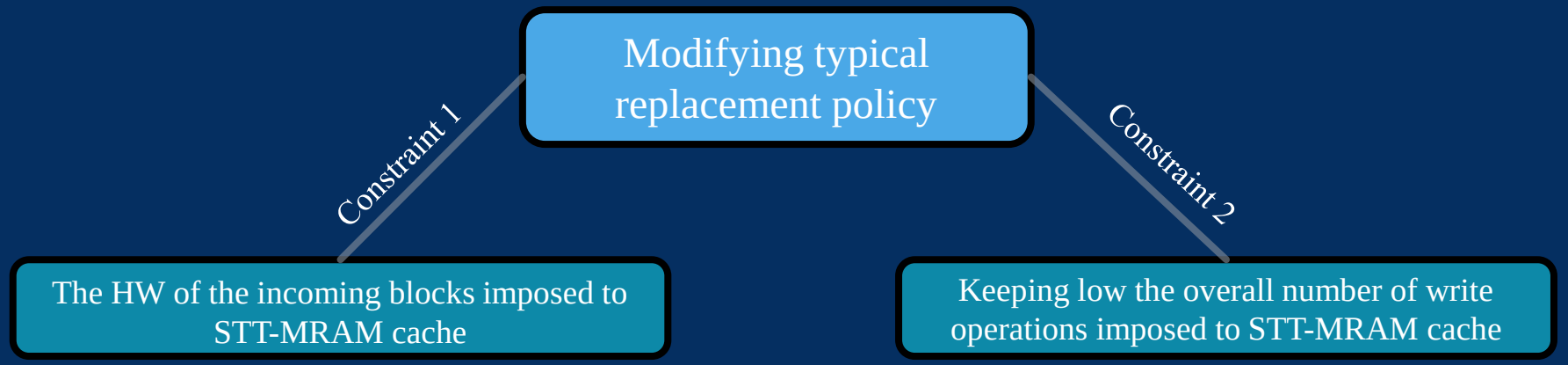


HW ratio of L2 cache incoming data blocks, normalized to LRU



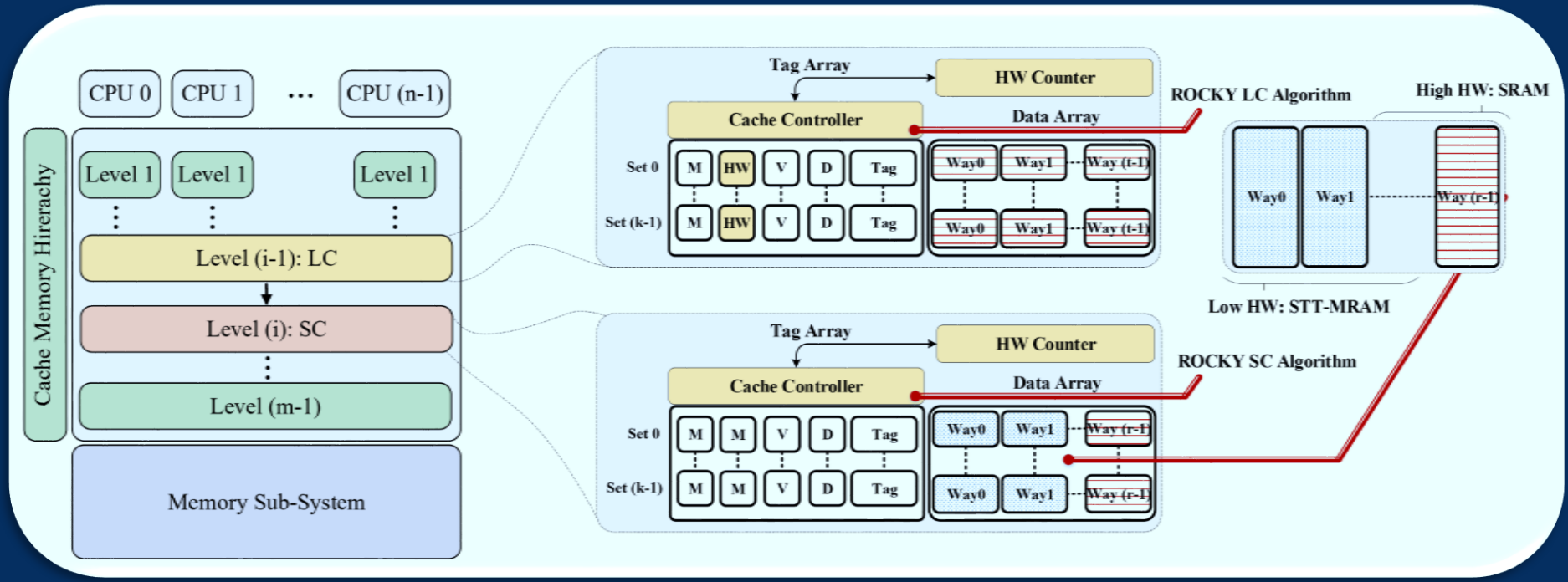
IPC of selected replacement policies for level one cache, normalized to LRU

gem5 full-system simulator

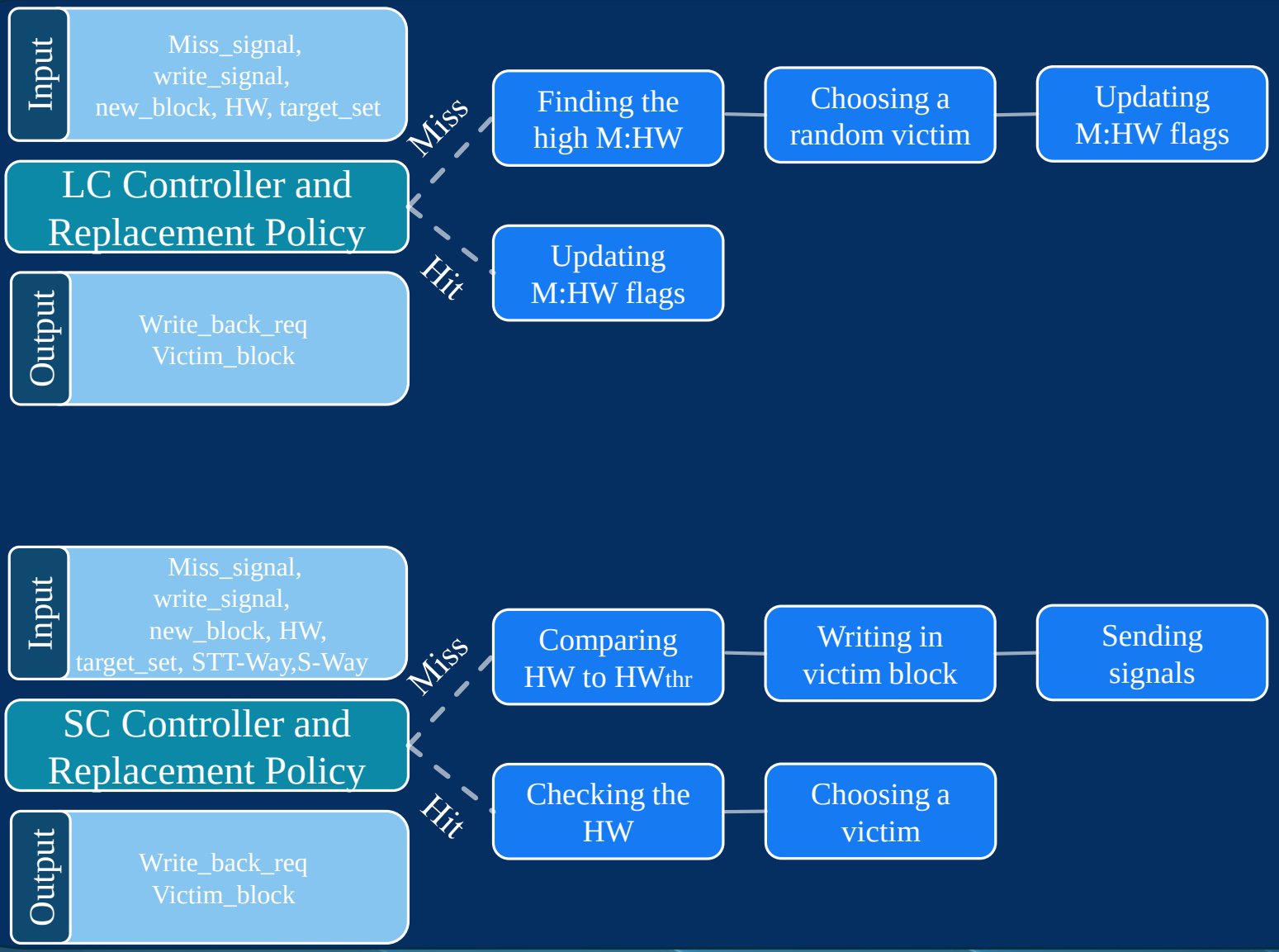


ROCKY

- ✓ A new replacement policy for upper-level (Leading Cache (LC))
- ✓ Decreasing the overall HW ratio of data blocks in STT-MRAM cache (SC)
- ✓ Augmenting the HW parameter with the traditional parameters used in LRU

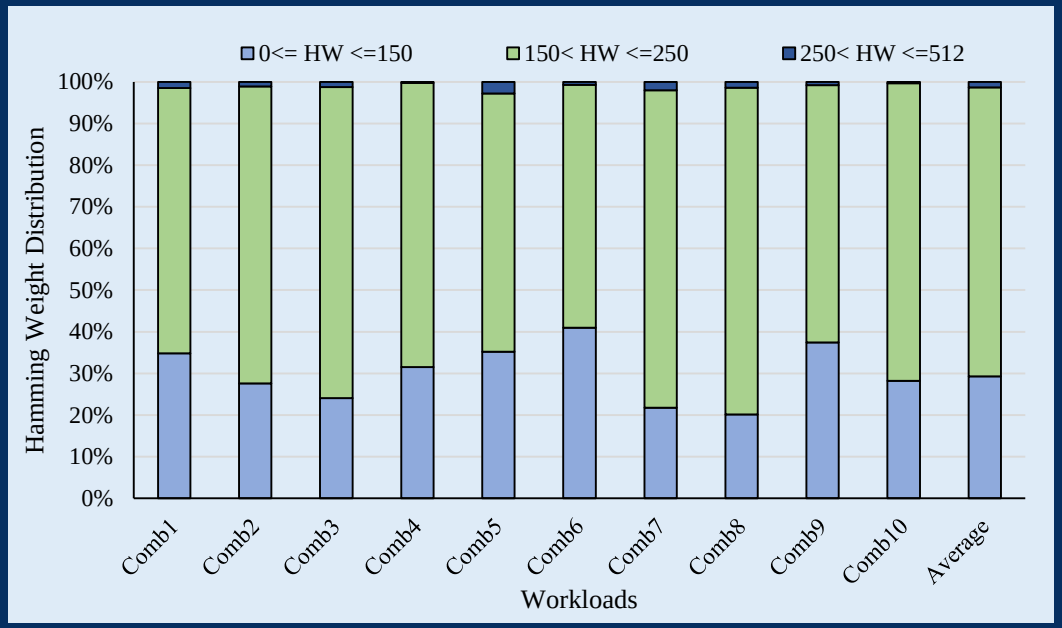


HW Counter	Counting the HW of each block
HW Flag	Keeping the basic information about the HW of block
Cache Controller	Implementing the ROCKY replacement policy



Simulations Configurations

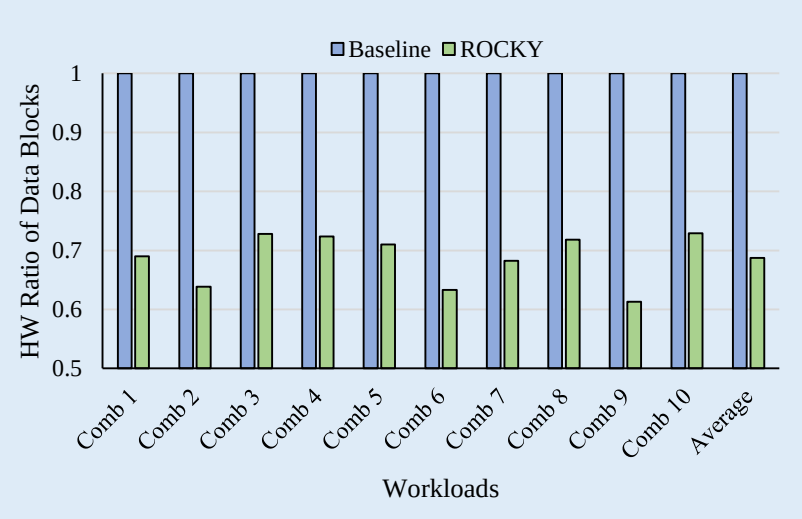
Component	Parameter	Value
Processor	Architecture	ARM
	Frequency	1GHz
	Number of Cores	4
	Type	Homogeneous
	Model	Detailed
	Warmup Cycles	100 millions cycles
L1 Cache	Memory Technology	SRAM
	Size	32KB
	Associativity	4
	Block Size	64B
	Access Time	2 cycles
	Replacement Policy	ROCKY LC
	Energy per Access	0.240 nJ
L2 Cache (Hybrid)	Memory Technology	STT-MRAM+SRAM
	Size	4MB (3584KB +512KB)
	Associativity	16 (14 way + 2 way)
	Block Size	64B
	STT-MRAM Access Time	Write 20 and Read 5 cycles
	SRAM Access Time	5 cycles
	Replacement Policy	ROCKY SC
	STT-MRAM Energy per Read	0.210 nJ
	STT-MRAM Energy per Write	1.01 nJ
	SRAM Energy per Access	0.240 nJ
Main Memory	Memory Technology	DRAM DDR3
	Size	4GB
	Access Time	100 cycles



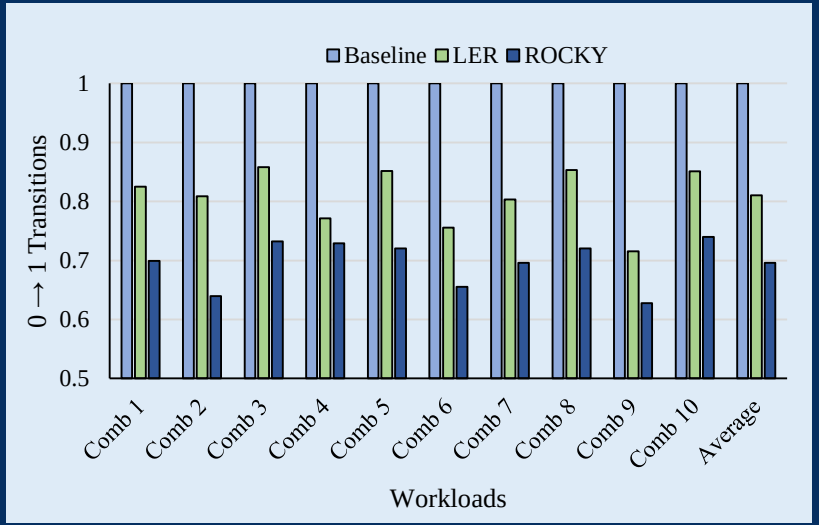
- 10 combinations from SPEC CPU2006 benchmarks
- HW_{thr} as a knob for ROCKY
- High HW_{thr} → High write error rate

IEEE Transactions on Computers

Reliability



HW ratio of L2 cache, normalized to the baseline

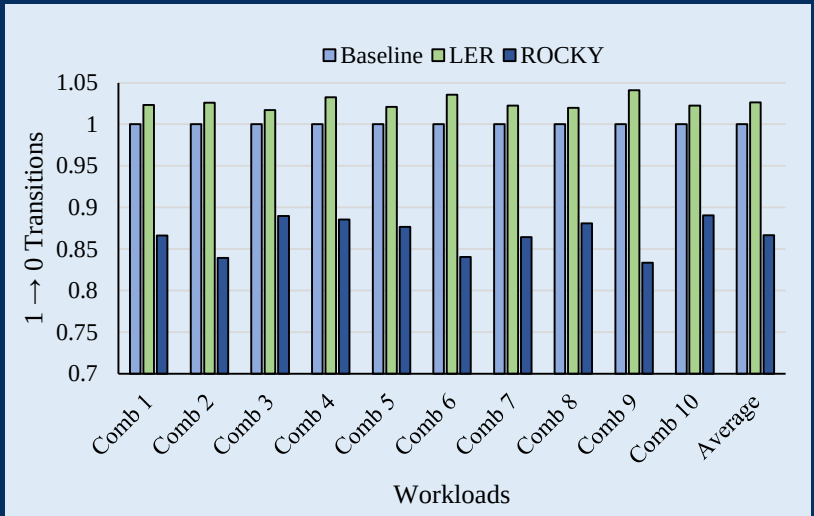


0 → 1 transitions in L2 cache, normalized to the baseline

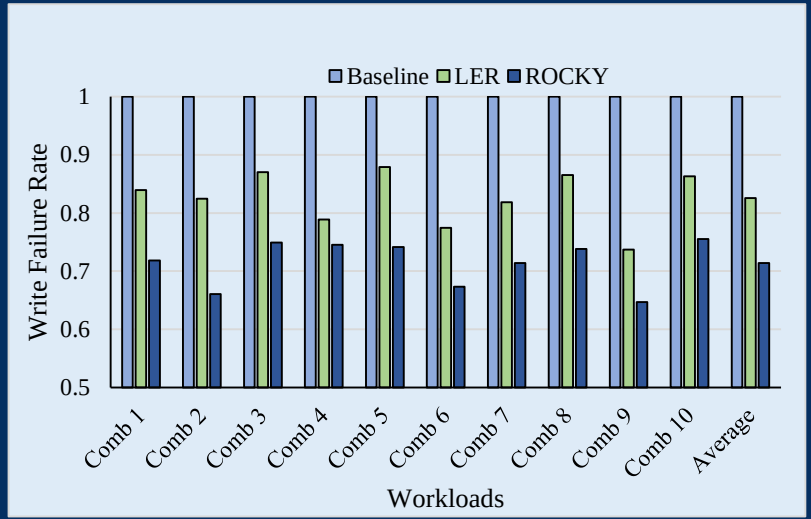
Reducing HW Ratio	
ROCKY	LER
31.3%	0%

Reducing 0→1 Transitions	
ROCKY	LER
30.5%	19%

Reliability



1 → 0 transitions in L2 cache, normalized to the baseline

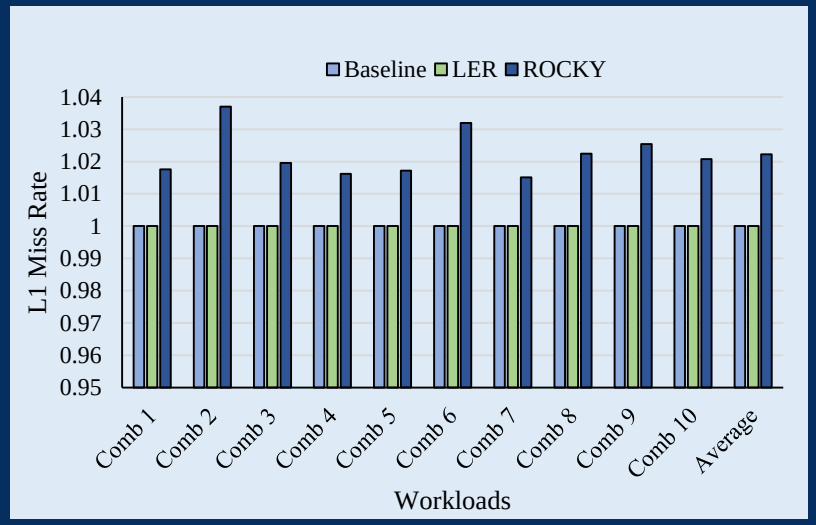


WER in L2 cache, normalized to the baseline

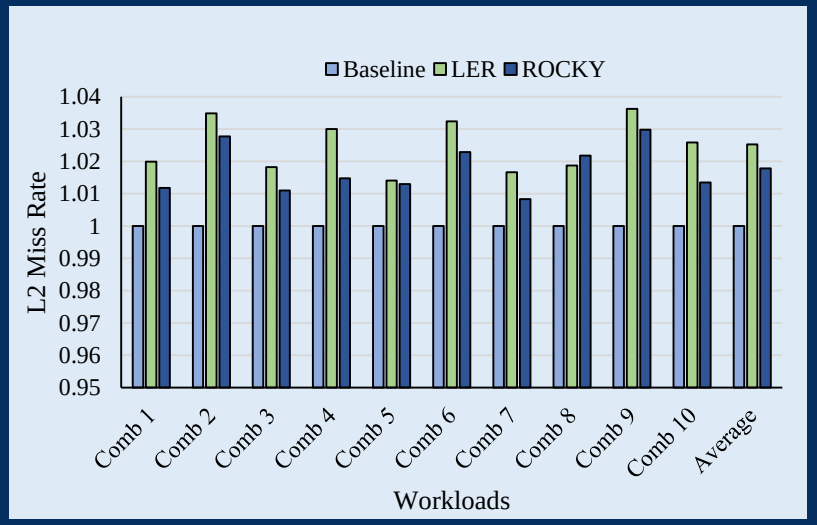
Reducing 1→0 Transitions	
ROCKY	LER
11.7%	-2.6%

Reducing WER	
ROCKY	LER
28.7%	17.5%

Performance



Miss rate of L1 cache, normalized to the baseline

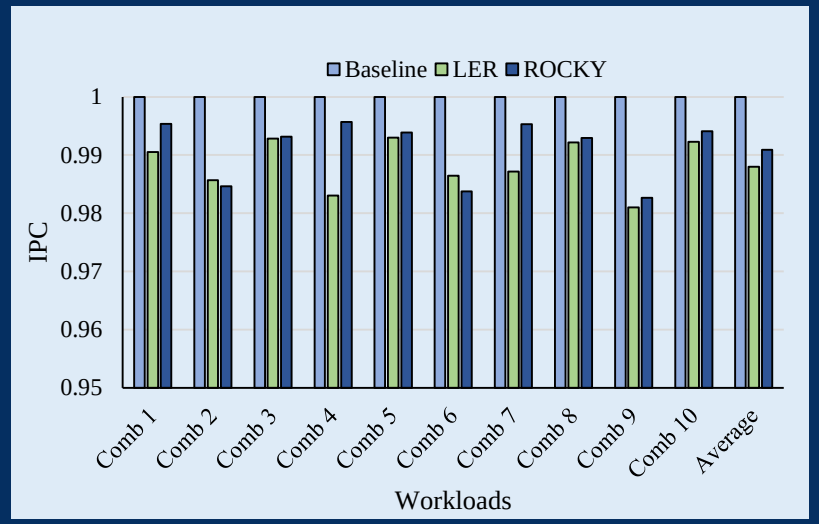


Miss rate of L2 cache, normalized to the baseline

Imposed Miss Rate	
ROCKY	LER
2.2%	0%

Imposed Miss Rate	
ROCKY	LER
1.8%	2.4%

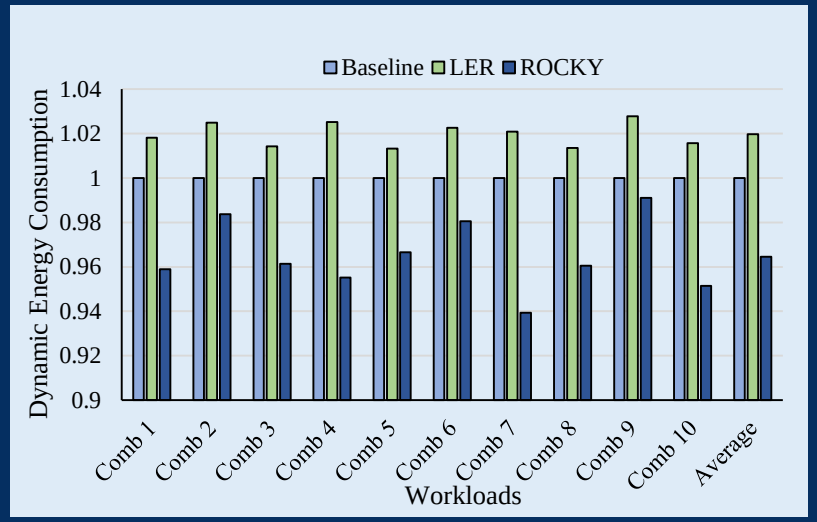
Performance



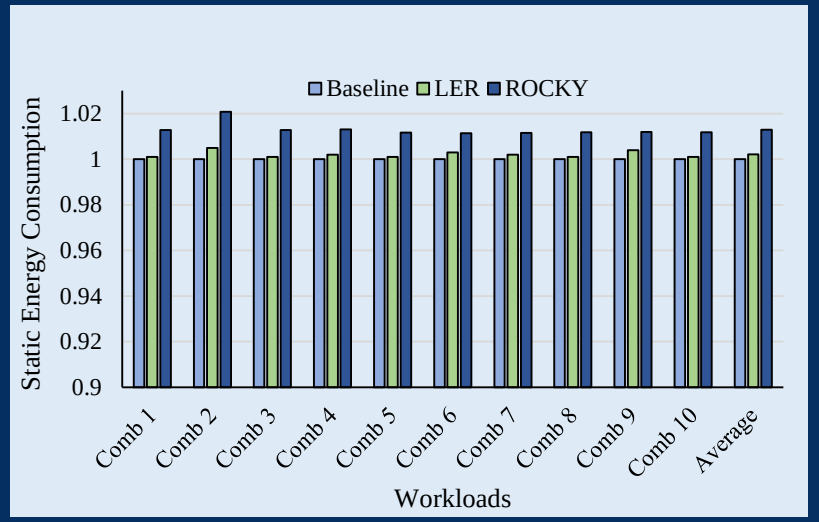
Normalized IPC to the baseline

IPC Overhead	
ROCKY	LER
0.9%	1.2%

Energy Consumption



Dynamic energy consumption of L2 cache, normalized to the baseline

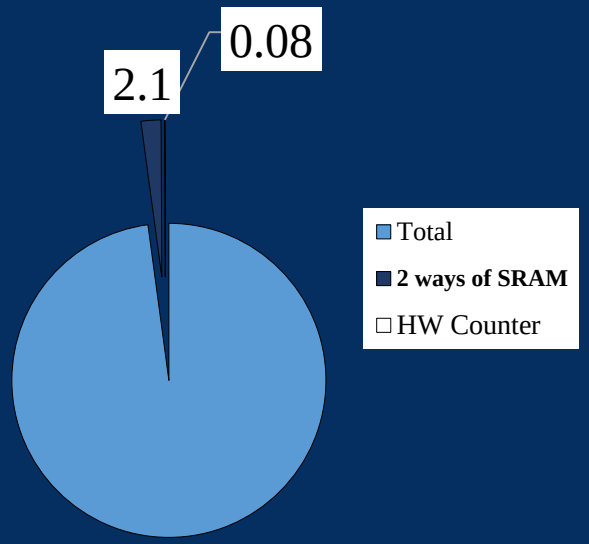


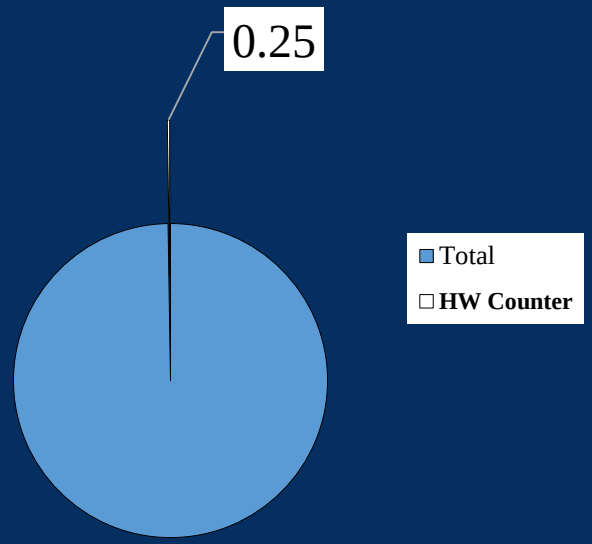
Static energy consumption of L2 cache, normalized to the baseline

Dynamic Energy Consumption	
ROCKY	LER
-3.55%	1.9%

Imposed Static Energy	
ROCKY	LER
1.3%	0.2%

Area





L2 Cache

- 2 Ways of SRAM→ 2.1%
- HW Counter→ 0.08%

L1 Cache

- HW Counter→ 0.25%

01 High Leakage power and low density of SRAM

02 Advantages of STT-MRAM

03 Reliability challenges of STT-MRAM

04 0→1 transitions in write patterns

05 Reducing HW ratio of data blocks using ROCKY

06 Reduction of write failure by 28.7%

THANK YOU

Room 121, Department of Computer Engineering, Iran University of Science and Technology, University Road, Hengam Street, Resalat Square, Narmak, Tehran, IRAN 16846-13114.

+98 (21) 73225350

cps.iust.ac.ir