RELIABILITY CHALLENGES IN COMPUTING-IN-MEMORY SYSTEMS

Yiorgos Tsiatouhas

VLSI Ssystems and Computer Architecture (VCAS) Lab

Dept. of Computer Science & Engineering – School of Engineering – University of Ioannina

VCAS

Outline

Computing in Memory and Reliability:

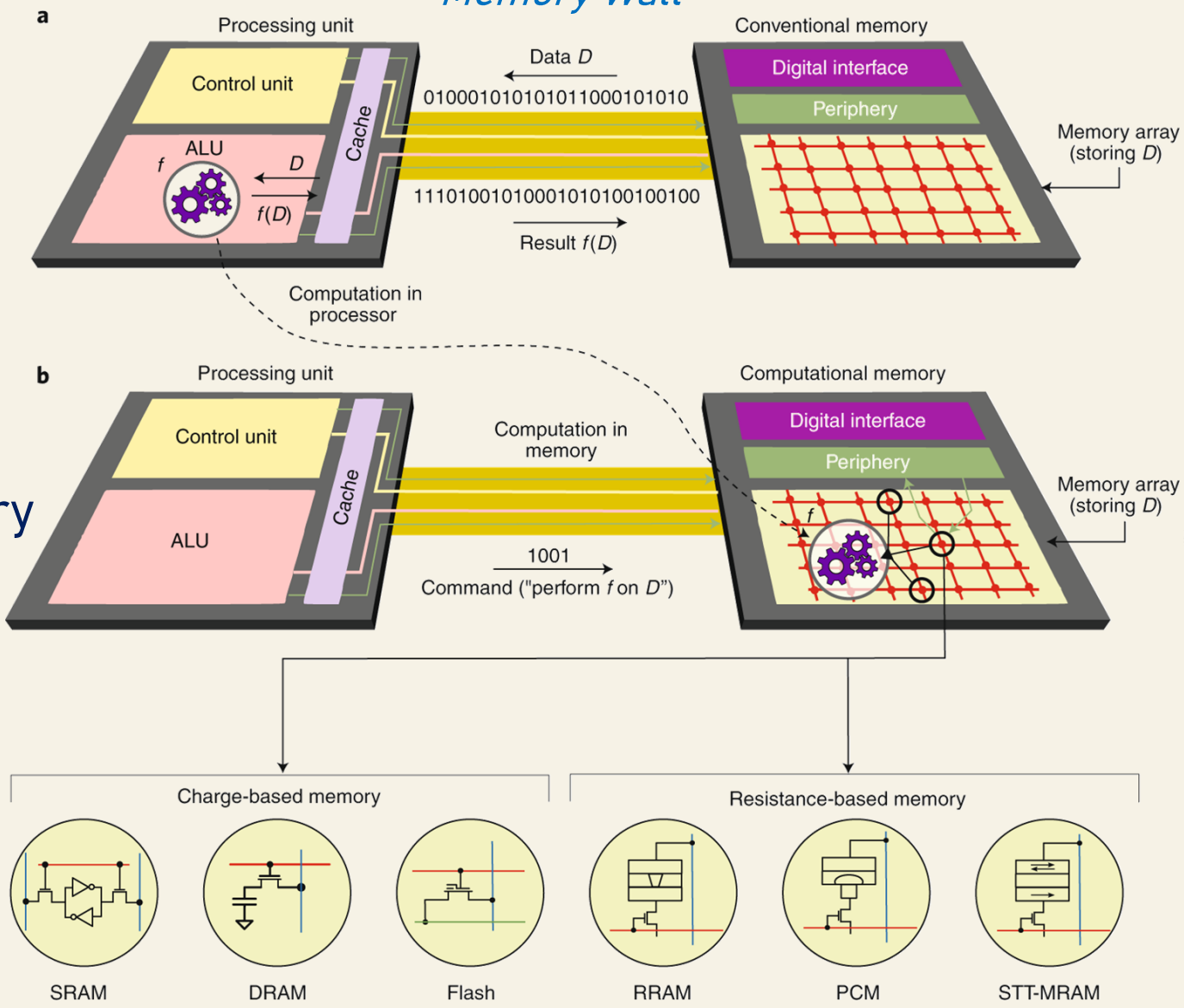
- Testing Challenges
  - Defect mechanisms should be defined
  - Fault Models should be developed
  - Test Generation tools are required
  - Testing Algorithms, Design for Testability & Design for Diagnosis techniques are mandatory
  - Testing Standards are needed
  - On-Line Testing solutions should be proposed
- Aging Challenges
  - Emerging Memories Aging Models should be developed
  - Aging Alleviation techniques should be explored
- Process Variation Challenges
- Radiation induced Single Event Effect (SEE) Challenges
  - Radiation Hardening techniques may be crucial for system reliability

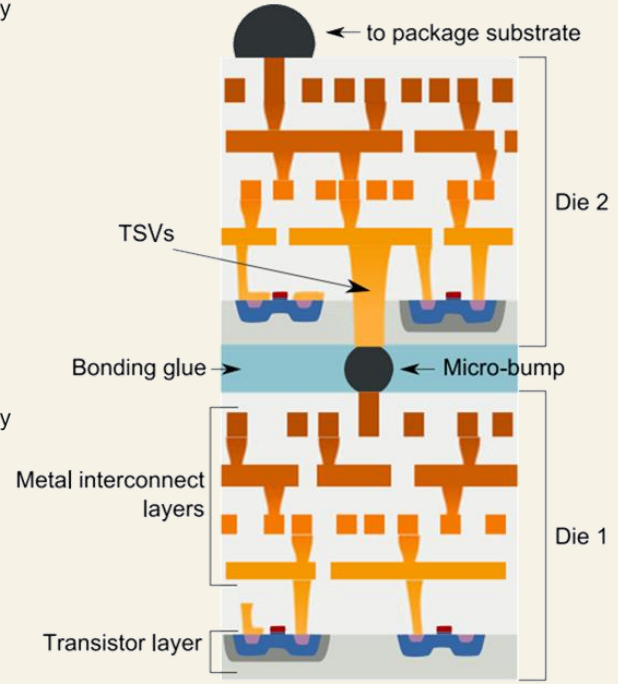
Computing In Memory – CIM

Von Neumann Architecture

Computing In Memory Architecture

Memory Wall



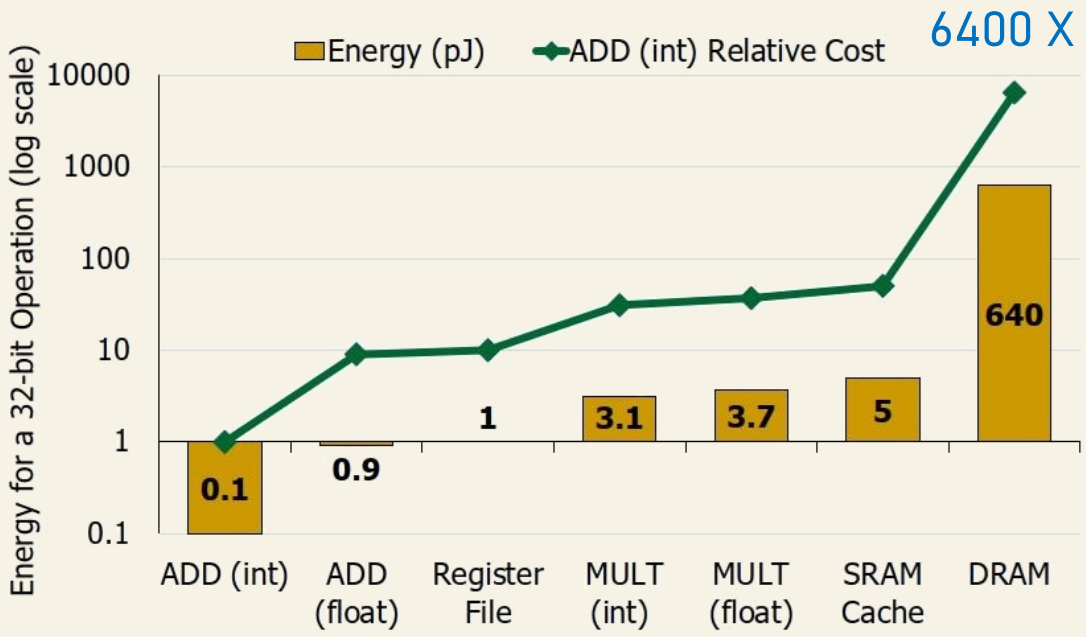


3D Integration

A. Sebastian, Nature Nano, 2020

Conventional vs CIM Computations

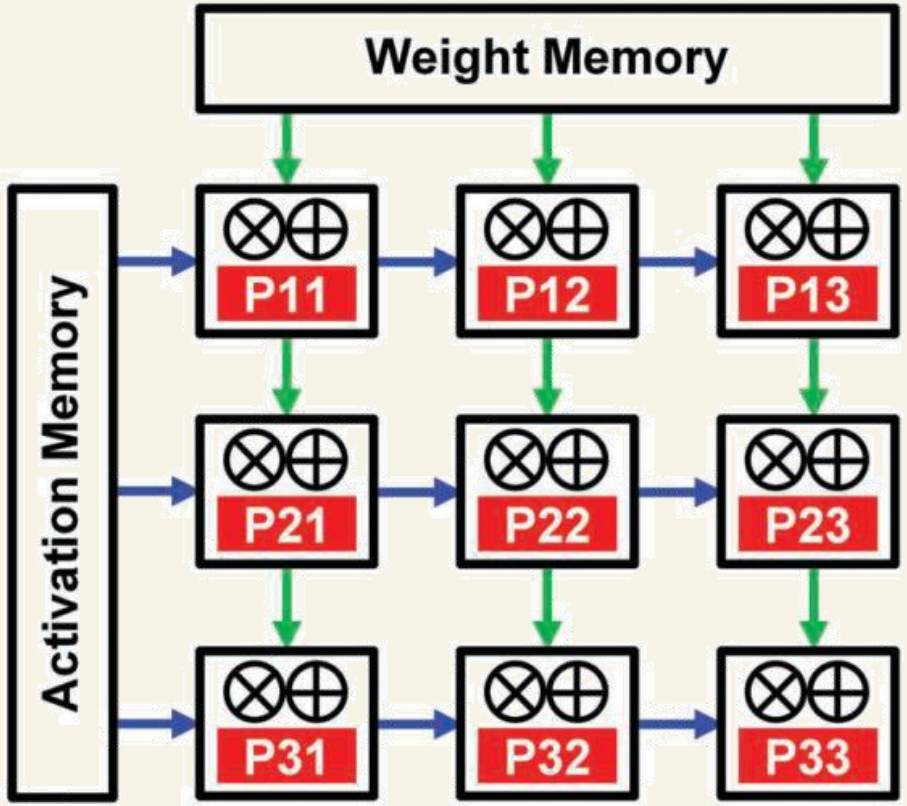
Aspect	Conventional (CPU+Memory)	Computing-in-Memory
Data movement	High	Very low
Energy efficiency	Lower	Higher
Memory bandwidth dependence	High	Low
Parallelism	Moderate	Very high
Latency	Higher	Lower
Programming ease	Mature	More challenging
Precision	High	Often lower
Flexibility	Excellent	Workload-specific
Hardware complexity	Moderate	Higher



O. Mutlu, EPFL, 2026

Digital vs Analog AI Accelerators

Digital AI Accelerators

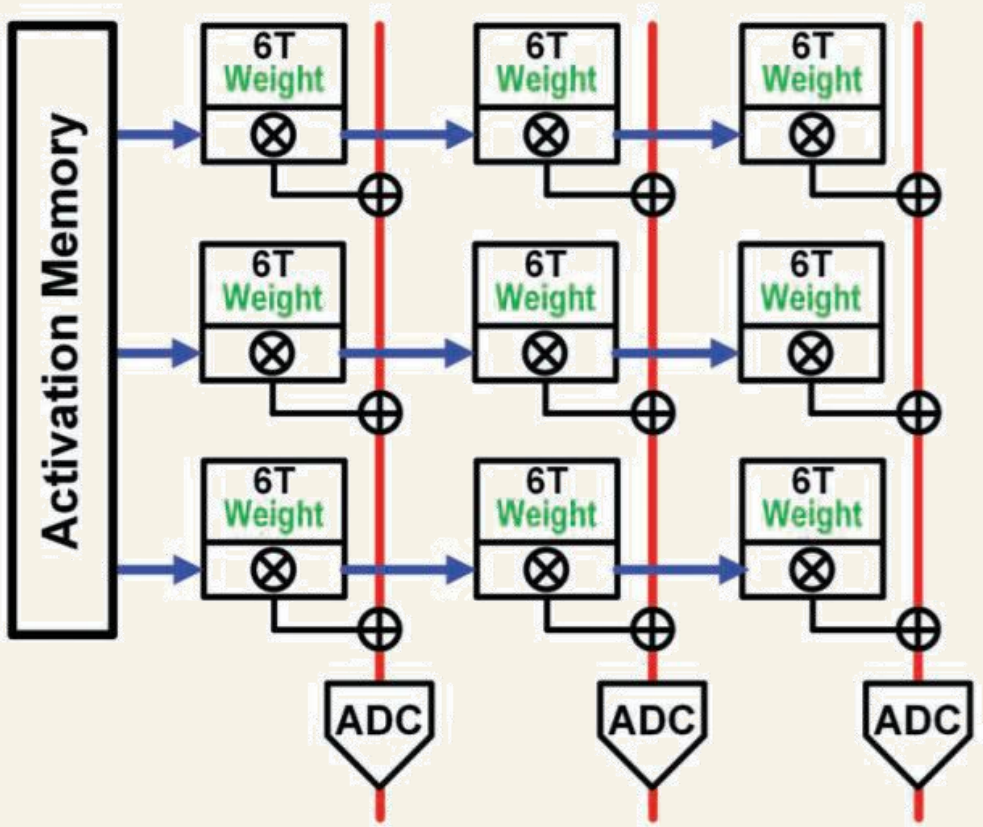


Input Activation

Weight

Partial Sum

Analog CIM AI Accelerators

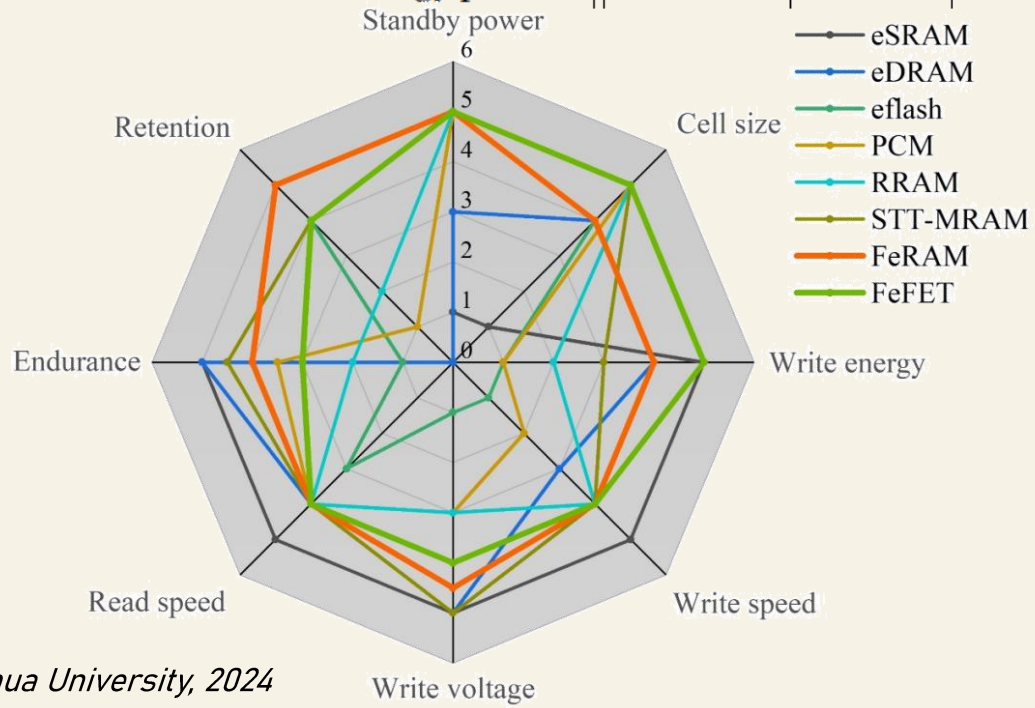


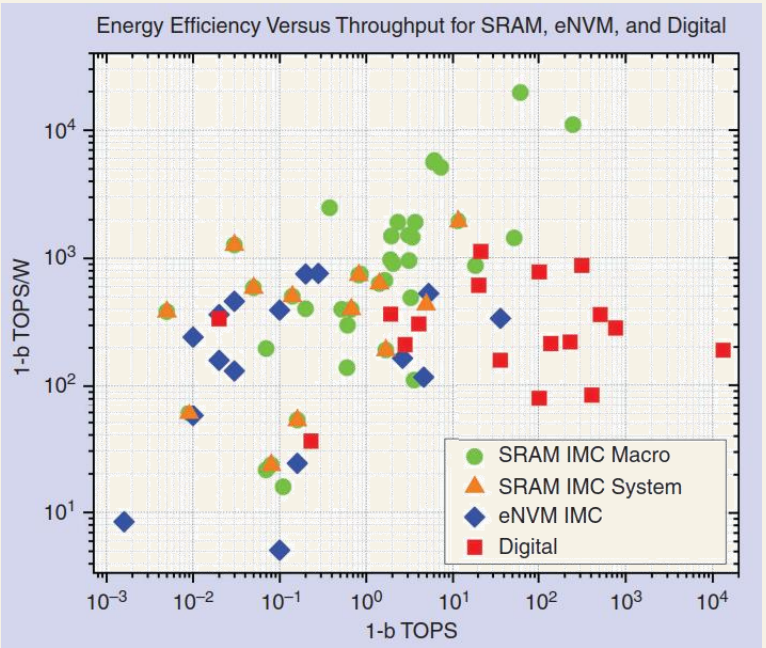
The most common computing operation is Multiply & Accumulate (MAC).

J.S Seo, IEEE SSC Mag, 2022

Memory Technologies for Analog CIM

	Mainstream Memories				Emerging Memories		
	SRAM	DRAM	NOR Flash	NAND Flash	PCRAM	STT-MRAM	RRAM
Cell area	$>100F^2$	$6F^2$	$10F^2$	$<4F^2$	$6-50F^2$	$4-30F^2$	$4-12F^2$
Multibit	1	1	2	3	2	1	2
Voltage	<1V	<1V	>10V	>10V	<3V	<1.5V	<3V
Read time	1ns	10ns	50ns	10µs	<10ns	<10ns	<10ns
Write time	1ns	10ns	10µs – 1ms	100µs – 1ms	50ns	<10ns	<10ns
Retention	None	ms	years	years	years	years	years
Endurance	$>10^{16}$	$>10^{16}$	$>10^5$	$>10^4$	$>10^9$	$>10^{15}$	$>10^6$
Write energy per bit	1 fJ	10 fJ	100 pJ	10 fJ	10 pJ	100 fJ	100 fJ

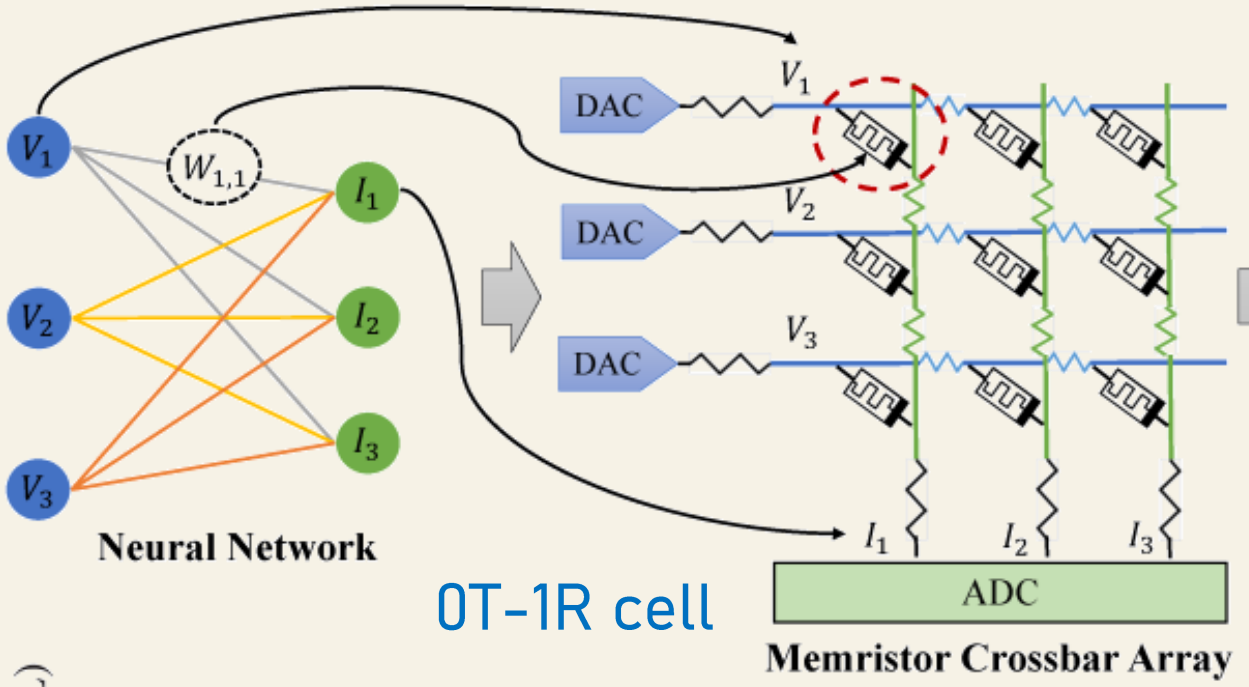




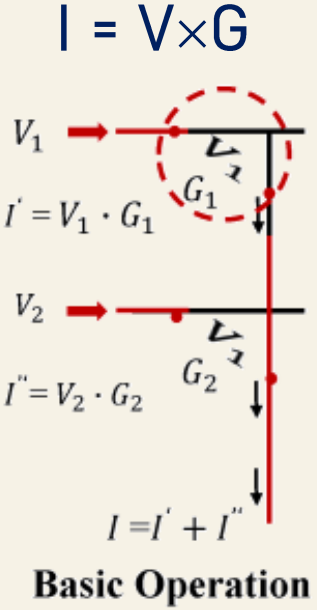
J.S Seo, IEEE SSC Mag, 2022

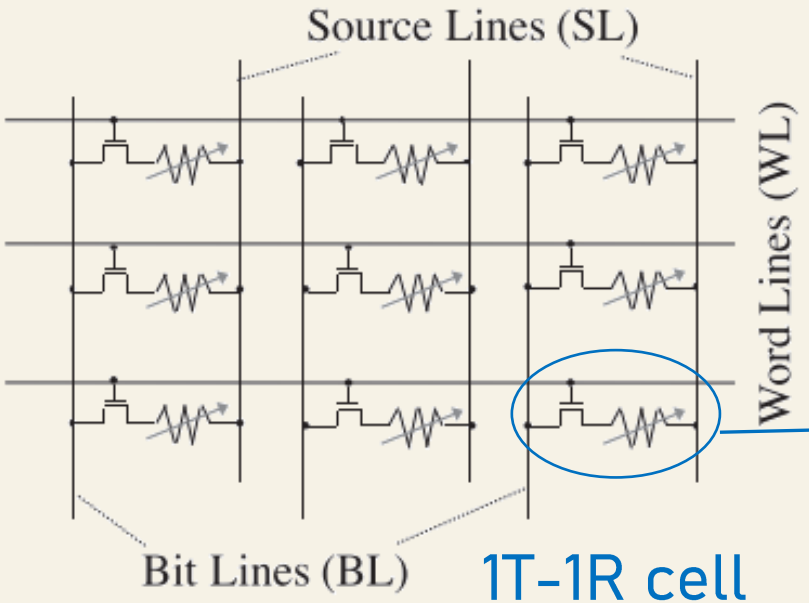
RRAM_PCM_STT-MRAM CIM Organization

Multiply & Accumulate (MAC) Operation

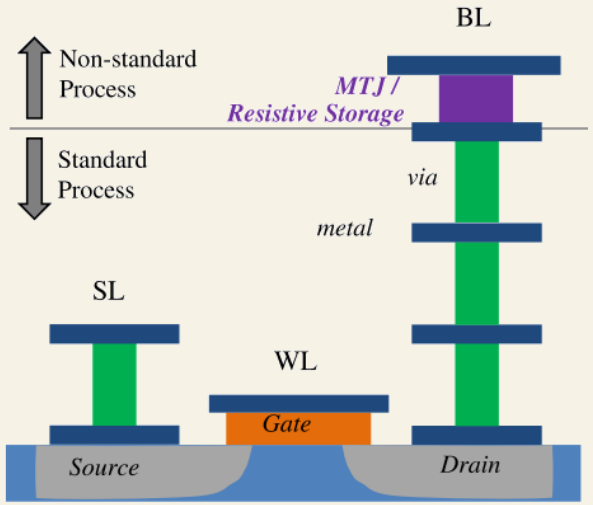


0T-1R cell

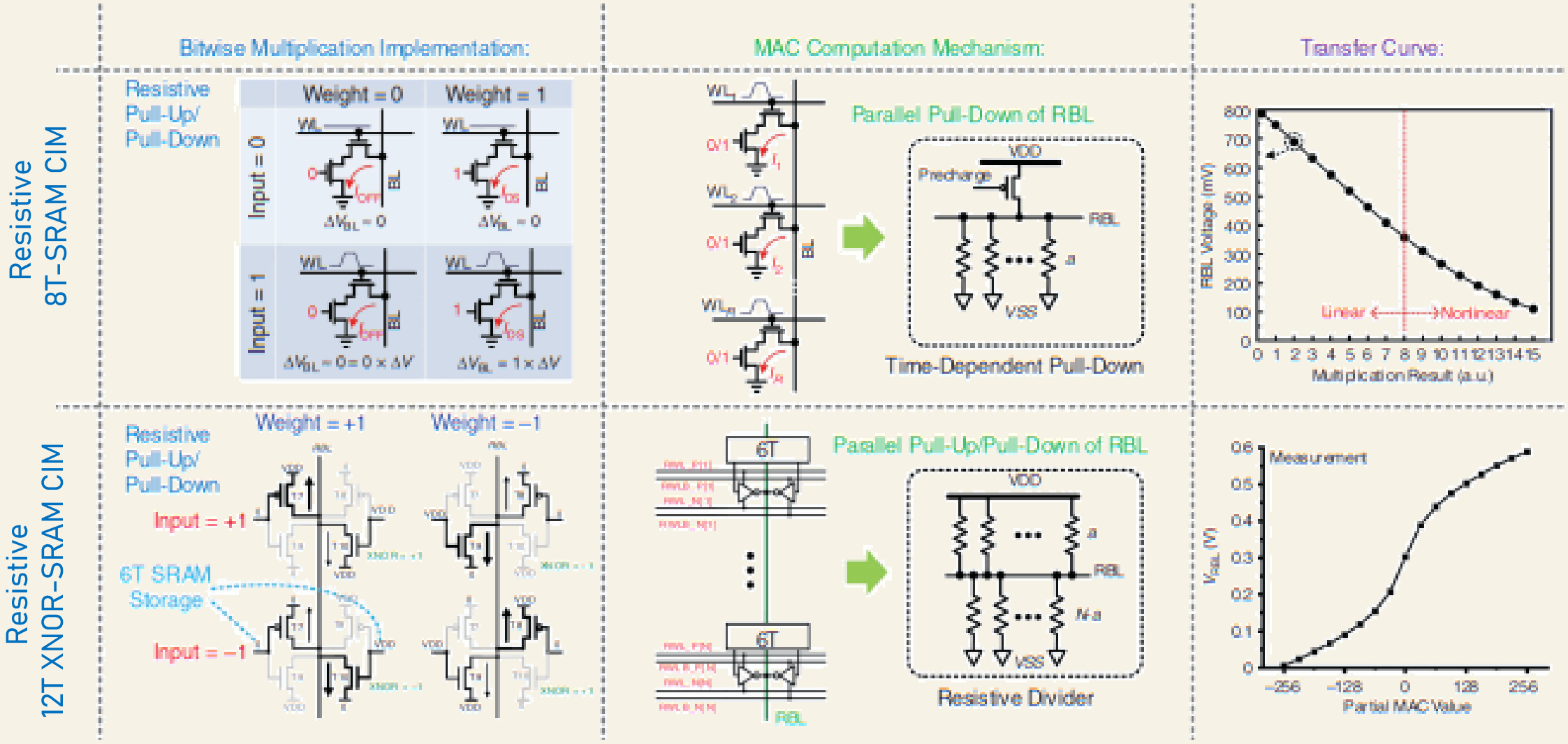




1T-1R cell

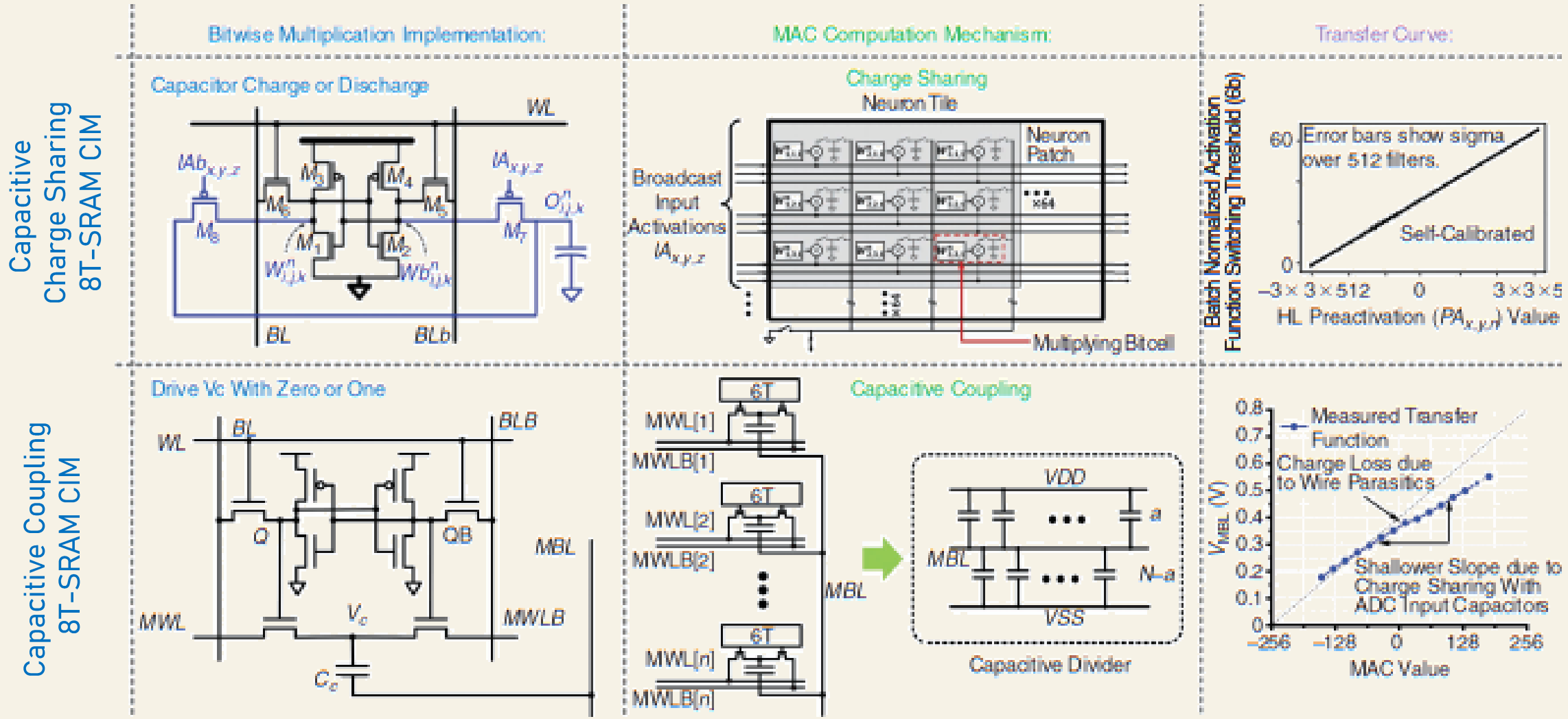


SRAM-based Resistive CIM Organization



J.S Seo, IEEE SSC Mag, 2022

SRAM-based Capacitive CIM Organization



J.S Seo, IEEE SSC Mag, 2022

Outline

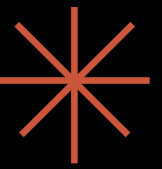
Computing in Memory and Reliability:

- Testing Challenges
  - Defect mechanisms should be defined
  - Fault Models should be developed
  - Test Generation tools are required
  - Testing Algorithms, Design for Testability & Design for Diagnosis techniques are mandatory
  - Testing Standards are needed
  - On-Line Testing solutions should be proposed
- Aging Challenges
  - Emerging Memories Aging Models should be developed
  - Aging Alleviation techniques should be explored
- Process Variation Challenges
- Radiation induced Single Event Effect (SEE) Challenges
  - Radiation Hardening techniques may be crucial for system reliability

Fault Tolerant Design for Memristor-based AI Accelerators

T. Spyrou, A. Zografou, S. Hamdioui and A. Grebregiorgis,

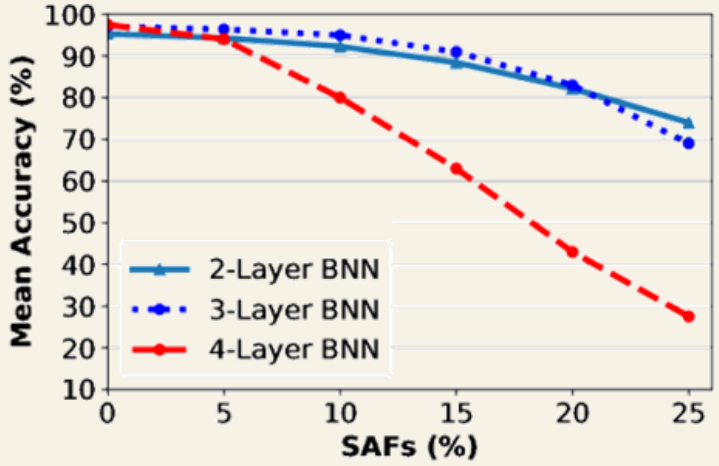
IEEE International Conference on Design, Test and Technology of Integrated Systems (DTTIS), 2024.



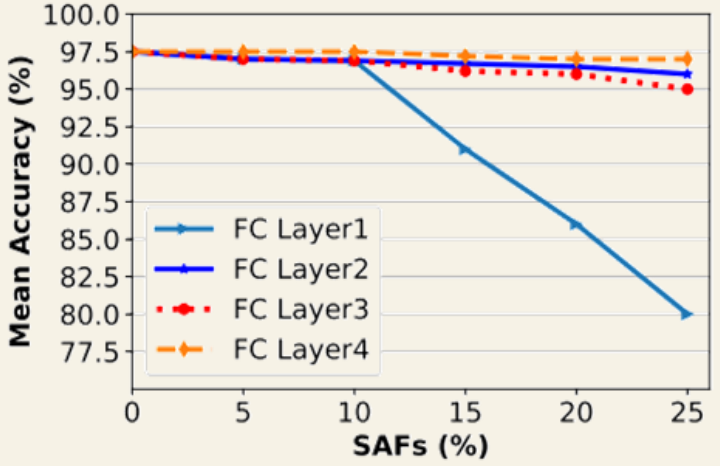
Fault Tolerant Design

Defect and Variation Impact Analysis

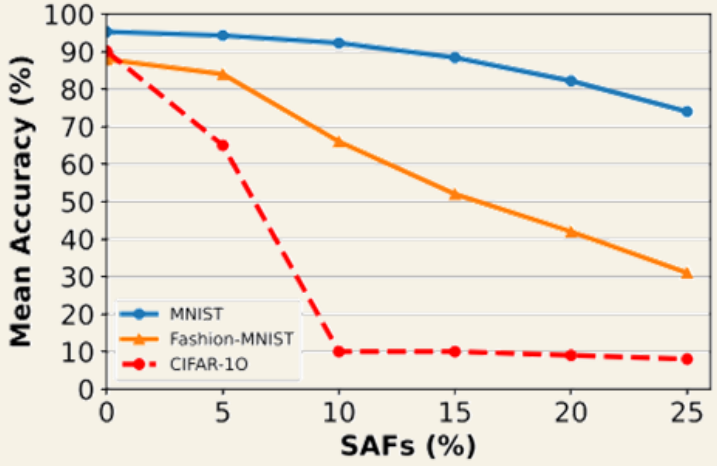
Stuck-at Faults



Impact on n-layer MNIST BNN

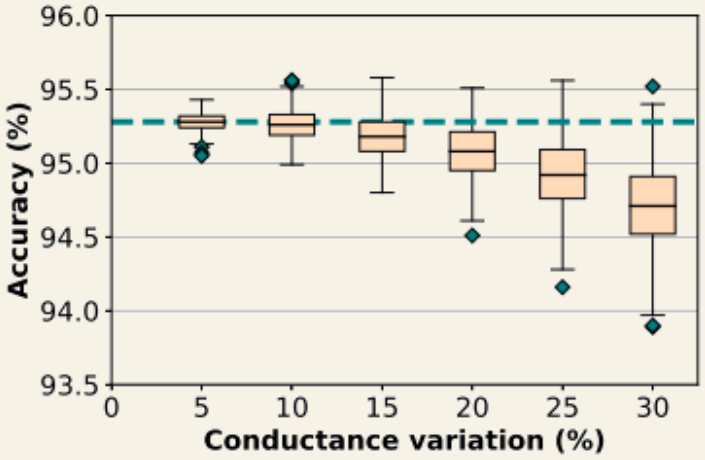


Layer-wise impact on MNIST BNN

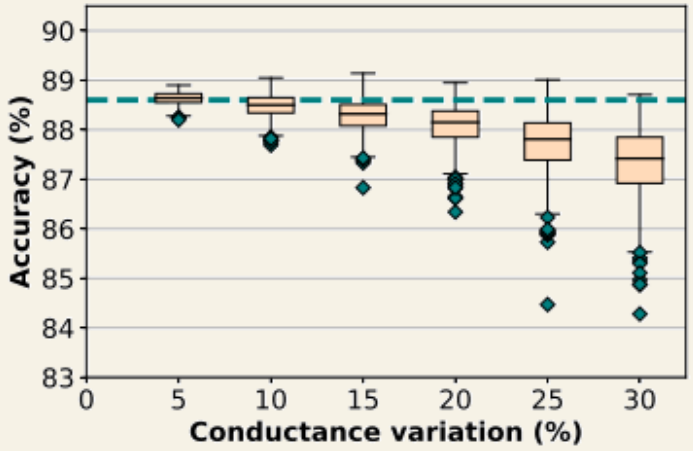


Impact on different BNN architectures

Conductance Variation

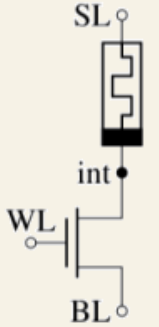


2-layer MNIST BNN



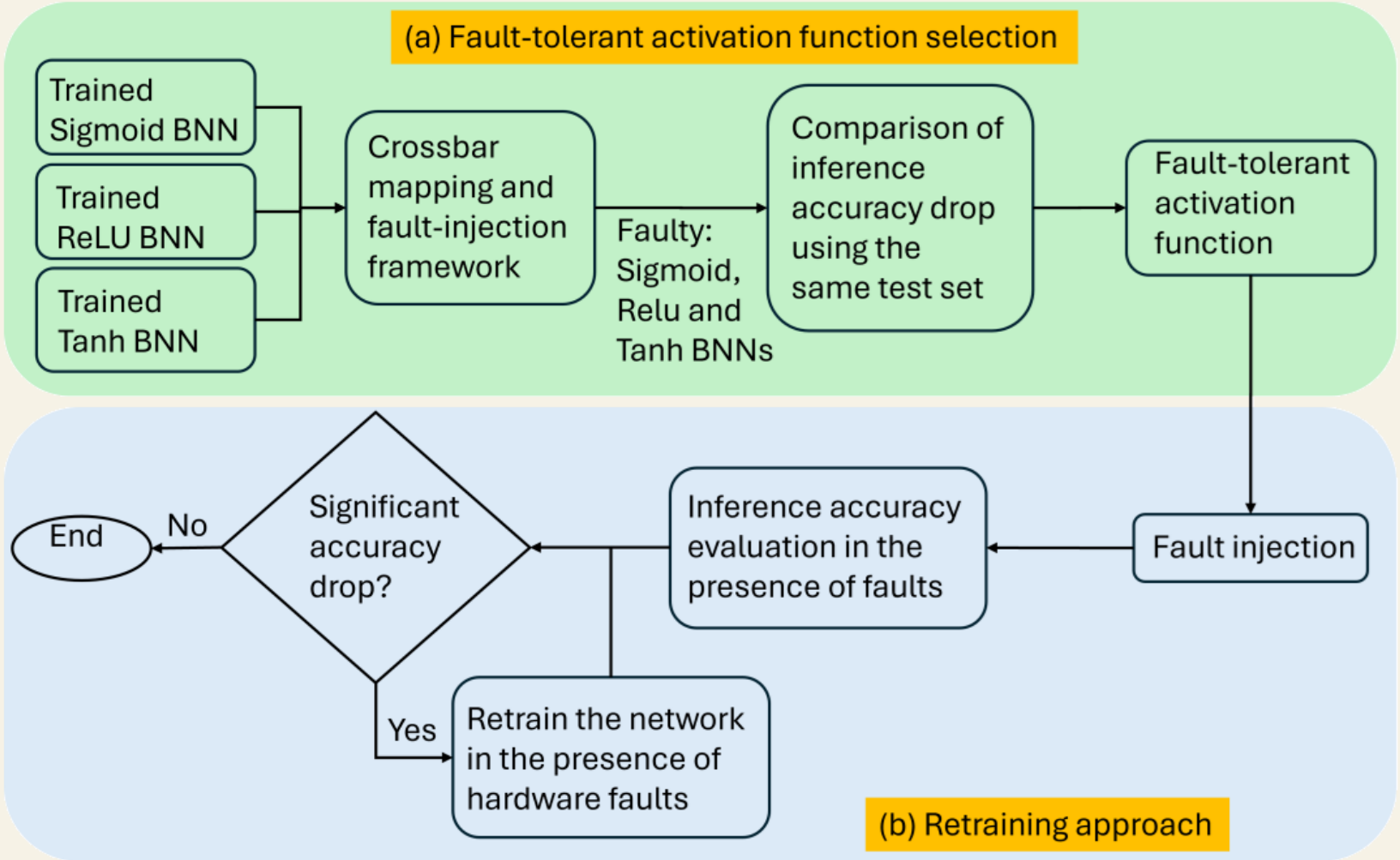
4-layer Fashion-MNIST BNN

- BNN: Binary Neural Networks
- SAF: Stuck-at Faults
  - Stuck-at LRS & Stuck-at HRS



1T-1R cell

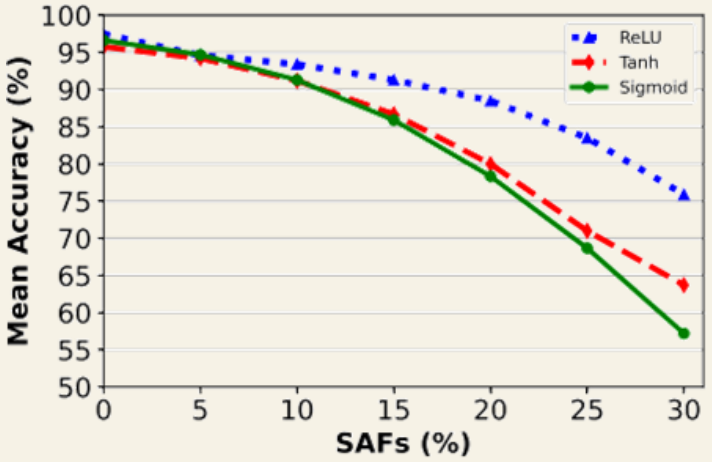
Fault Tolerant Training Flow



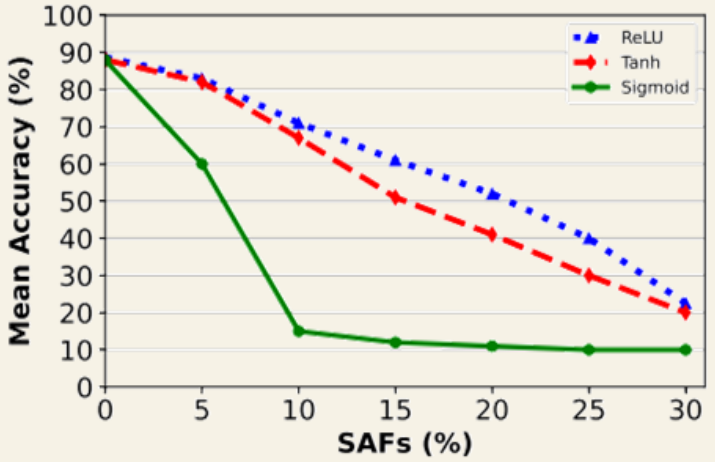
- Experimentation for the selection of the proper activation function: **sigmoid**, **ReLU**, **tanh**.

- Fault-aware retraining of a network to enhance its fault tolerance capability by learning around the faulty components.
- Faulty locations are excluded from the training algorithm.

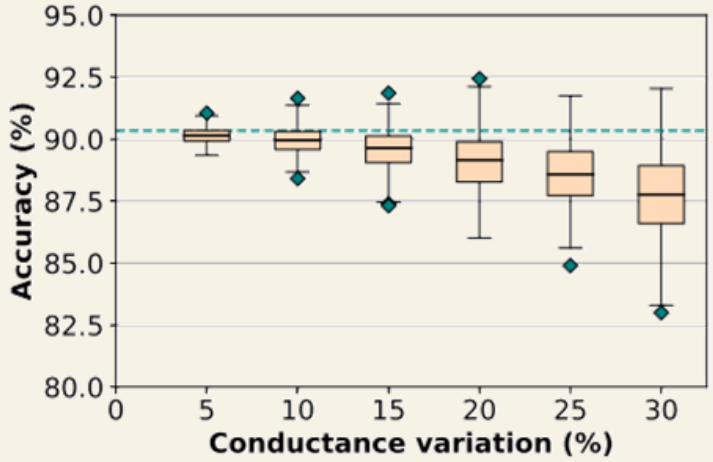
Impact for Different Activation Functions



2-layer MNIST BNN



Fashion-MNIST BNN



ReLU 2-layer MNIST BNN, 15% SAFs

Three activation functions under consideration: ReLu, Tanh and Sigmoid.

The ReLU activation function achieves the least accuracy drop when SAFs are injected in the crossbar array. ReLU achieves improvements in the accuracy of around 10%.

SAF Aware Retraining Results

ReLU activation function

	2-layer BNN (MNIST)		4-layer BNN (Fashion-MNIST)		VGG (CIFAR-10)	
SAF (%)	Classification Accuracy (%)					
	Baseline	Retrained	Baseline	Retrained	Baseline	Retrained
0	97.3	-	88.2	-	90.1	-
5	96.0	97.3	84.0	87.1	65.0	89.9
10	95.4	97.0	71.0	88.0	10.0	89.7
15	93.0	97.2	62.0	87.9	10.0	89.6
20	89.0	97.1	53.0	88.0	9.0	89.0
25	84.0	97.2	40.0	87.8	8.0	88.7
30	71.0	96.8	22.0	88.5	8.0	89.0

- SAF-aware retraining can almost fully recover the accuracy back to the (fault-free) baseline.

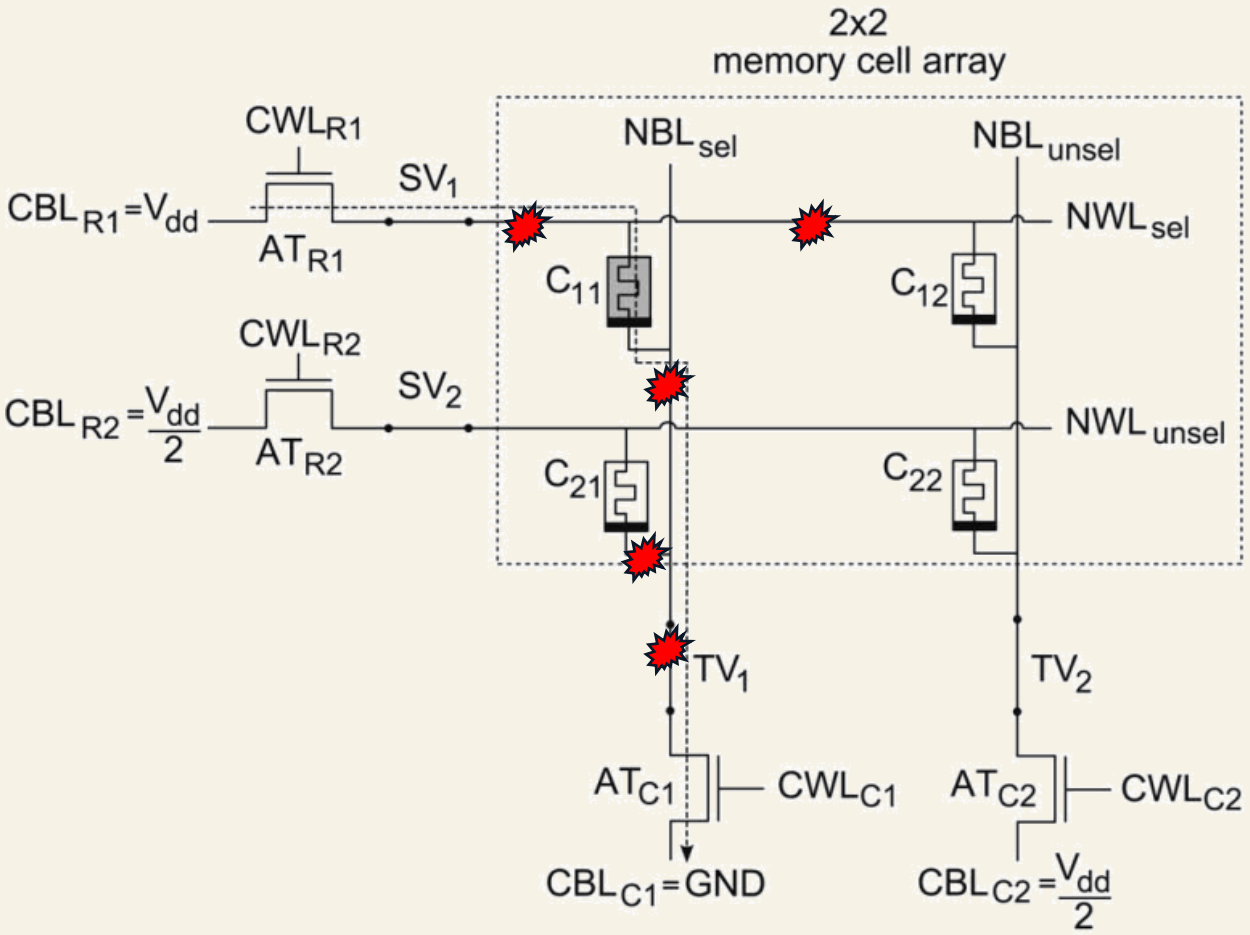
Testing Open Defects in Memristor-Based Memories

S. Hamdioui, M. Taouil and N.Z. Haron

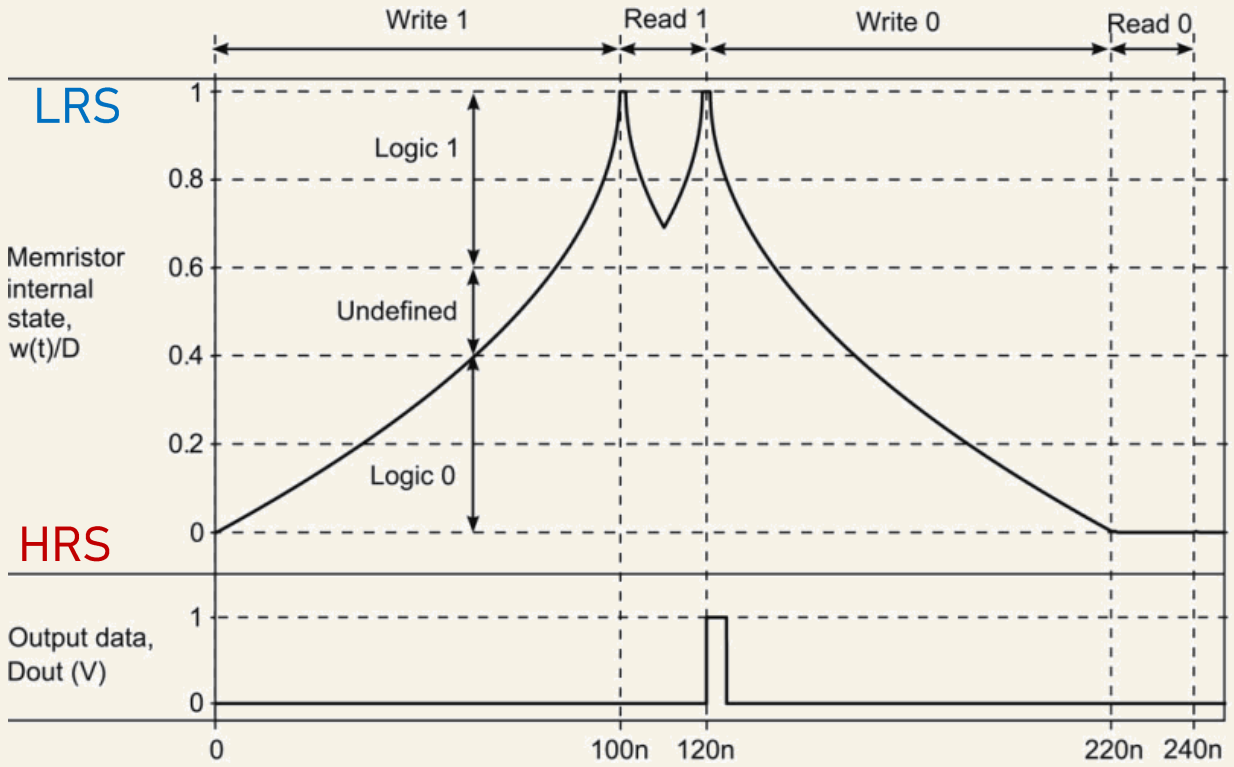
IEEE Transaction on Computers, vol. 64, no. 1, 2015.

DfT Support

Defect Free Write & Read Operation

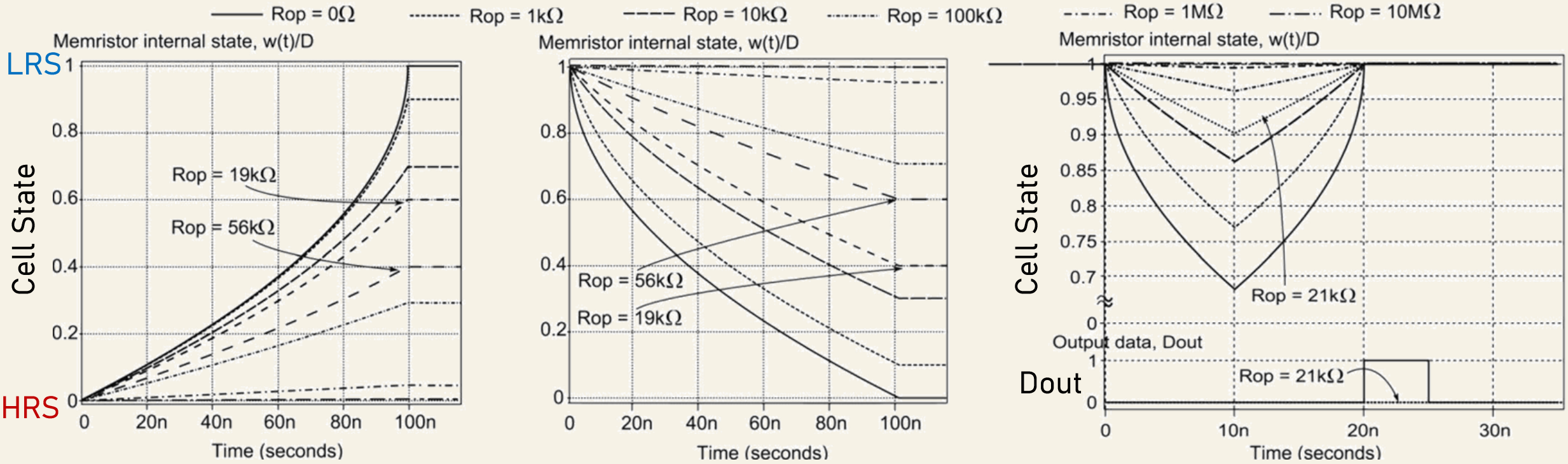


Titanium oxide memristor



Resistive Open Defect Influence

Cell Open Defect



- Conventional Faults - Detectable by March Tests
  - Transition Faults (TF0, TF1), Stuck-At Faults (SAF0, SAF1), Incorrect Read Faults (IRF1)
- Unique Faults – Hard to detect faults – Random responses – Special DfT is required !
  - Undifined State Faults (USF0, USF1).

Design for Testability (DfT)

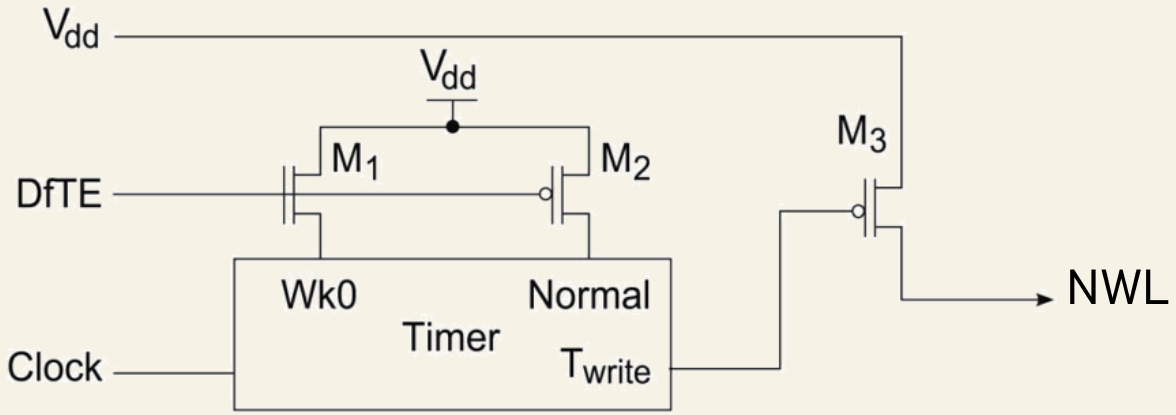
A possible way to improve fault coverage of march tests and detect the Undefined State Faults is to stress the cells, aiming to:

- to shift an undefined state to an incorrect state, and in parallel
- to avoid overtesting that leads to yield loss.

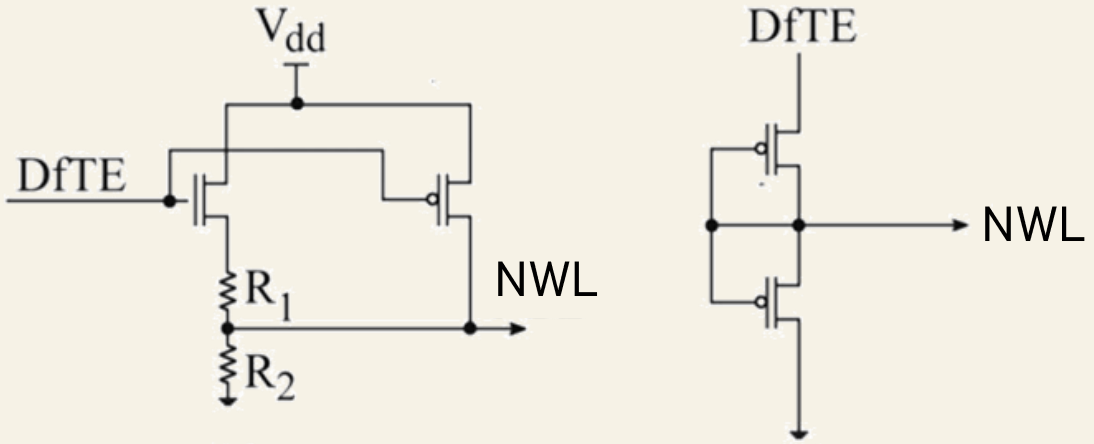
Two stress test mechanisms adopted, which lead to corresponding DfT schemes:

- short write time (SWT), and
- low voltage write (LVW).

Both mechanisms are defined as weak write operations.

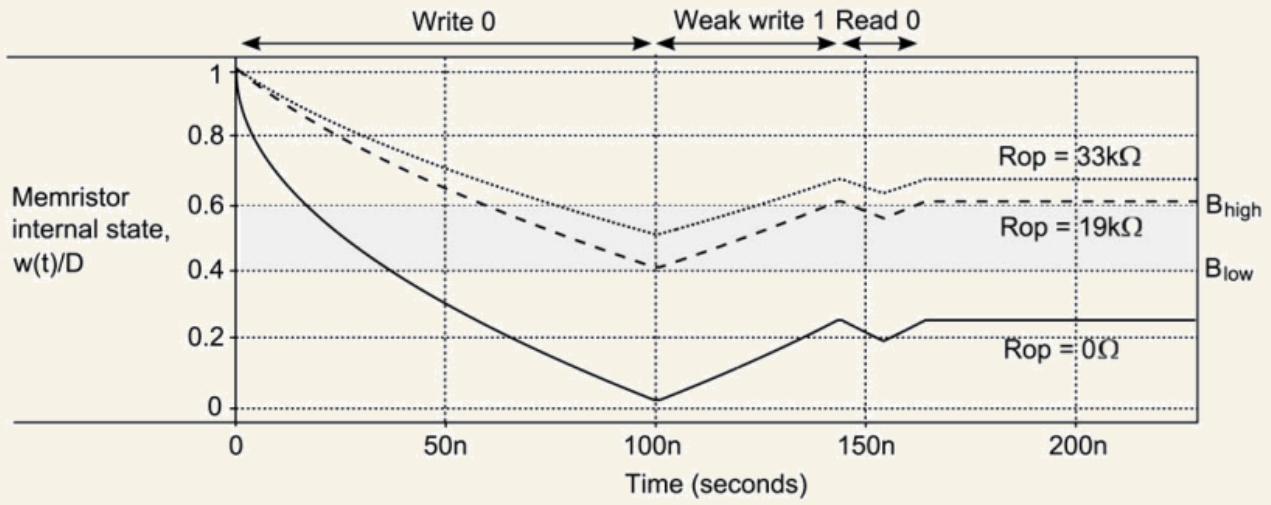


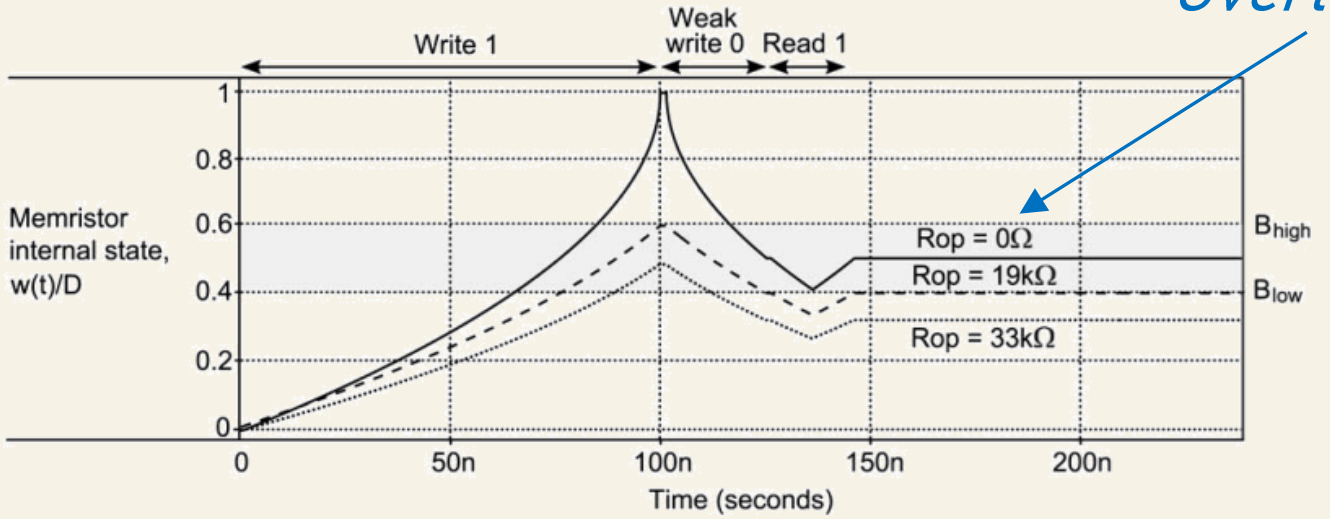
SWT-DfT



LVW-DfT

Short Write Time DfT





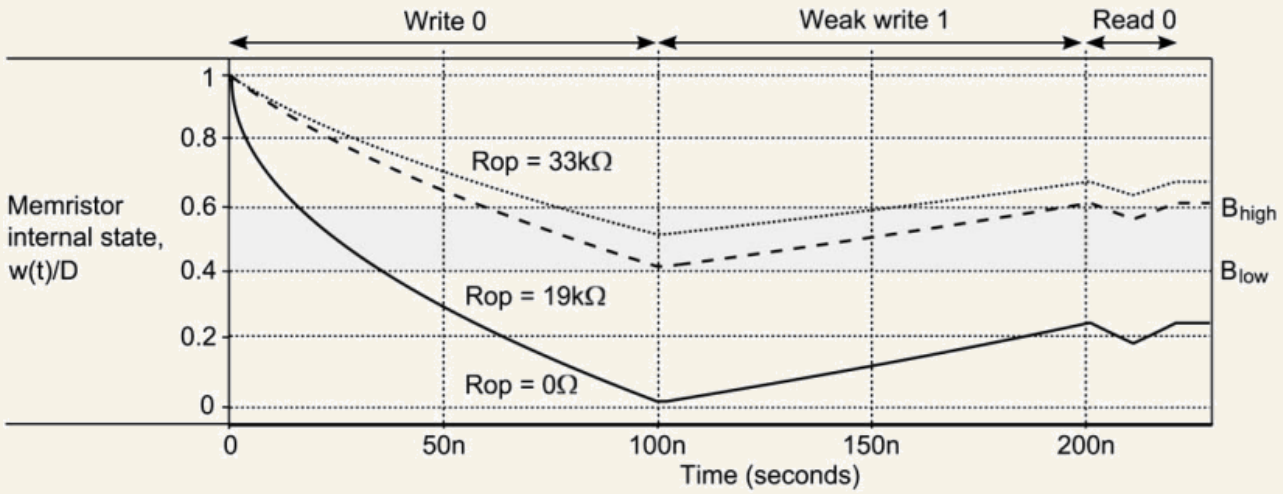
Overtesting

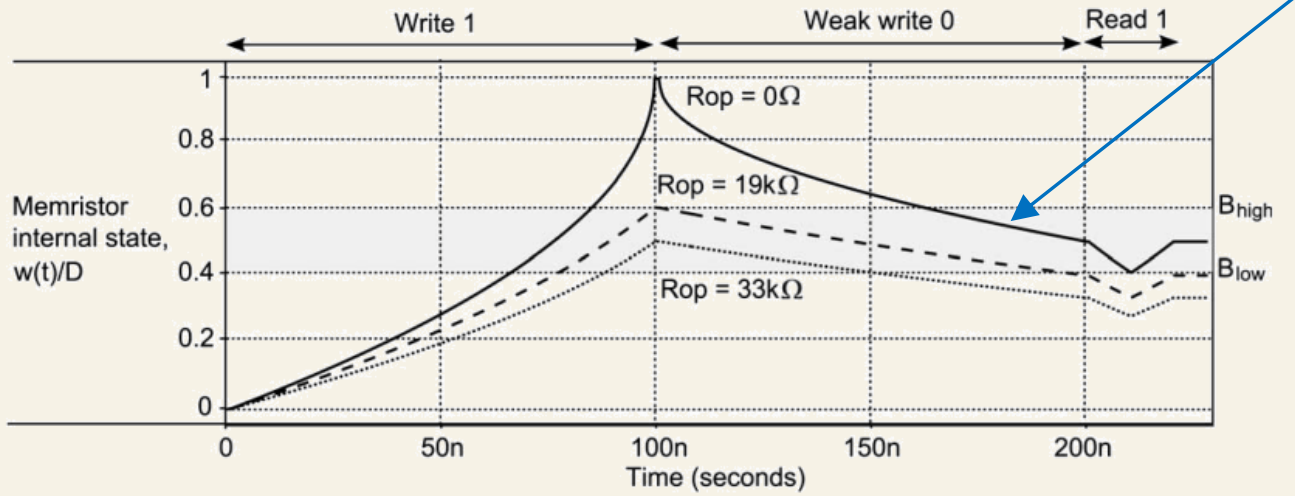
$\{\Uparrow (w1);\ \Uparrow (w0\ \hat{w}1r0)\}$

Weak Write

$\{\Uparrow (w0);\ \Uparrow (w1\ \hat{w}0r1)\}$

Low Voltage Write DfT





Overtesting

$\{\Uparrow (w1);\ \Uparrow (w0\ \hat{w}1r0)\}$

Weak Write

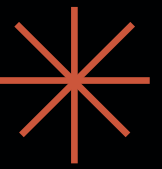
$\{\Uparrow (w0);\ \Uparrow (w1\ \hat{w}0r1)\}$

To avoid overtesting issues, programmable SWT and LVW operations are required.

Structural Testing of a RRAM-based AI Accelerator Core

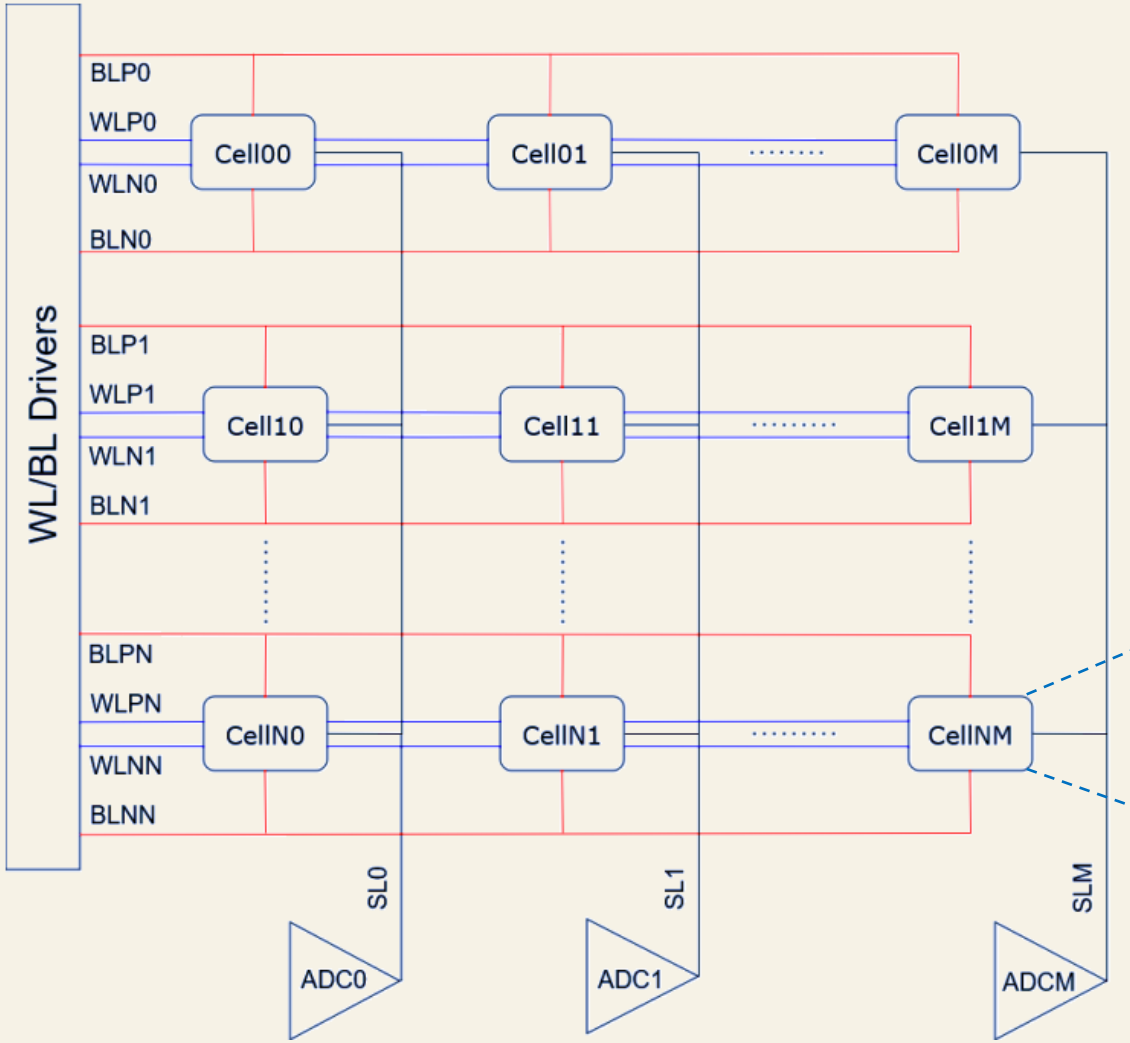
E.A. Serlis, H. Xun, M. Taouil, S. Hamdioui, and M. Fieback

IEEE European Test Symposium (ETS), 2025.



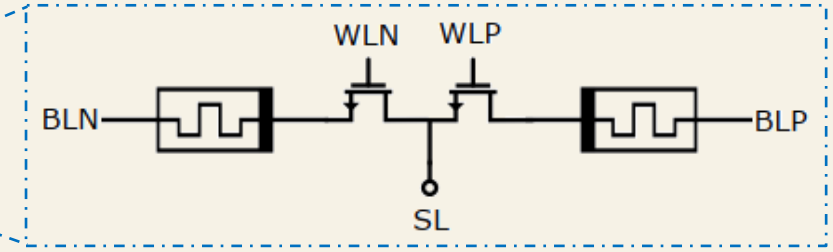
ATPG Tool

Analog CIM Organization



Mapping RRAM states and BL voltage to numerical values

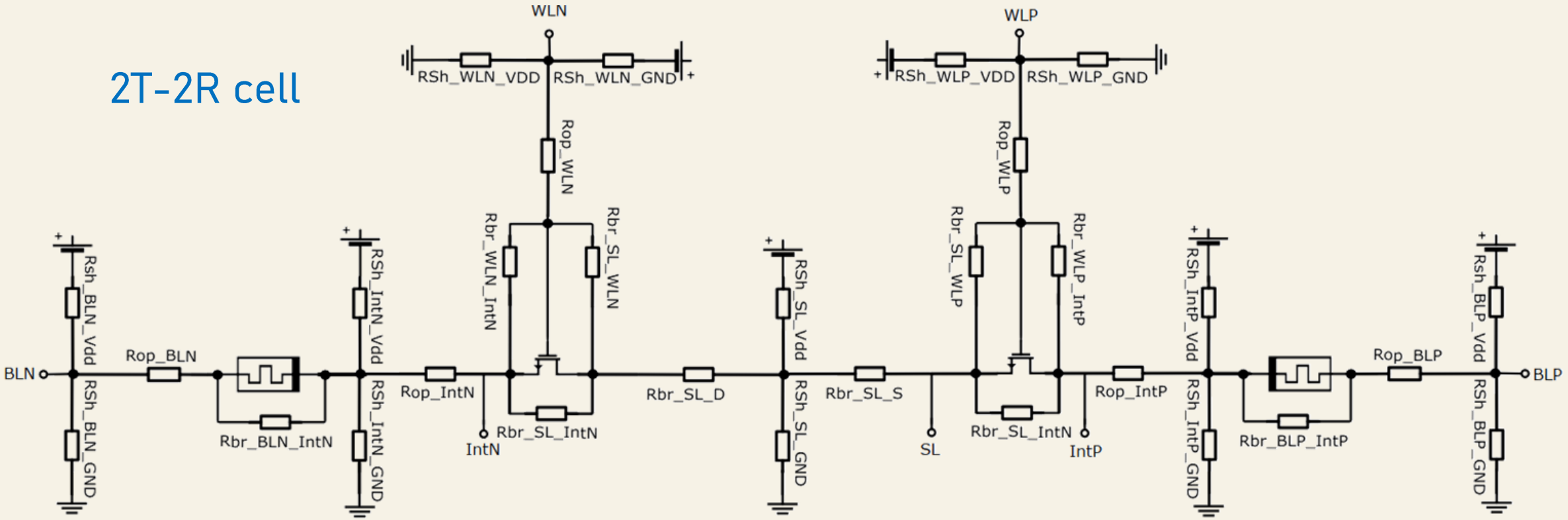
	Positive Side	Negative Side	Value
Weights (W)	HRS LRS LRS	LRS LRS HRS	-1 0 1
Inputs (IN)	250mV 550mV 850mV	850mV 550mV 250mV	-1 0 1



2T-2R cell

Cell Resistive Open & Short Defects

2T-2R cell



Resistive defect strength is swept between 10 Ω and 5 MΩ in 10 log-spaced steps.
Single cell, single fault assumption.

Structural Testing Procedure

During testing the cell under test along with four neighboring cells in the same column are active.

The test vectors are defined as: (R_P R_N :IN_P IN_N :N_i).

An ATPG tool developed for test vector generation with the highest fault coverage and the minimum test length by combining input signals and weights. The tool minimizes the amount of column inputs as well as the number of write operations necessary (due to the high write delay), while still sensitizing all potential defects.

The most optimal sequence of test vectors is: {(-1+1:-1+1:-1-1-1-1), (-1+1:+1-1:+1+1-1-1), (+1+1:+1-1:-1-1-1-1)}.

Required Read / Write operations 3N / 1N (N is the number of rows). Fault coverage: 85%.

A Dynamic Testing Scheme for Resistive-Based Computation-In-Memory Architectures

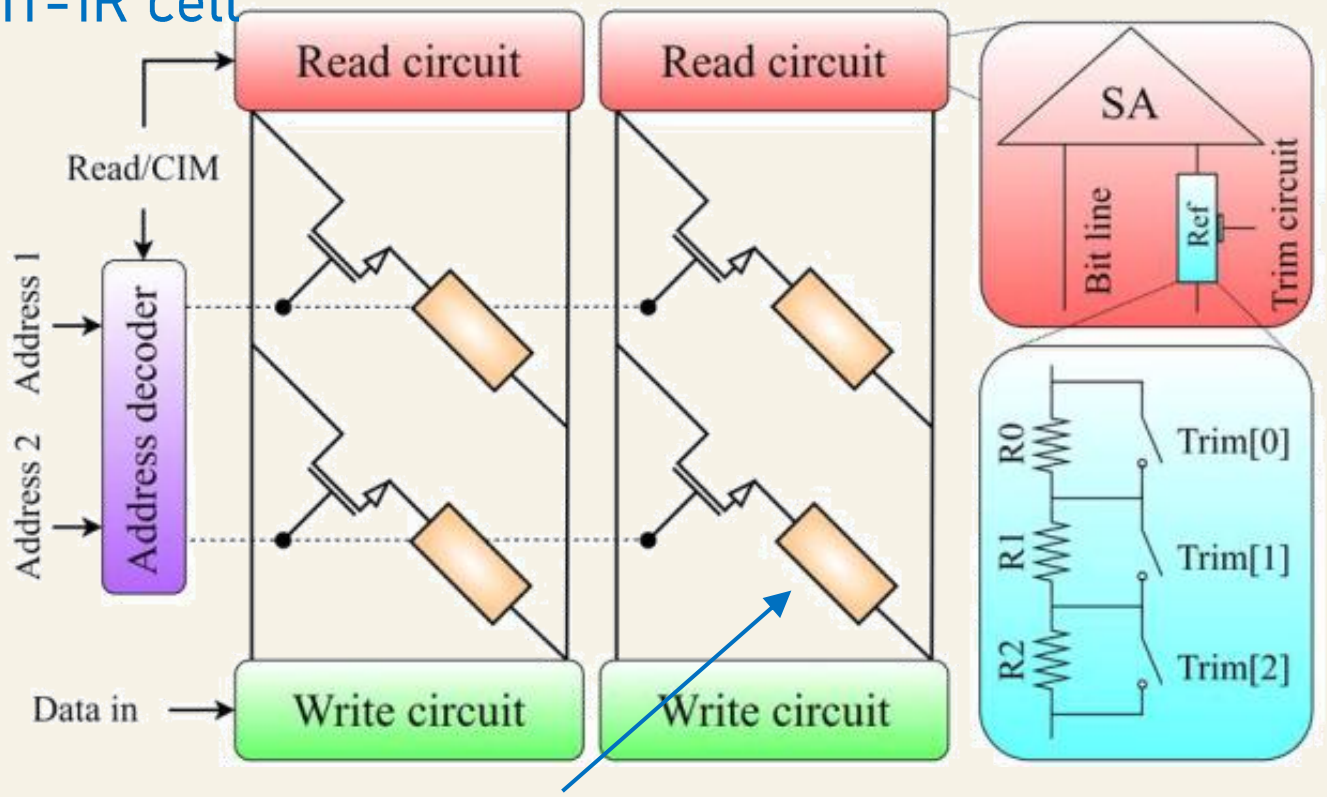
S.B. Mamaghani, P. Pal, M.B. Tahoori

IEEE Asia and South Pacific Design Automation Conference (ASP-DAC), 2024.

BIST Support – Test & Diagnosis

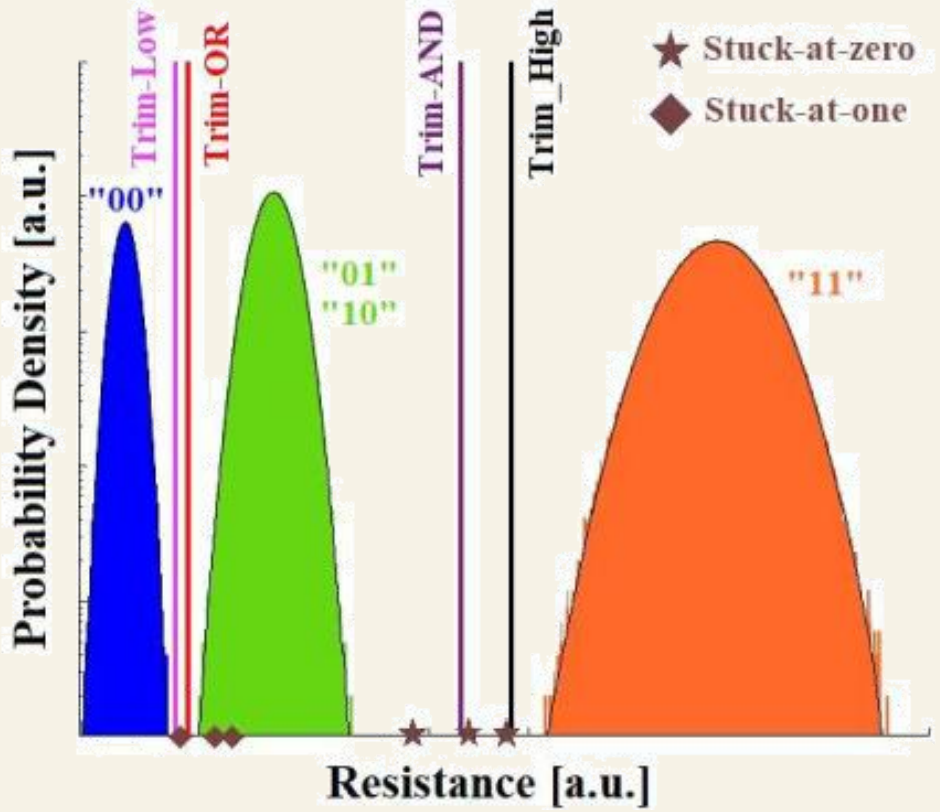
Testing for STT-MRAM & RRAM CIMs

1T-1R cell



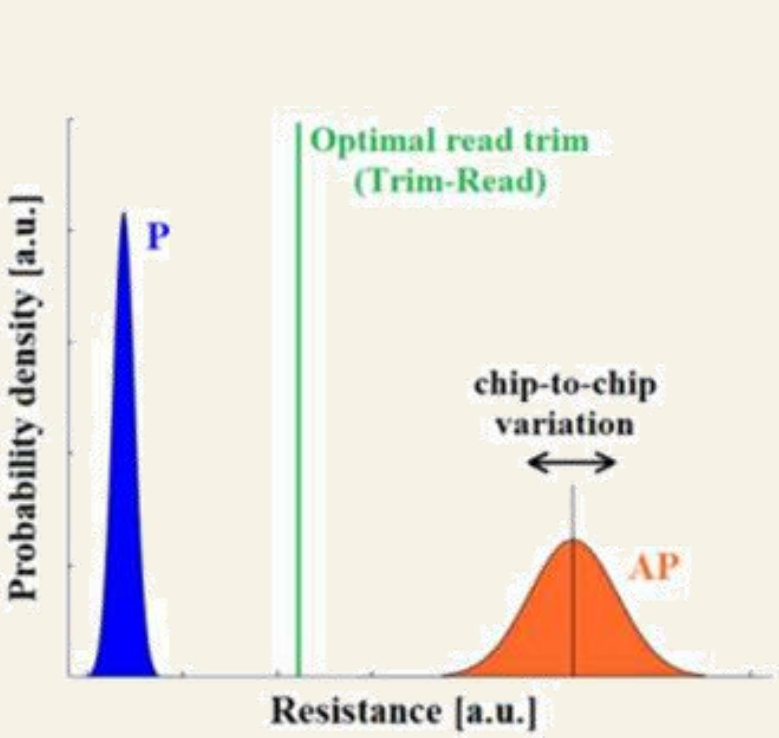
Spin Transfer Torque Magnetic RAM (STT-MRAM)
or
Resistive RAM (RRAM)

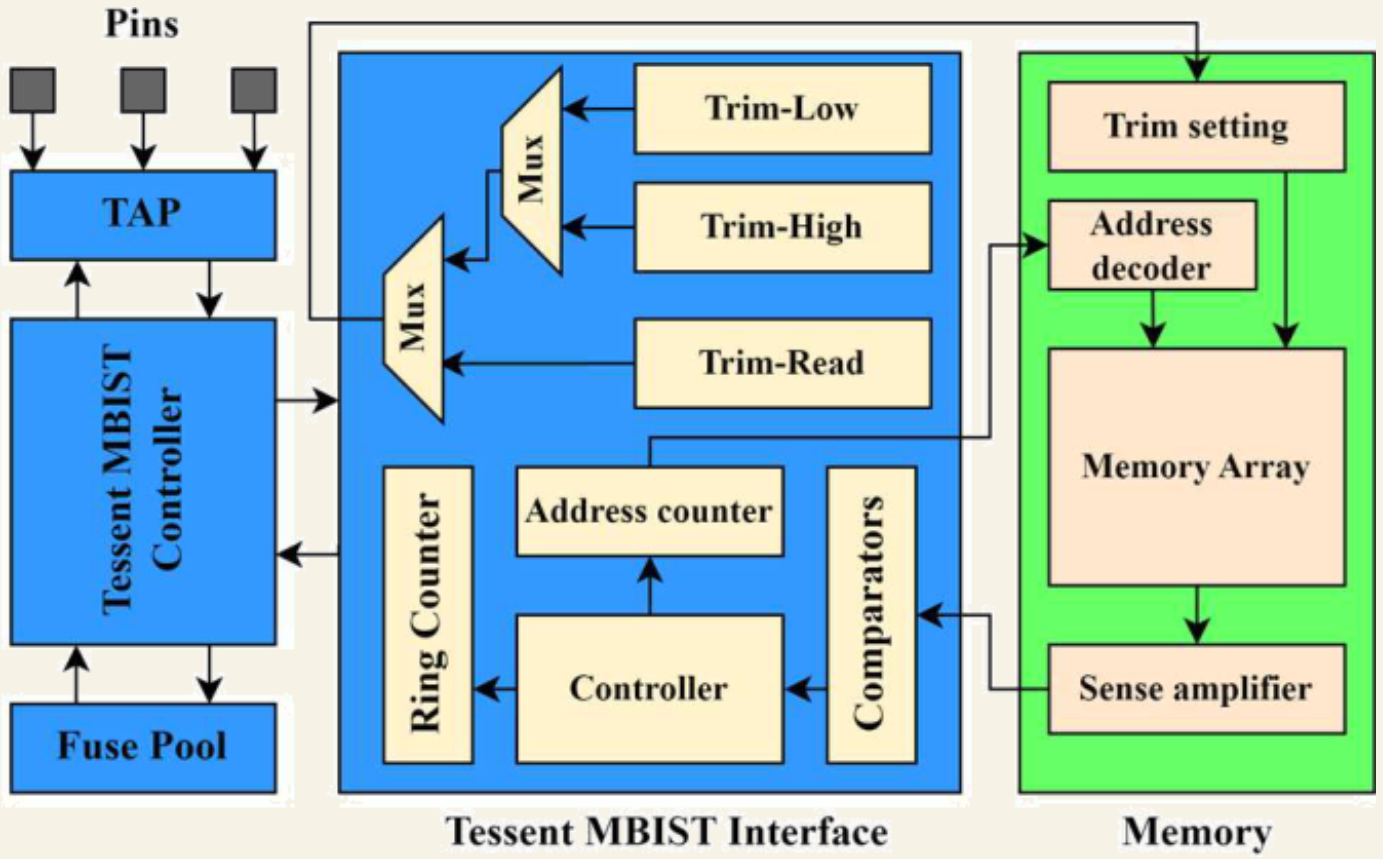
Two-phase Test



- Back-Hopping defects for STT-MRAMs.
- Under-Formation and Over-Formation defects for RRAMs.

Built-In Self Test (BIST) Scheme





	Proposed	Case 1	Case 2	Case 3	Case 4	March C
Test time order	$2\left(N+\frac{N}{m}\right)+(d\times m)$	$10N+4N$	$10N+5.5N$	$12N$	$10N+3kN$	$10N$

N = # cells, m = # of CIM inputs, d = of batches that fail with Trim-Low or Trim-High settings

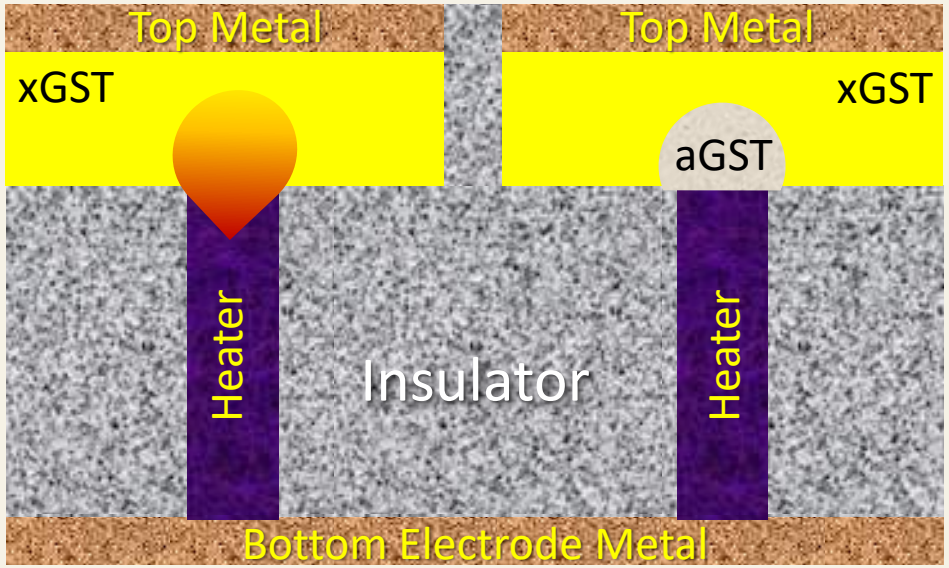
Testing Algorithms for Hard to Detect Thermal Crosstalk Induced Write Disturb Faults in Phase Change Memories

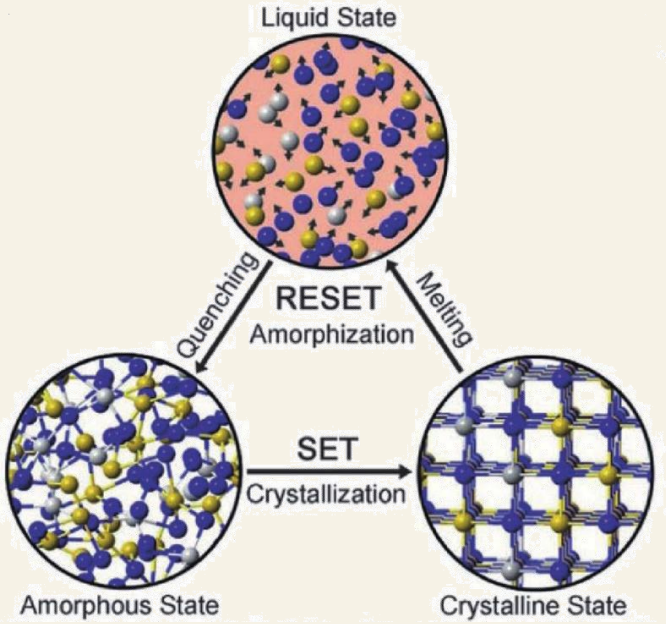
S. Spyridonos, Y. Tsiatouhas

Design Automation and Test in Europe Conference (DATE), 2024.

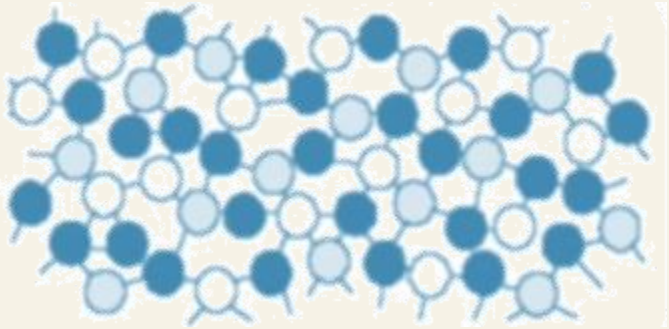
Testing Algorithms

Phase-Change Memory



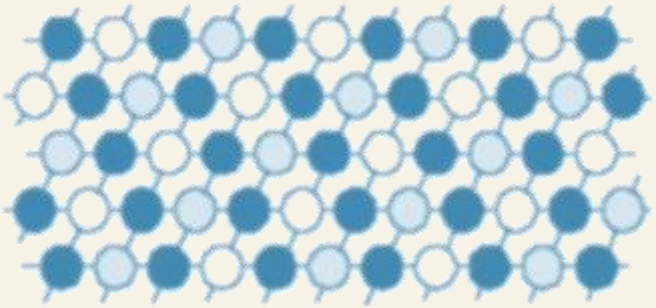


Amorphous Phase

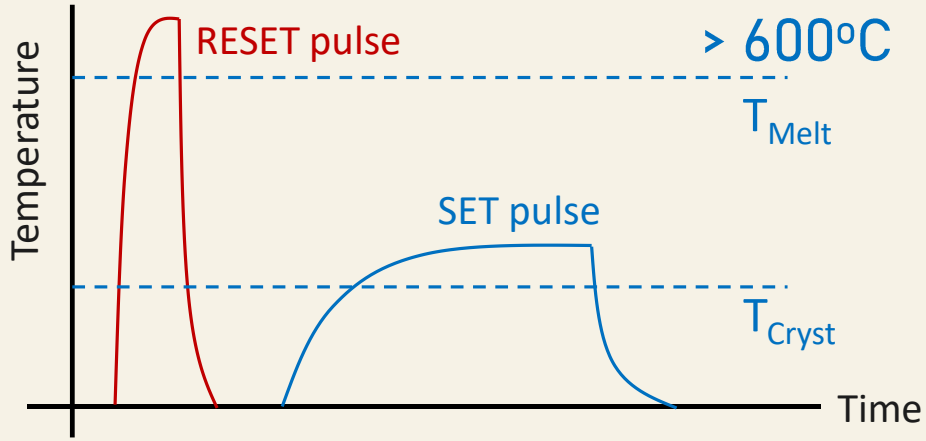


HRS – RESET state

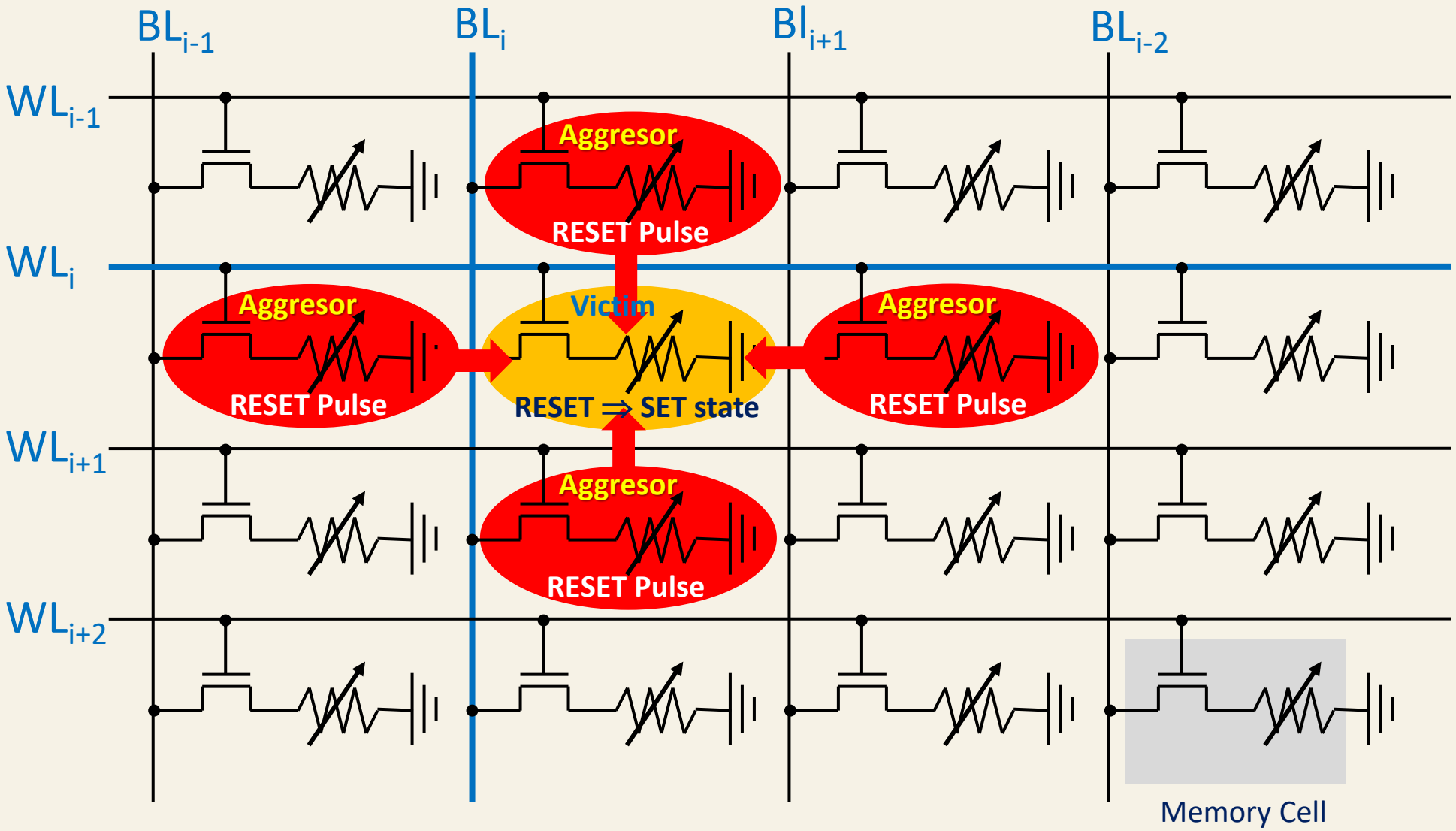
Crystalline Phase

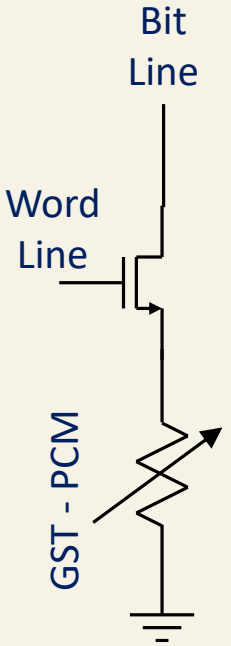


LRS – SET state



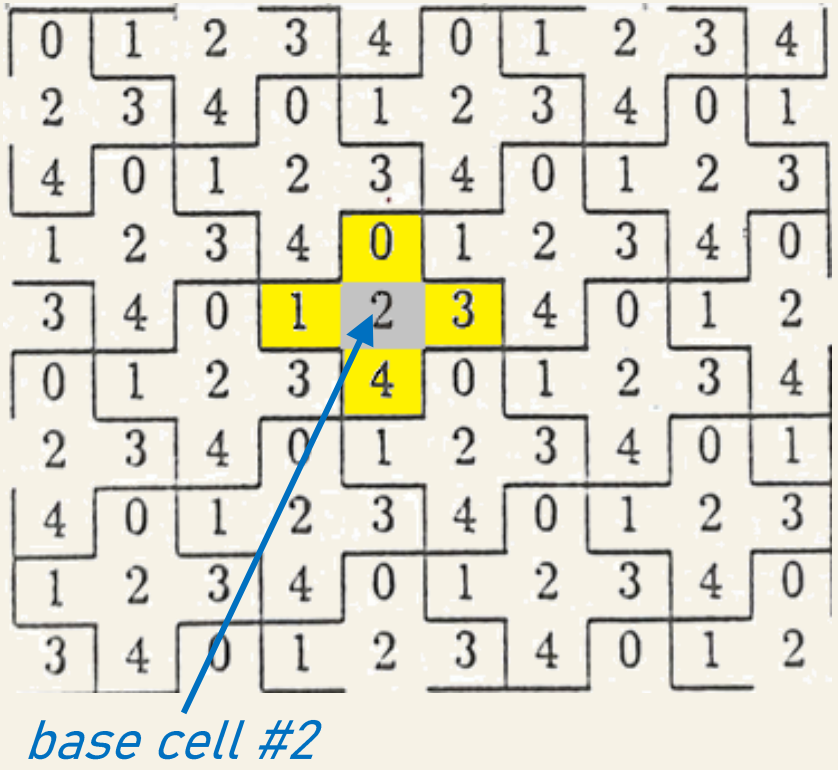
Thermal Crosstalk Write Disturb Faults





Neighborhoods for Testing Operations

Typical Type-1 tiling neighborhoods



The two-group tiling method for the Type-1 neighborhood

A	b	B	b	A	b	B	b
b	C	b	D	b	C	b	D
B	b	A	b	B	b	A	b
b	D	b	C	b	D	b	C
A	b	B	b	A	b	B	b
b	C	b	D	b	C	b	D
B	b	A	b	B	b	A	b
b	D	b	C	b	D	b	C

b=base cell

b	A	b	B	b	A	b	B
C	b	D	b	C	b	D	b
b	B	b	A	b	B	b	A
D	b	C	b	D	b	C	b
b	A	b	B	b	A	b	B
C	b	D	b	C	b	D	b
b	B	b	A	b	B	b	A
D	b	C	b	D	b	C	b

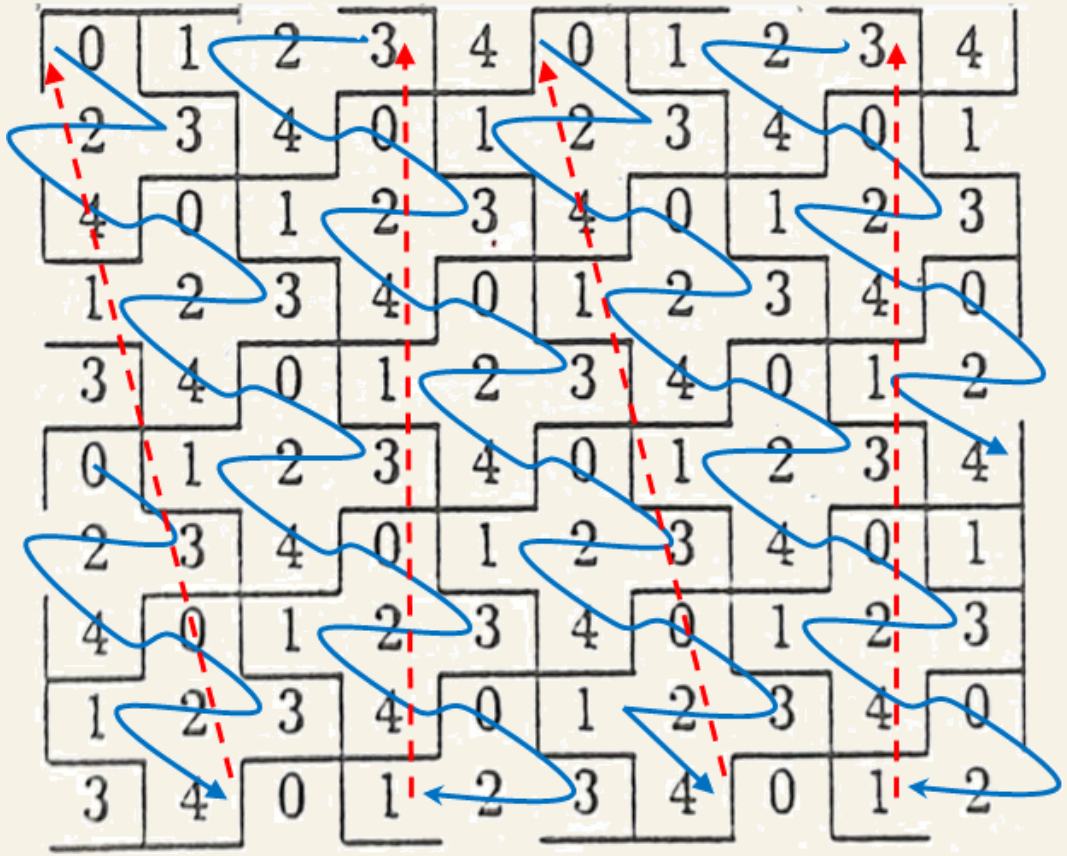
The Thunderbolt Testing Algorithm

Sequential RESET operations on **4 cells** in each neighborhood.

A total of 10MxN operations for the whole algorithm application.

By adopting the test strategy where RESET operations are performed on cells already being in the RESET state, the test cost is reduced to 6MxN operations.
Fast Thunderbolt algorithm.

base cell #2



Array of M rows and N columns

The Wave Testing Algorithm

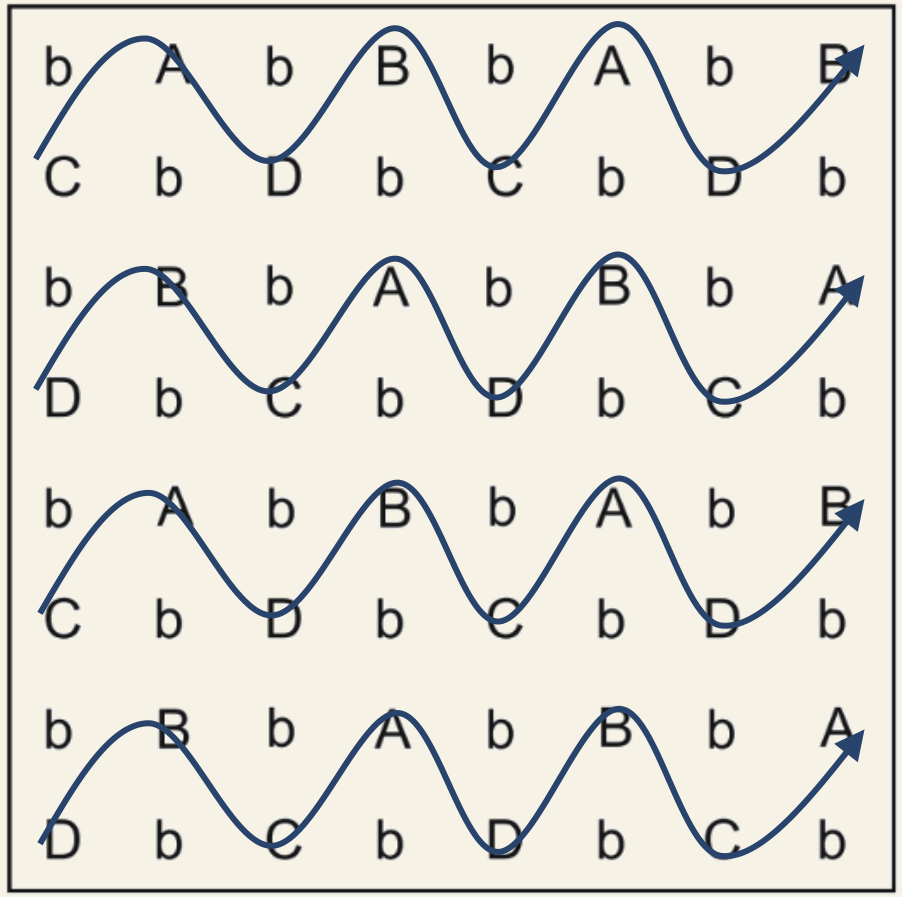
Sequential RESET operations on 3 cells in each neighborhood.

A total of 4MxN operations for the whole algorithm application.

By adopting the previous test strategy, the test cost is reduced to 3MxN operations. Fast Wave algorithm.

First Group

b=base cell



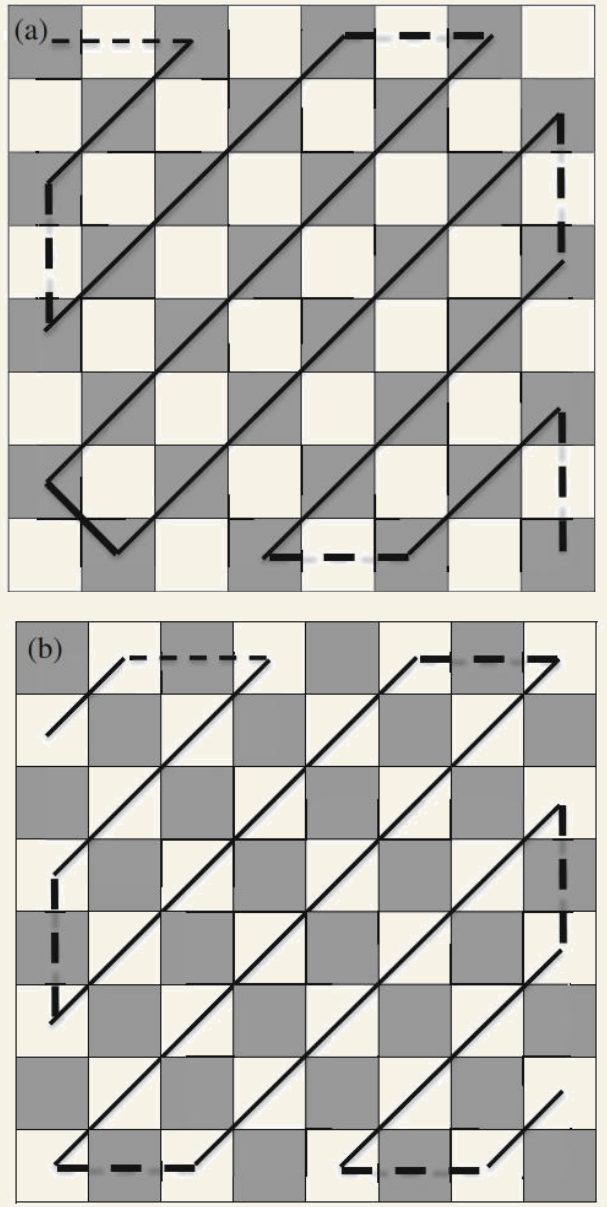
Array of M rows and N columns

Comparisons

		# of adjacent cells sequentially activated for the coverage of hard to detect TC-WDFs		
Algorithm	Complexity	2 cells	3 cells	4 cells
March-PC	11MxN	–	–	–
March-PCM	8MxN	–	–	–
March-PDF	3MxN	✓	–	–
Thunderbolt	10MxN	✓	✓	✓
Fast Thunderbolt	6MxN	✓	✓	✓
Wave	4MxN	✓	✓	–
Fast Wave	3MxN	✓	✓	–

Thunderbolt and Wave based addressing algorithms are capable to cover hard to detect thermal crosstalk faults.

The two-group tiling method

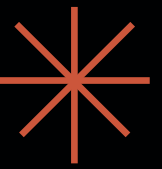


Snake based addressing mode.

DfT-Assisted Testing of Resistive Short Defects in SRAM-Based In-Memory Computing Architectures

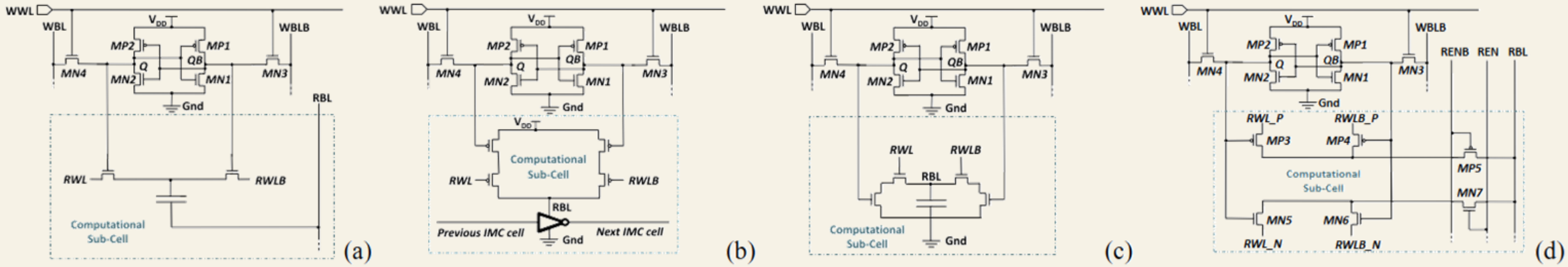
D.S. Lampridis, Y. Tsiatouhas, M. K. Michael and T. Theocharides

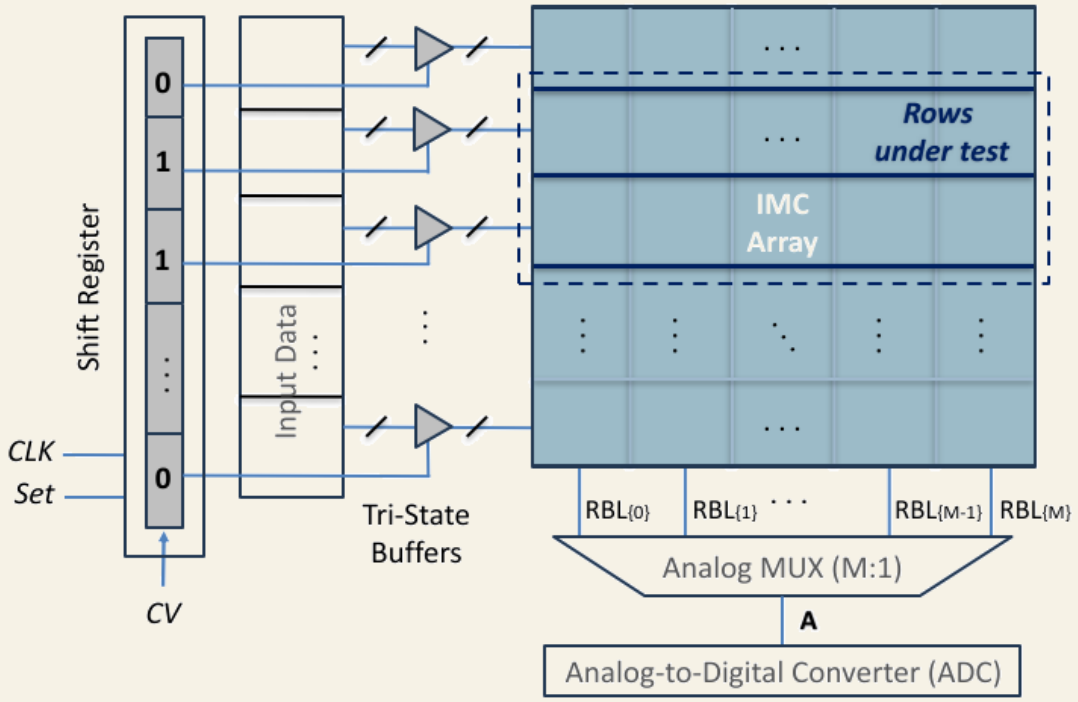
Under review, 2026.



DfT Support – Test & Diagnosis

DfT for SRAM-based CIMs

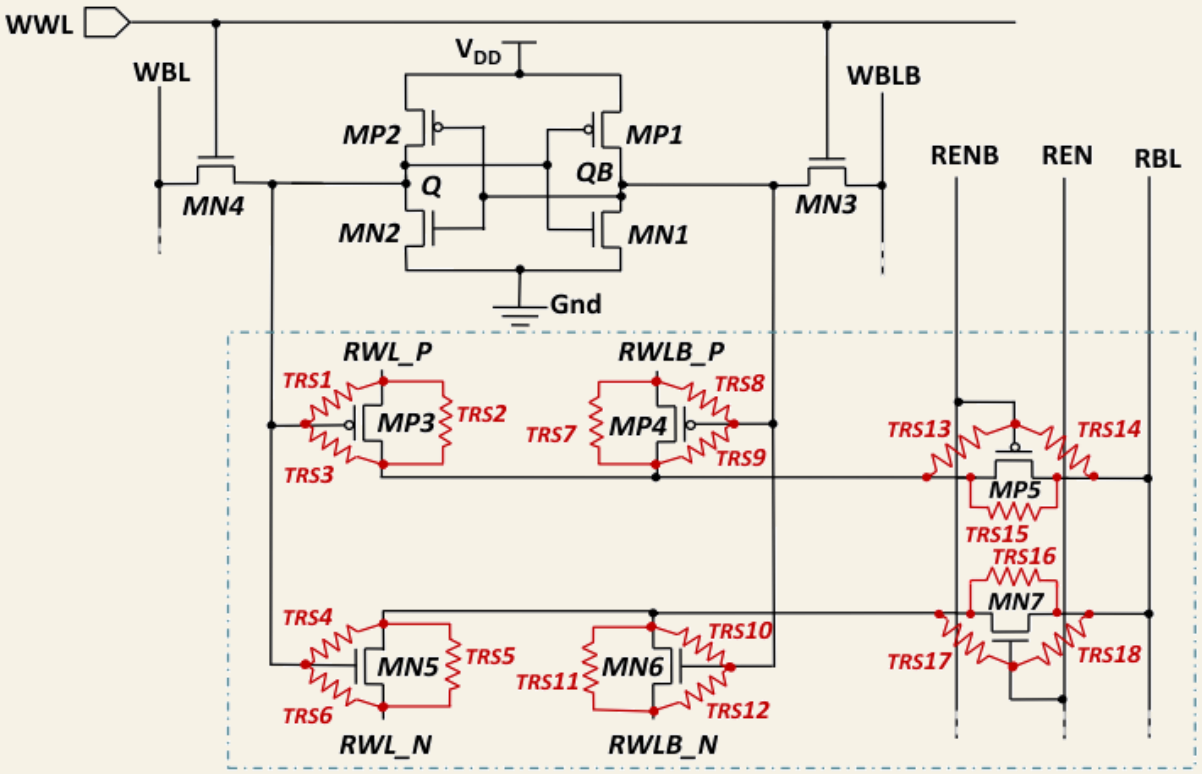




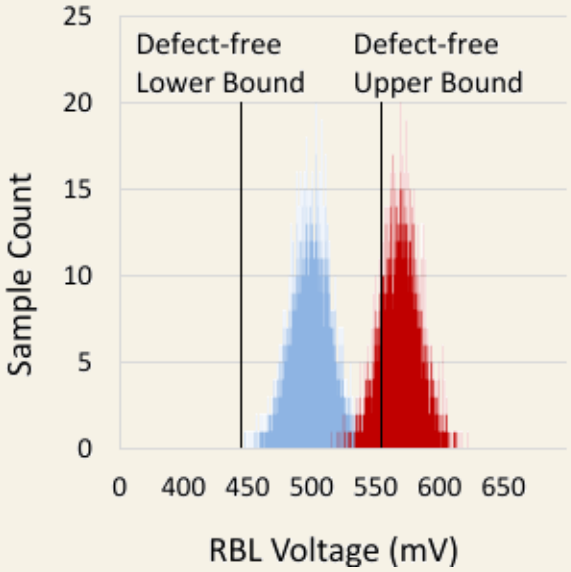
According to the Design for Testability (DfT) technique, in test mode only two memory cells per column (i.e. two rows) remain active each time, whereas in normal CIM operation all rows are activated simultaneously (all high register state).

Differential Testing & Diagnosis Technique

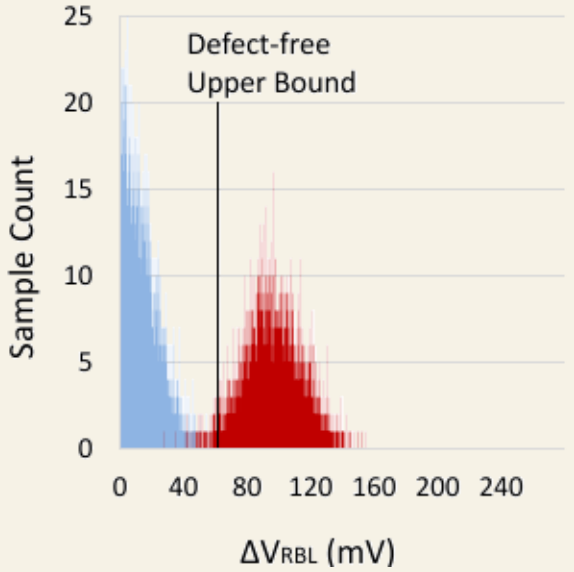
Transistor resistive short (TRS) defects



12T SRAM-based CIM



16.3% overlap



2.3% overlap

85.9% overlap reduction

The differences between V_{RBL} responses of adjacent pairs within the same column are calculated and differential comparisons are performed.

Analog JTAG (std. 1149.4) support.

DfT for SRAM-based CIMs

Equivalent Defects	*Max Detectable Resistance*	*Localization*
TRS2, TRS5, TRS7, TRS11	7.5 kΩ	Yes
TRS1, TRS6, TRS8, TRS12	8.7 kΩ	Yes
TRS3, TRS4, TRS9, TRS10	1.8 kΩ	Yes
TRS13, TRS17	11.4 kΩ	Yes
TRS14, TRS18	25 kΩ	No

The DfT assisted differential Test and Diagnosis methodology achieves 100% defect detection and 87.5% defect localization coverage.

Outline

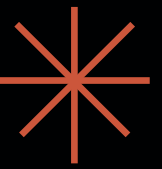
Computing in Memory and Reliability:

- Testing Challenges
  - Defect mechanisms should be defined
  - Fault Models should be developed
  - Test Generation tools are required
  - Testing Algorithms, Design for Testability & Design for Diagnosis techniques are mandatory
  - Testing Standards are needed
  - On-Line Testing solutions should be proposed
- Aging Challenges
  - Emerging Memories Aging Models should be developed
  - Aging Alleviation techniques should be explored
- Process Variation Challenges
- Radiation induced Single Event Effect (SEE) Challenges
  - Radiation Hardening techniques may be crucial for system reliability

Lifetime Enhancement for RRAM-based Computing-In-Memory Engine Considering Aging and Thermal Effects

S. Zhang, G.L. Zhang, B. Li, H. Li and U. Schlichtmann

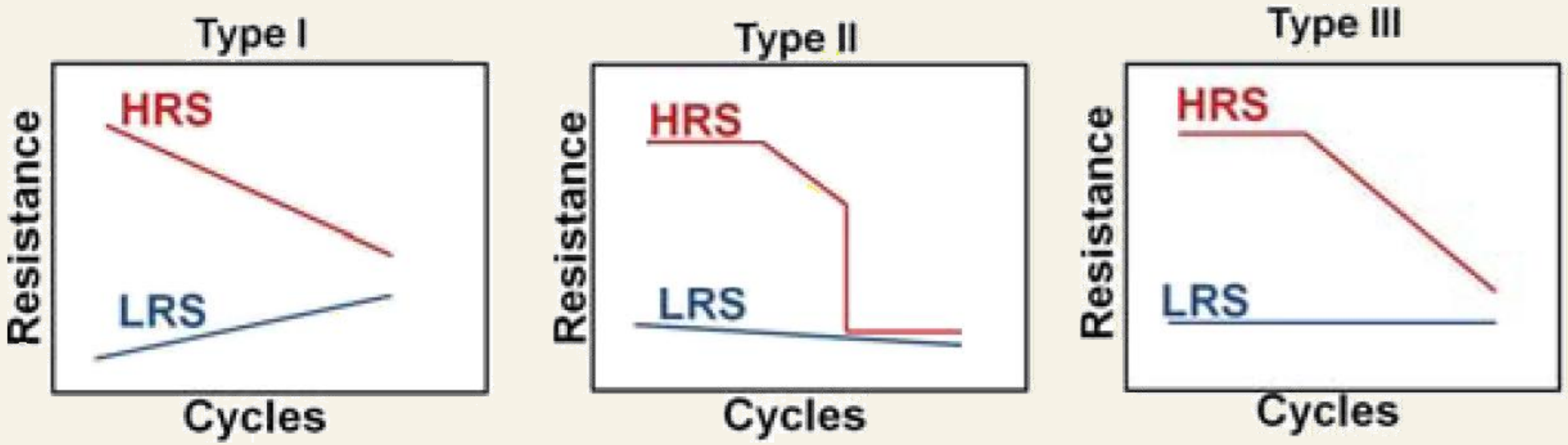
IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), 2020.



Aging & Thermal Effects Alleviation

Aging Effects in RRAM-based CIM

RRAM Cell Aging Types

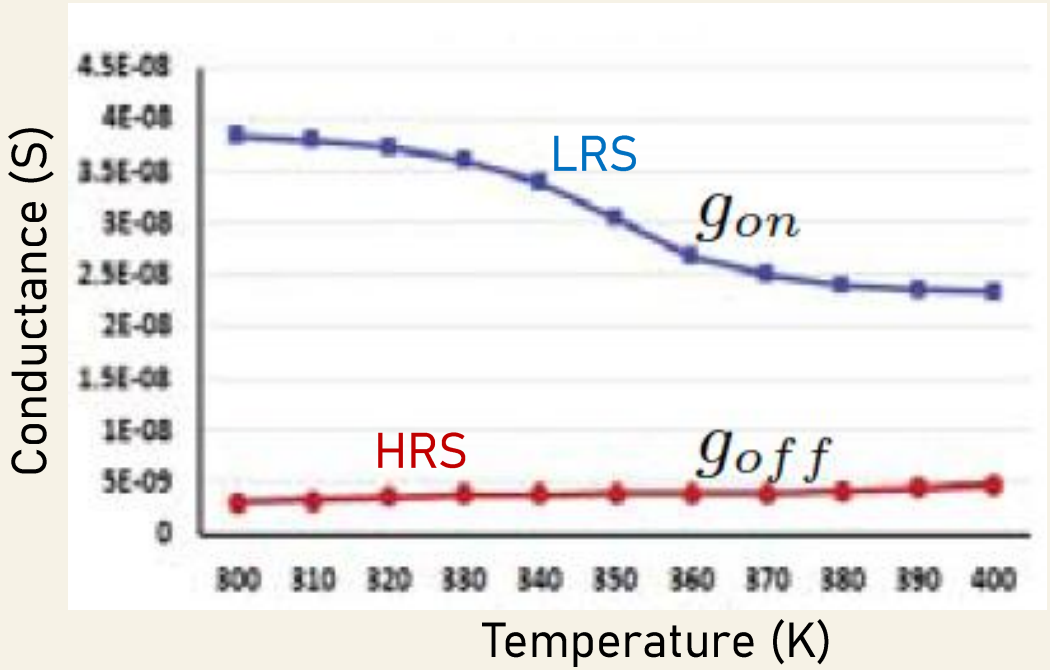


During programming, a high voltage is applied to RRAM cells for many cycles, so that the resistance bounds change and the usable resistance range decreases !

Programming an aged RRAM cell assuming the fresh resistance range the target resistance may fall outside of the aged resistance range. Consequently, the inference accuracy may be lower than expected.

Thermal Effects in RRAM Crossbars - I

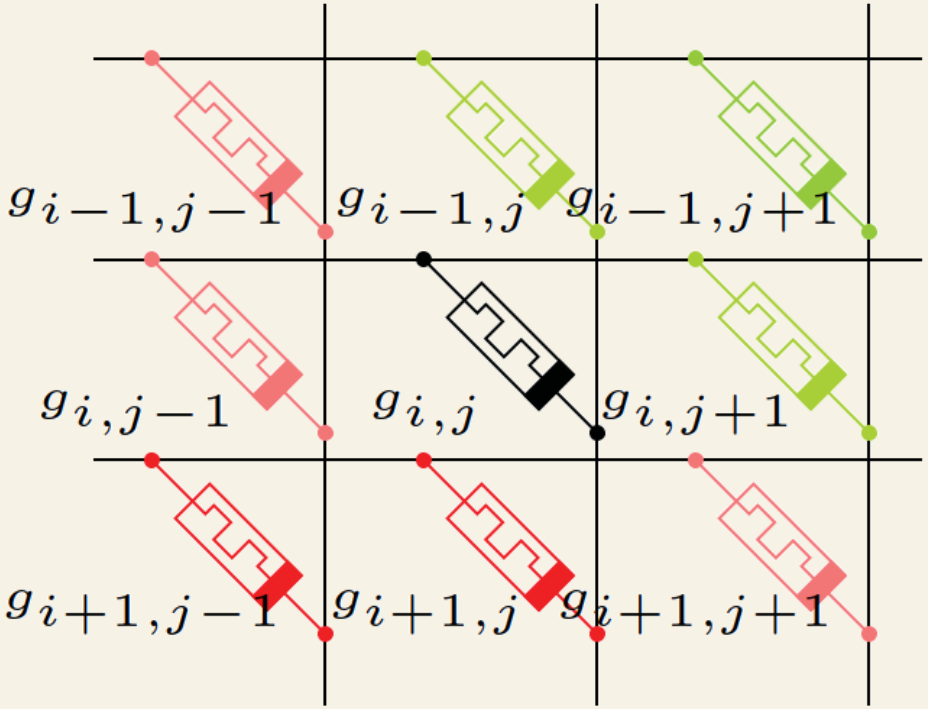
The high voltages applied onto RRAM cells during programming and tuning generate large currents though RRAM cells, so that their internal temperature may rise, resulting in conductance changes.



Due to the linear relationship between conductances of RRAM cells and the weights in the neural network, the conductances may differ significantly from the target one under different temperatures, resulting in accuracy degradation.

Thermal Effects in RRAM Crossbars - II

During online tuning, the temperature of RRAM cells may be high and uneven. As a result, the conductances of RRAM cells drift.



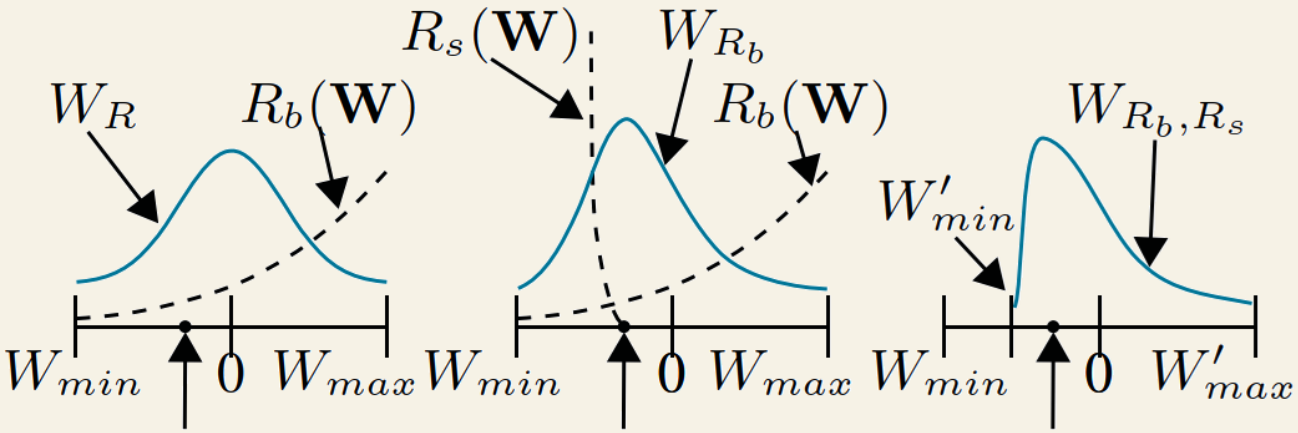
At the end of tuning, the accuracy is actually evaluated with conductances under high temperature, whose values and distribution may differ significantly from the scenario when the crossbar is used for inference, where the applied voltage is normally much lower.

Consequently, the accuracy of inference also deviates from that during tuning, once again resulting in accuracy degradation.

Thermal issues aggravate aging, which in turn reduce the overall lifetime of a RRAM crossbar.

Mitigation of Aging & Thermal Effects - I

Balancing aging stress among RRAM cells



preselected reference weight W_r

To deal with the aging and thermal effects, it is essential to control RRAM cells' conductances, which is equivalent to manage neural network weights before mapping. A cost function is used:

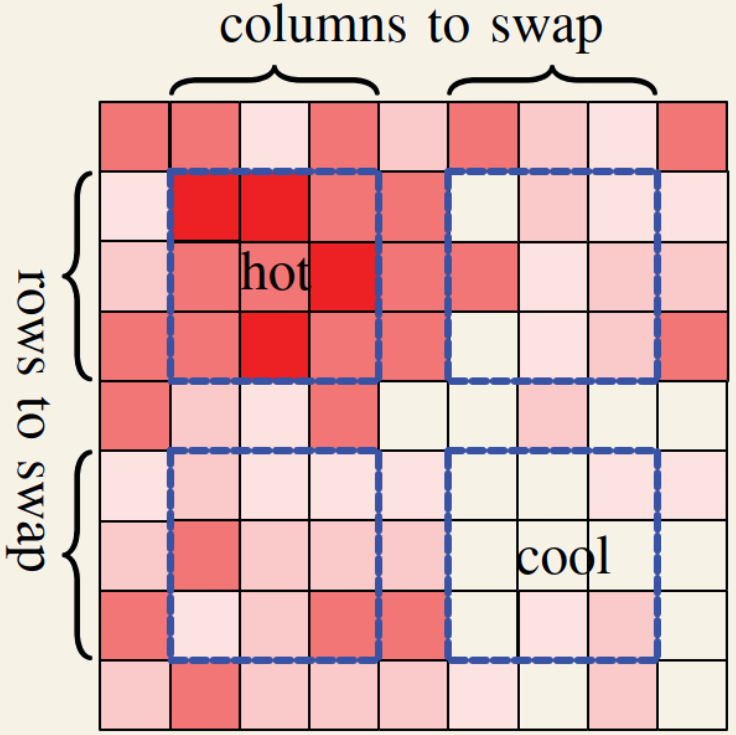
$$Cost = C(W) + R_b(W) + R_s(W)$$

After penalization $R_b(W)$, RRAM cells with aging and thermal issues have smaller weights, which will be mapped to lower conductance, reducing aging and thermal effects. Overall, the distribution of the weights are shifted towards a preselected reference weight, W_r.

With $R_s(W)$ penalty, small weights are concentrated towards W_r. This mapping can lead to less aging due to small currents flowing through RRAM cells during programming-tuning.

Mitigation of Aging & Thermal Effects - II

Row-Column Swapping



To deal with the residual unevenness of aging and thermal effects, a remapping of the weights according to the aging status and temperature of the RRAM cells across the crossbar with row-column swapping is proposed.

An RRAM cell's temperature is estimated by adding a given proportion of the internal temperature values of neighboring RRAM cells onto its own internal temperature. The internal temperature of a cell is evaluated from the applied voltage and its conductance.

To balance aging and thermal effects, the hottest sub-matrix and the coolest sub-matrix are mixed in such a way that the temperature distribution on the crossbar becomes even and the large conductances are mapped onto less-aged RRAM cells.

Accuracy and Lifetime Comparisons

		Accuracy Comparison		Lifetime Comparison		
	Dataset	NT	MT	NT+T	MT+T	MT+RCST
SNN	MNIST	91.40%	89.44%	1×	12×	45×
LeNet-5	Cifar10	76.08%	72.61%	1×	7×	89×
ResNet-20	Cifar100	67.35%	63.12%	1×	26×	453×

- NT = normal training
- MT = modified training
- NT+T = normal training and on-line tuning
- MT+T = modified training and on-line tuning
- MT+RCST = modified training and row-column swapping and on-line tuning

- Maximal accuracy drop is around 4%.

An Aging-Aware CMOS SRAM Structure Design for Boolean Logic In-Memory Computing

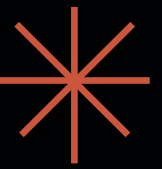
W. Chang, Y-G. Chen, P-Y. Huang, J-F. Li

IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT), 2021.

Aging alleviation technique for 8T IMC SRAMs

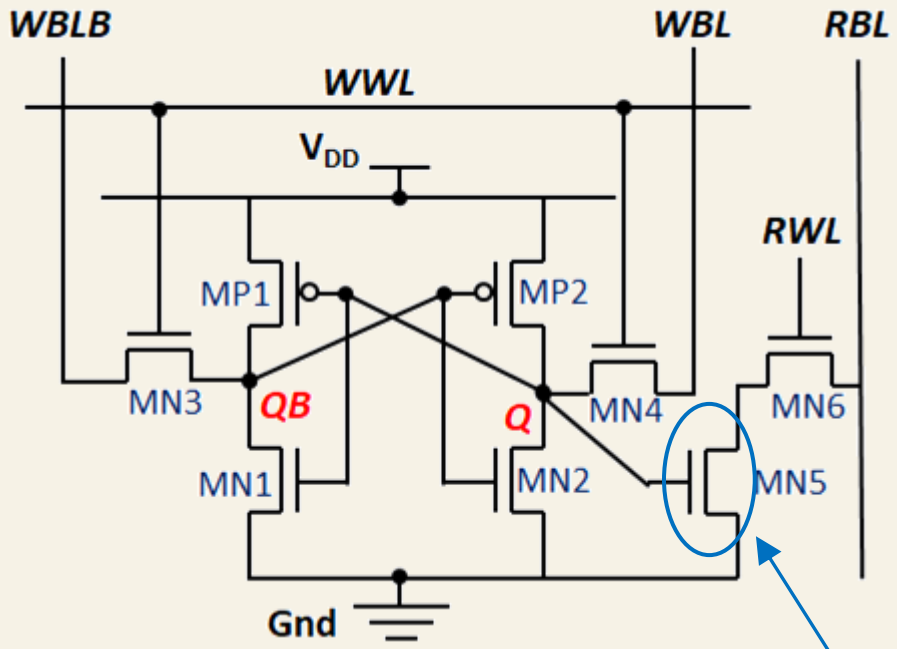
H.M. Dounavi, Y. Tsiatouhas

Emerging Technology Conference: Edge Intelligence (ETCEI), 2023.



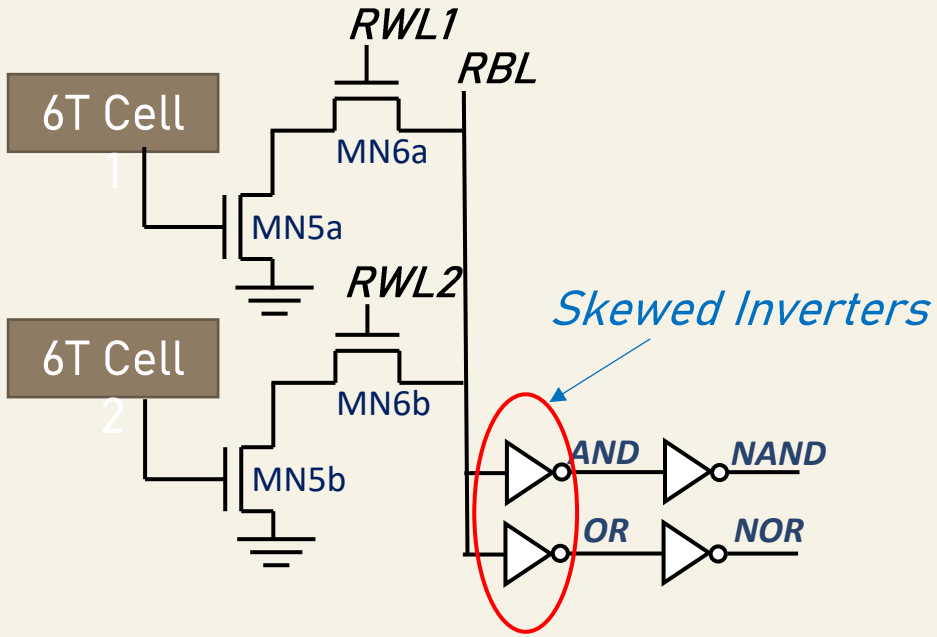
Aging Alleviation

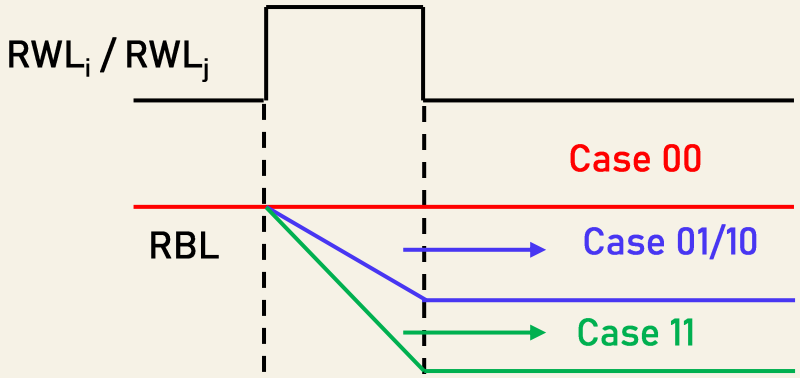
8T SRAM CIM for Logic Operations



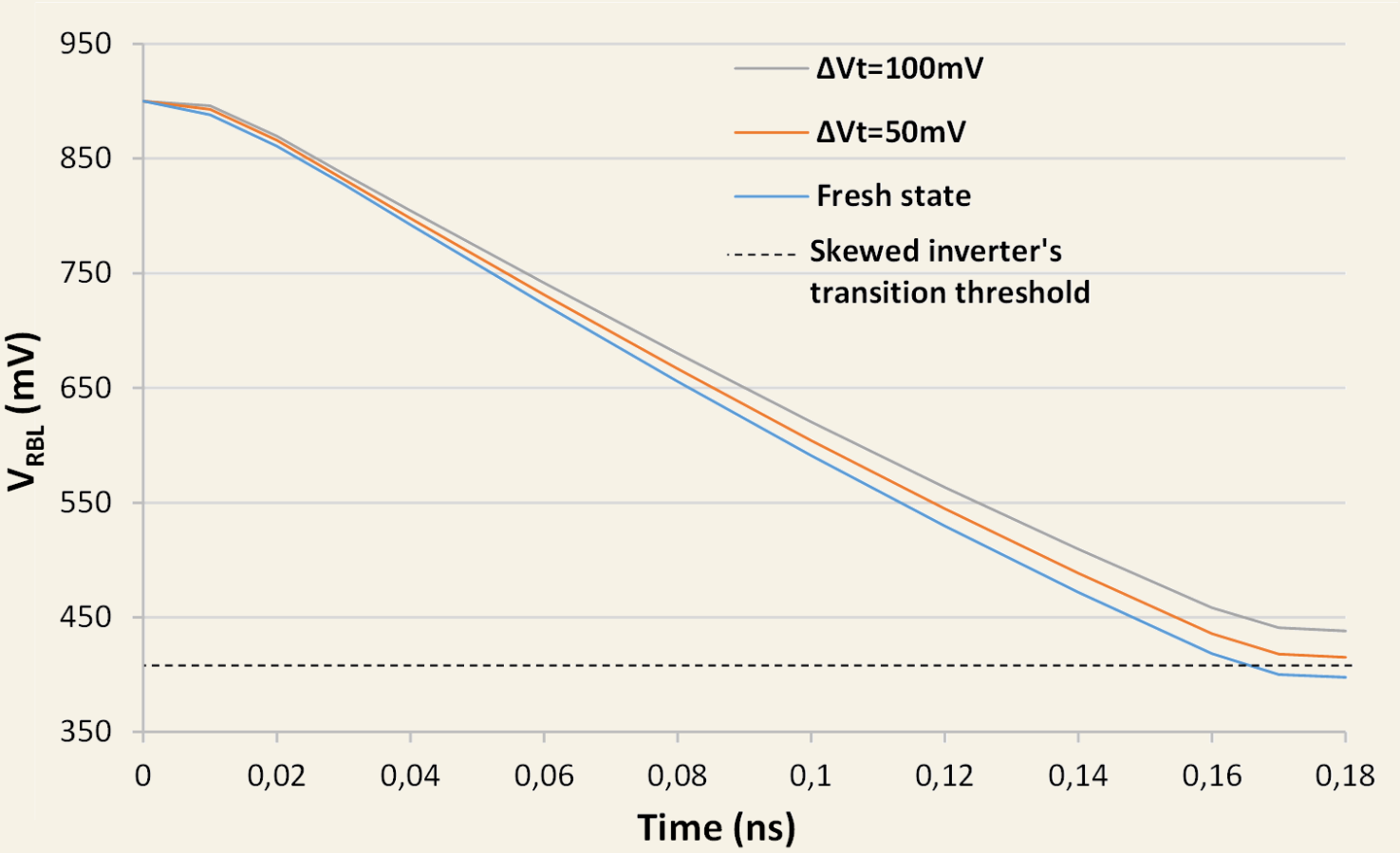
8T SRAM-based logic CIM

Aging vulnerable and computation critical transistor



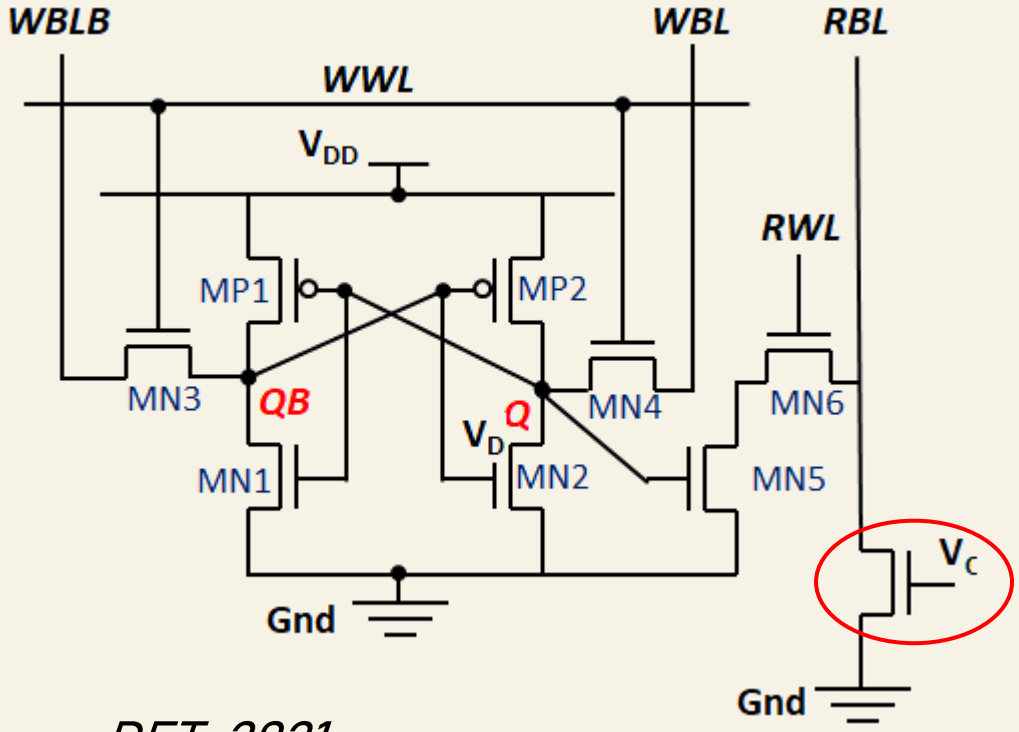


BTI Aging Impact on CIM Operation



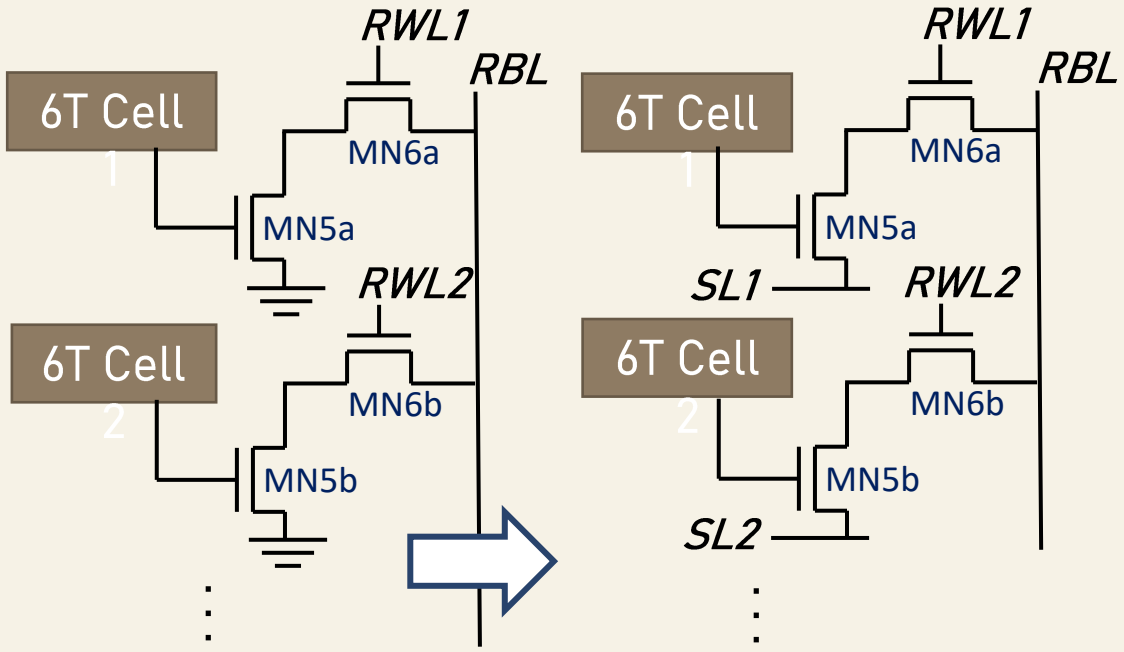
Under aging, V_{RBL} violates the inverter's transition threshold voltage, and an error is generated.

BTI Aging Alleviation Techniques



Chang, DFT-2021

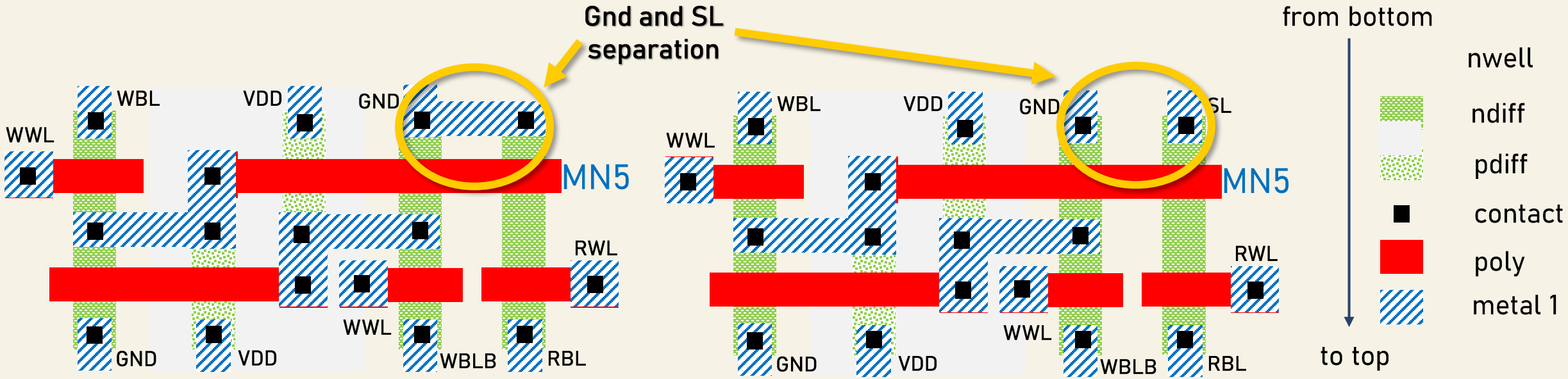
- Addition of a clamp transistor.
- Aging is not uniform in the array.
- Aging monitoring circuitry is required.



Dounavi, ETCEI-2023

- Addition of an extra source line SL.
- Application of positive bias V_{SL}>0 (eg. V_{dd}) on SL lines when meory is inactive. So, V_{GS}< V_{dd}.

Design for Reliability



- Each SL line is shared by all MN5 transistors of two consecutive rows.
- No additional silicon area cost at memory cell level.
- Up to 14% silicon area overhead at memory array level due to routing requirements of higher-level power supply metal lines.
- Dynamic power consumption is increased by 16.9%.
- No speed performance degradation.

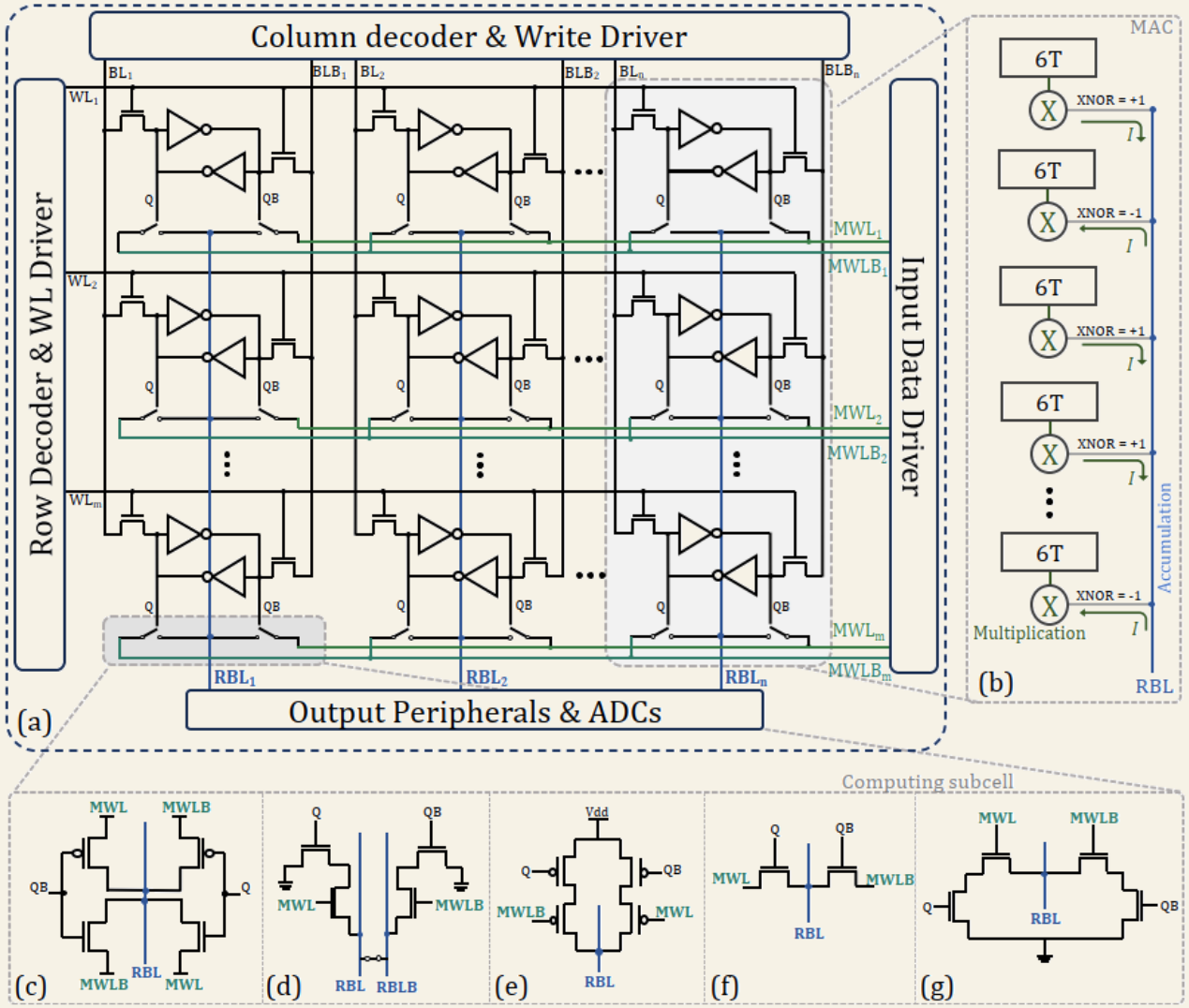
BTI Aging Analysis and Mitigation for Differential Input In-Memory Computing SRAMs

C. Dilopoulou and Y. Tsiatouhas

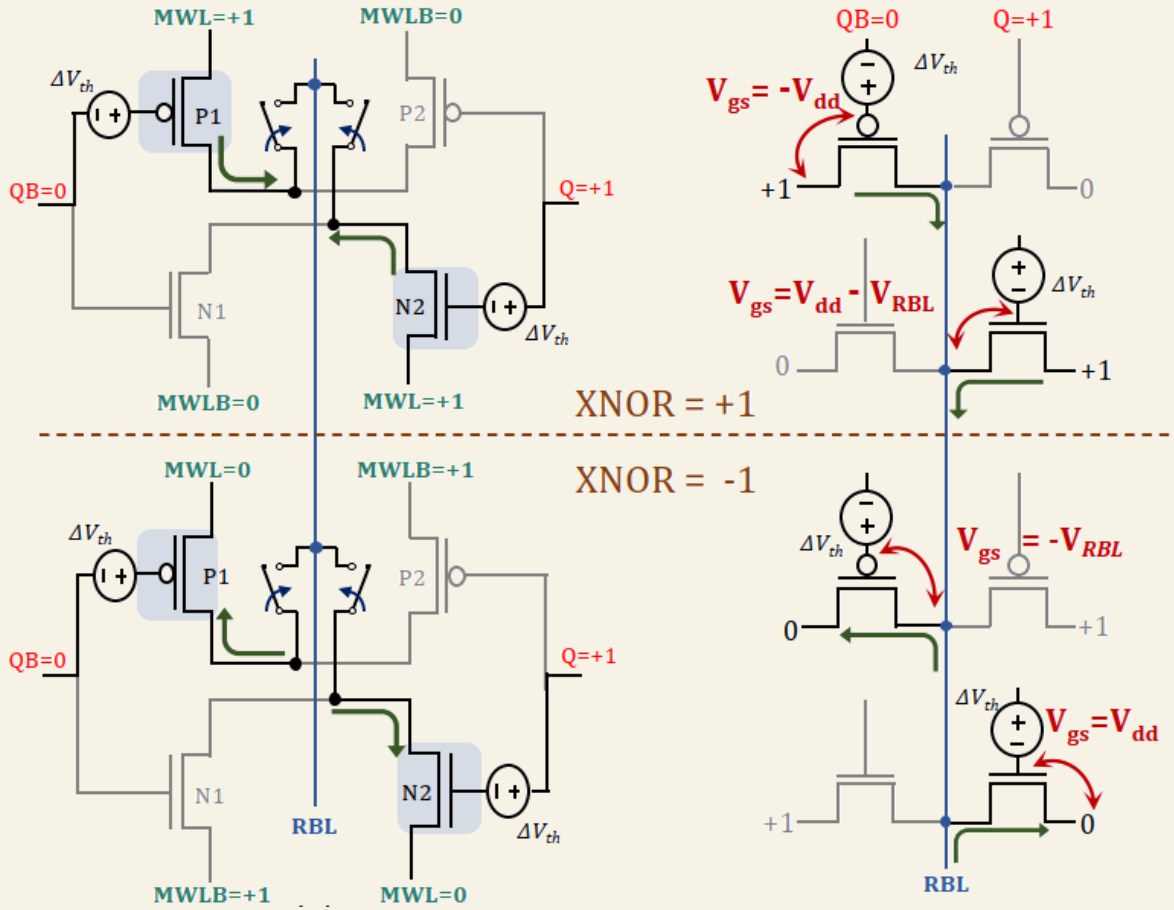
IEEE Transactions on VLSI Systems, vol. 33, no. 9, pp. 2570-2579, 2025.

Aging Alleviation under Process Variations

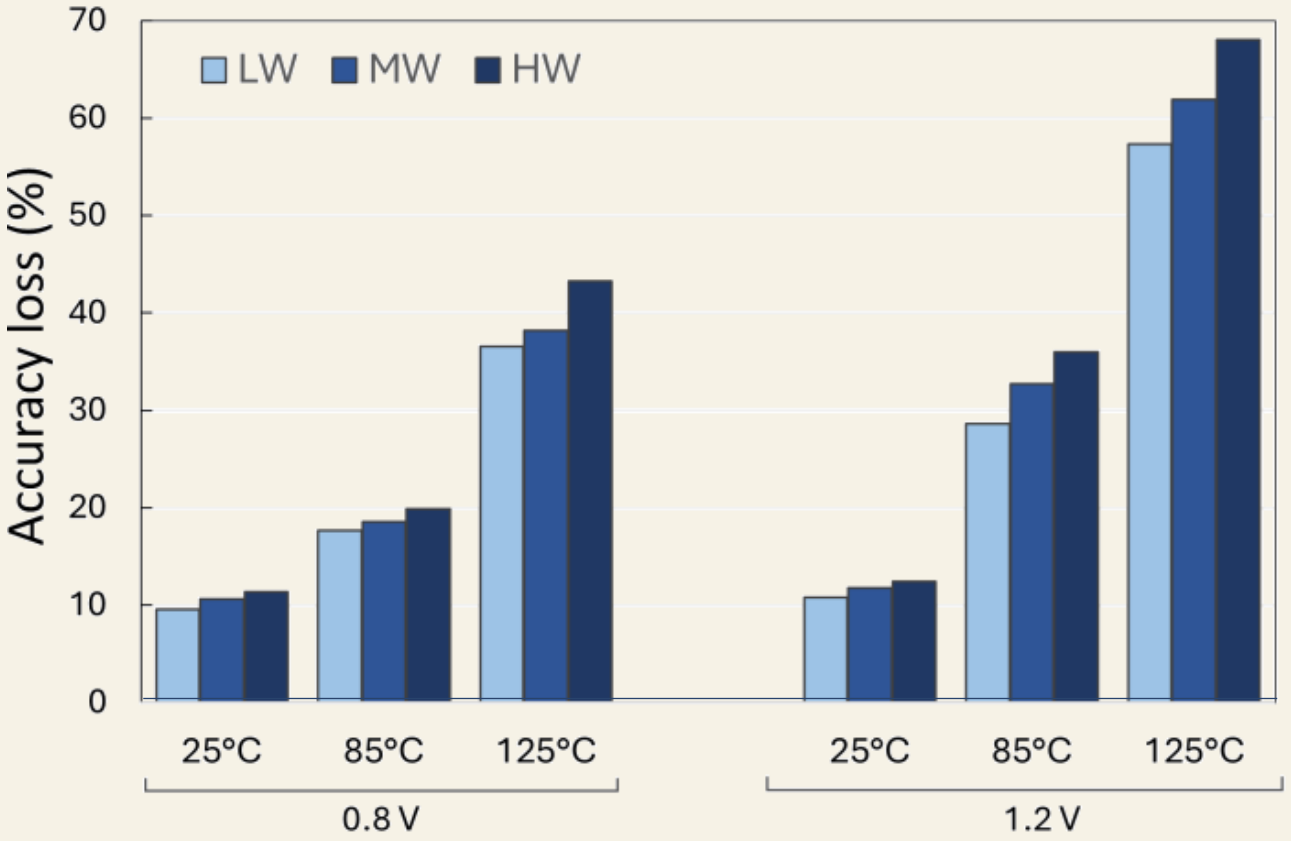
SRAM-based CIM



12T SRAM-based CIM

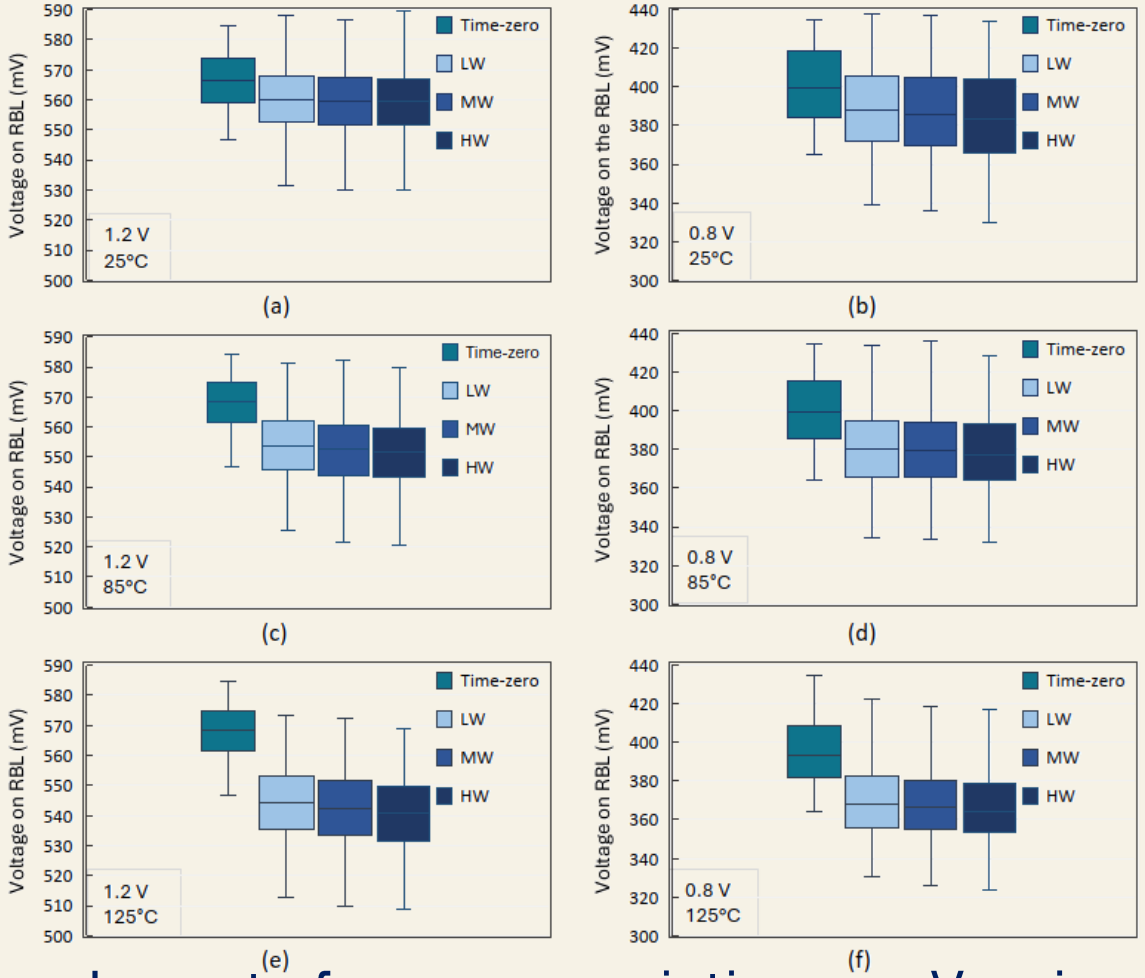


BTI Aging Influence



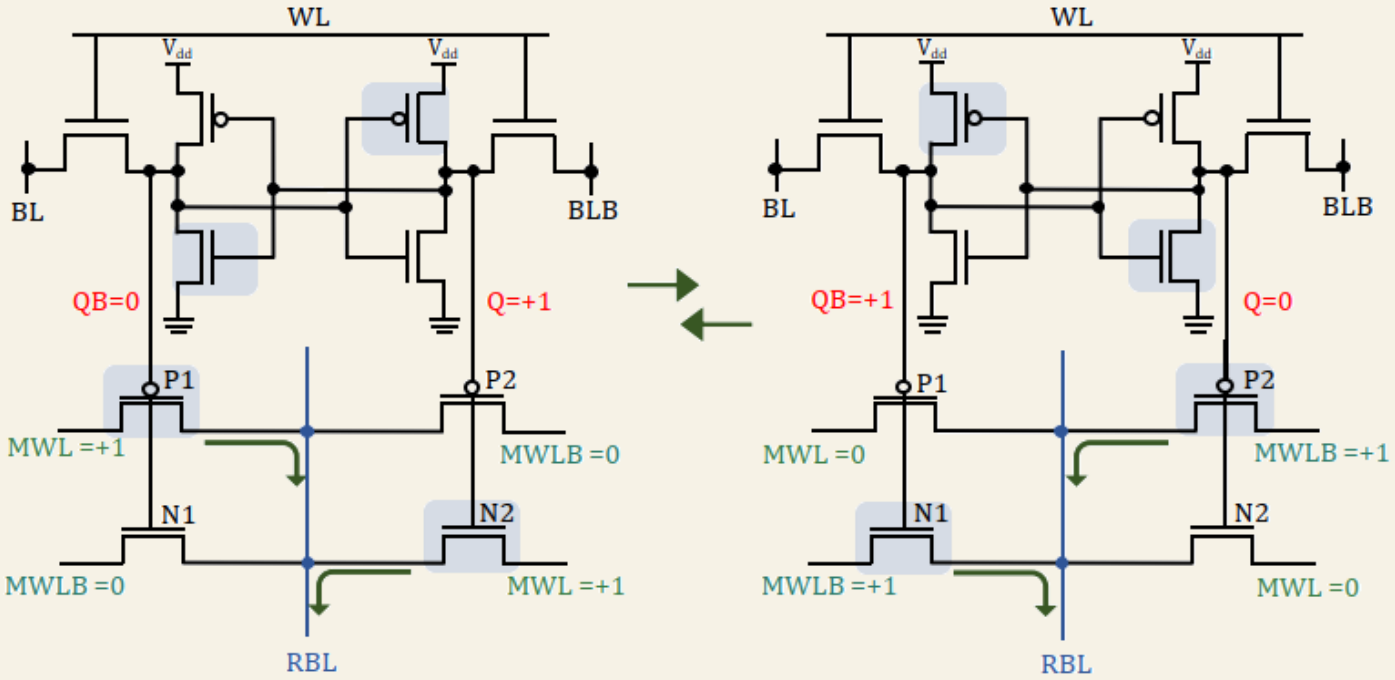
Aging impact on accuracy under different operating conditions, after 3 years of operation.

V_{RBL} Distribution



Impact of process variations on V_{RBL} in time-zero and after 3 operating years

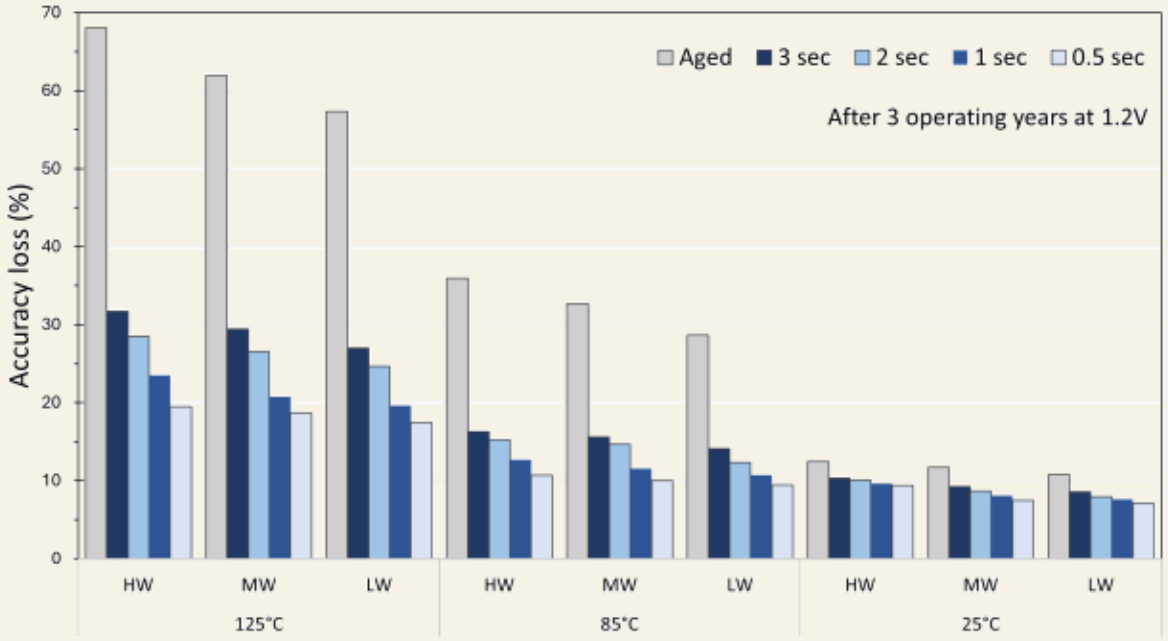
Stress Mitigation for Aging Alleviation

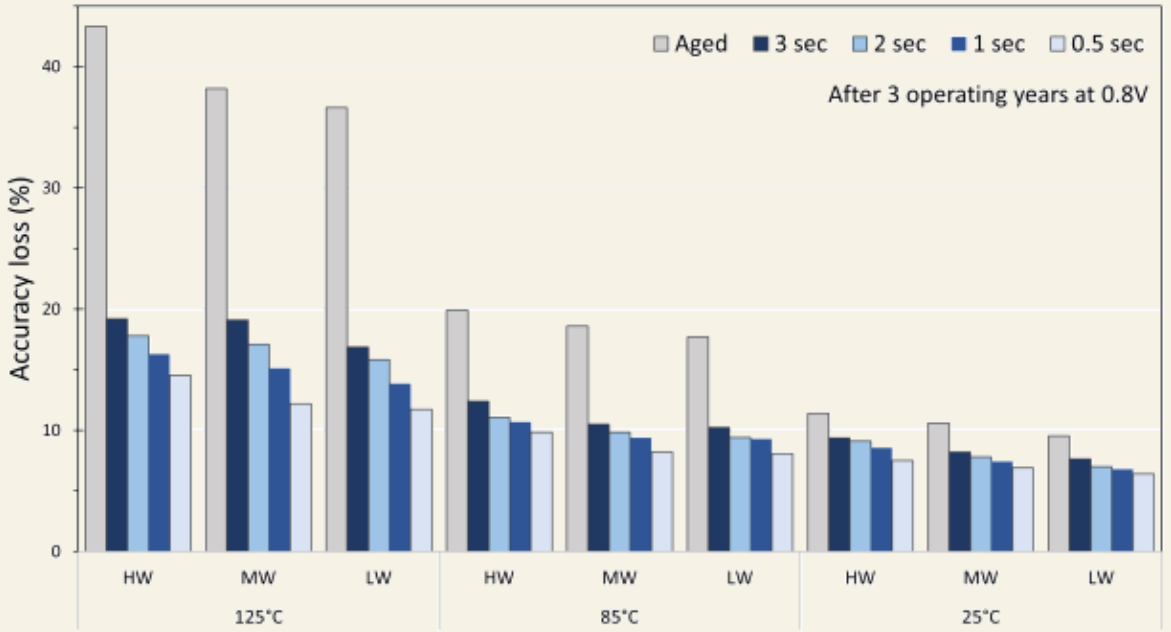


Periodical alternation of weights and data with their complementary values, ensures that the initially active transistors under permanent stress will enter a recovery state, while transistors in a relaxation state are activated for the computations instead.

The same stands for the transistors of the memory subcells. Hence, both memory and computing subcells will benefit by the application of successive stress-recovery cycles, since the transistors will undergo substantially reduced levels of aging over time with respect to the permanently activated transistors in the unmitigated aging case.

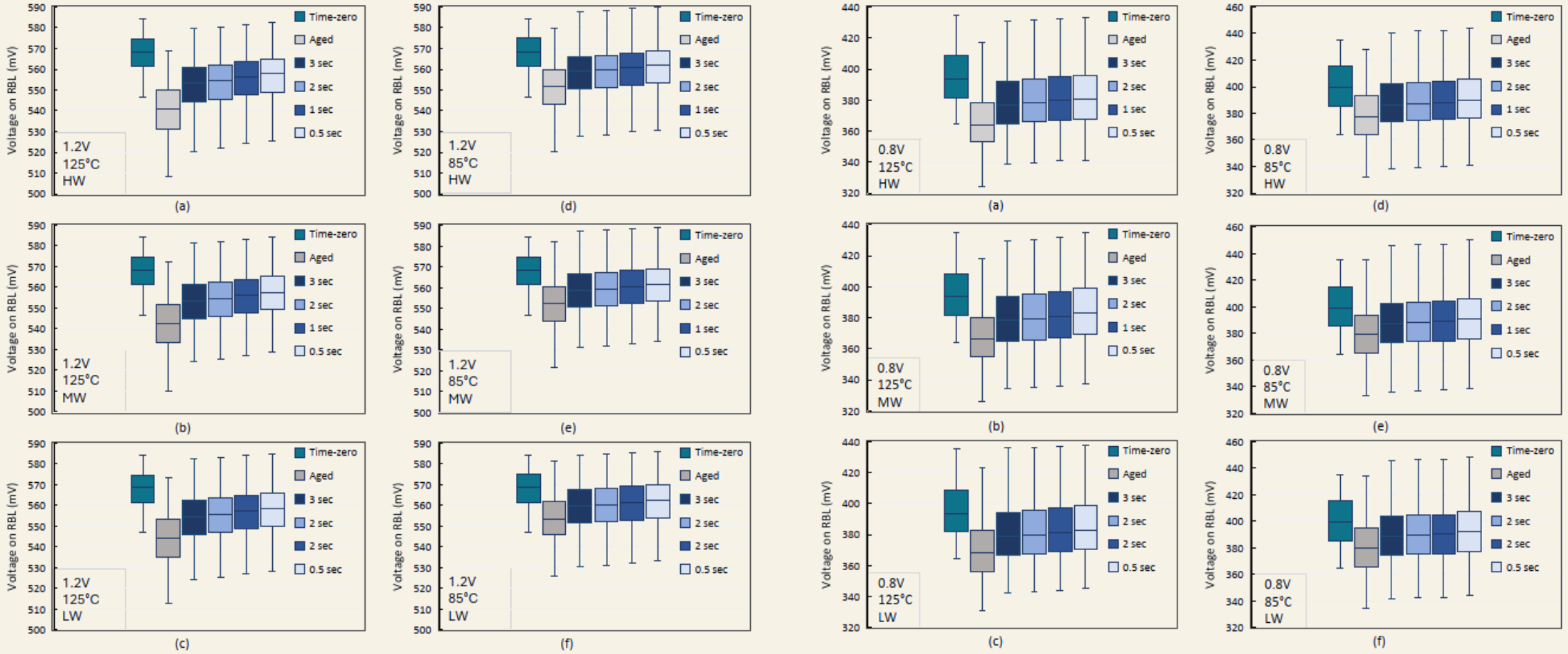
Accuracy Recovery by Aging Mitigation





Reduction in accuracy loss with the mitigation technique applied at various time periods for different power supplies, temperatures and workloads, for 3 years of operation.

V_{RBL} Voltage Recovery by Aging Mitigation



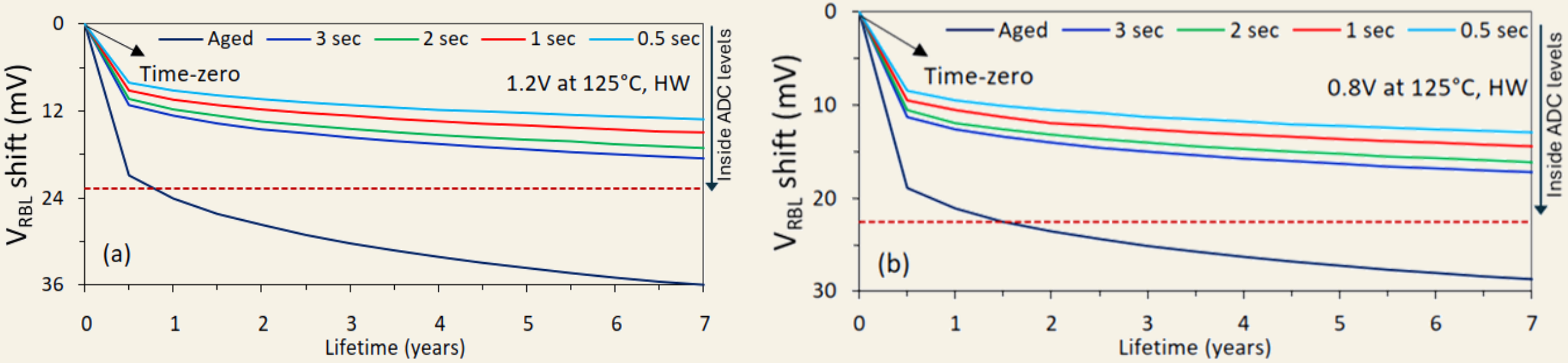
V_{RBL} distribution in time-zero, and after 3 operating years without and with mitigation applied in various time periods, for different power supplies, temperatures and workloads.

Reliable Operating Lifetime Extencion

Accuracy loss over 7 years, without mitigation & by applying the mitigation technique every 0.5s.

HW case

Years	1.2V						0.8V					
	125°C		85°C		25°C		125°C		85°C		25°C	
	Aged	Mitigated	Aged	Mitigated	Aged	Mitigated	Aged	Mitigated	Aged	Mitigated	Aged	Mitigated
1	52.6%	14.3%	28.0%	8.6%	10.3%	8.4%	33.8%	11.4%	15.8%	9.0%	9.7%	6.9%
3	68.0%	19.4%	35.9%	10.7%	12.4%	9.3%	43.3%	14.5%	19.9%	9.8%	11.4%	7.5%
5	77.5%	22.2%	40.2%	12.3%	16.8%	9.5%	49.2%	15.3%	22.4%	10.1%	12.6%	7.6%
7	82.7%	23.9%	43.4%	13.1%	17.8%	9.6%	51.9%	16.3%	23.9%	10.8%	13.5%	7.9%



Seven years V_{RBL} shift evolution under aging without and with mitigation for (a) 1.2V and (b) 0.8V in the typical process conditions and high impact workload (HW) at 125°C.

Outline

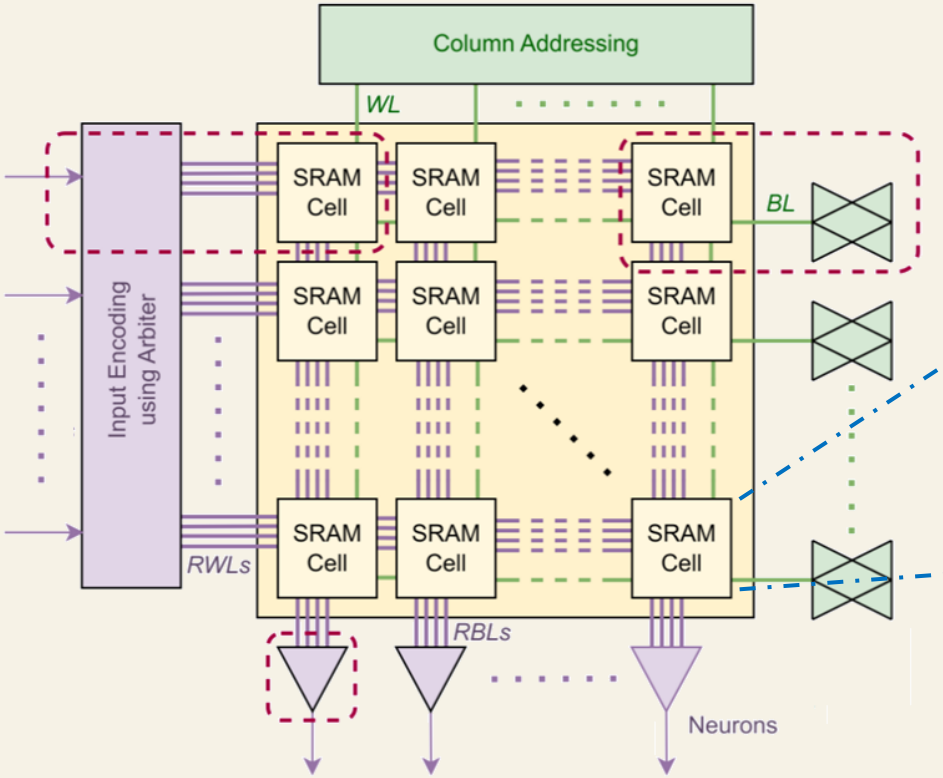
Computing in Memory and Reliability:

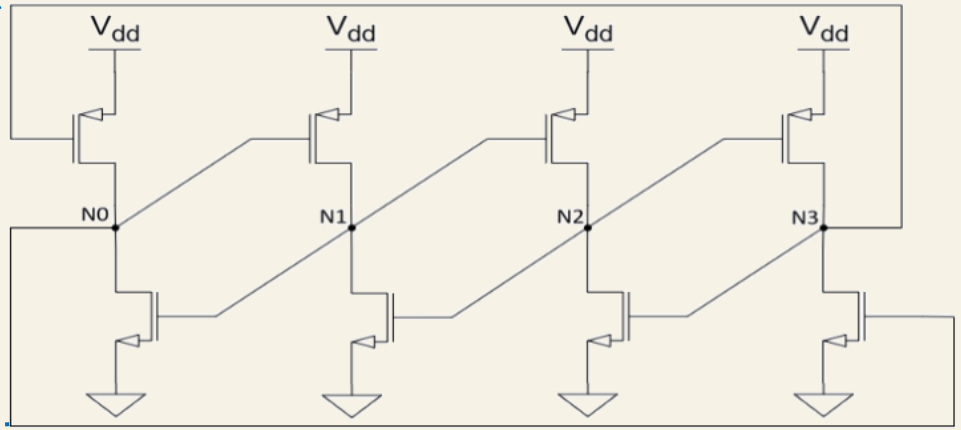
- Testing Challenges
  - Defect mechanisms should be defined
  - Fault Models should be developed
  - Test Generation tools are required
  - Testing Algorithms, Design for Testability & Design for Diagnosis techniques are mandatory
  - Testing Standards are needed
  - On-Line Testing solutions should be proposed
- Aging Challenges
  - Emerging Memories Aging Models should be developed
  - Aging Alleviation techniques should be explored
- Process Variation Challenges
- Radiation induced Single Event Effect (SEE) Challenges
  - Radiation Hardening techniques may be crucial for system reliability

Single Event Effects Tolerance

In emerging memories based CIM architectures, SEUs and SETs are not a serious problem.
For SRAM or DRAM based CIM architectures:

- ECC and TMR does not seem to be efficient solutions.
- Periodic weights refreshing can be a viable SEU Mitigation approach.
- Alternatively, for SRAM-based CIMs, SEU hardened memory cells (like DICE) can be exploited for radiation hardening.





DICE as the memory sub-cell

Conclusions

- CIM Testing
  - Defects Definition
  - Fault Models (Stuck-at-Faults, Write Disturb Faults, Read Disturb Faults, Slow Write Faults, Undefined Write Faults, Unknown Read Faults, Incorrect Read Faults, Retention Faults)
  - Testing Algorithms
  - Design for Testability – Testing Standards (e.g. IEEE P1832 3D IC oriented testing)
  - Design for Diagnosis
  - Built-In Self Test Techniques
  - On-Line Testing Techniques
- CIM Aging
  - Emerging Memories Aging Models
  - Aging Alleviation Techniques
- CIM Process Variations
- CIM SEE Influence & Mitigation

References – I

- T. Spyrou, A. Zografou, S. Hamdioui and A. Grebregiorgis, "*Fault Tolerant Design for Memristor-based AI Accelerators*," IEEE International Conference on Design, Test and Technology of Integrated Systems (DTTIS), 2024.
- S. Hamdioui, M. Taouil and N.Z. Haron, "*Testing Open Defects in Memristor-Based Memories*," IEEE Transaction on Computers, vol. 64, no. 1, 2015.
- E.A. Serlis, H. Xun, M. Taouil, S. Hamdioui, and M. Fieback "*Structural Testing of a RRAM-based AI Accelerator Core*," IEEE European Test Symposium (ETS), 2025.
- S.B. Mamaghani, P. Pal, M.B. Tahoori, "*A Dynamic Testing Scheme for Resistive-Based Computation-In-Memory Architectures*," IEEE Asia and South Pacific Design Automation Conference (ASP-DAC), 2024.
- S. Spyridonos, and Y. Tsiatouhas, "*Testing Algorithms for Hard to Detect Thermal Crosstalk Induced Write Disturb Faults in Phase Change Memories*," Design Automation and Test in Europe Conference (DATE), 2024.
- S. Zhang, G.L. Zhang, B. Li, H. Li and U. Schlichtmann, "*Lifetime Enhancement for RRAM-based Computing-In-Memory Engine Considering Aging and Thermal Effects*," IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS), 2020.

References – II

- W. Chang, Y-G. Chen, P-Y. Huang, J-F. Li, "*An Aging-Aware CMOS SRAM Structure Design for Boolean Logic In-Memory Computing*" IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems (DFT), 2021.
- H.M. Dounavi, Y. Tsiatouhas, "*Aging alleviation technique for 8T IMC SRAMs*," Emerging Technology Conference: Edge Intelligence (ETCEI), 2023.
- C. Dilopoulou and Y. Tsiatouhas, "*BTI Aging Analysis and Mitigation for Differential Input In-Memory Computing SRAMs*," IEEE Transactions on VLSI Systems, vol. 33, no. 9, pp. 2570-2579, 2025.

THANK YOU FOR YOUR ATTENTION