Flash Memory Summit

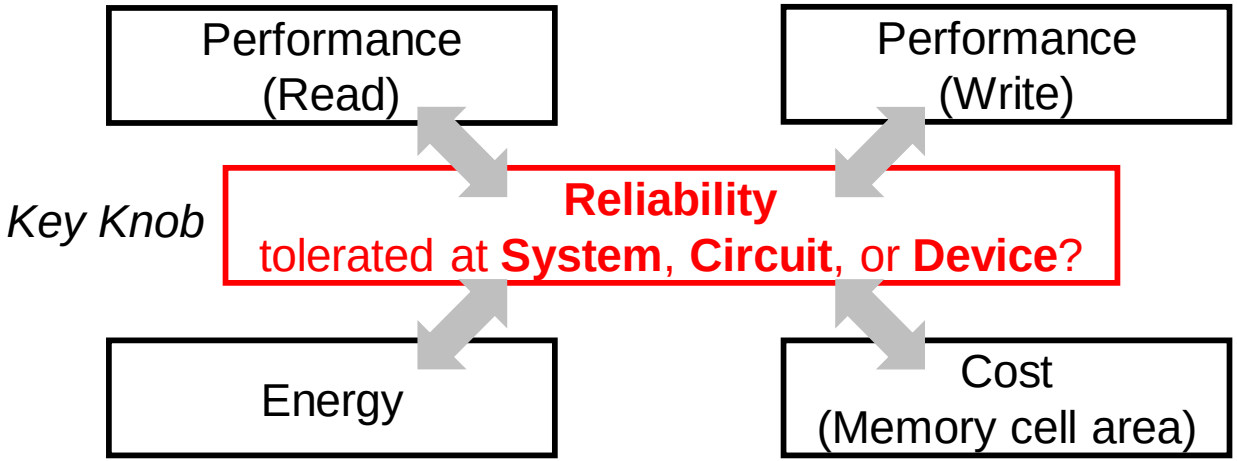
TaO_X-based ReRAM for Variability-Aware Approximate Computing

Chihiro Matsui, Shouhei Fukuyama,
Atsuna Hayakawa, and Ken Takeuchi

Chuo University, Tokyo, Japan

Flash Memory Summit

Reliability-Aware Approximate Computing in Storage



K. Takeuchi, *IEDM* 2017.

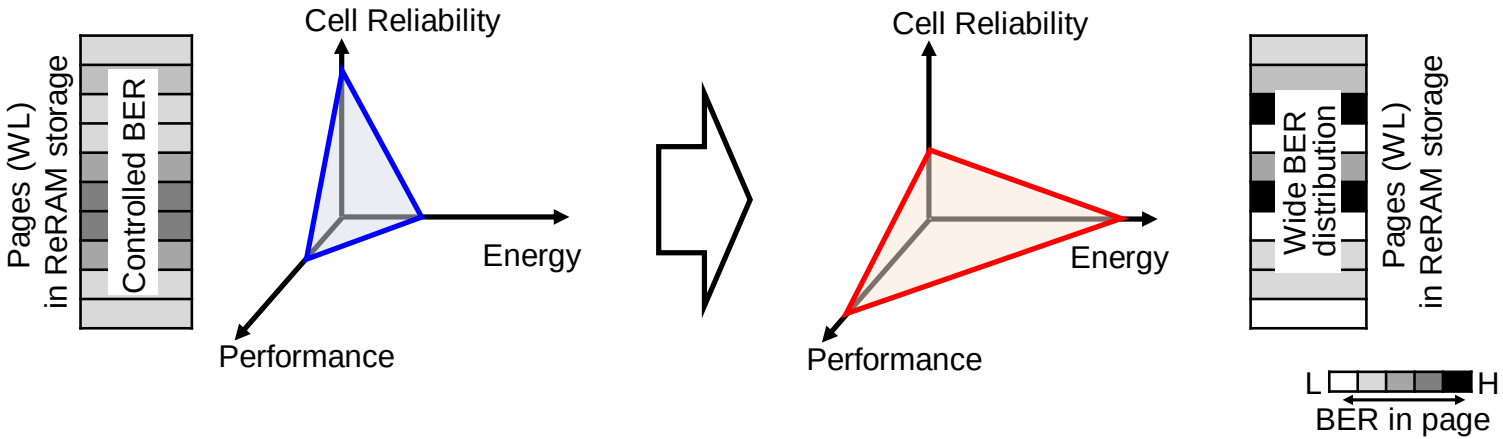
Which hierarchy of ReRAM storage has Error Toleration techniques?

How to Relax Reliability for Approximate Computing?

Outline

- Variability-Aware Approximate Computing (V-AC)
- Application-Induced Variability of TaO_X ReRAM Cell Errors and V-AC Evaluation Platform
- V-AC Strategies of System, Circuit and Device Co-Design (SCDCD)
- Conclusions

Variability-Aware Approximate Computing (V-AC)

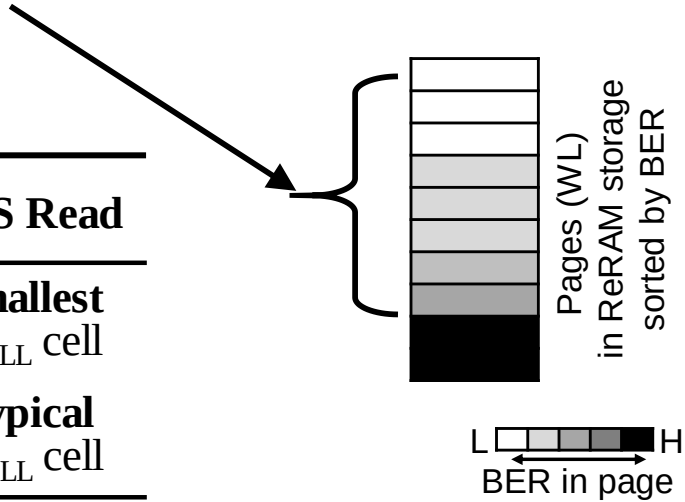


- System, Circuit, and Device in ReRAM-based storage have Variabilities in nature
- By tolerating variability, Performance, Energy, and Cost gain

Typical Cell Target Strategy of V-AC in ReRAM Storage

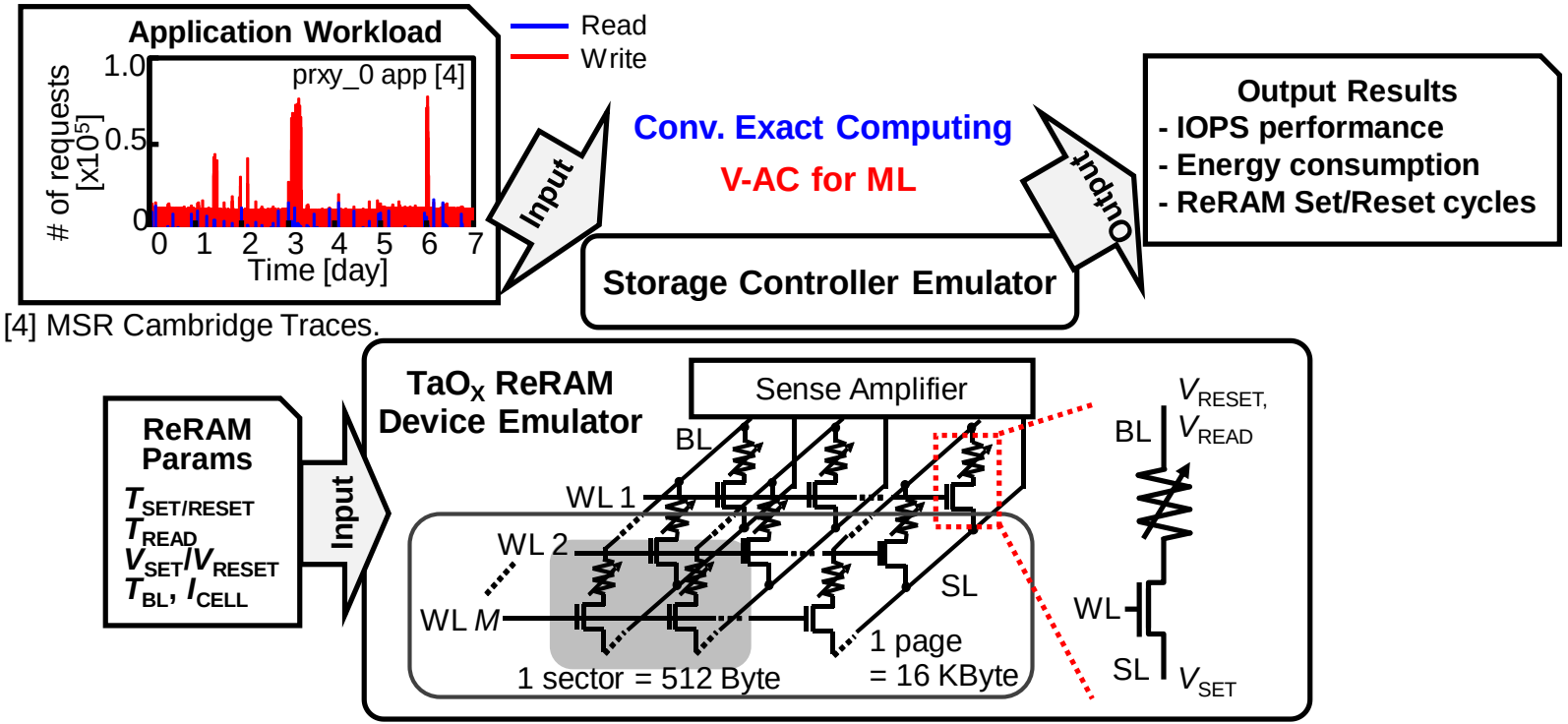
Target Typical ReRAM Cell, NOT Worst Cell

	Error tolerance	ECC	Set/Reset	LRS Read
Conv. Computing	**No** error	**Worst** cell	**Slowest** cell	**Smallest** I_{CELL} cell
V-AC	**1-10 %** errors [*]	**Typical** cell	**Typical** cell	**Typical** I_{CELL} cell



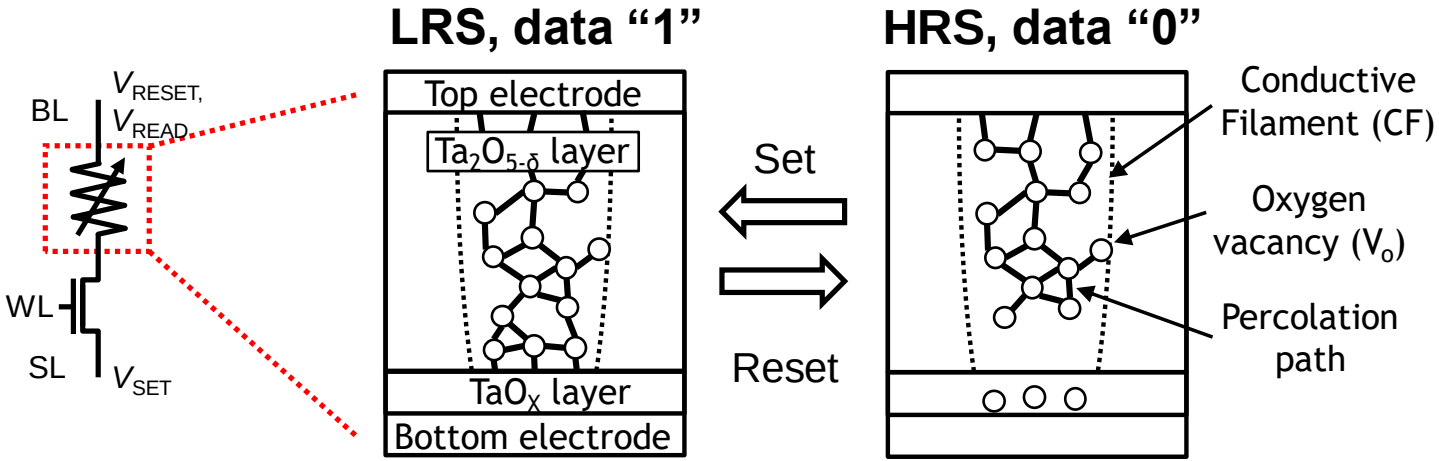
[*] Y. Yamaga et al., *VLSI Tech* 2018.

System, Circuit and Device Co-Design (SCDCD) Platform [*]



[*] C. Matsui et al., *VLSI Technology* 2019.

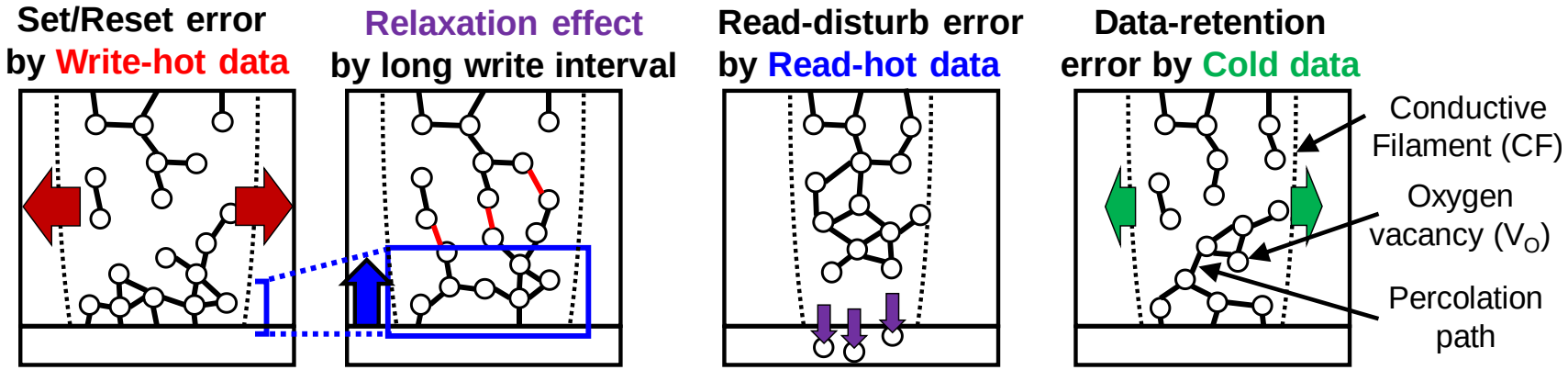
Set/Reset in TaO_X-based ReRAM Cell [*]



- In Set (Reset) operation, LRS (HRS) is formed by moving O^{2-} to TaO_X layer (CF)
- Percolation paths connect (disconnect) between V_Os in LRS (HRS)

[*] Z. Wei et al., *IEDM* 2008.

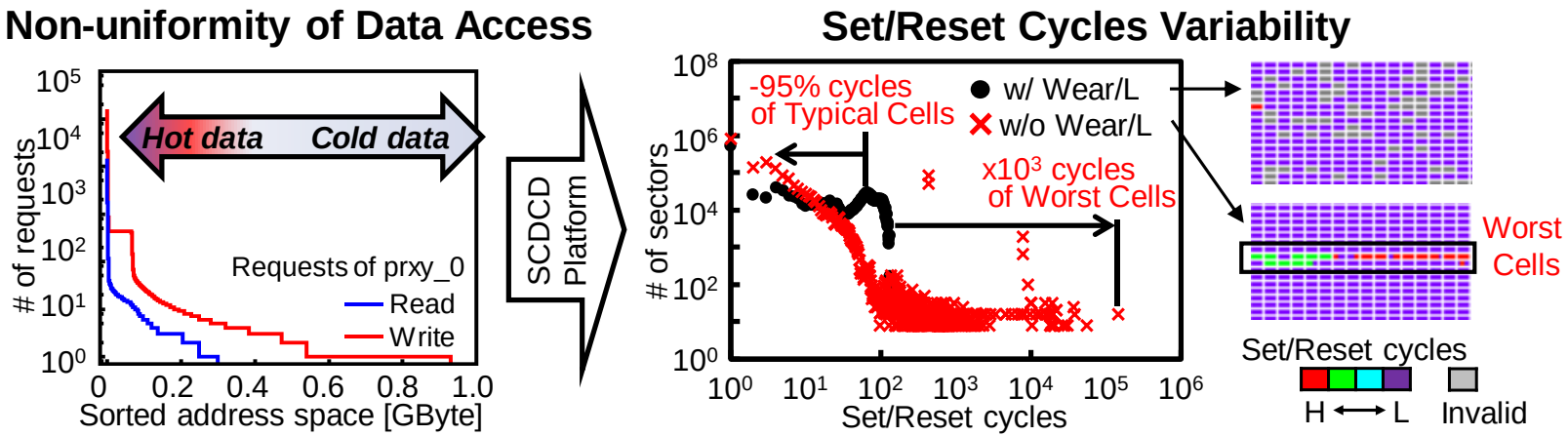
TaO_X ReRAM Conductive Filament (CF) Model [*]



- Write-hot data decrease Vo density in CF by horizontal diffusion
- Relaxation effect reconnects percolation paths by interface Vo diffusion to CF
- Read-hot data cause weak reset by vertical Vo diffusion
- Data retention of cold data causes horizontal Vo diffusion

[*] S. Fukuyama et al., *IRPS* 2019.

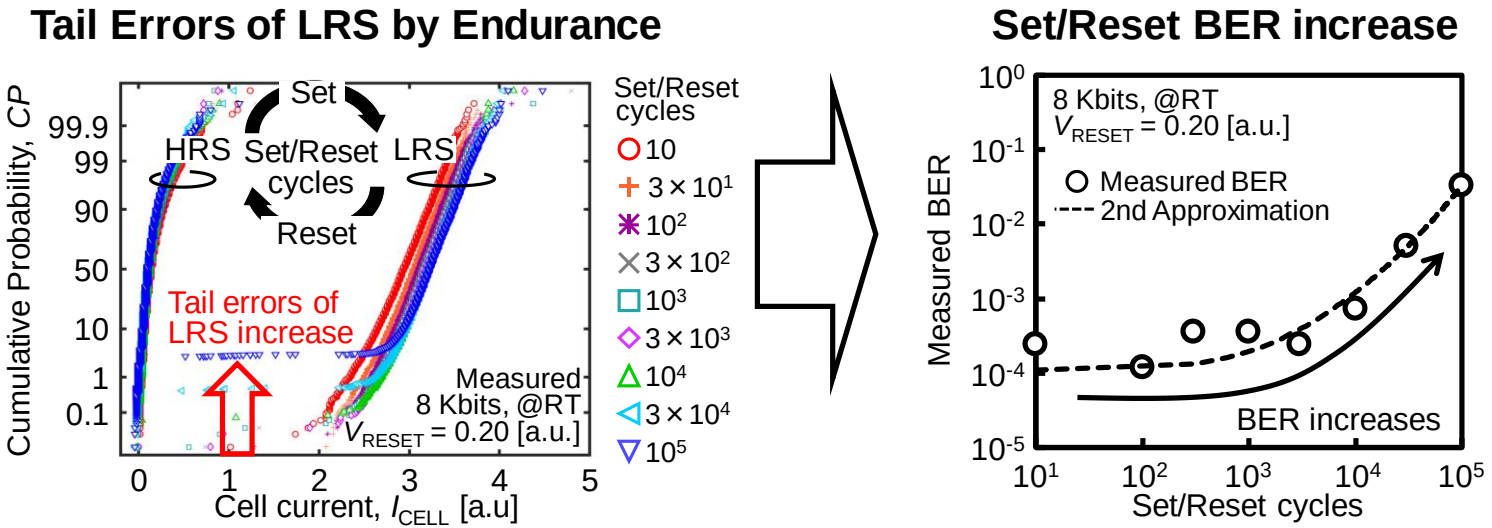
Storage System Variability



- Write-hot data induce large Set/Reset cycle difference in cells without smoothing by Wear-leveling (Wear/L) [*]
- Set/Reset cycles of Typical Cells reduce by 95% while those of Worst Cells increase by x10^3

[*] T. Onagi et al., *SSDM* 2014.

ReRAM Device-induced BER Variability

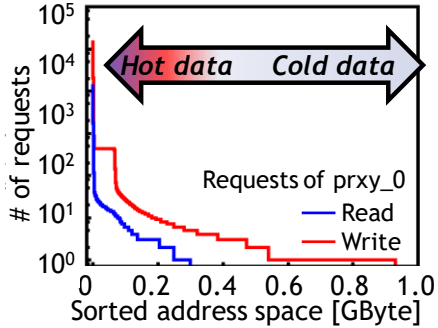


- Measured LRS show tail error cells at high Set/Reset cycles [*]
- BER increases with Set/Reset cycles

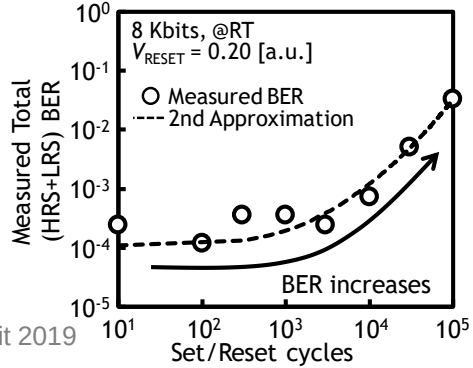
[*] K. Maeda et al., *IRPS* 2017.

System-induced BER Variability

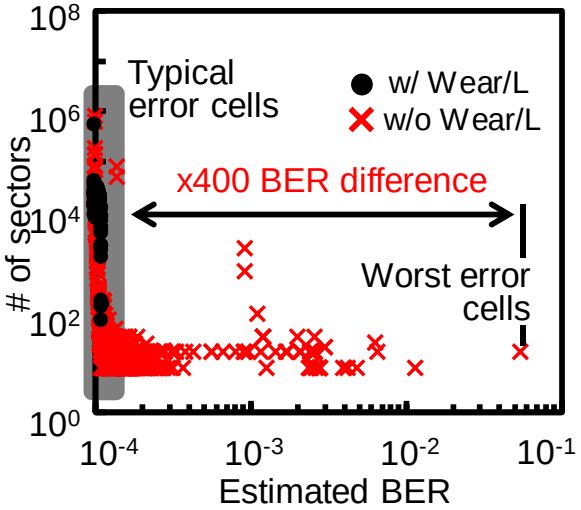
Application-induced Variability



Device-induced Variability



System-induced ReRAM Cell BER Variability



For Variability-Aware Approximate Computing (V-AC)

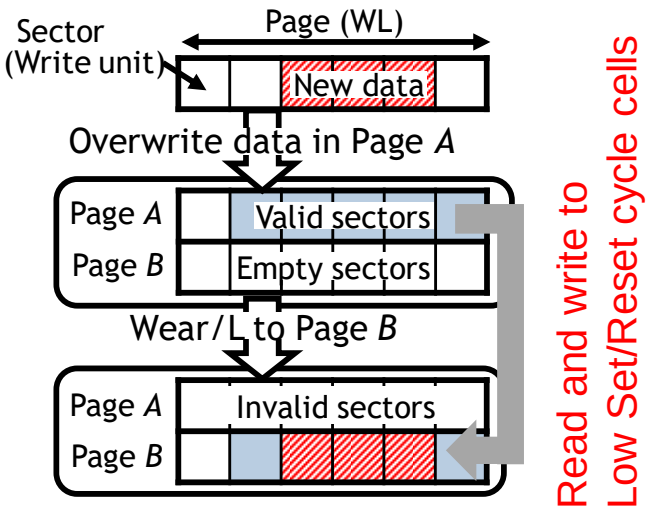
V-AC Error Toleration Strategies [*]

Hierarchy	Operation	Conv. computing	Proposed V-AC Strategy	Proposed V-AC Technique	Data characteristic *Write-hot*	*Read-hot*	*Cold*
System	Wear-leveling (Wear/L)	w/ Wear/L	**I**	**w/o Wear/L**	✓	✓	
	ECC	Worst-error target (35-bit correction)	**II**	**Typical-error target (5-bit correction)**	✓	✓	
Circuit	Read	NA	**III**	**Adaptive Read**		✓	
Device	Set/Reset	Verify	**IV**	**w/o Verify**	✓		
		NA		**Lower V_{SET}/V_{RESET}**	✓		✓

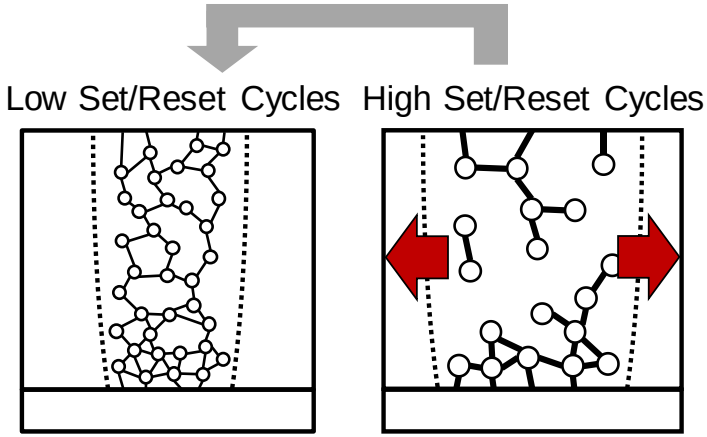
[*] C. Matsui et al., *VLSI Technology* 2019.

Strategy I: Wear-Leveling (Wear/L) Elimination

Wear/L Operation [*]



Endurance Error Reduction by Wear/L

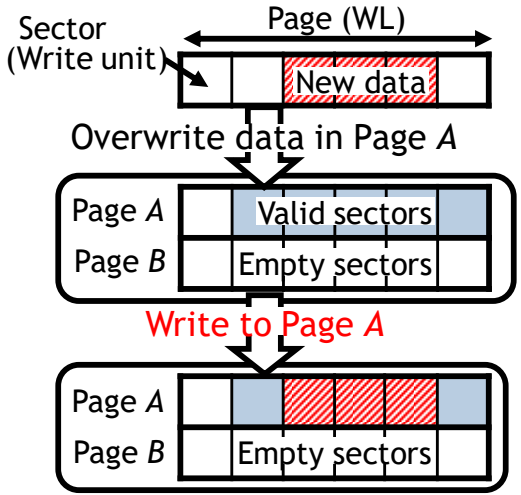


- Wear/L reduces BER of ReRAM storage by smoothing Set/Reset cycles. However, Total Set/Reset cycles increase by extra data copy

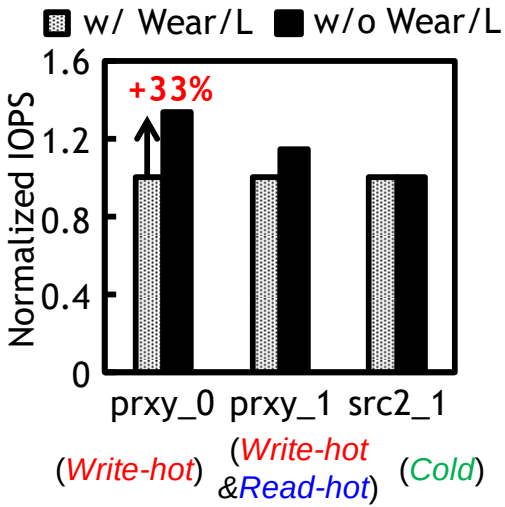
[*] T. Onagi et al., *SSDM* 2014.

Strategy I: Wear-Leveling (Wear/L) Elimination

Wear/L Elimination



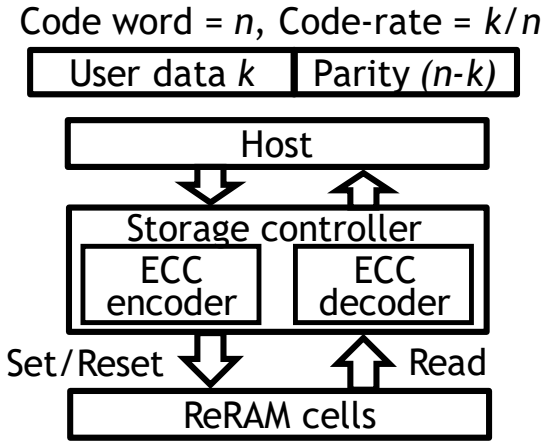
ReRAM Storage Performance



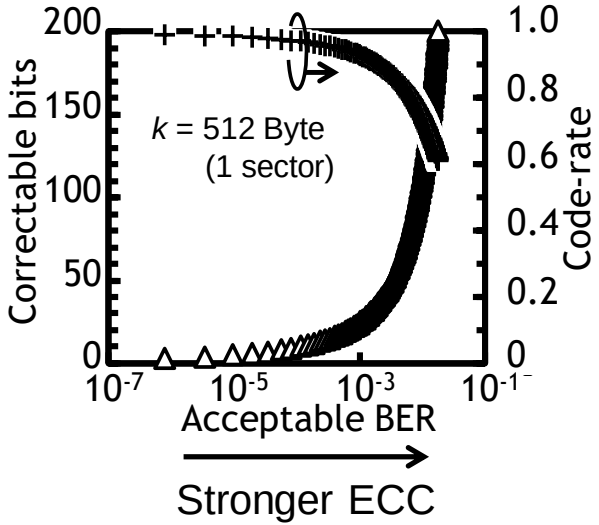
- Strategy I eliminates Wear/L to remove extra data copy and improves storage performance by 33%

Strategy II: Typical-Error Target ECC

ECC Structure

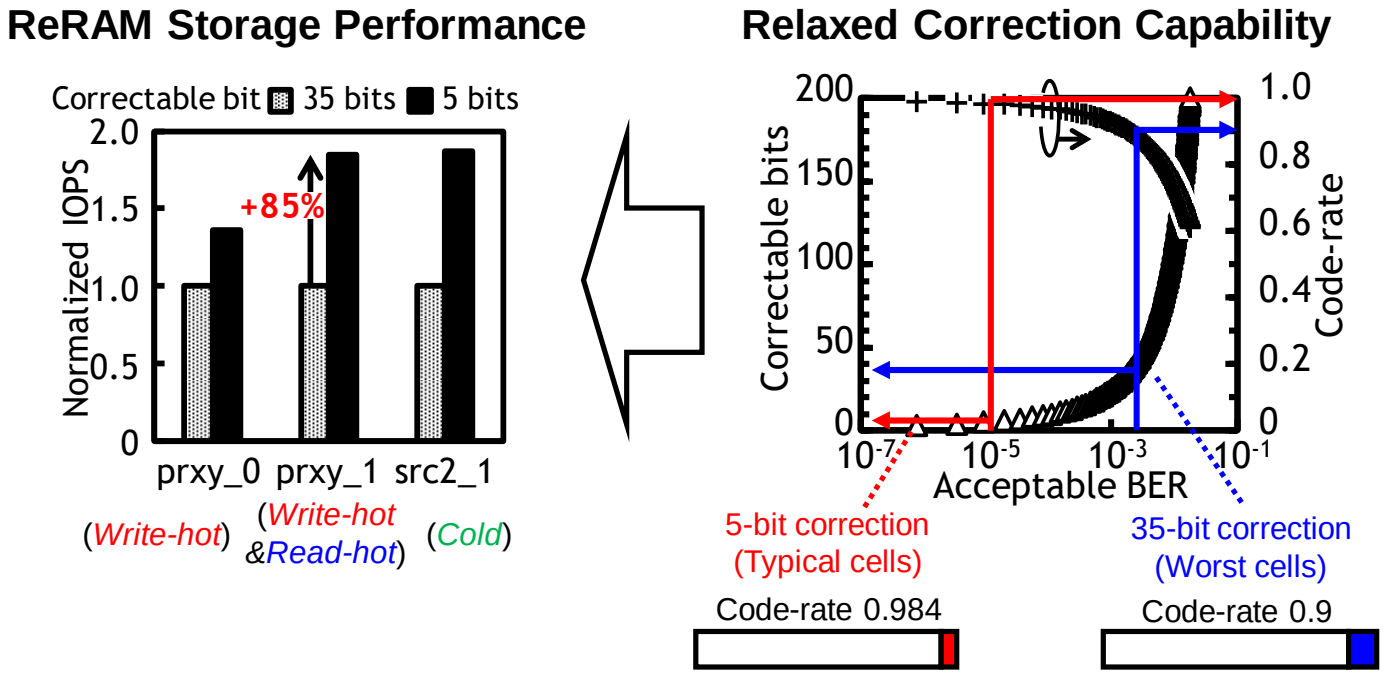


BCH ECC Trade-off



- ECC has trade-off between error correction capability, decoding time and code-rate (cell area)

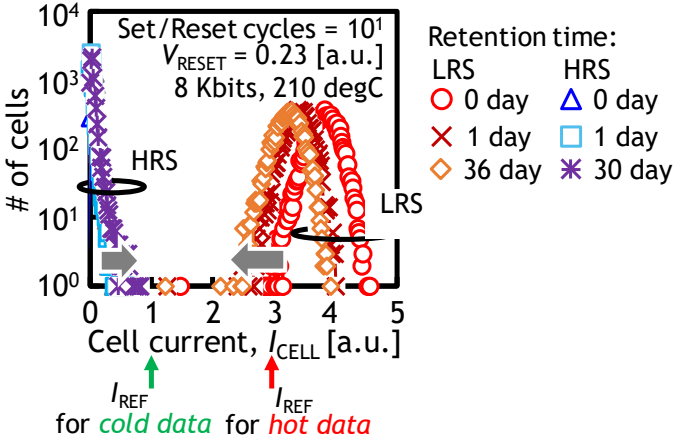
Strategy II: Typical-Error Target ECC



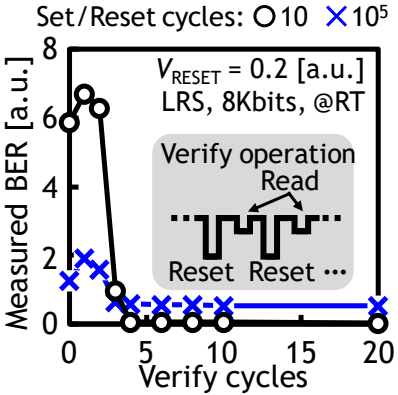
- ECC code-rate increase and performance improves by 85%

Strategy III & IV: Error Toleration in Circuit & Device

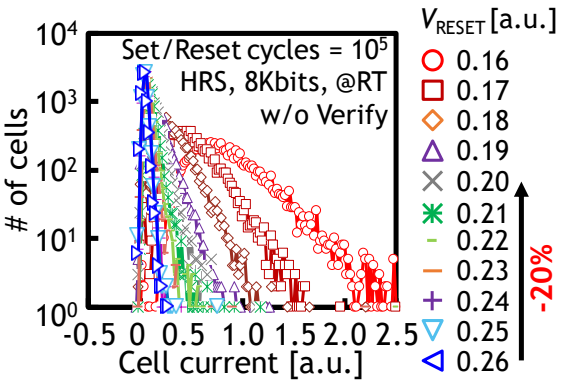
Strategy III (Circuit): Adaptive Read



Strategy IV (Device): Verify Elimination



Strategy IV (Device): Lower V_{SET}/V_{RESET}

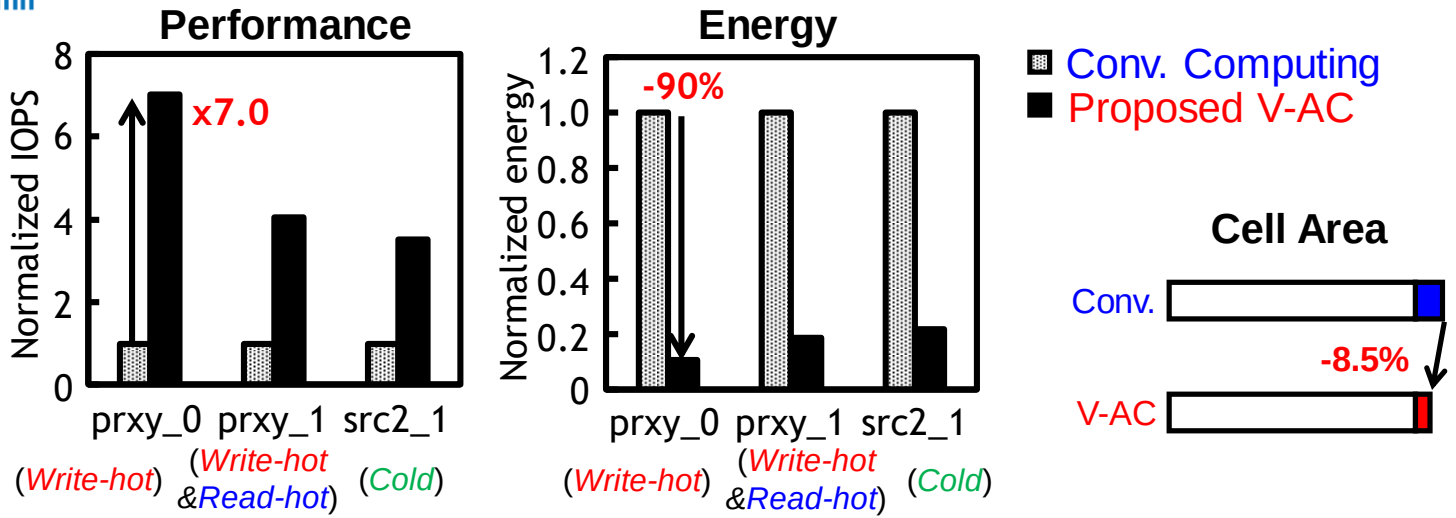


Read Performance

Write Performance

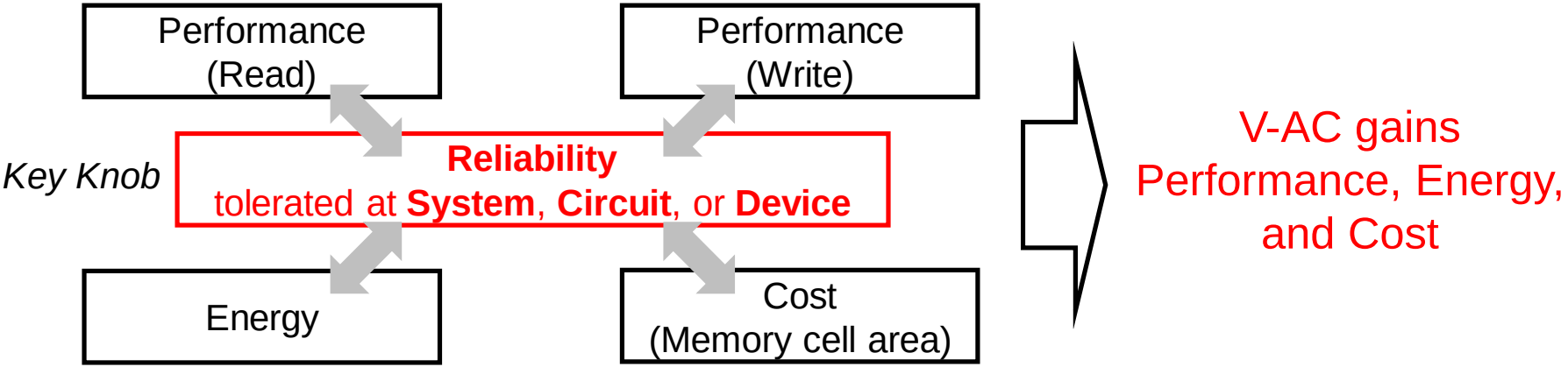
Selector Scalability

Conclusions



- Application-induced Variability-aware Approximate Computing (V-AC) is proposed with System, Circuit and Device Co-Design (SCDCD)
- Performance, Energy, and Cell Area of ReRAM storage improve by x7.0, 90%, and 8.5%

Conclusions



Thank you for your attention

This presentation is based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO)