imw
2022
International Memory Workshop

Tutorial

14th International Memory Workshop (IMW)
May 15th 2022
Dresden, Germany

Platinum Sponsors

Gold Sponsors

Silver Sponsors

Agenda

Tutorials 09:00AM – 04:00PM

Organizers: Dirk Wouters (RWTH Aachen) and Thomas Mikolajick (NaMLab/TU Dresden)

Part 1 – Ferroelectric Memories 09:00AM – 12:00PM

Chairs: *Dirk Wouters (RWTH Aachen)*
Katherine Chiang (TSMC)

Laurent Grenouillet **CEA-Leti** 09:00AM – 10:00AM

“Ferroelectric Random Access Memory (FeRAM)”

Halid Mulaosmanovic, **GlobalFoundries** 10:00AM – 11:00AM

“Ferroelectric Field Effect Transistors (FeFET)”

Shosuke Fujii **Kioxia** 11:00AM – 12:00PM

“Ferroelectric Tunnel Junction (FTJ)”

Part 2 – 3D Memories – Security Aspects of Memories 02:00PM – 04:00PM

Chairs: *Antonio Arreghini (imec)*
Thomas Mikolajick (Namlab/TU Dresden)

Onur Mutlu **ETH Zurich** 02:00PM – 03:00PM

“Security aspects of DRAM”

Swaroop Ghosh **Penn State University** 03:00PM – 04:00PM

“Security Aspects in Nonvolatile Memories”

Laurent Grenouillet
CEA-Leti

Laurent Grenouillet received the Engineer degree in physics in 1998 from the National Institute of Applied Sciences (INSA) in Lyon, France, and the PhD degree in electronic devices in 2001 for his work on the optical spectroscopy of diluted nitrides grown on GaAs substrates. After a post-doctoral position in the field of Molecular Beam Epitaxy, he joined CEA-Leti in 2002 and worked on GaAs-based VCSELs emitting in the 1.1-1.3µm range and single photon sources with quantum dots. In 2006, he joined the Silicon Photonics group where he developed CMOS compatible hybrid III-V on silicon lasers. In 2009, he joined IBM Alliance in Albany as a Leti assignee to contribute to the development of FDSOI technology. Within Albany state-of-the-art facilities, he extensively worked on device integration to improve performance of FDSOI devices (28nm and 14nm node). Back in France at CEA-LETI in 2013, he focused on the performance boosters for the 10nm node FDSOI technology, and took part to the FDSOI technology transfer to Global Foundries (22FDX) in 2015. During that period he joined the Advanced Memory Device Laboratory at CEA-Leti. His current research interests include resistive switching memory devices and selectors, and ferroelectric HfO_2-based memories. Laurent Grenouillet authored or co-authored over 80 papers (conferences and journals) and has filed over 40 patents. He serves as committee member of Solid-State Devices and Materials (SSDM) conference.

1T-1C FERROELECTRIC RAM

Dr. L. Grenouillet, CEA-Leti | 14th International Memory Workshop, 2022 | Tutorial | 2022-05-15
laurent.grenouillet@cea.fr

OUTLINE

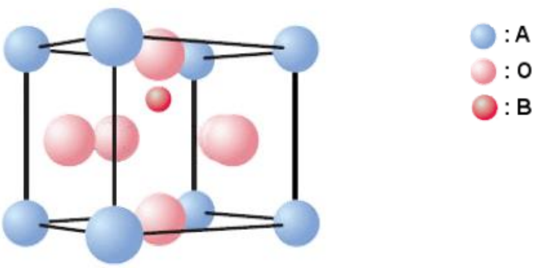
- Ferroelectricity basics
- Ferroelectric HfO_2: a change of paradigm for NVM
- 1T-1C FeRAM arrays: basics
- HfO_2-based MFM capacitors integrated above CMOS
- HfO_2-based 1T-1C FeRAM arrays: performance overview
- Scalability: challenges and perspectives

OUTLINE

- **Ferroelectricity basics**
- **Ferroelectric HfO_2: a change of paradigm for NVM**
- **1T-1C FeRAM arrays: basics**
- **HfO_2-based MFM capacitors integrated above CMOS**
- **HfO_2-based 1T-1C FeRAM arrays: performance overview**
- **Scalability: challenges and perspectives**

FERROELECTRIC MATERIAL: SOME BASICS

CRYSTALLOGRAPHY

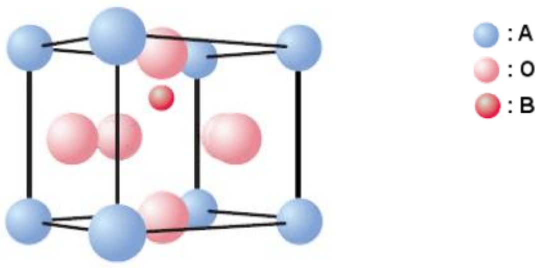


FERROELECTRICS:

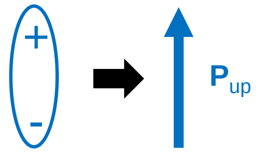
- Crystallize in a non-centrosymmetric phase → ELECTRIC DIPOLE

FERROELECTRIC MATERIAL: SOME BASICS

CRYSTALLOGRAPHY



ELECTRIC DIPOLE

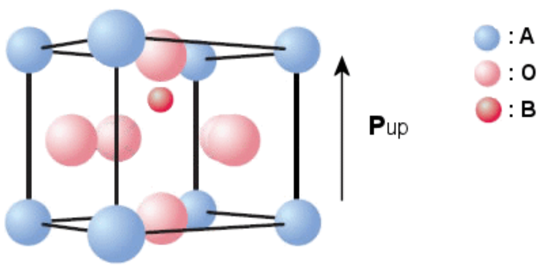


FERROELECTRICS:

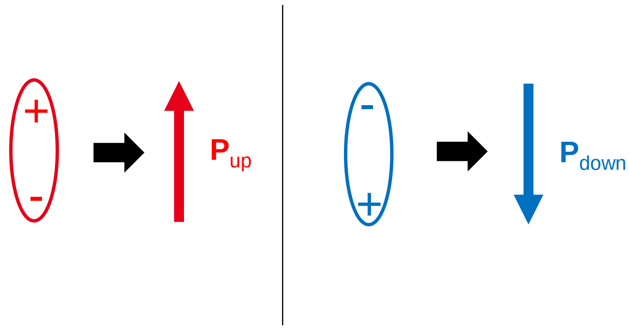
- Crystallize in a non-centrosymmetric phase → ELECTRIC DIPOLE

FERROELECTRIC MATERIAL: SOME BASICS

CRYSTALLOGRAPHY



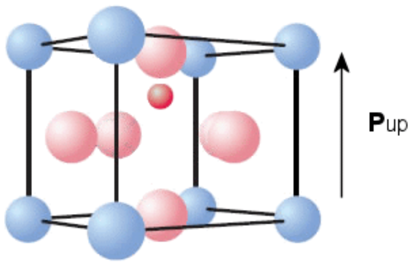
ELECTRIC DIPOLE



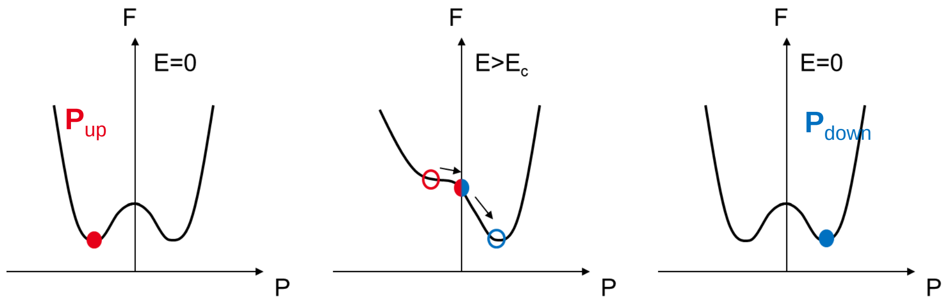
FERROELECTRICS:

- Crystallize in a non-centrosymmetric phase → ELECTRIC DIPOLE
- Show a SPONTANEOUS & SWITCHABLE **ELECTRIC POLARIZATION**

CRYSTALLOGRAPHY



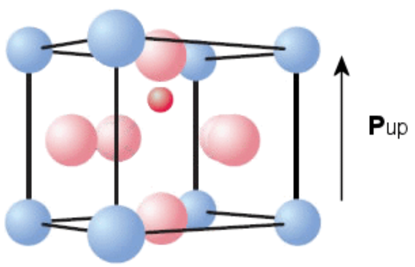
THERMODYNAMICS (Landau-Ginzburg-Devonshire)



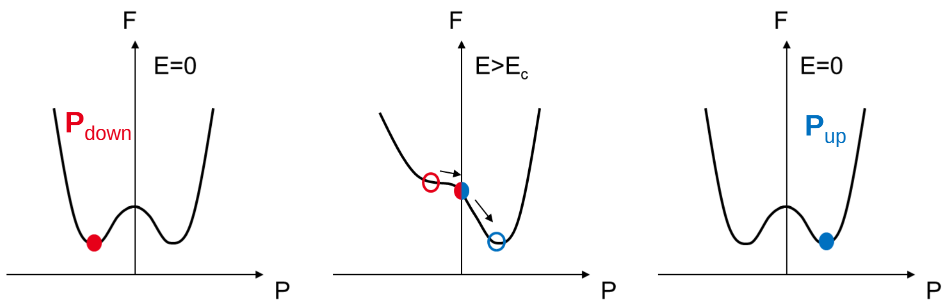
FERROELECTRICS:

2 **STABLE POLARIZATION STATES** that can be switched by an **ELECTRIC FIELD**

CRYSTALLOGRAPHY



THERMODYNAMICS (Landau-Ginzburg-Devonshire)



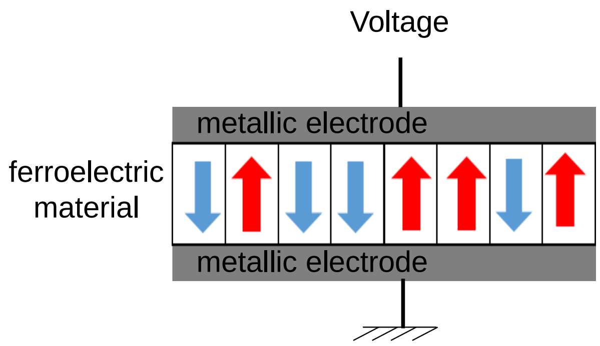
non-volatile memory

ultra low power

FERROELECTRICS:

2 **STABLE POLARIZATION STATES** that can be switched by an **ELECTRIC FIELD**

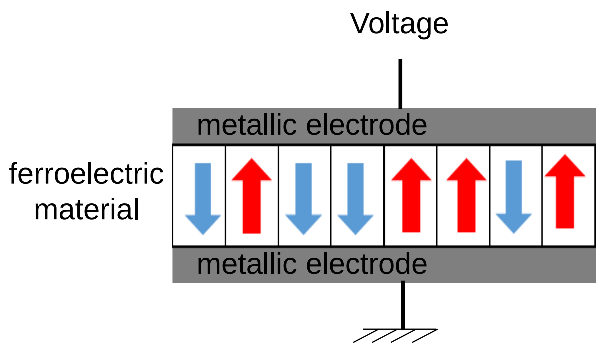
FERROELECTRIC MATERIAL: SOME BASICS

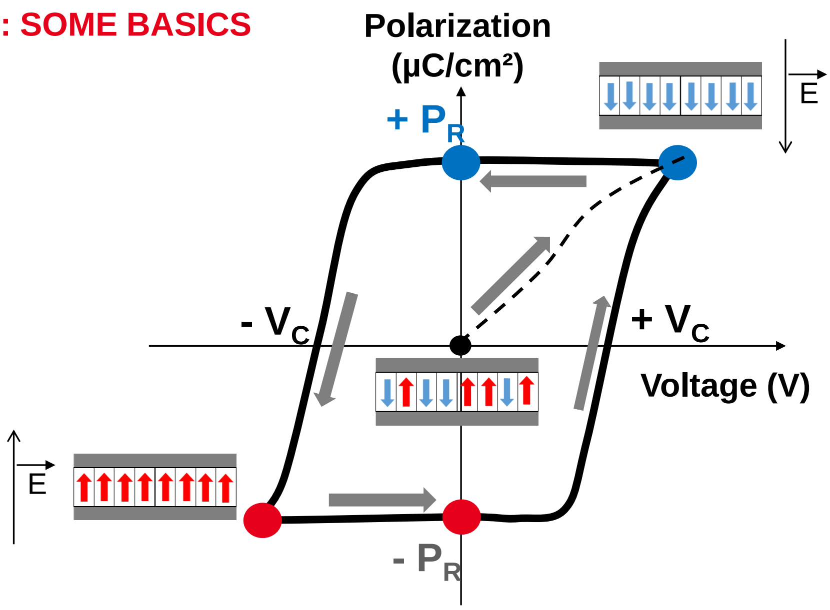


FERROELECTRICS:

2 **STABLE POLARIZATION STATES** that can be switched by an **ELECTRIC FIELD**

FERROELECTRIC MATERIAL: SOME BASICS





FERROELECTRICS:

2 **STABLE POLARIZATION STATES** that can be switched by an **ELECTRIC FIELD**

FERROELECTRICS: KEY DATES

- **1921: discovery of the elegant, fundamental physics of ferroelectricity by J. Valasek**

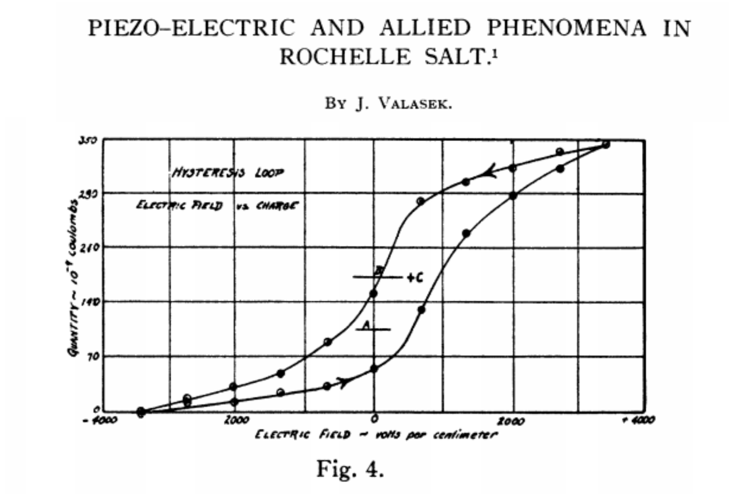
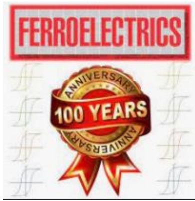


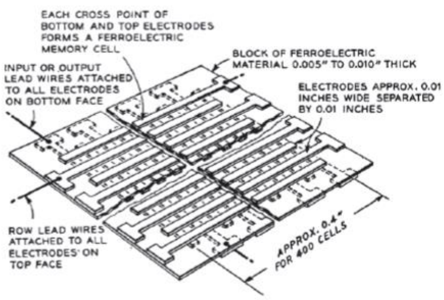
Fig. 4.



J. Valasek, Phys. Rev. 17, 475 (1921)

FERROELECTRICS: KEY DATES

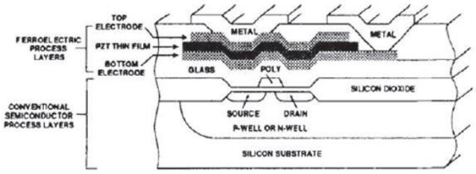
- **1921: discovery of the elegant, fundamental physics of ferroelectricity by J. Valasek**
- **1952: ferroelectric memory invented by Dudley Allen Buck (crosspoint arrays)**

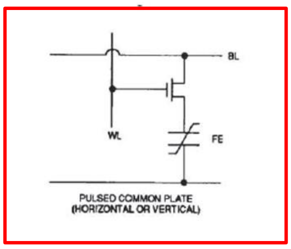


ferroelectric material: $BaTiO_4$

FERROELECTRICS: KEY DATES

- 1921: discovery of the elegant, fundamental physics of ferroelectricity by J. Valasek
- 1952: ferroelectric memory invented by Dudley Allen Buck (crosspoint arrays)
- 1990's: disturb issue solved by adding a select transistor (**1T-1C bitcell**)





4096-Bit FRAM® Memory

ferroelectric material:
$Pb(Zr_xTi_{1-x})O_3$ (PZT)
$Sr_{1-x}Bi_{2+y}Ta_2O_9$ (SBT)

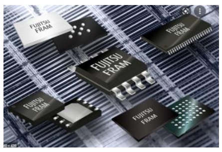
D. Bondurant, Ferroelectrics, (1990)
"The FeRAM is the world's first integrated 'nonvolatile static RAM'."

FERROELECTRICS: KEY DATES

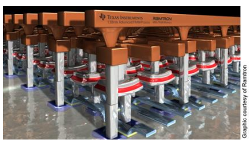
- 1921: discovery of the elegant, fundamental physics of ferroelectricity by J. Valasek
- 1952: ferroelectric memory invented by Dudley Allen Buck (crosspoint arrays)
- 1990's: disturb issue solved by adding a select transistor (1T-1C cell)
- Today: main Ferroelectric Random Access Memory (FeRAM) players

INFINEON

LAPIS (ROHM)



FUJITSU



TEXAS INSTRUMENT

SKhynix

Fujitsu has sold more PZT FRAM chips than any emerging memory in history – over 4 billion units since 1999!

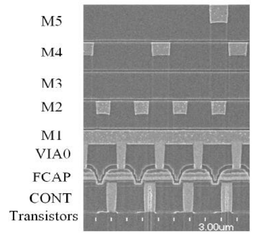
FERROELECTRIC MEMORY (FERAM) MAIN INTERESTS

ATTRACTIVE FEATURES:
- **Ultra low power**: < 10fJ/bit
- **Fast**: < 100ns
- **Low voltage**: < 3V
- **High endurance**: Up to 10^{15} cycles

Reliability of Ferroelectric Random Access Memory Embedded within 130nm CMOS

J. Rodriguez, K. Remack, J. Gertas, L. Wang, C. Zhou, K. Boku, J. Rodriguez-Latorre, K. R. Udayakumar, S. Summerfelt, T. Moise
Texas Instruments Inc.
Dallas, Texas USA
jrz@ti.com

D. Kim, J. Groat, J. Eliason, M. Depner, F. Chu
RAMTRON International Corporation
Colorado Springs, Colorado USA



BUT

DOWNSIDES:
- PZT is **not CMOS friendly** (lead)
- PZT is **not scalable**

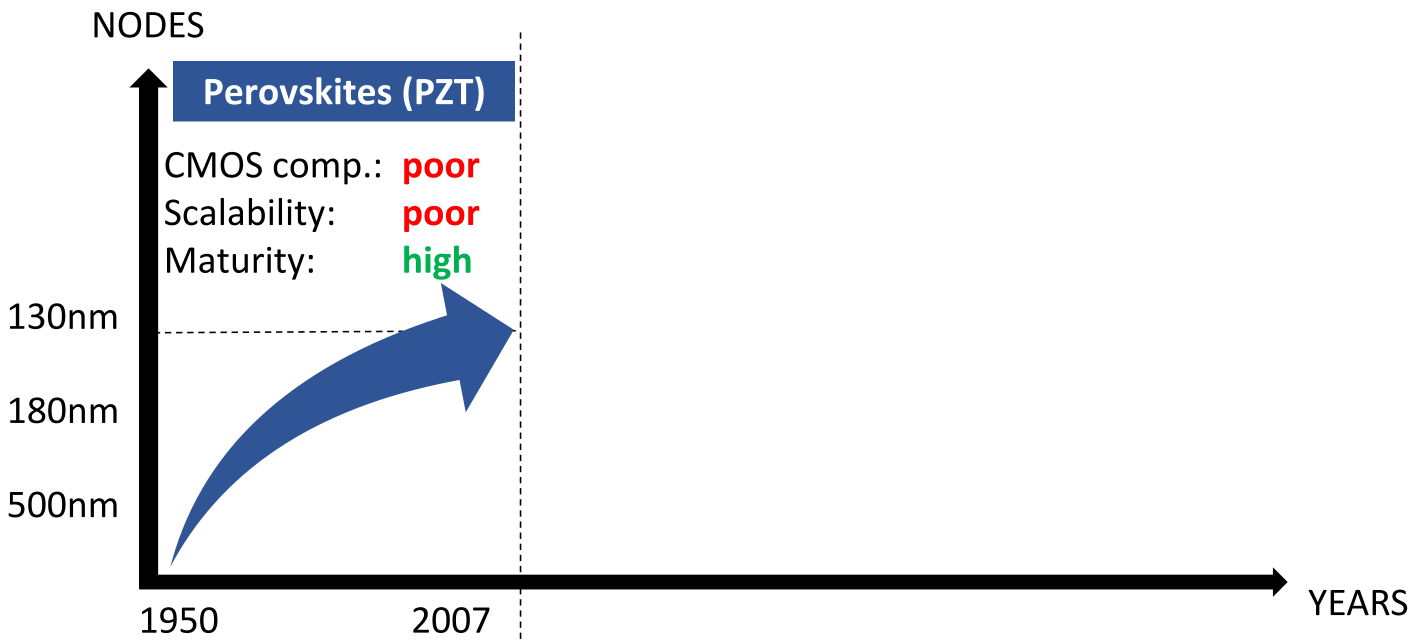
➔ PZT-based FeRAM limited to niche applications & relaxed nodes (130nm)

OUTLINE

- Ferroelectricity basics
- ➡ Ferroelectric HfO_2: a change of paradigm for NVM
- 1T-1C FeRAM arrays: basics
- HfO_2-based MFM capacitors integrated above CMOS
- HfO_2-based 1T-1C FeRAM arrays: performance overview
- Scalability: challenges and perspectives

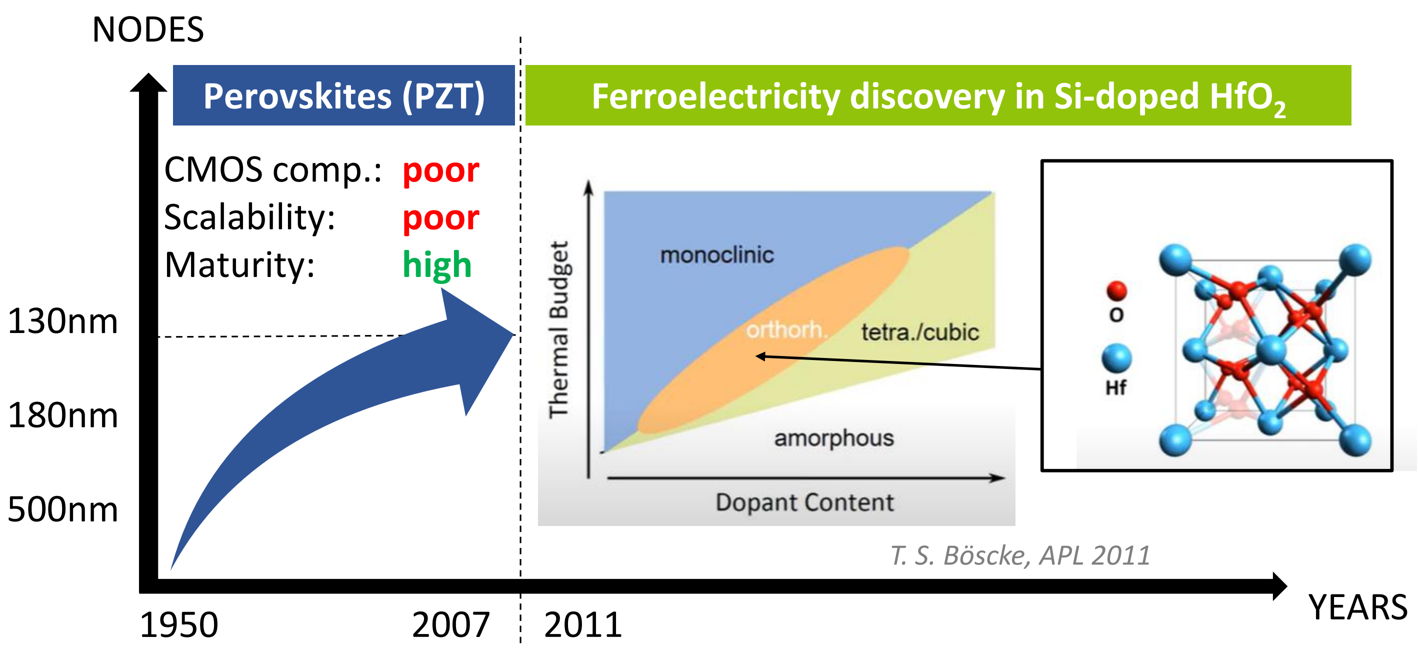
FERROELECTRIC HfO_2

› a change of paradigm for NVM



FERROELECTRIC HfO_2

› a change of paradigm for NVM

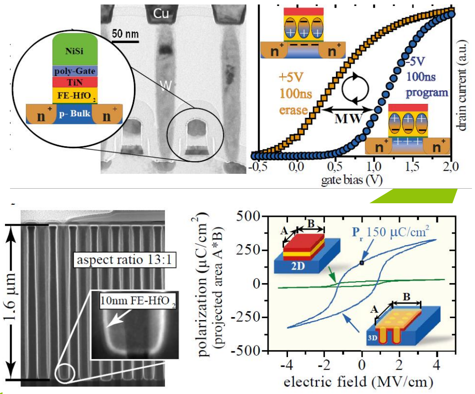


FERROELECTRIC HfO_2

› a change of paradigm for NVM

NODES

Perovskites (PZT)		Ferroelectric HfO_2
CMOS comp.:	poor	→ excellent
Scalability:	poor	→ excellent
Maturity:	high	→ low



J. Müller et al., IEDM'13

130nm

180nm

500nm

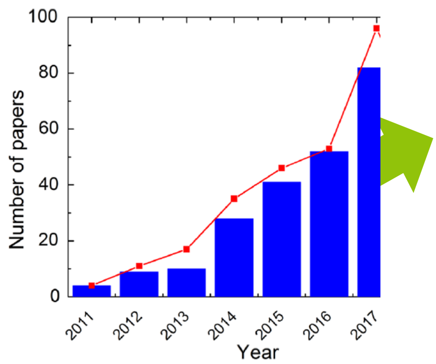
1950 2007 2011 2013 YEARS

FERROELECTRIC HfO_2

› a change of paradigm for NVM

NODES

Perovskites (PZT)		Ferroelectric HfO_2
CMOS comp.:	poor	→ excellent
Scalability:	poor	→ excellent
Maturity:	high	→ low



M.H. Park et al., MRS'18

130nm

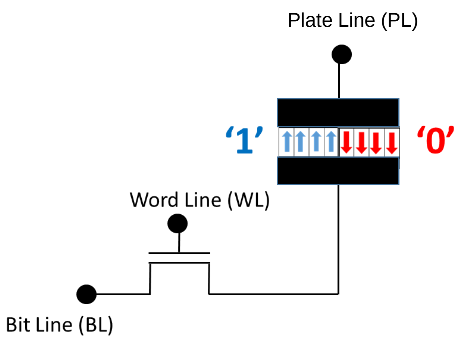
180nm

500nm

1950 2007 2011 2013 2018 YEARS

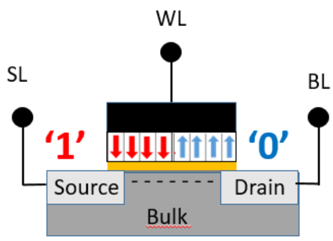
VARIOUS FERROELECTRIC MEMORIES

FeRAM



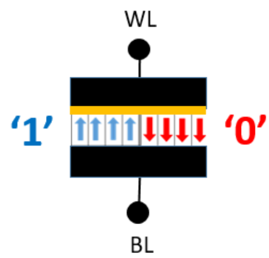
Low / High Displacement Current

FeFET



Low / High V_T Transistor

FTJ

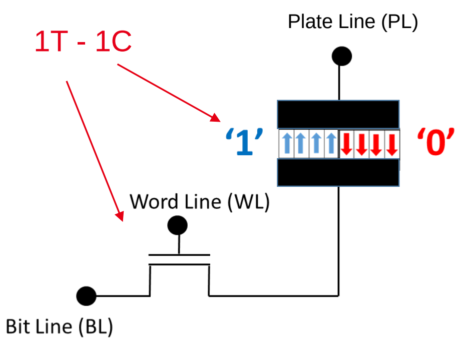


Low / High Resistive State

Embedded memory arrays ← **DIGITAL** — **APPLICATION SPACE** — **ANALOG** → **Artificial synapse (neuromorphics)**
Vector Matrix Multiplications

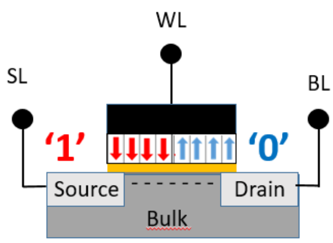
VARIOUS FERROELECTRIC MEMORIES

FeRAM



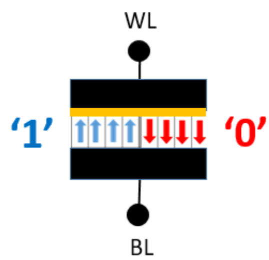
Low / High Displacement Current

FeFET



Low / High V_T Transistor

FTJ



Low / High Resistive State

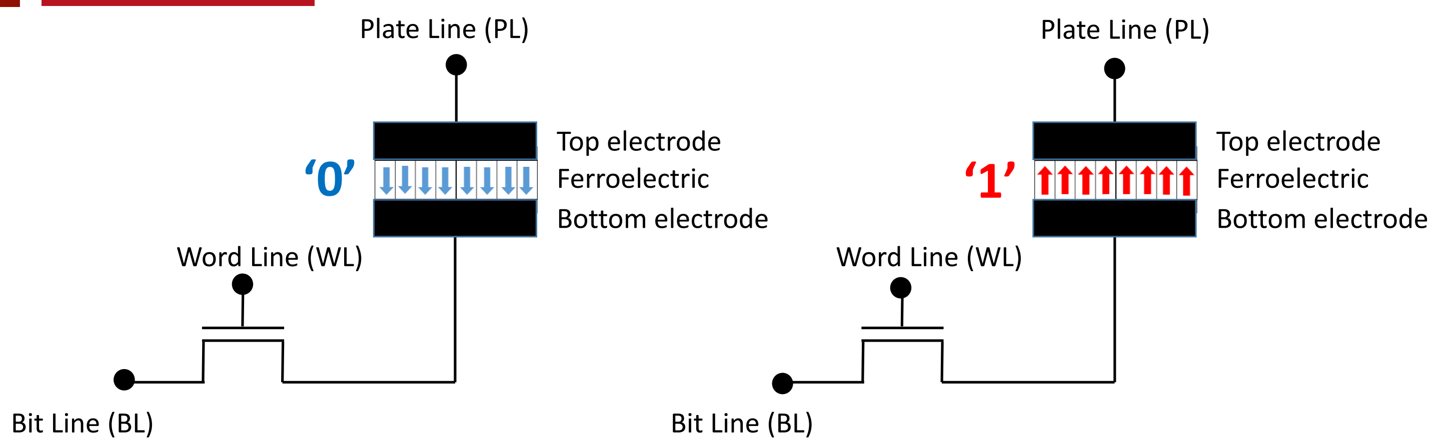
Embedded memory arrays ← **DIGITAL** — **APPLICATION SPACE** — **ANALOG** → **Artificial synapse (neuromorphics)**
Vector Matrix Multiplications

OUTLINE

- Ferroelectricity basics
- Ferroelectric HfO_2: a change of paradigm for NVM
- 1T-1C FeRAM arrays: basics
- HfO_2-based MFM capacitors integrated above CMOS
- HfO_2-based 1T-1C FeRAM arrays: performance overview
- Scalability: challenges and perspectives

1T-1C FeRAM

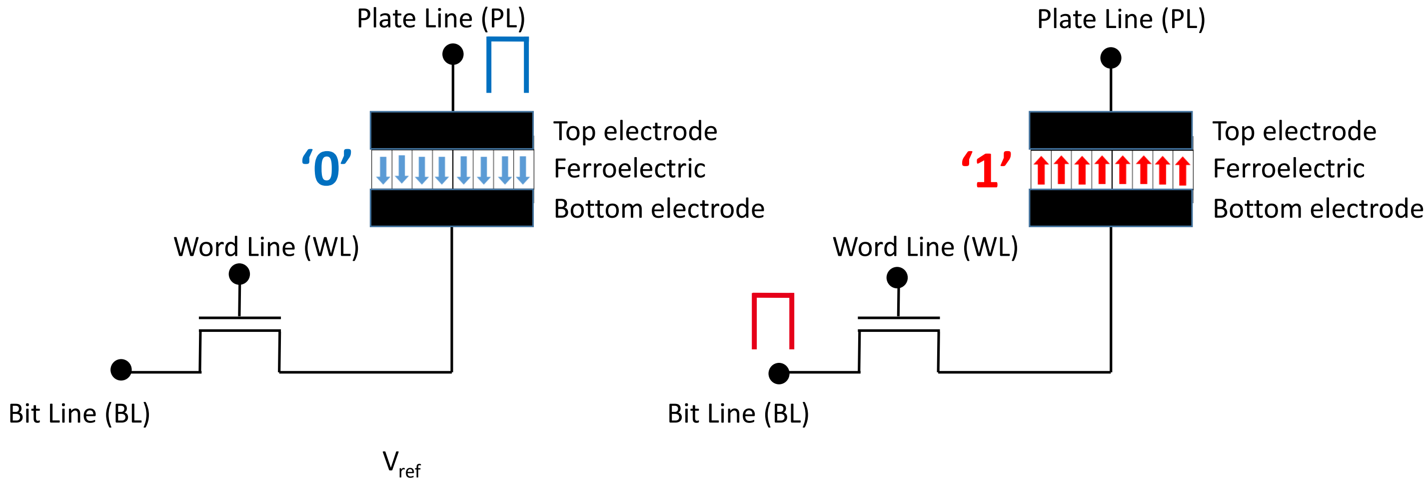
› ‘0’ and ‘1’ states



- State ‘0’ ➔ Pdown
- State ‘1’ ➔ Pup

1T-1C FeRAM

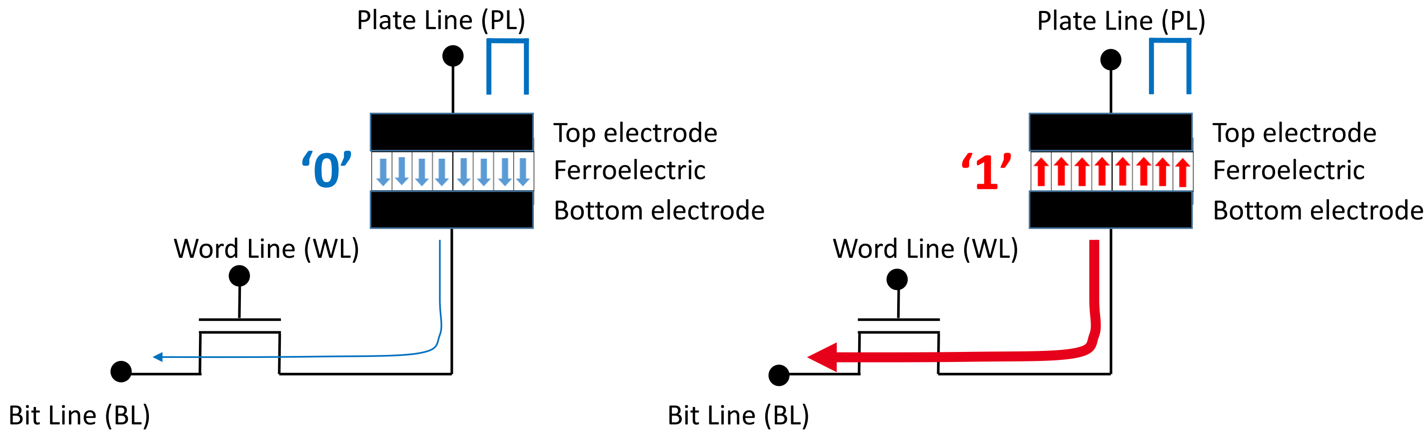
› WRITE ‘0’ and ‘1’ states



- WRITE ‘0’ :
  - precharge BL to GND
  - pulse WL > Vth
  - pulse PL > Vc
- WRITE ‘1’ :
  - PL to GND
  - pulse WL > Vth
  - pulse BL > Vc

1T-1C FeRAM

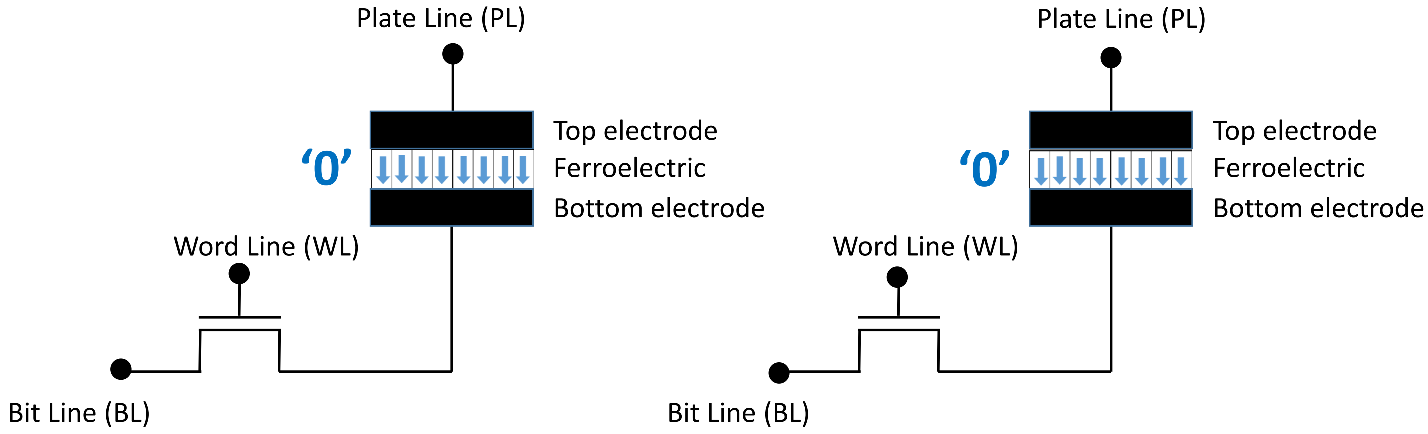
› READ ‘0’ and ‘1’ states



- READ ‘0’ :
  - BL floating
  - pulse WL > Vth
  - pulse PL > Vc ➔ small V_{BL} elevation (displacement current)
- READ ‘1’ :
  - BL floating
  - pulse WL > Vth
  - pulse PL > Vc ➔ high V_{BL} elevation (displacement current + **switching current**)

1T-1C FeRAM

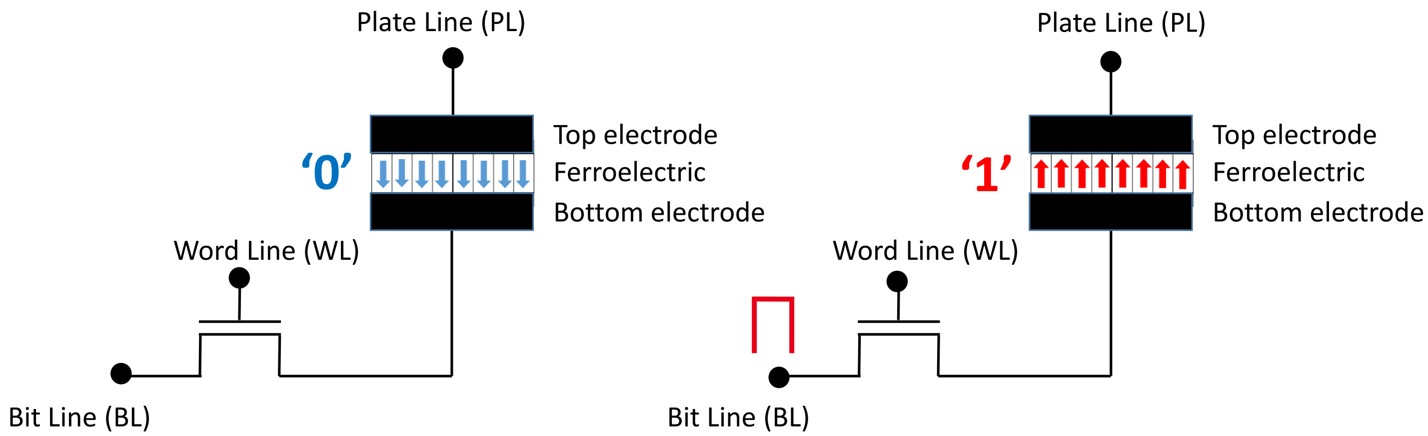
› Destructive READ



- READ operation is destructive!

1T-1C FeRAM

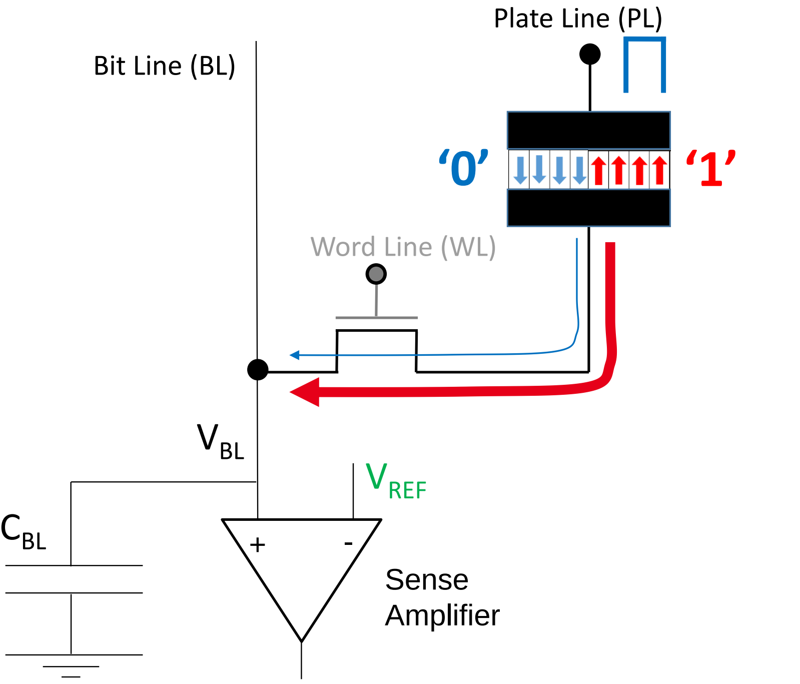
› Destructive READ



- READ operation is destructive! ➔ WRITE BACK necessary

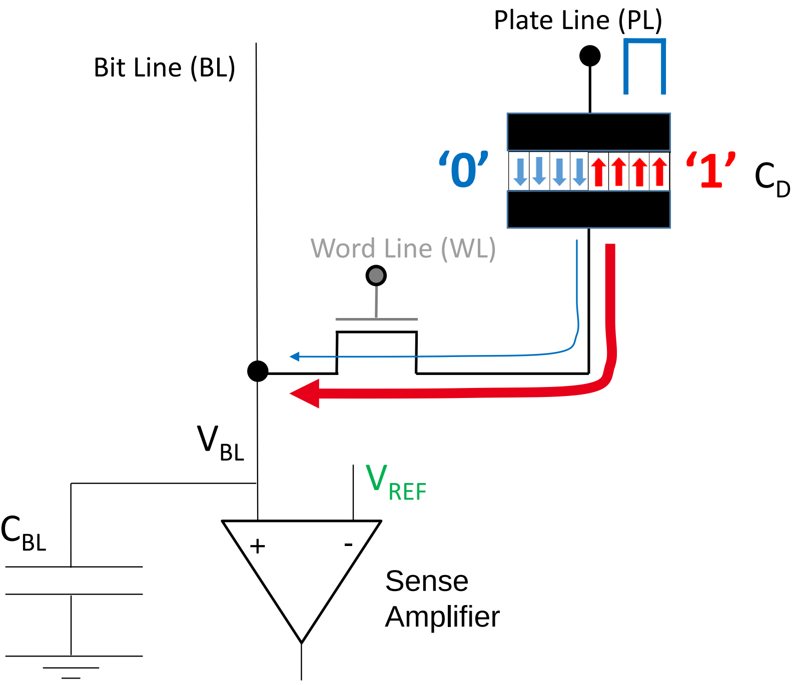
1T-1C FeRAM

› READ ‘0’ and ‘1’ states



1T-1C FeRAM

› BL voltages



$$V_{BL}^{NSW} = \frac{C_D}{C_D + C_{BL}} \times V_{PL}$$

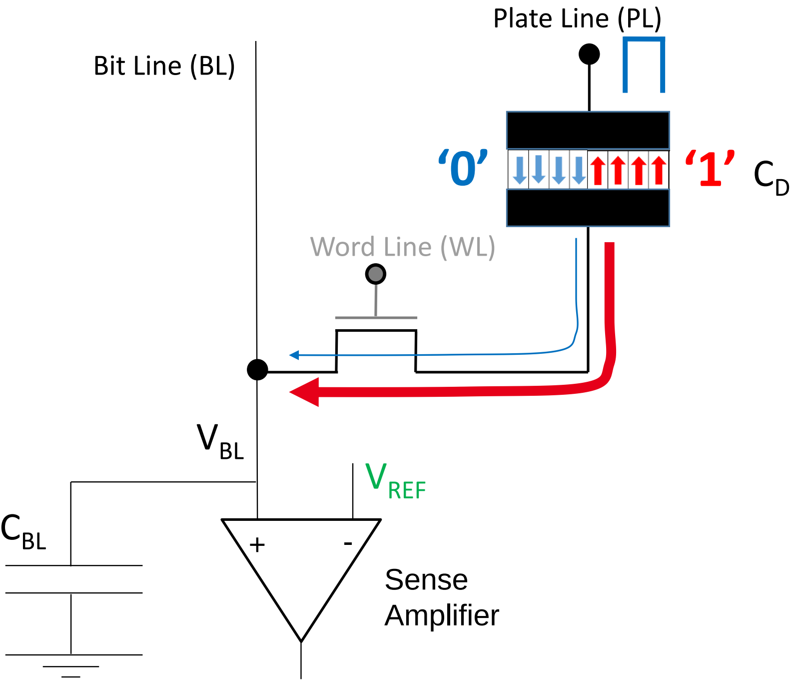
$$V_{BL}^{SW} = V_{BL}^{NSW} + \frac{2.P_R \times S}{C_D + C_{BL}}$$

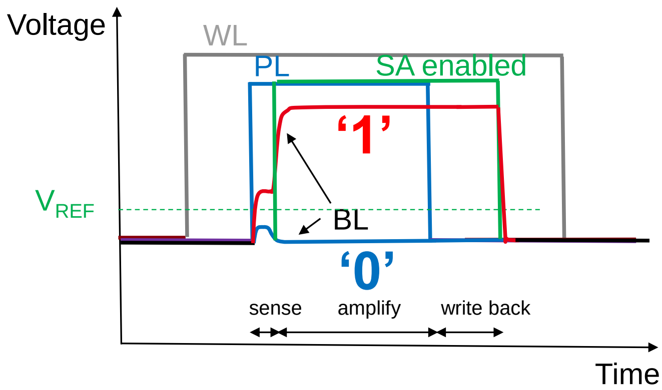
1T-1C FeRAM

› Sense, amplify, and write back

$$V_{BL}^{NSW} = \frac{C_D}{C_D + C_{BL}} \times V_{PL}$$

$$V_{BL}^{SW} = V_{BL}^{NSW} + \frac{2.P_R \times S}{C_D + C_{BL}}$$



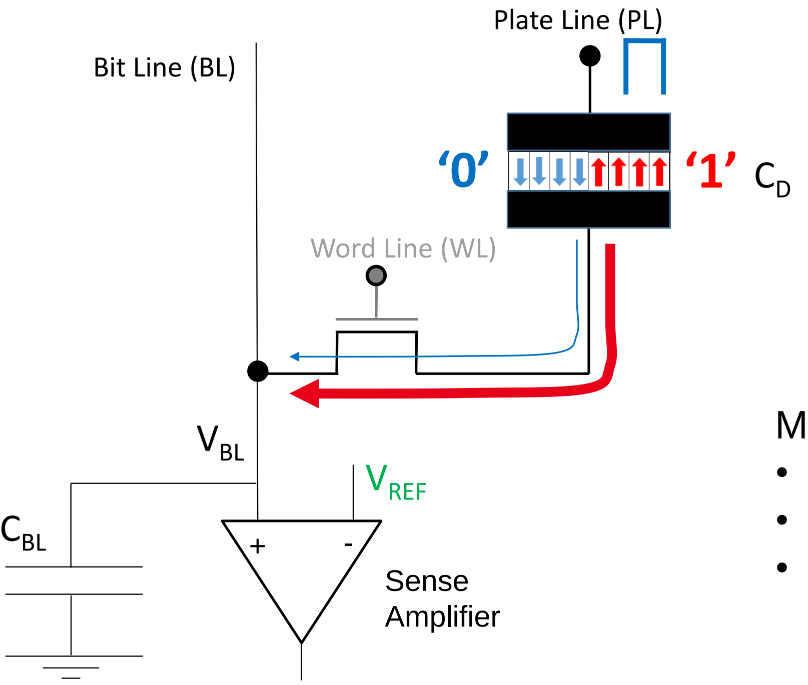


1T-1C FeRAM

› Memory Window

$$V_{BL}^{NSW} = \frac{C_D}{C_D + C_{BL}} \times V_{PL}$$

$$V_{BL}^{SW} = V_{BL}^{NSW} + \boxed{\frac{2.P_R \times S}{C_D + C_{BL}}}$$

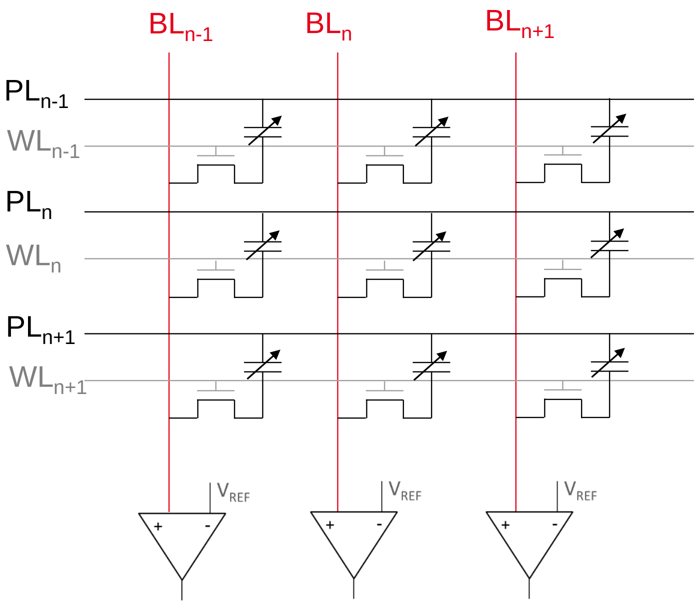


Memory Window (MW):

- increases when FeCAP area increases
- increases when $2.P_R$ increases
- decreases when C_{BL} increases

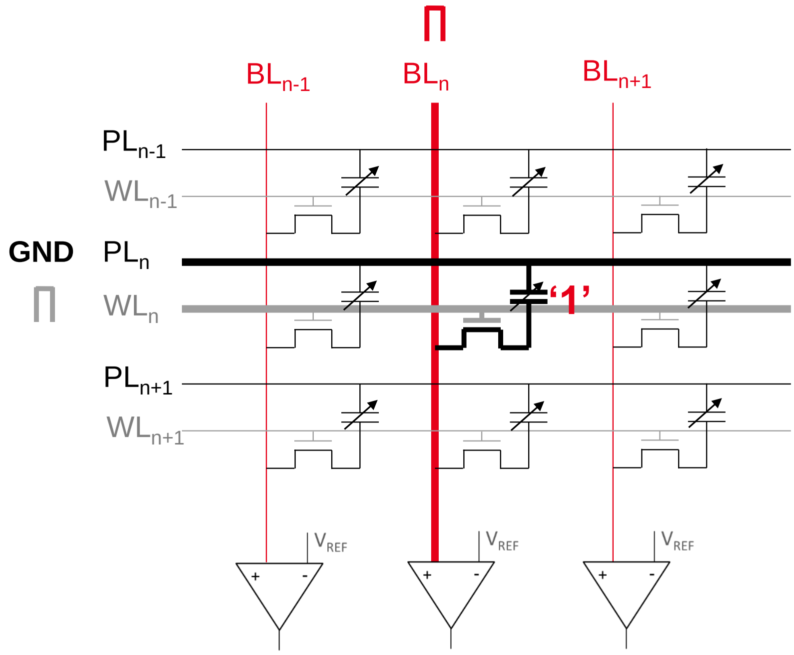
1T-1C FeRAM

› Arrays



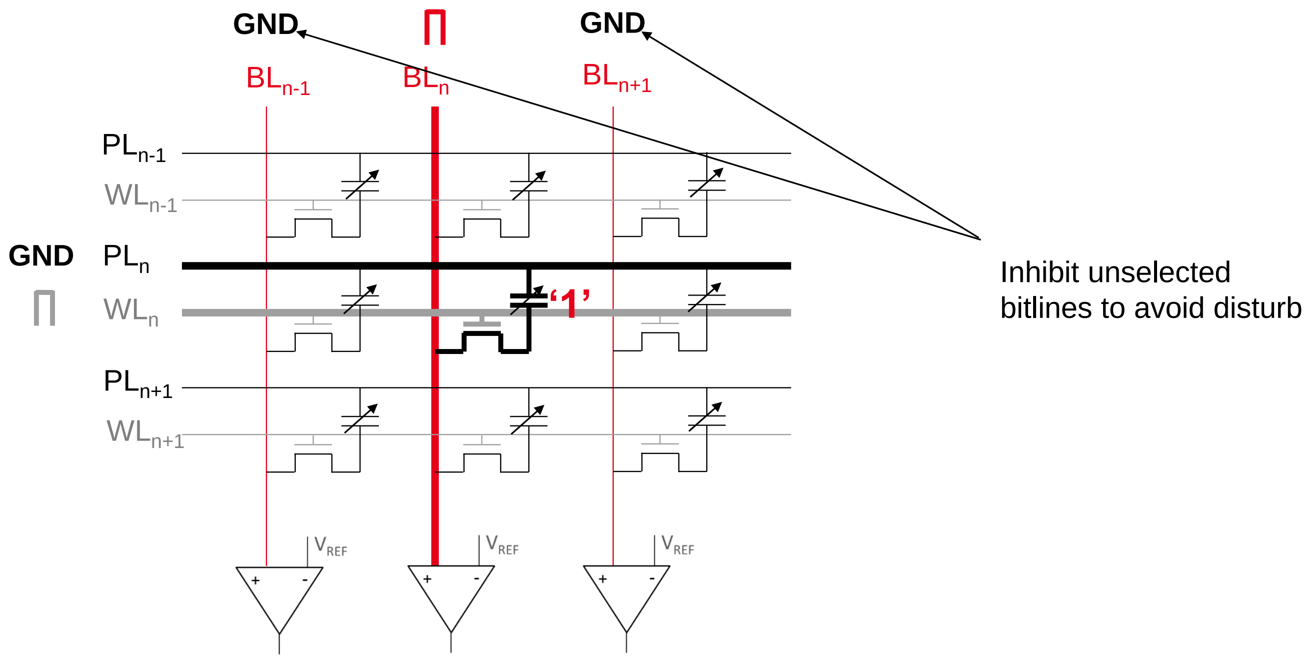
1T-1C FeRAM

› Arrays: bitcell programming



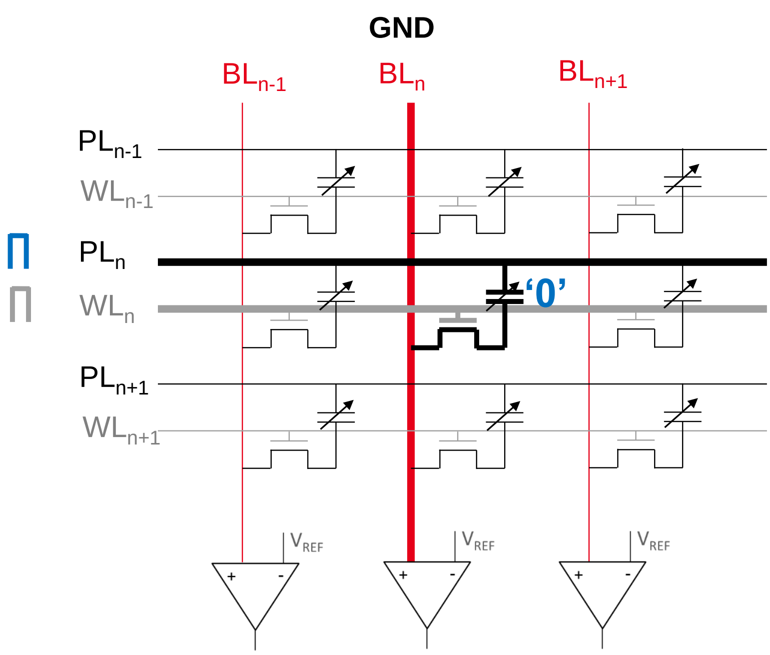
1T-1C FeRAM

› Arrays: bitcell programming



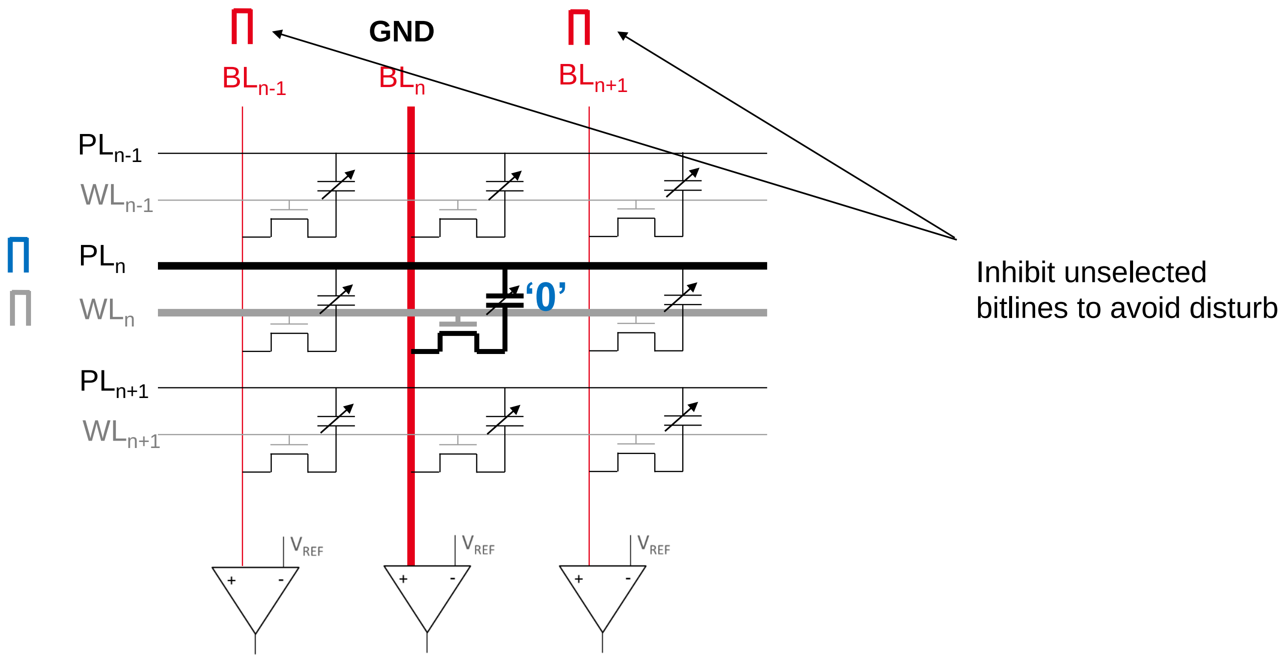
1T-1C FeRAM

› Arrays: bitcell programming



1T-1C FeRAM

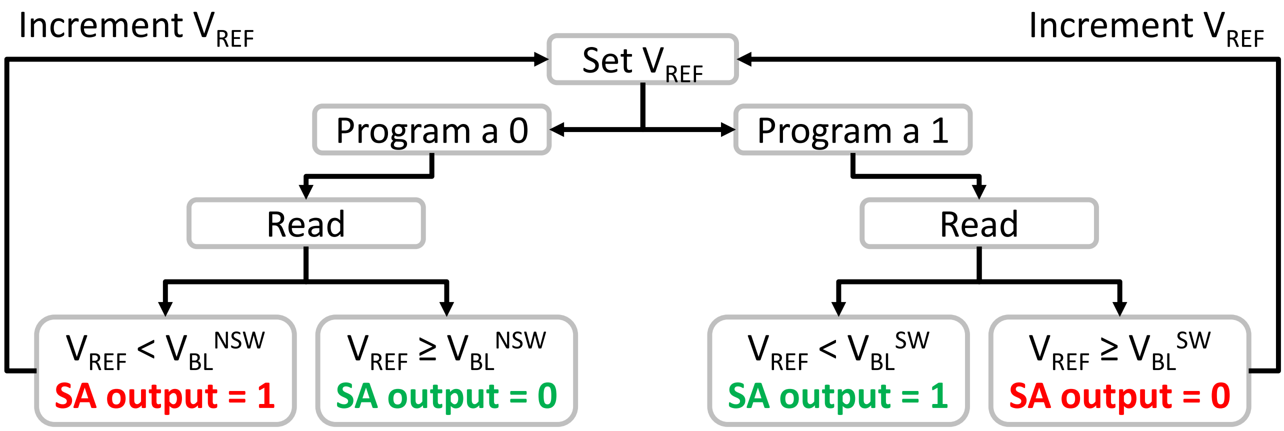
› Arrays: bitcell programming



1T-1C FeRAM

› Arrays: analog-like characterization with destructive READ operation

▪ **Specific methodology to achieve analog-like characterization**

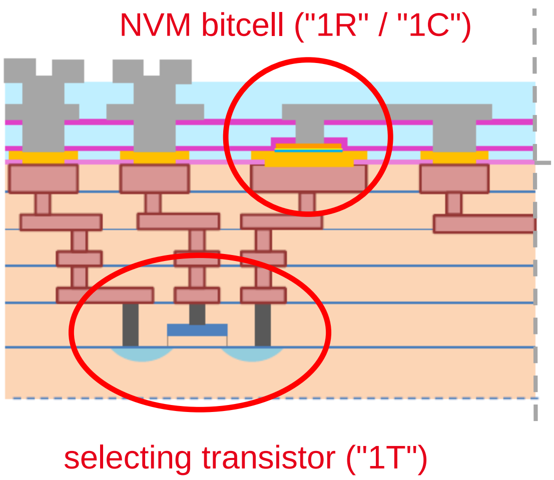


Iterative READ and WRITE operations with scanning V_{REF} allows to reconstruct '0' and '1' distributions in 1T-1C FeRAM arrays

OUTLINE

HfO_2-BASED MFM CAPACITORS INTEGRATION

› BEOL integration in MAD200 test vehicle (130nm node)



BEOL NVM

FEOL CMOS
→ *mature foundry technology*
→ *from single NVM bitcells up to complex circuits*

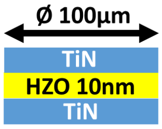
life.augmented

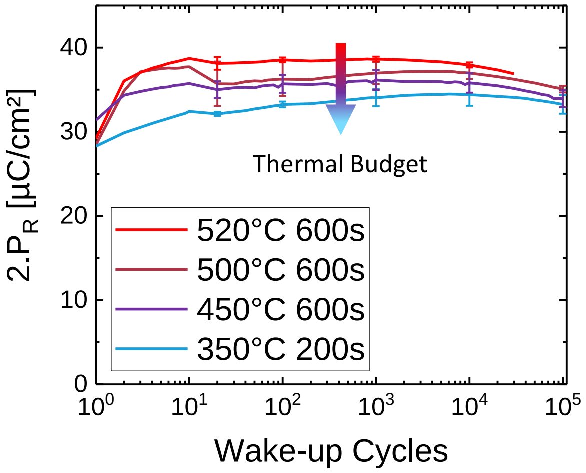
HCMOS9 (130nm node, 200mm)

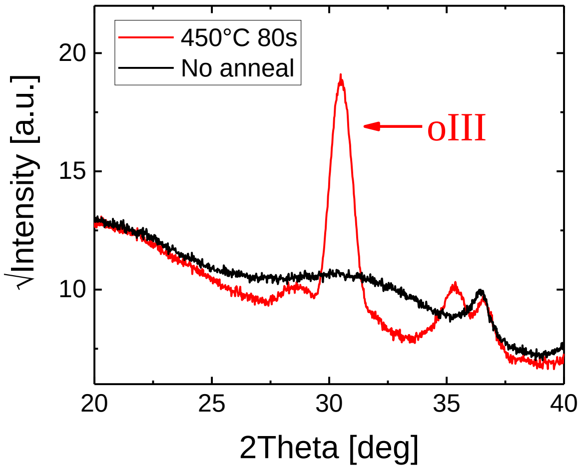
MAD200: a versatile platform for assessing BEOL-NVMs (OxRAM, PCRAM… and FeRAM)

HfO_2-BASED MFM CAPACITORS INTEGRATION

› Thermal budget for crystallization: prerequisite for BEOL integration





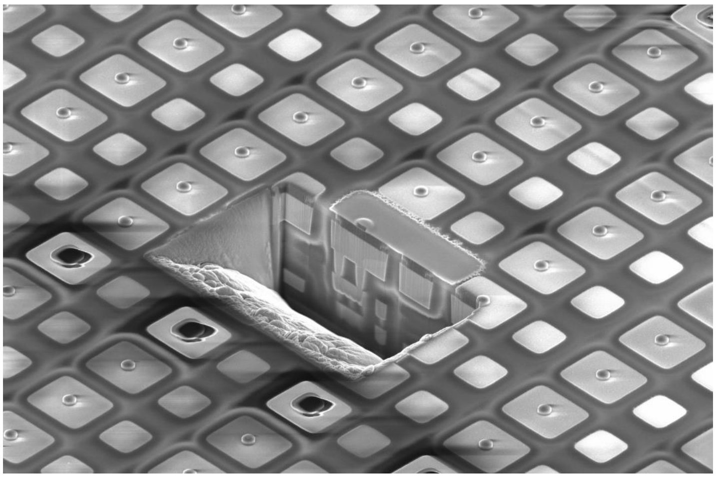


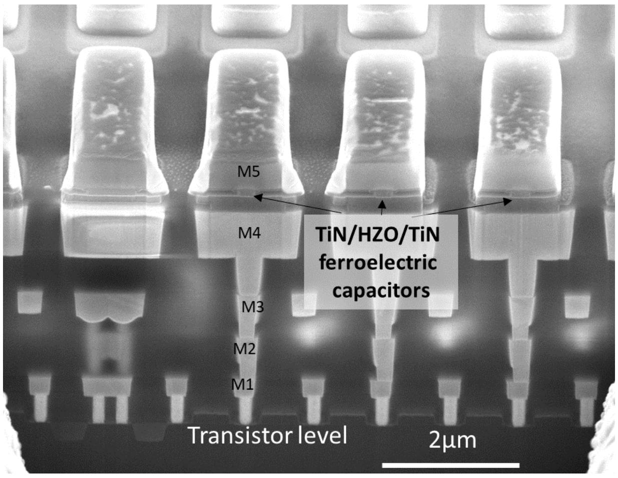
[T. Francois *et al.*, IEDM 2019]

$2.P_R > 30\mu C/cm^2$ demonstrated down to a thermal budget of 350°C

HfO_2-BASED MFM CAPACITORS INTEGRATION

› BEOL integration at 130nm node – Morphological results

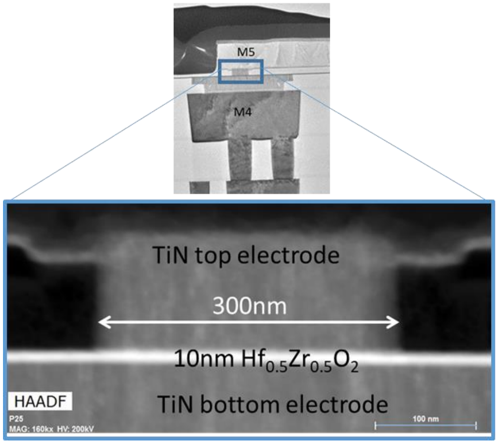


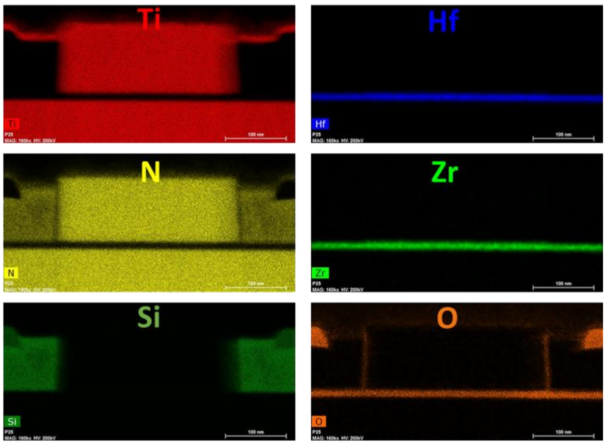


[T. Francois *et al.*, IEDM 2019]

HfO_2-BASED MFM CAPACITORS INTEGRATION

› BEOL integration at 130nm node – Morphological results

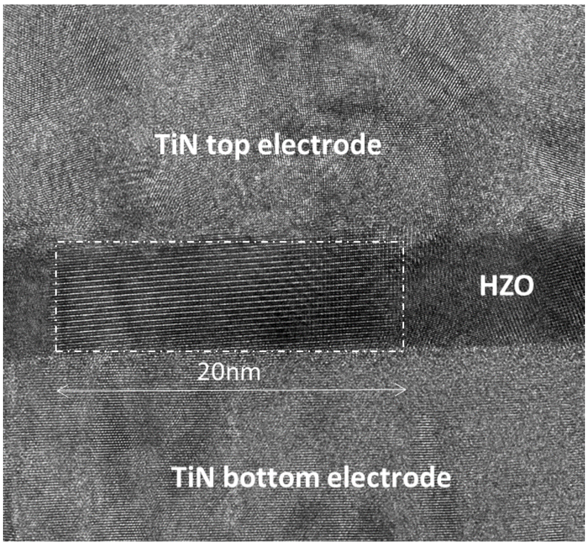




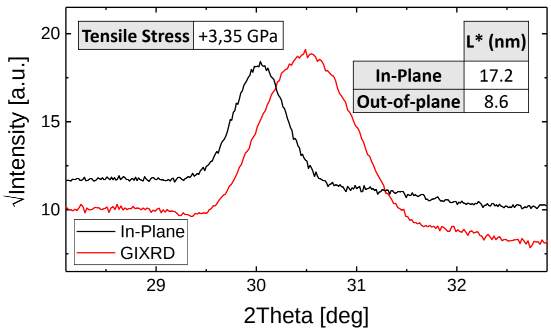
[T. Francois *et al.*, IEDM 2019]

HfO_2-BASED MFM CAPACITORS INTEGRATION

› BEOL integration at 130nm node – Morphological results



GIXRD characterization demonstrating oIII ferroelectric phase, with ~10nm-thick / 20nm-large cristallites

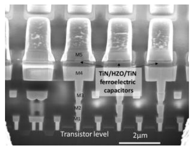


[T. Francois *et al.*, IEDM 2019]

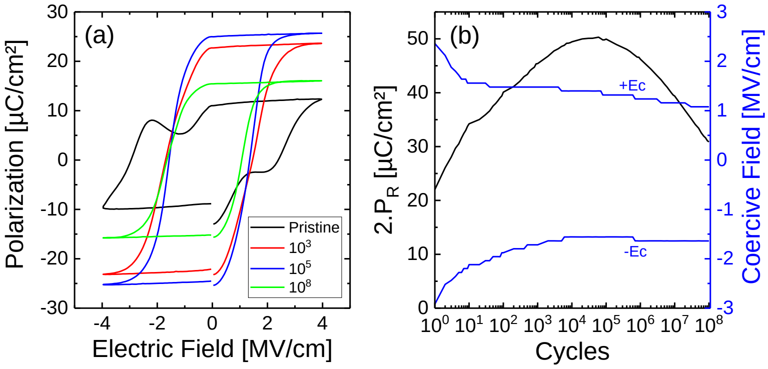
2019: demonstration of ferroelectric $Hf_{0.5}Zr_{0.5}O_2$ in BEOL-integrated sub-µm² FeCap (NaMLab/Leti)

HfO_2-BASED MFM CAPACITORS INTEGRATION

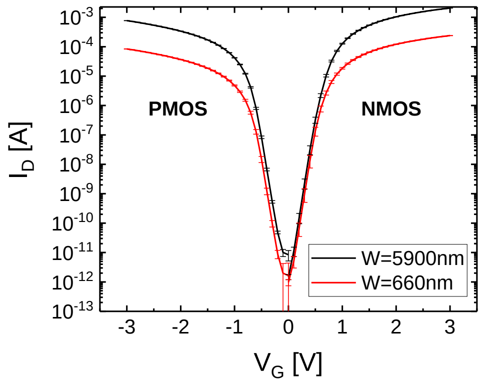
› BEOL integration at 130nm node – Electrical results



Integration of ⌀550nm HZO FeCap in BEOL…



… while preserving FEOL CMOS

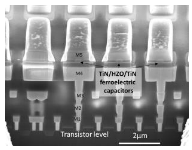


[T. Francois *et al.*, IEDM 2019]

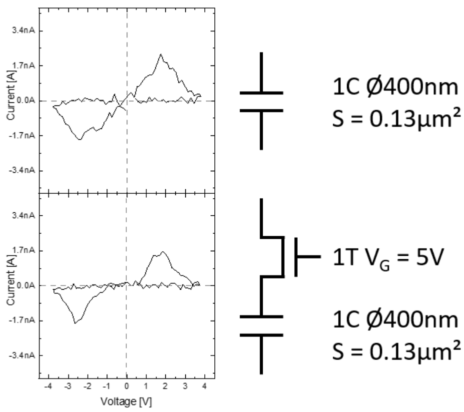
electrical demonstration of scaled FeRAM capacitors integrated in BEOL without impacting FEOL

HfO_2-BASED MFM CAPACITORS INTEGRATION

› BEOL integration at 130nm node – Electrical results

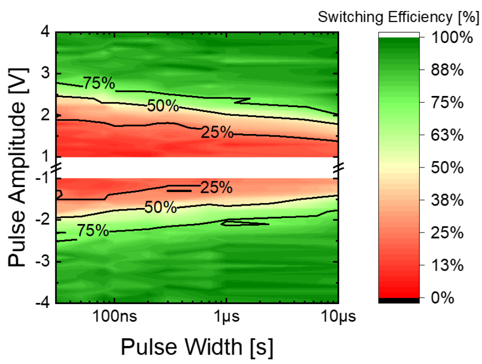


Switching kinetics on single scaled capacitors (10nm HZO)



After 10^3 wake-up cycles

Single Ø600nm (0.28µm²) capacitor



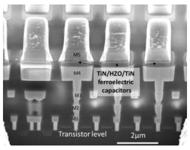
[T. Francois *et al.*, IEDM 2019]

Same behavior on 1C and 1T-1C

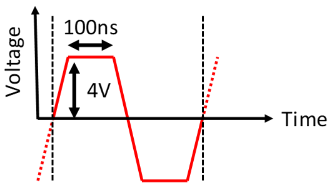
down to 30ns switching capability
>50% even at low voltages (2V)
Suitable for memory application

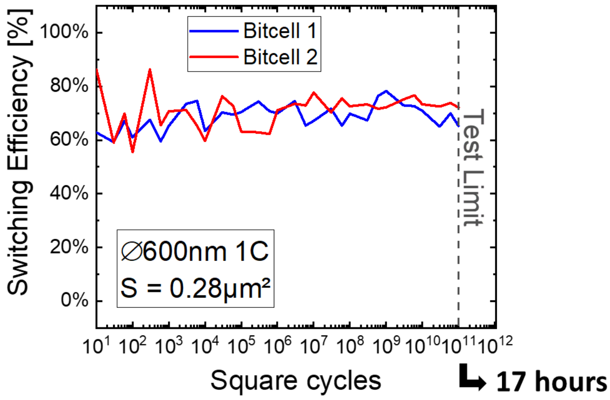
HfO_2-BASED MFM CAPACITORS INTEGRATION

› BEOL integration at 130nm node – Electrical results



Endurance measurement on single scaled capacitors (10nm HZO)



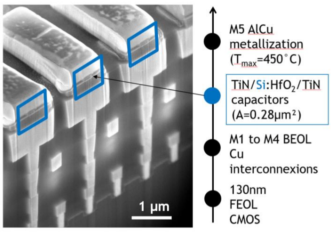


[T. Francois *et al.*, IEDM 2019]

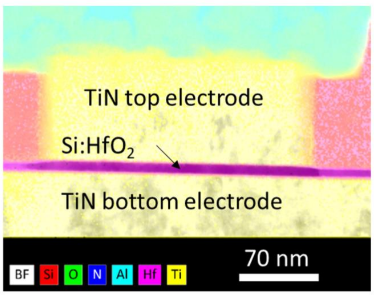
Endurance $>10^{11}$ cycles (4V) for a BEOL-integrated HZO FeCap

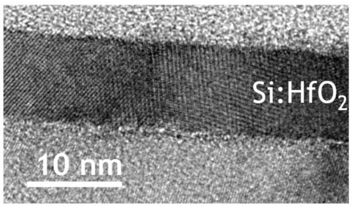
HfO_2-BASED MFM CAPACITORS INTEGRATION

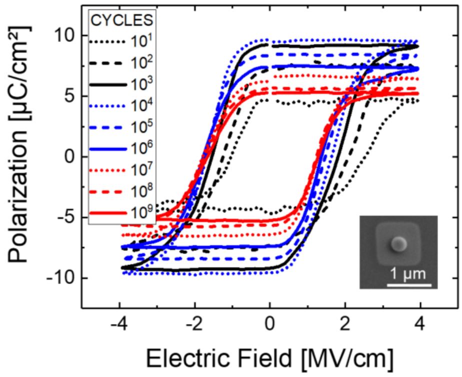
› BEOL integration at 130nm node– Si-implanted HfO_2



$Si:HfO_2$ (HSO) FeCap in BEOL (thermal budget 450°C)





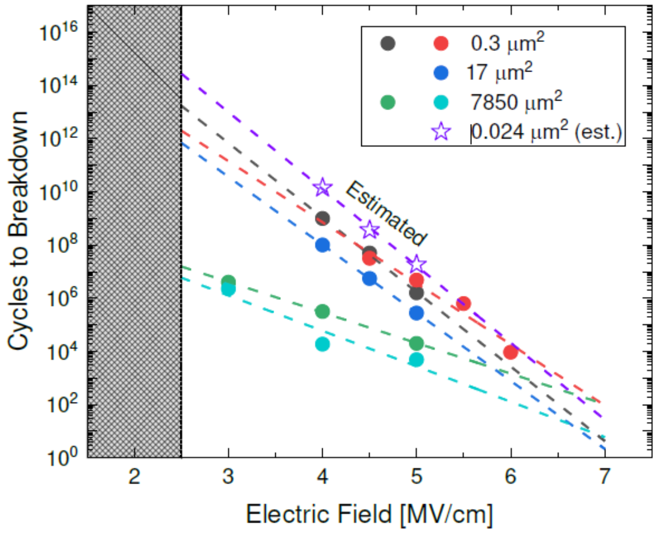


[L. Grenouillet *et al.*, VLSI 2020]

$Si:HfO_2$ as a BEOL compatible ferroelectric material. Si doping by ion implantation

HfO_2-BASED MFM CAPACITORS INTEGRATION

› BEOL integrated FeCap: area scaling

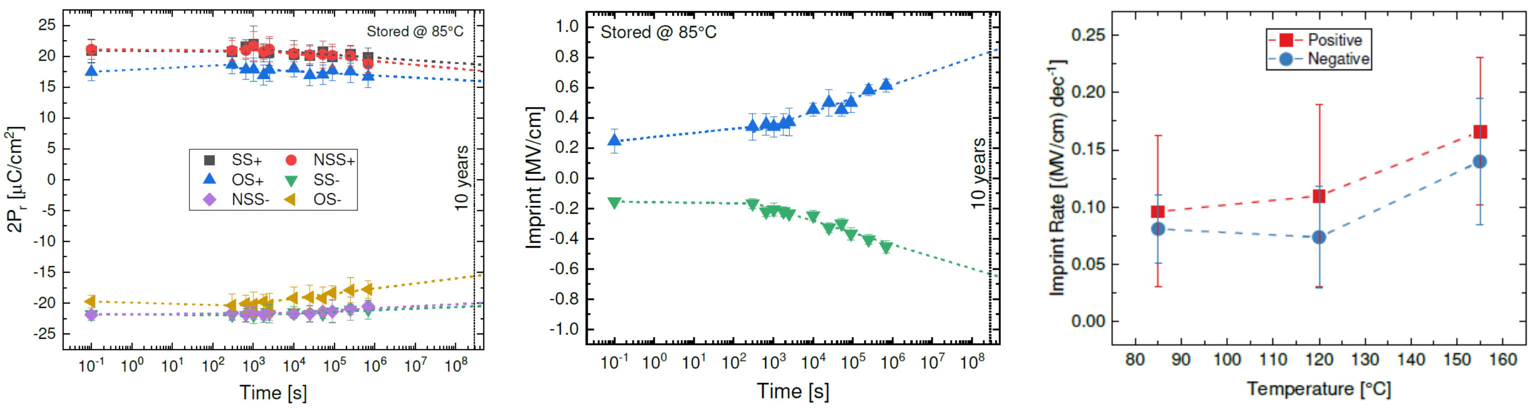


[R. Alcala *et al.*, EDTM 2022]

Scaling down FeRAM to sub-μm² sizes has a positive impact on reliability: promising for NVM application

HfO_2-BASED MFM CAPACITORS INTEGRATION

› Data retention & Imprint



[R. Alcala *et al.*, EDTM 2022]

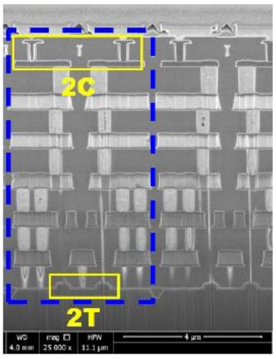
Good data retention at 85°C even for Opposite State. Imprint rate increases above 125°C

HfO_2-BASED MFM CAPACITORS INTEGRATION

› BEOL integration of 3D FeCap

3D Scalable, Wake-up Free, and Highly Reliable FRAM Technology with Stress-Engineered $HfZrO_x$

Y.D. Lin[1,3], H.Y. Lee[1,*], Y.T. Tang[2], P.C. Yeh[1], H.Y. Yang[1], P.S. Yeh[1], C.Y. Wang[1], J.W. Su[1,3], S.H. Li[1], S.S. Sheu[1], T.H. Hou[1], W.C. Lo[1], M. H. Lee[4], M.F. Chang[3], Y.C. King[3] and C.J. Lin[3]
[1]EOSL, Industrial Technology Research Institute, Hsinchu, Taiwan, *email: hengyuan@itri.org.tw
[2] Taiwan Semiconductor Research Institute, Hsinchu, Taiwan,
[3]Institute of Electronics Engineering, National Tsing Hua University, Hsinchu, Taiwan
[4]Institute of Electro-Optical Science and Technology, National Taiwan Normal University, Taipei, Taiwan



Device \ Ref.	This work	
1C stacks	(2D) IL/HfZrO 10nm/IL	(3D) IL*/HfZrO 10nm/IL*
Cycling speed (kHz)	625	625
Applied field (MV/cm)	2.5	2.2
P_r (uC/cm²)	20~32	18~20
Endurance	10^{10}	10^{9}
Retention	5x10⁴. (105˚C)	5x10⁴. (85˚C)

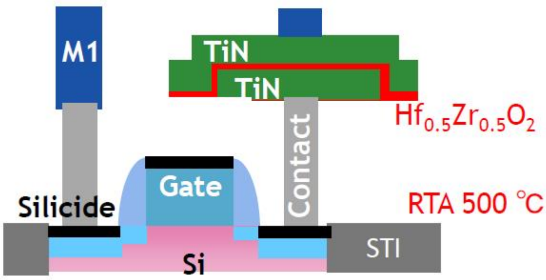
* 170 uC/cm² : N

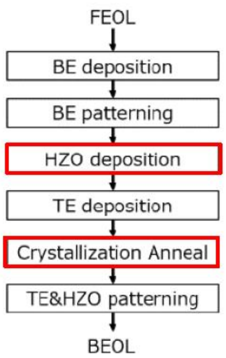
[Y.D. Lin *et al.*, IEDM 2019]

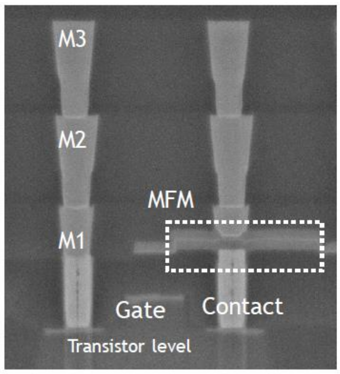
2D & 3D 10nm HZO-based FeCap integrated between M6 & M7. Endurance > 10^9 cycles

HfO_2-BASED MFM CAPACITORS INTEGRATION

› Middle Of Line integration







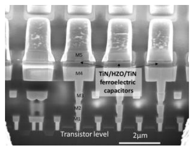
[J. Okuno *et al.*, VLSI 2020]

Advantage of MOL integration: higher thermal budget allowed for crystallization

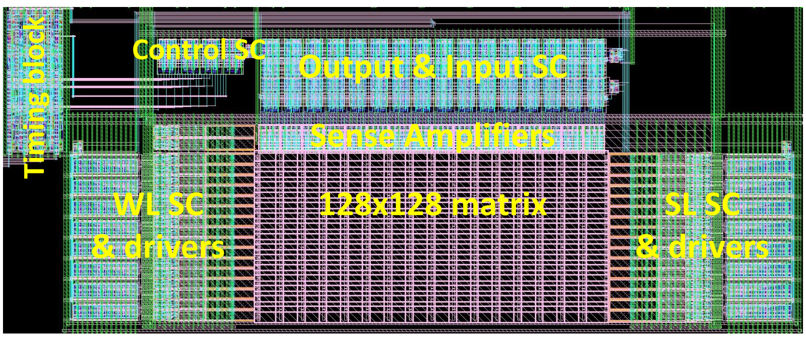
OUTLINE

HfO_2-BASED FERAM ARRAYS

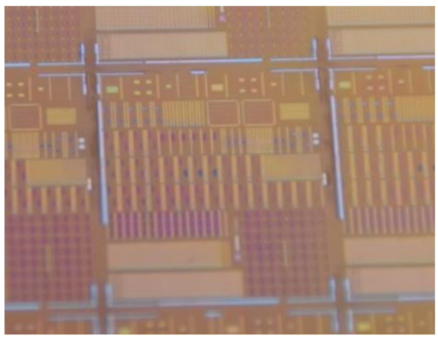
› 16 kbit FeRAM test vehicle (Leti - MAD200v3)



16 kbit 1T-1C FeRAM layout



16 kbit 1T-1C FeRAM chip view



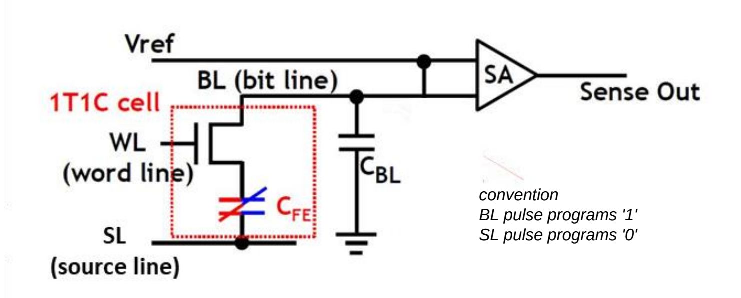
- 3x circuit versions, with FeCap areas = 0.36 / 0.24 / 0.16µm²
- Scan chains for bitcell addressing, circuit control and buffering out data
- Sense Amplifiers for (destructive) reading operations
- Internal Pulse Generators for sub-ns programming

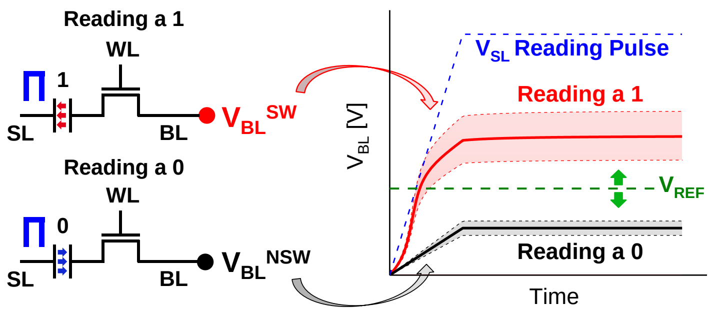
HfO_2-BASED FERAM ARRAYS

› 16 kbit FeRAM test vehicle (Leti - MAD200v3)

FeRAM bitcell & read operation

- bitcell = 1T FEOL + 1C BEOL
  Array = 128 Word Lines x 128 Bit Lines
- Polarization state of a bitcell is not directly measurable
- Read operation = attempt to program a '0'
  - if bitcell = '1' → ferro switch detected
  - if bitcell = '0' → no ferro switch detected
  - program back data

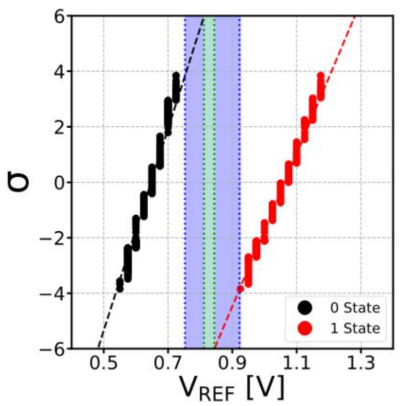




HfO_2-BASED FERAM ARRAYS

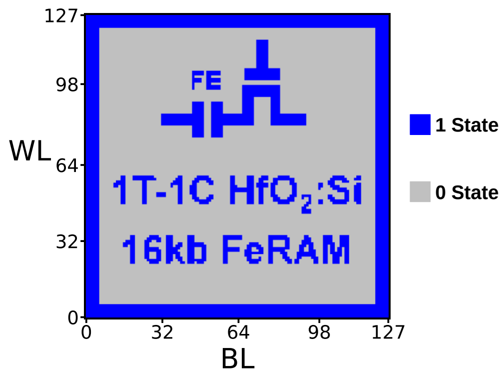
› 16 kbit FeRAM electrical results (Leti - MAD200v3)

Distributions on Si:HfO_2-based 16 kbit FeRAM (0.36 µm² FeCap, 4.8V/2µs pulses, after wake-up)



Measured
MW 16kbit = 170mV
Extrapolated
MW 6σ = 32mV

Nominal memory operation at V_{REF}=0.85V



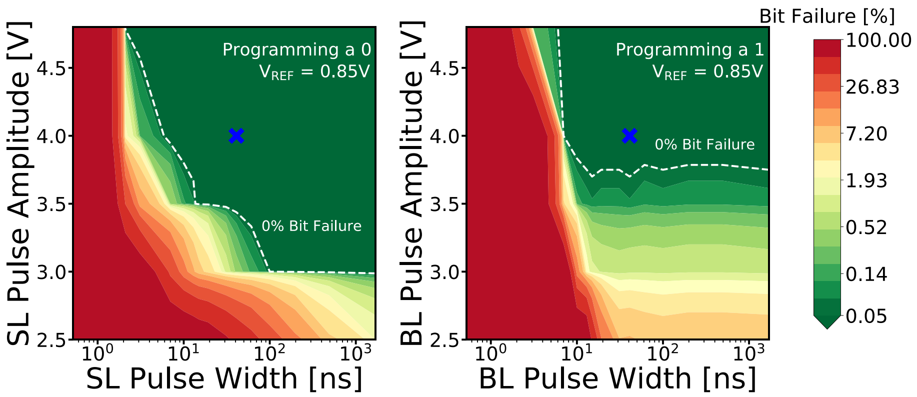
[T. Francois *et al.*, IEDM 2021]

- 0 bitfail, large Memory Window

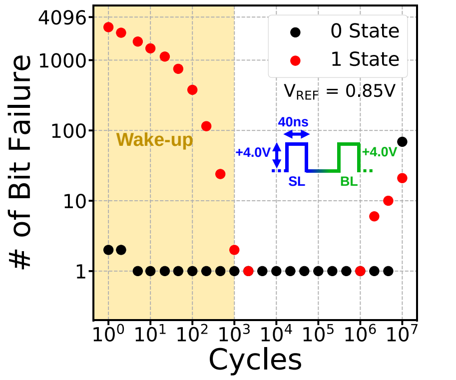
HfO_2-BASED FERAM ARRAYS

› 16 kbit FeRAM electrical results (Leti - MAD200v3)

Memory speed



Endurance Si:HfO_2-based 16 kbit FeRAM



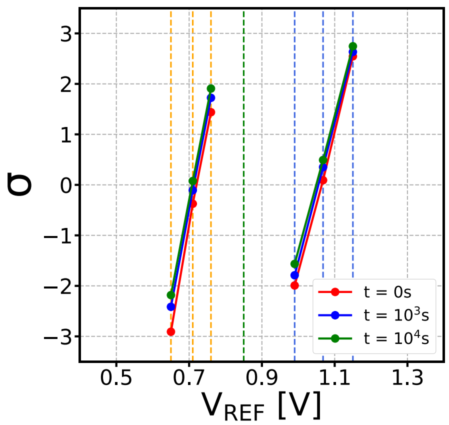
[T. Francois *et al.*, IEDM 2021]

- Excellent switching speed down to 10ns
- Endurance > 10^7 cycles using high cycling field

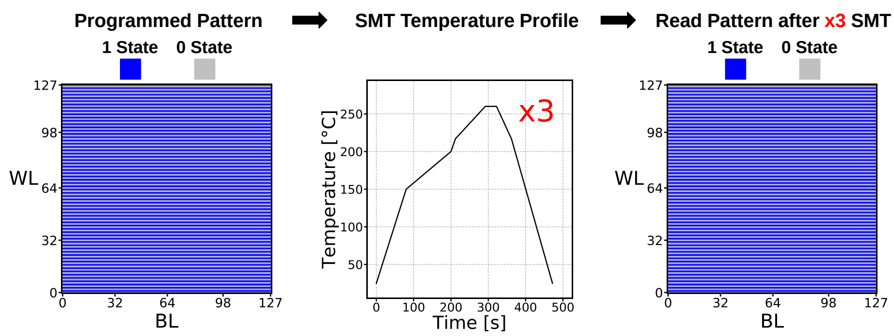
HfO_2-BASED FERAM ARRAYS

› 16 kbit FeRAM electrical results (Leti - MAD200v3)

Data retention at 125°C



0 bitfail after 3x solder reflow test (T_{max}=260°C)

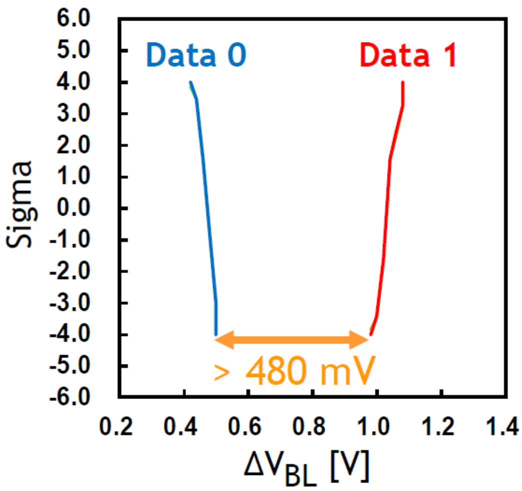


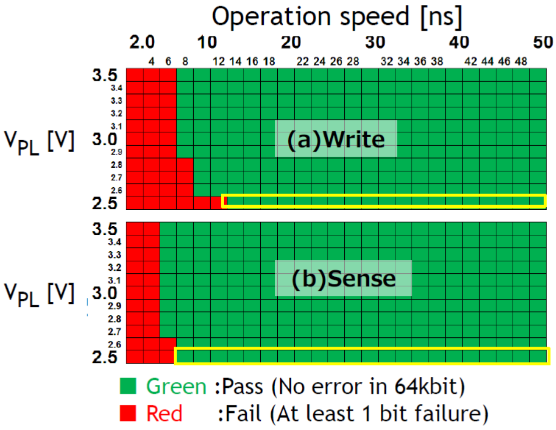
[T. Francois *et al.*, IEDM 2021]

- MW open after 10^4s @125°C with V_{REF} = 0.85V
- Solder reflow compatibility demonstrated for the first time

HfO_2-BASED FERAM ARRAYS

› 64 kbit FeRAM electrical results (SONY)



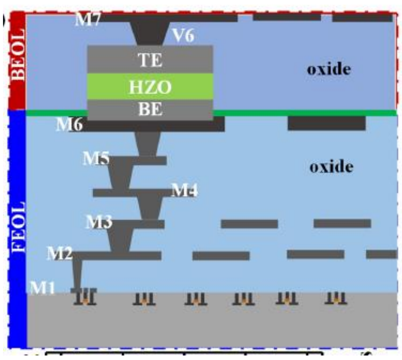


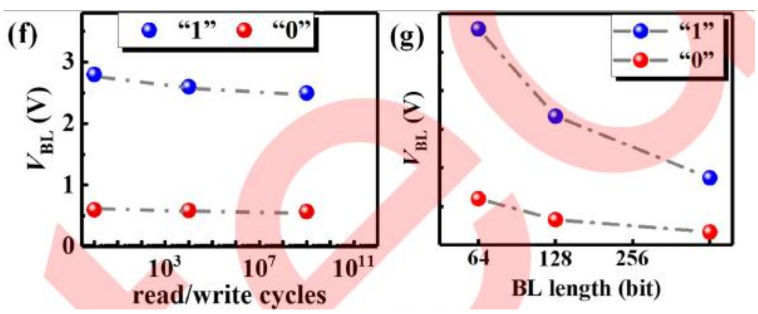
[J. Okuno *et al.*, VLSI 2020]

100 % bit functionality and 480 mV read margin (1µm² FeCap)
Sub 10 ns operation speed and < 2.5 V operating voltage

HfO_2-BASED FERAM ARRAYS

› 16 kbit FeRAM electrical results (Xidian University, Xi'an UniIC Semiconductor)





[W. Xiao *et al.*, Science China 2022]

Integration between M6 & M7 of TaN/HZO/TaN 64µm² FeCap and larger
30 ns switching speed, >10^4 s data retention, and >10^{11} cycling capability.
> 10^9 write/read cycling for the 1T–1C cell is achieved for the first time at array level.

leti ceatech

BENCHMARK

› HfO_2-based 1T-1C FeRAM

	Sony (VLSI 2020)	Sony (IMW 2021, EDTM2022)	Leti (IEDM 2021)	Xidian University, Xi'an UniIC Semiconductor
Node	130nm	130nm	130nm	130nm
Circuit size	64 kbit	64 kbit	16 kbit	16 kbit
Ferro material-stack	TiN / HZO 10nm / TiN	TiN / HZO 8nm / TiN	TiN / HSO 10nm / TiN	TaN / HZO 20nm / TaN
Integration, thermal budget	MOL (M1) > 500°C	MOL (M1) > 500°C	BEOL (M4-M5) < 500°C	BEOL (M6-M7) < 500°C
Min. write voltage	2.5V	2V	2.5V	3.5V
Write speed	14ns at 2.5V (one state)	**16ns at 2V** (one state)	4ns at 4.8V (both state)	30ns
Endurance	-	>10^{15} at 2V (extrapolated) > 10^{10} at 2V/2.8V(measured)	> 10^7 at 4V (measured)	**> 10^9 at 3.5V** (measured)
Retention	-	-	**125°C 10^4s**	10^4s (RT)
Solder reflow	-	-	**Yes**	-

leti ceatech

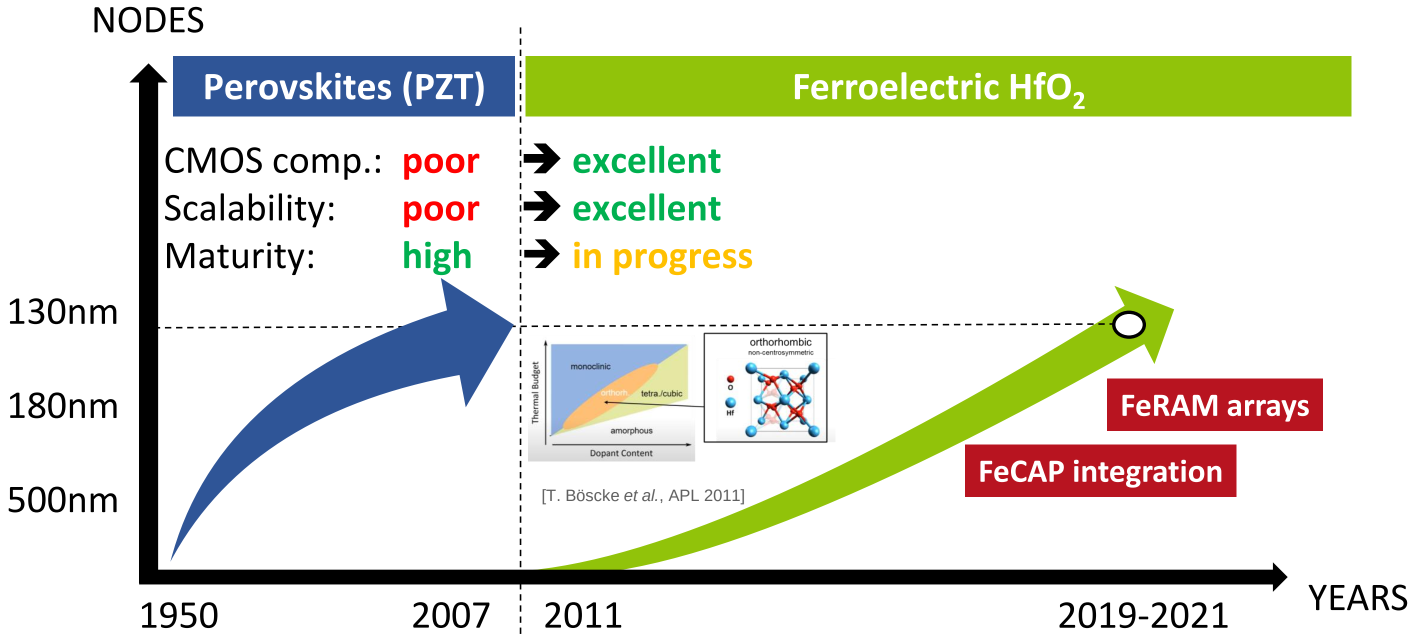
BENCHMARK

› NVM

	NOR FLASH	MRAM	PCRAM	OxRAM	FeRAM (PZT)	FeRAM (HfO_2)
Programming power	~200pJ/bit	~20pJ/bit	~300pJ/bit	~100pJ/bit	**~100fJ/bit**	**~100fJ/bit**
Write speed	20 µs	**20 ns**	**10-100 ns**	**10-100 ns**	<100ns	**14ns @ 2.5V (Sony) 4ns @ 4.8V (Leti)**
Endurance	10^5 - 10^6	**10^6-10^{15}**	10^8	10^5 – 10^6 on 16 kbit	**> 10^{15}**	**> 10^{11} single device 10^9 on 16 kbit**
Retention	> 125°C	85°C - 165 °C	**165°C**	> 150°C	125°C	125°C – SMT compliant
Extra masks	Very high (>10)	Limited (3-5)	Limited (3-5)	**Low (2)**	**Low (2)**	**Low (2)**
Process flow	Complex	Medium	Medium	**Simple**	**Simple**	**Simple**
Multi-Level Cell	Yes	No	**Yes**	**Yes**	No	No
Scalability	Bad	Medium	**High**	**High**	Poor (130nm)	**Poor (2D) High (3D)**

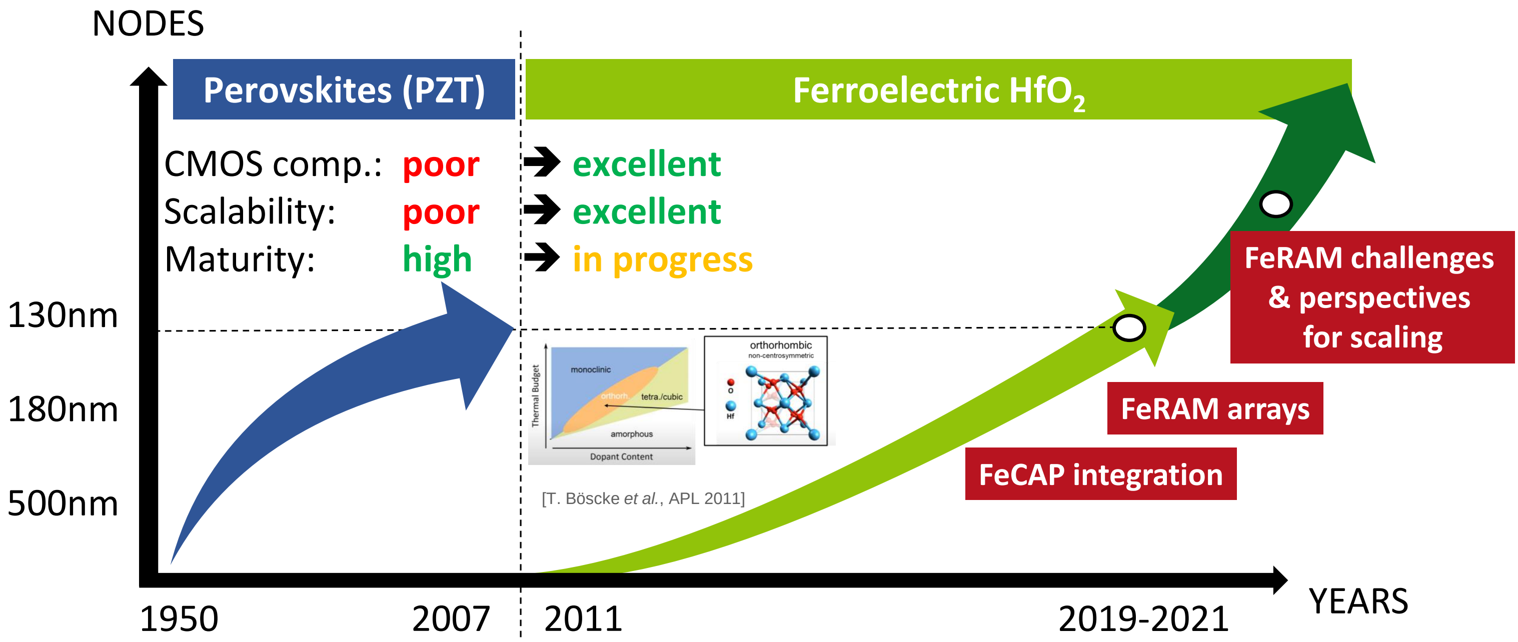
HfO_2-BASED FERAM ARRAYS

› 130nm demonstrations since 2020



HfO_2-BASED FERAM ARRAYS

› Scalability towards more advanced nodes?



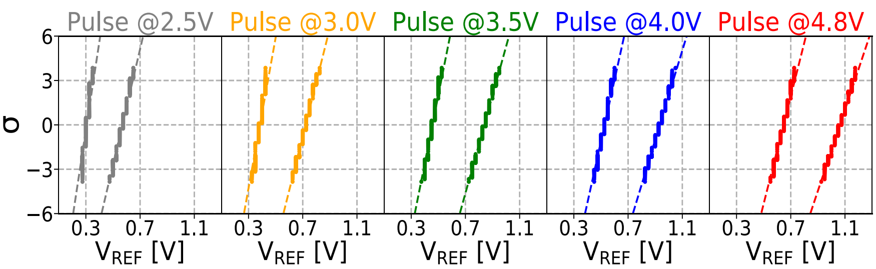
OUTLINE

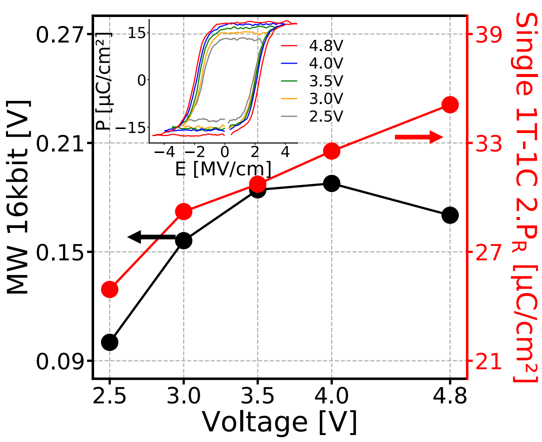
- Ferroelectricity basics
- Ferroelectric HfO_2: a change of paradigm for NVM
- 1T-1C FeRAM arrays: basics
- HfO_2-based MFM capacitors integrated aboveCMOS
- HfO_2-based 1T-1C FeRAM arrays: performance overview
- **Scalability: challenges and perspectives**

CHALLENGES AND PERSPECTIVES FOR FERAM SCALING

› Voltage scaling

Si:HfO_2 16 kbit FeRAM
4.8V ➔ 2.5V programming voltage





[T. Francois *et al.*, IEDM 2021]

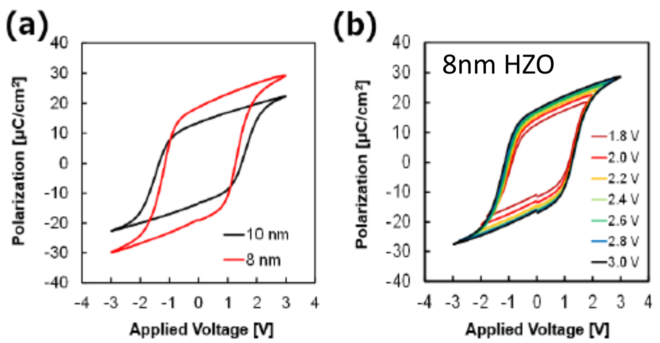
Low voltage operation reduces MW

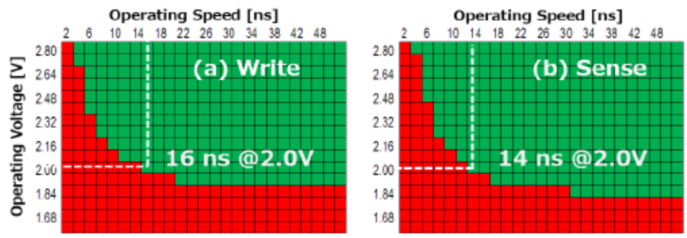
leti cea tech

CHALLENGES AND PERSPECTIVES FOR FERAM SCALING

› Voltage scaling

10nm HZO → 8nm HZO





[J. Okuno *et al.*, IMW 2021]

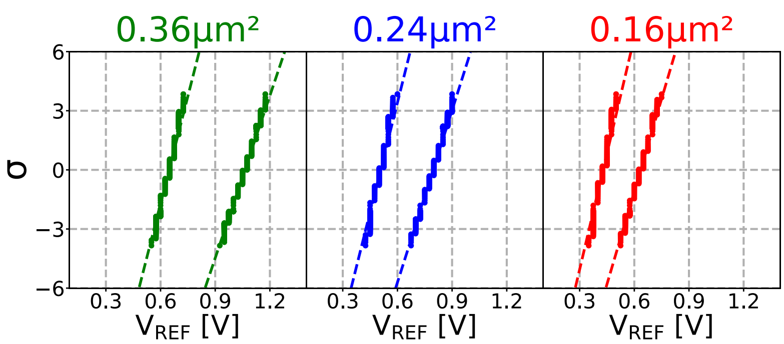
Thickness scaling of ferroelectric layer from 10nm down to 8nm enables operation down to 2V

leti cea tech

CHALLENGES AND PERSPECTIVES FOR FERAM SCALING

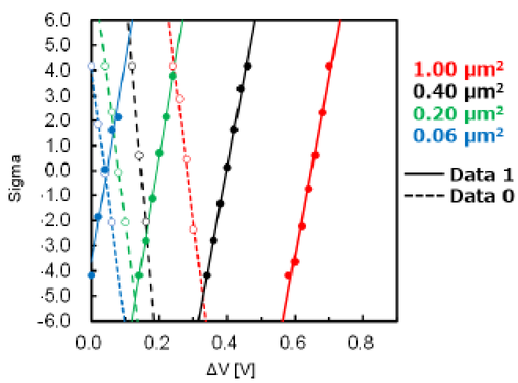
› Area scaling

10nm Si:HfO$_2$ 16 kbit FeRAM
0.36µm² → 0.16µm² FeCap



[T. Francois *et al.*, IEDM 2021]

8nm HZO 64 kbit FeRAM
1µm² → 0.06µm² FeCap



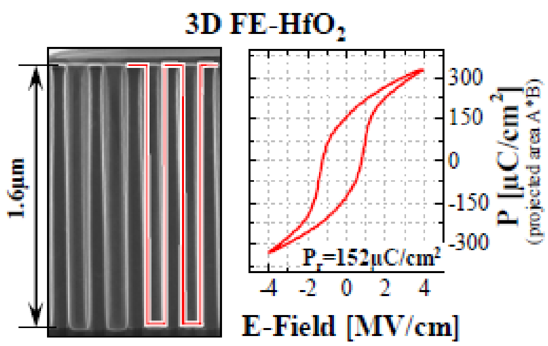
[J. Okuno *et al.*, IMW 2021]

Scaling FeCap area reduces MW

CHALLENGES AND PERSPECTIVES FOR FERAM SCALING

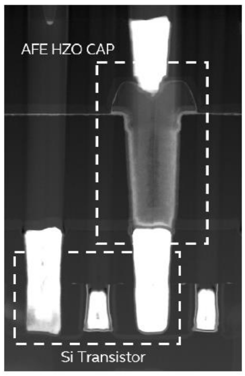
› Area scaling

3D Al:HfO_2 FeCap, aspect ratio = 13:1



[P. Polakowski *et al.*, IMW 2014]

3D HZO anti-ferroelectric capacitor above FinFET

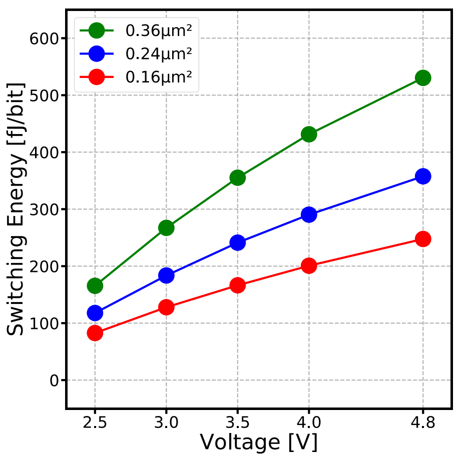


[S. -C. Chang *et al.*, IEDM 2021]

Maintaining large FeCap area while reducing footprint is possible using 3D capacitors

CHALLENGES AND PERSPECTIVES FOR FERAM SCALING

› Voltage & Area scaling



[T. Francois *et al.*, IEDM 2021]

Voltage scaling & FeCap area scaling results in switching energy lower than 100 fJ/bit

CHALLENGES AND PERSPECTIVES FOR FERAM SCALING

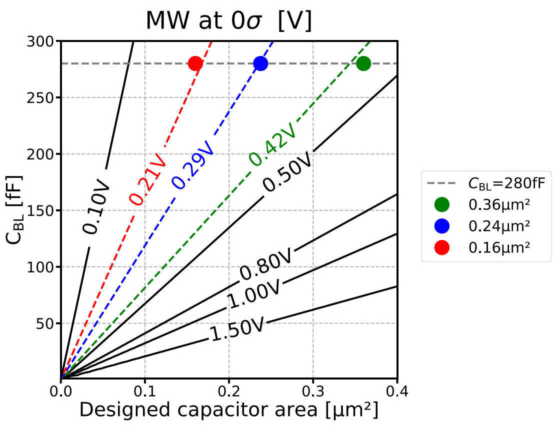
› Area scaling – 1T-1C bitcell

In FeRAM : total charge Q_{diel} + Q_{ferro} that needs to flow through the 1C during switching corresponds to < 100µA (for 1ns)
➔ 1T footprint can be small

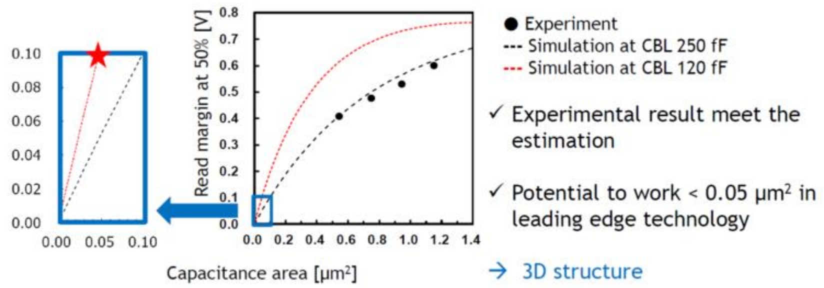
Main diffentiator w.r.t. resistive memories (OxRAM, PCM, …) for which selector needs to drive > 100 µA

CHALLENGES AND PERSPECTIVES FOR FERAM SCALING

› Bitline Capacitance scaling



[T. Francois *et al.*, IEDM 2021]



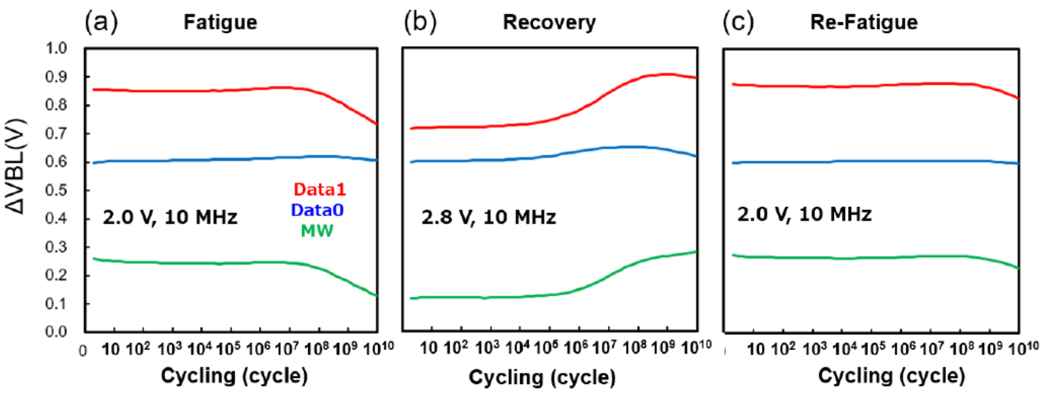
[J. Okuno *et al.*, VLSI 2020]

C_{BL} expected to decrease for nodes beyond 130nm
➔ allows to improve MW at fixed capacitor area
➔ allows to decrease capacitor area at fixed MW

CHALLENGES AND PERSPECTIVES FOR FERAM SCALING

› Endurance

HZO-based FeRAM arrays (4 kbit)



[J. Okuno *et al.*, EDTM 2022]

Encouraging endurance results reported at array level (> 10^{10} cycles) with recovery phase
Better understanding of role of defects needed at the fundamental level, to reduce wake-up and fatigue.

OUTLINE

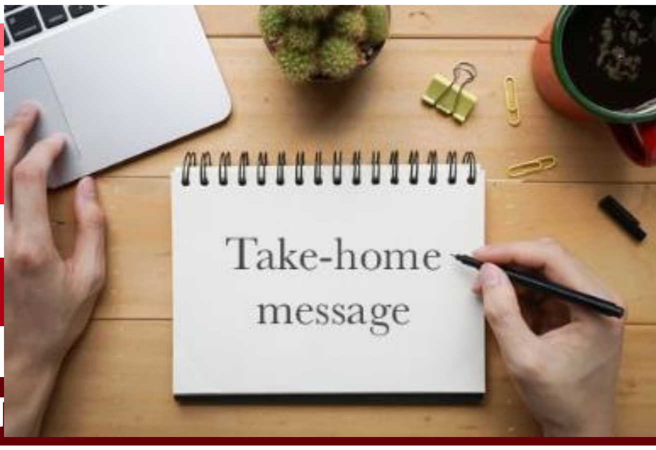
Ferroelectricity basics

Ferroelectric H… …M

1T-1C FeRAM

HfO_2-based M… …IOS

HfO_2-based 1T… …verview

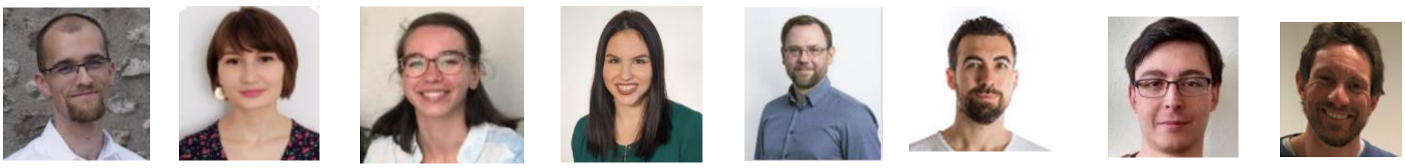


Scalability: challenges and perspectives

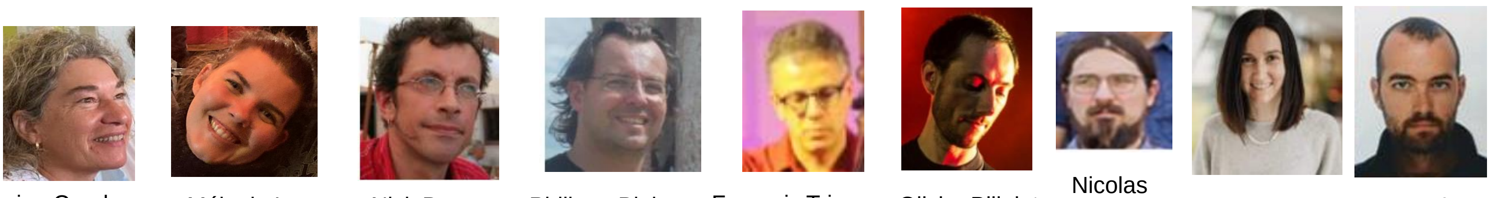
FERROELECTRIC HfO_2 : TAKE-AWAY MESSAGES FOR FeRAM

- **Ferroelectricity in HfO_2 was unveiled 10 years ago**
- **Ferroelectric HfO_2 currently attracts a lot of attention for non-volatile memory applications:**
  - CMOS compatibility
  - Scalability
  - Ultra low power
  - Easy integration
- **HfO_2-based ferroelectric ramdom access memories (FeRAM) are already demonstrating excellent performances at the array level (16kbit – 64 kbit) at 130nm**
  - High speed operation < 20ns
  - Low voltage operation < 2.5V
  - Ultra low energy < 100 fJ/bit
  - Endurance > 10^9 cycles
  - Data retention at 125°C
  - Solder Reflow Compatilibity
- **Material and stack improvement and better understanding needed (wake-up, fatigue, imprint …)**
- **Destructive reading requires very high endurance and prevents multi-level capability**
- **HfO_2-based FeRAM scaling to node < 130 nm is envisioned**

SPECIAL THANKS TO CEA TEAM …



Terry François, Justine Barbot, Julie Laguerre, Jennifer Izaguirre, Adam Makosiej, Bastien Giraud, Simon Martin, Niccolo Castellani



Catherine Carabasse, Mélanie Louro, Nick Barrett, Philippe Blaise, François Triozon, Olivier Billoint, Nicolas Vaxelaire, Elisa Vianello, Jean Coignus

&

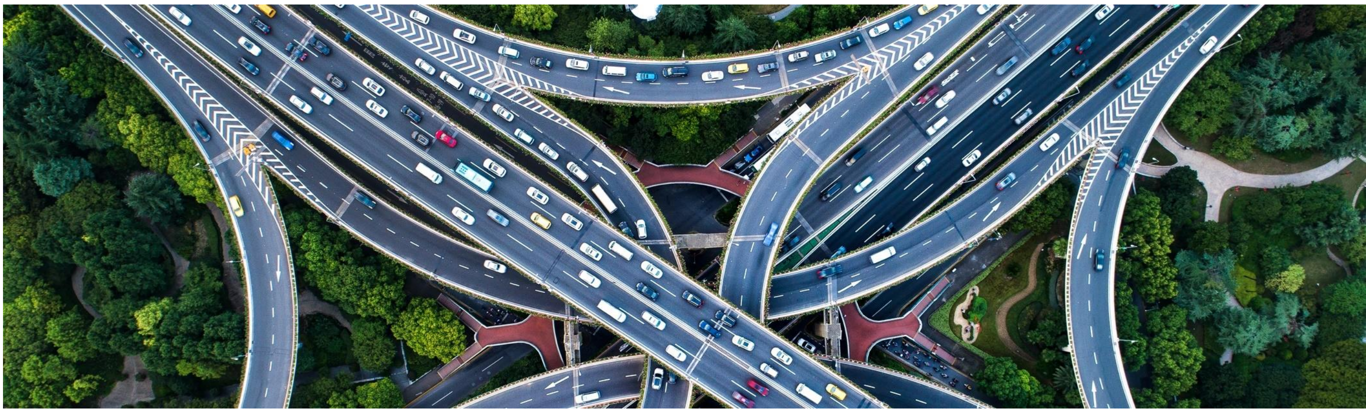
Thomas Magis, Catherine Pellissier, Massoud Bedjaoui, Virginie Beugin, Philippe Rodriguez, Guillaume Rodriguez, Sébastien Kerdilès, Hélène Grampeix, Virginie Loup, Pierre-Marie Deleuze, Tifenn Hirtzlin, Filippo Moro, Wassim Hamouda, François Aussenac, Chiara Sabbione, Magali Tessaire, Fred Mazen, Marianne Coig, Olivier Renault, Etienne Nowak, Caroline Coutier, François Andrieu, Thierry Poiroux, Julien Arcamone

… AND TO YOU FOR YOUR ATTENTION!

QUESTIONS?

Halid Mulaosmanovic
GlobalFoundries

Halid Mulaosmanovic received the Ph.D. in Information Technology at Politecnico di Milano, Italy, in 2016. He worked as a research fellow at NaMLab, Dresden, Germany, from 2016 to 2021, where his research interests included ferroelectric materials and devices, with a particular focus on ferroelectric field-effect transistors for memory and unconventional applications. Now, he is with GlobalFoundries Inc., Germany, and is involved in ferroelectric memory projects among others.



Ferroelectric FETs

Halid Mulaosmanovic

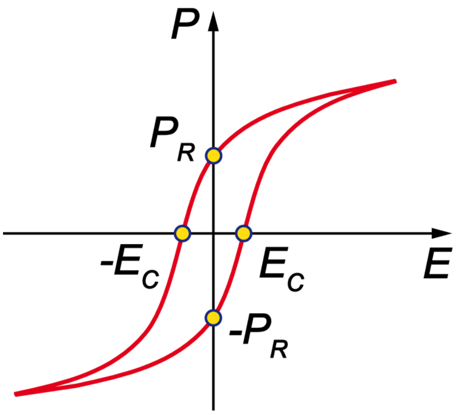
imw 2022

GlobalFoundries™

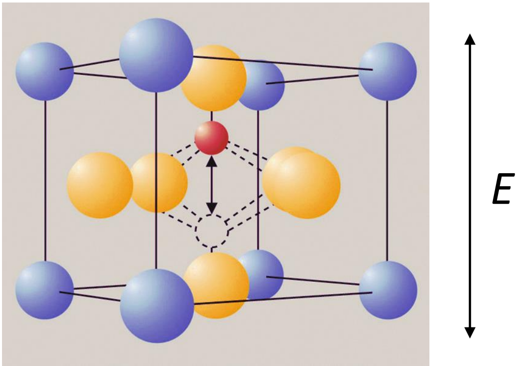
- **Introduction to Ferroelectric FETs**
- **Ferroelectric HfO_2**
- **Device characteristics**
  - **Memory window**
  - **Switching kinetics**
  - **Size dependence**
  - **Reliability**
  - **(Co)-Integration**
- **Ferroelectric FETs beyond memory**
- **Conclusions**

- Introduction to Ferroelectric FETs
- Ferroelectric HfO_2
- Device characteristics
  - Memory window
  - Switching kinetics
  - Size dependence
  - Reliability
  - (Co)-Integration
- Ferroelectric FETs beyond memory
- Conclusions

Ferroelectricity



perovskite oxide (ABO_3):
PZT, SBT, BTO ...

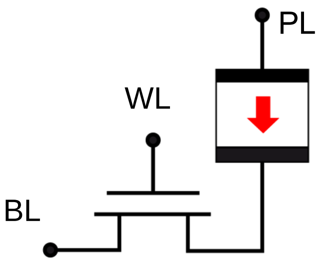


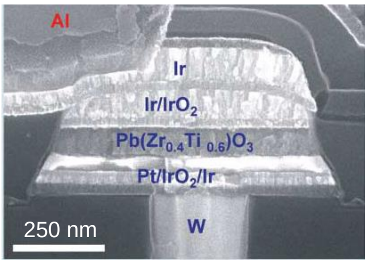
Courtesy of Fujitsu Semiconductor

- Polarization is switched between two equivalent states by an external electric field
- Reversibly switchable permanent dipoles appealing for information storage

Ferroelectric memory

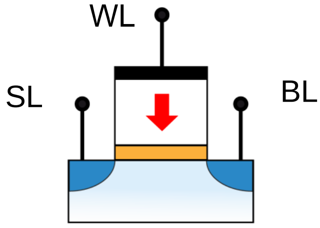
1T1C FeRAM

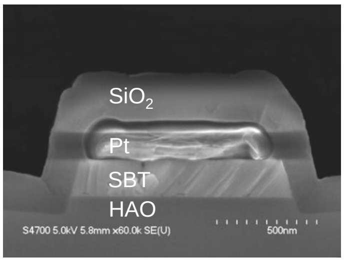




J. F. Scott, Science, 2007

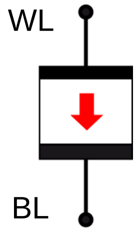
1T FeFET

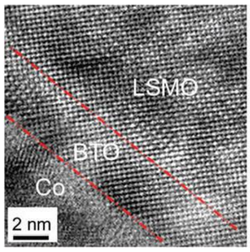




L. V. Hai, Semi. Sci. Tech., 2010

1C FTJ

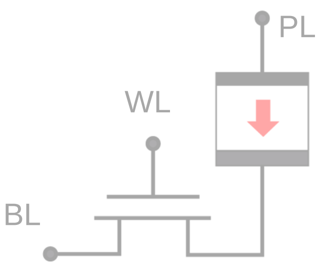


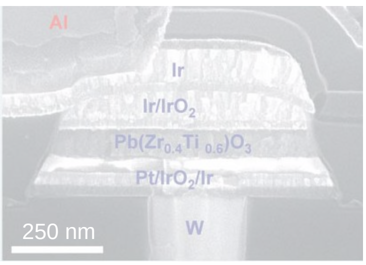


H. J. Mao, RCS PCCP, 2015

Ferroelectric memory

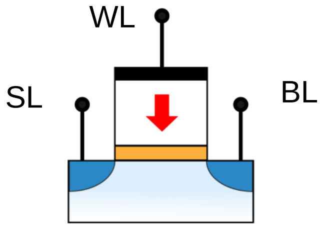
1T1C FeRAM

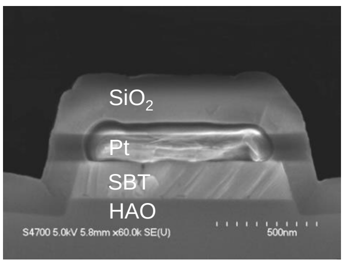




J. F. Scott, Science, 2007

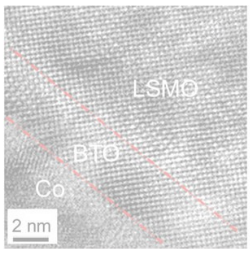
1T FeFET





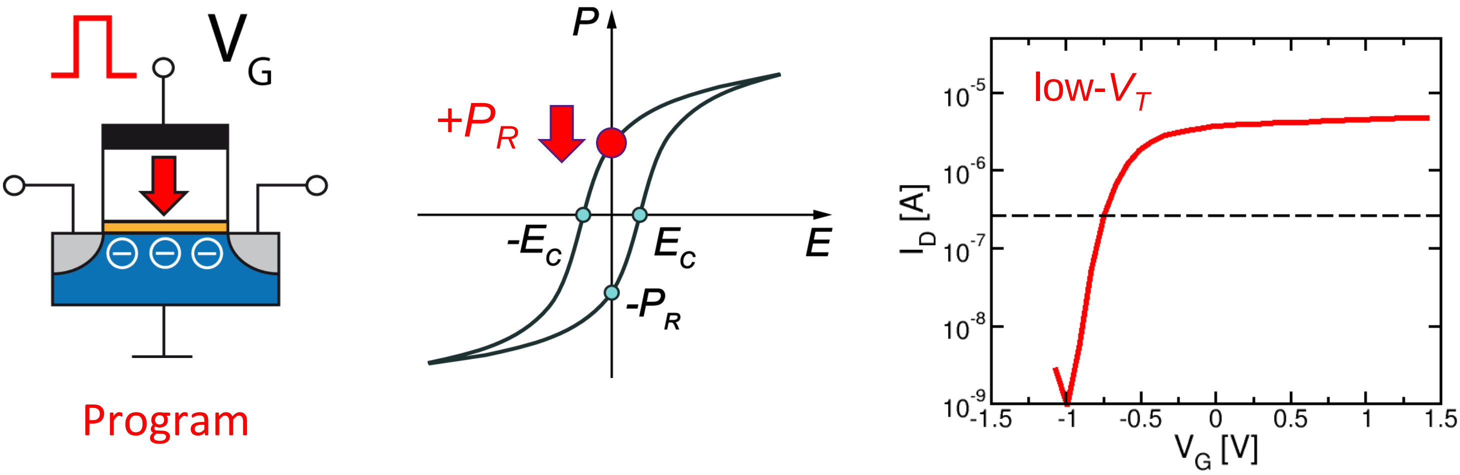
L. V. Hai, Semi. Sci. Tech., 2010

1C FTJ



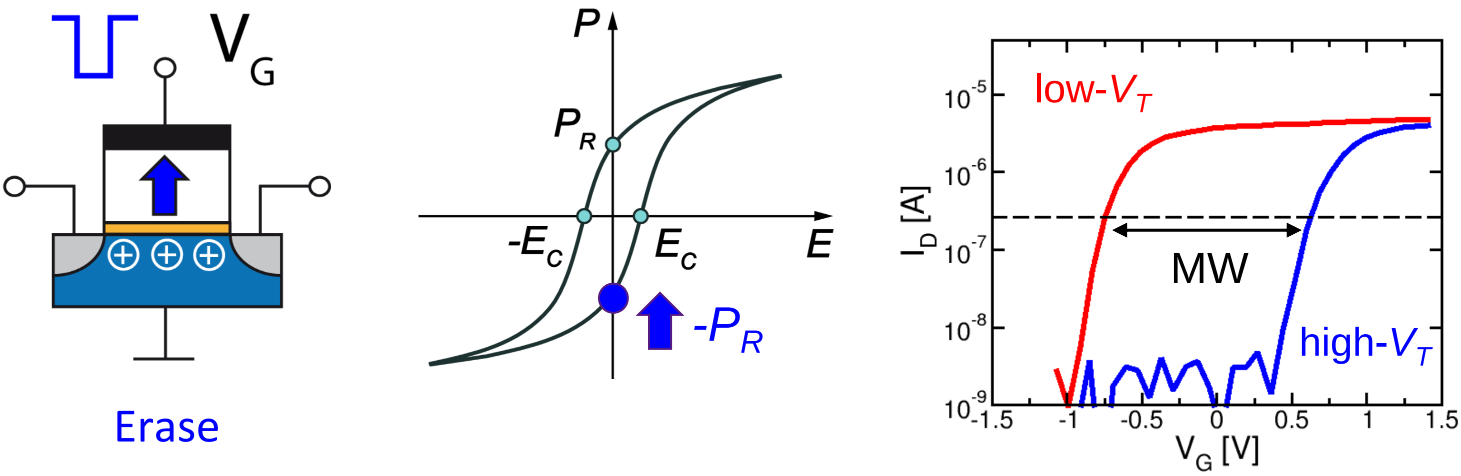
H. J. Mao, RCS PCCP, 2015

Ferroelectric FET



- Write: permanent reversal of polarization under $V_G = V_G$ ($|E_F| > |E_C|$)
  → e.g. n-type FeFET: $P\downarrow$ results in **low-V_T**
- Read is nondestructive

Ferroelectric FET



- Write: permanent reversal of polarization under $V_G = V_G$ ($|E_F| > |E_C|$)
  → e.g. n-type FeFET: $P\uparrow$ results in **high-V_T**
- Read is nondestructive

An old idea

United States Patent Office

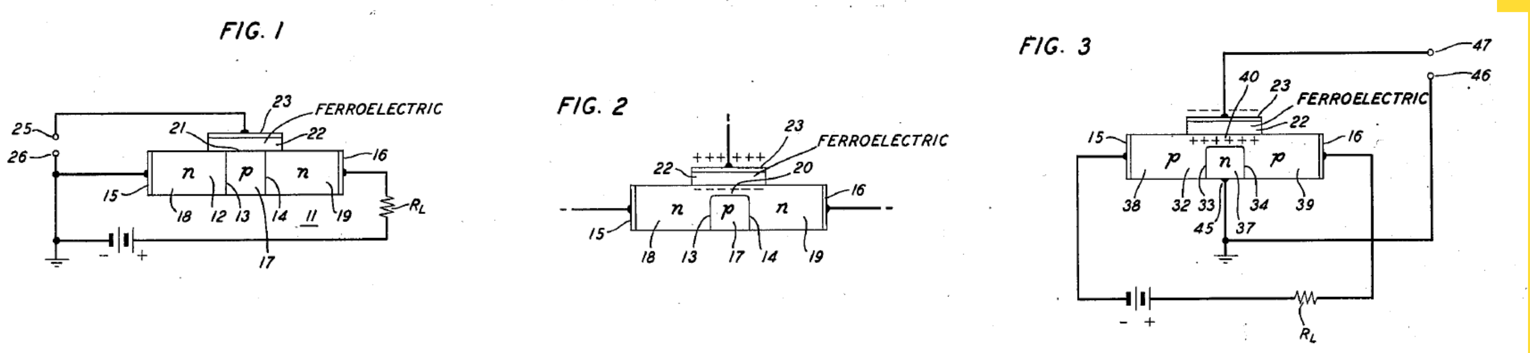
2,791,760

SEMICONDUCTIVE TRANSLATING DEVICE

Ian M. Ross, New Providence, N. J., assignor to Bell Telephone Laboratories, Incorporated, New York, N. Y., a corporation of New York

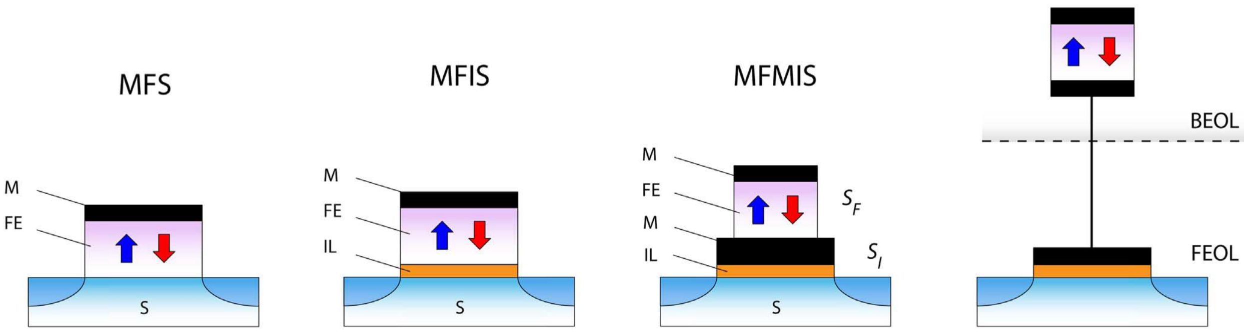
Application February 18, 1955, Serial No. 489,223

9 Claims. (Cl. 340—173)



"...one feature of this invention comprises altering the conductivity of a path through a semiconductive body by polarizing a ferroelectric maintained in proximity to the body to alter the surface charge on a portion of that body."

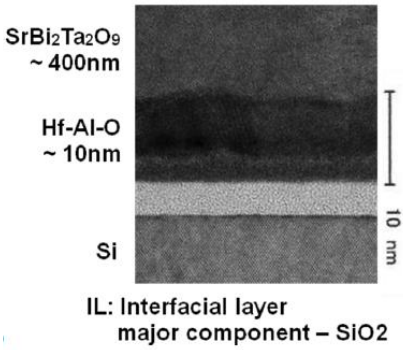
FeFET structures

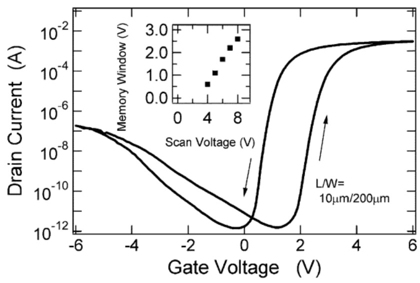


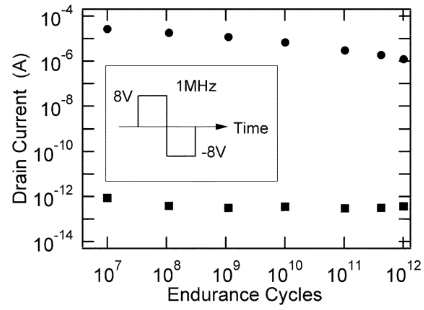
- Each of the structures has its particular advantages:
  - MFS: direct contact between FE and S → optimal voltage control of the device
  - MFIS: buffer layer between FE and S → interface quality tailoring; no interdiffusion phenomena
  - MFMIS: S_F/S_I tailoring to improve the operation voltage and memory window
- Also full-BEoL FeFETs have been demonstrated

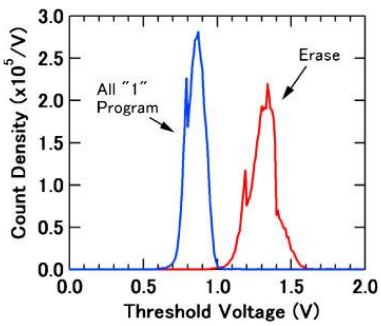
Mulaosmanovic et al., Nanotechnology, 2021

Perovskite FeFETs







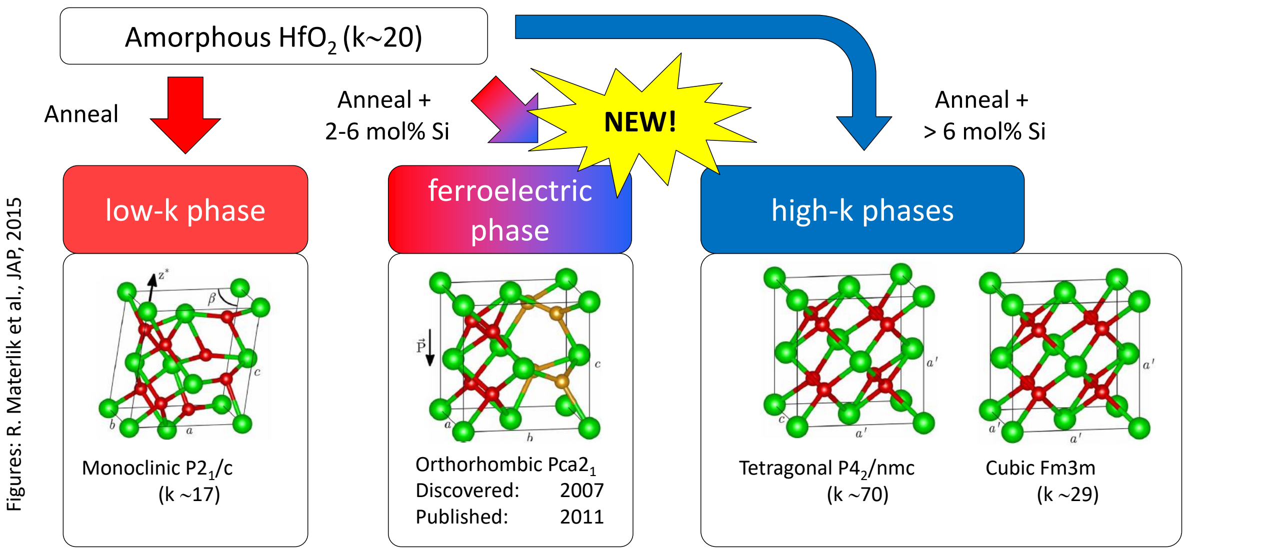


- ➢ Excellent endurance and retention
- ➢ 64kbit NAND functionality
- ➢ However, scaling and integration concerns!

Sakai et al., JJAP, 2004
Sakai et al., IEEE EDL, 2004
Zhang et al., JJAP, 2011

Ferroelectricity in HfO_2

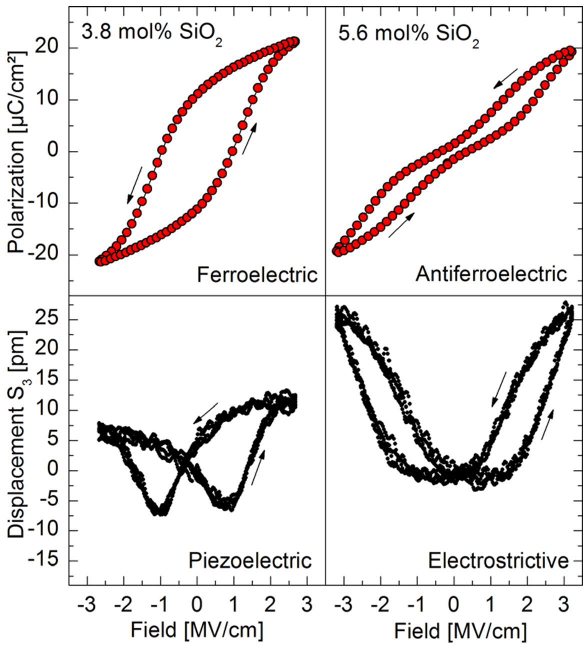
Courtesy of FMC



- ➢ Many stabilization knobs for the ferroelectric phase: doping, stress, annealing, film thickness …

Ferroelectric HfO_2

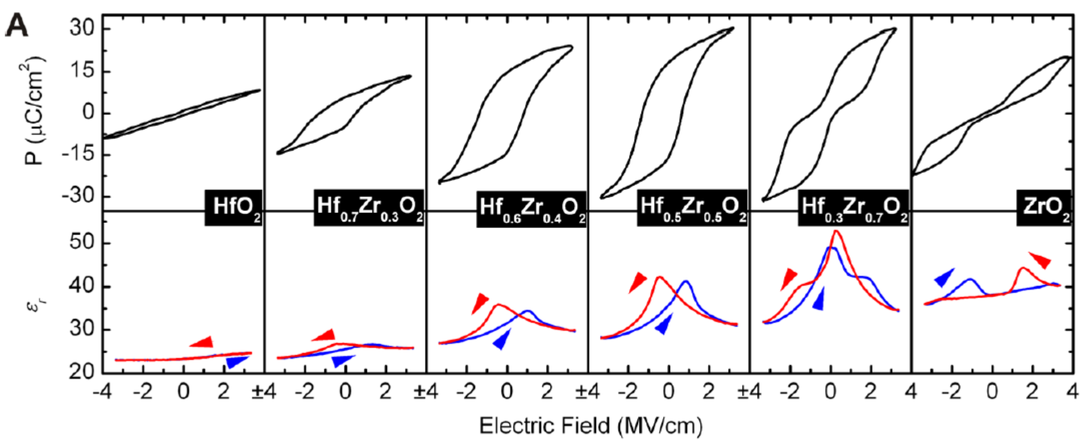
- HfO_2 is a simple binary oxide → various mature deposition techniques available
- Large bandgap (5.3-5.7 eV) → reduced leakage
- Well known high-k material in semiconductor industry → CMOS compatible
- Robust ferroelectricity even upon aggressive vertical and lateral scaling



T. S. Böscke et al., APL, 2011

Ferroelectric $Hf_{1-x}Zr_xO_2$

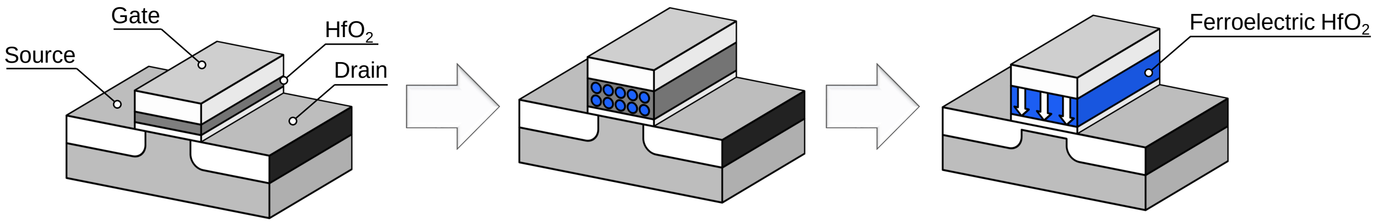
J. Müller, Nano Lett., 2012



- HfO_2 and ZrO_2 have very similar physical and chemical properties
- FE and AFE behavior has been confirmed in HfO_2 - ZrO_2 solid solution and in pure ZrO_2 as well

From MOSFET to HfO_2 FeFET

Courtesy of FMC

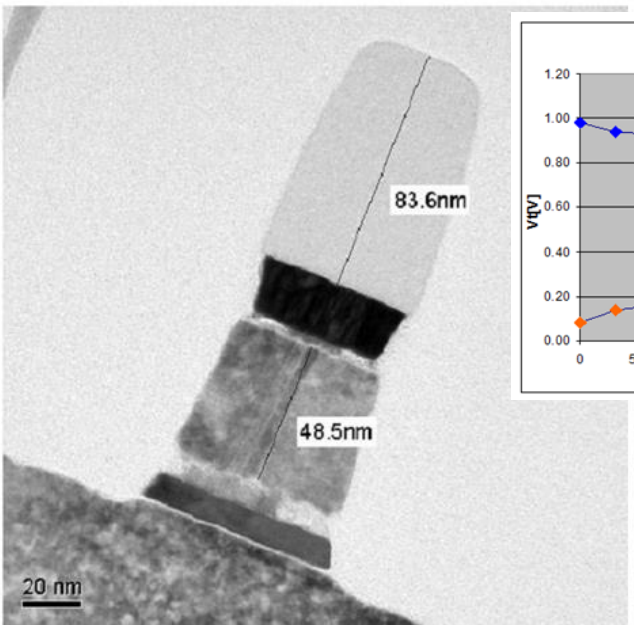


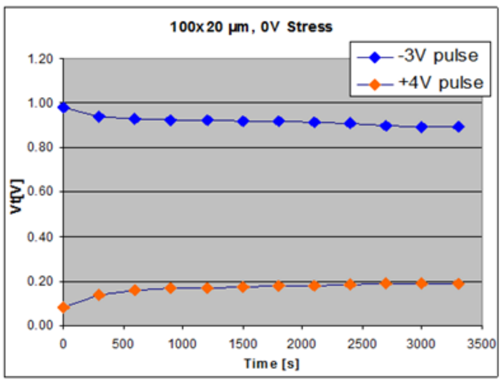
Take standard high-k metal-gate (HKMG) transistor… | … increase thickness, dope it and anneal… | … obtain nonvolatile transistor (FeFET)!

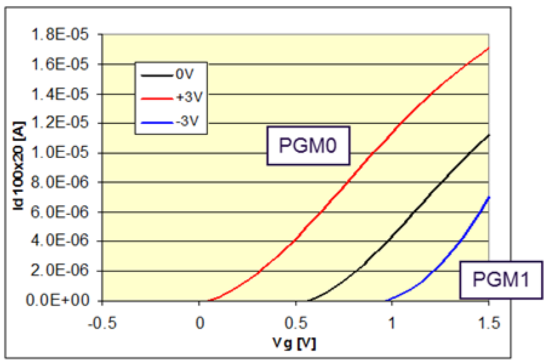
- Only a few additional masks needed for FeFET fabrication
- Full front-end CMOS compatibility make HfO_2 FeFETs attractive

2008

Ferroelectric HfSiO in a transistor



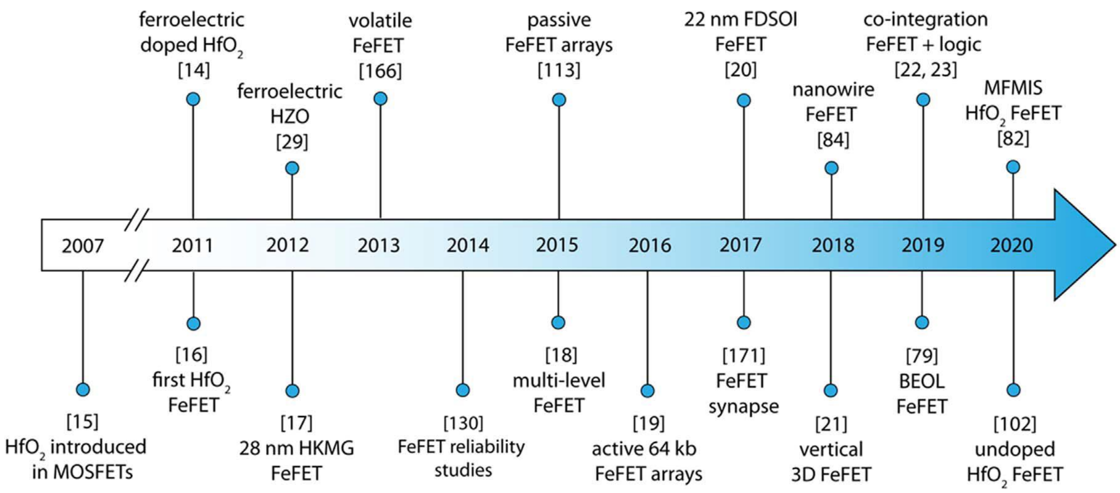




Courtesy of U. Schröder, NaMLab

- First HfO_2 based FeFET realized in 65 nm technology

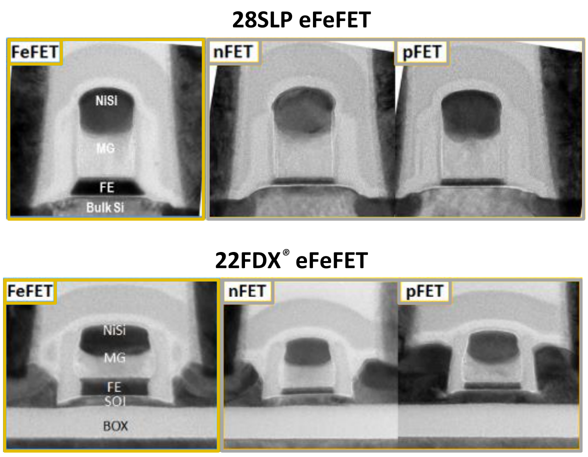
FeFET evolution



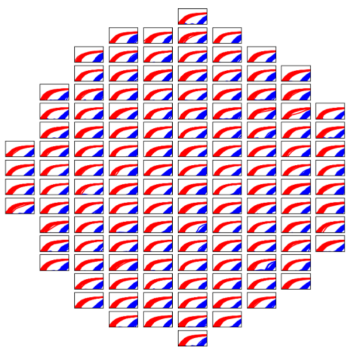
➢ Rapid progress in material and device development

Mulaosmanovic et al., Nanotechnology, 2021

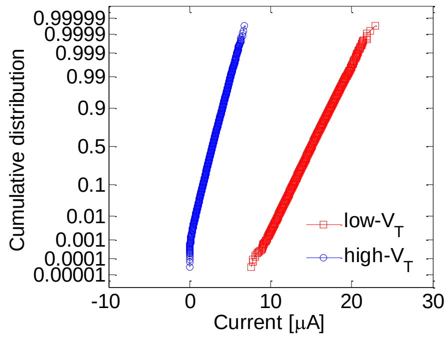
HKMG FeFET at GF



mini array devices full wafer map
W/L=450nm/450nm



64 kbit AND array
W/L=450nm/450nm



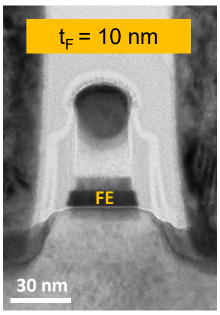
H. Mulaosmanovic, IEDM, 2015
M. Trentzsch, IEDM, 2016
S. Dünkel, IEDM, 2017
S. Beyer, IMW, 2020

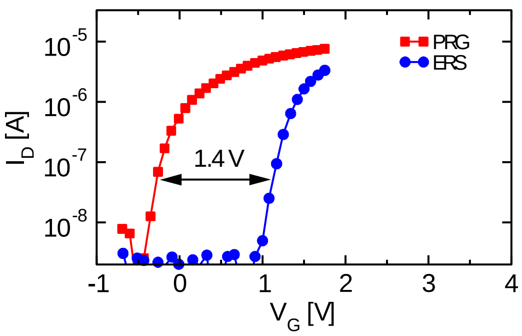
- Memory window > 1.5V
- Retention > 10 years
- Endurance: $10^4 - 10^5$
- Low operation voltages (< 4V)
- Fast access time (ns-regime)

- Full FEoL CMOS compatibility
- Fine-grained co-integration with CMOS transistors
- Only 2 structural DUV mask adder
- High scalability (L_G = 20 nm)

- Introduction to Ferroelectric FETs
- Ferroelectric HfO_2
- **Device characteristics**
  - **Memory window**
  - **Switching kinetics**
  - **Size dependence**
  - **Reliability**
  - **(Co)-Integration**
- Ferroelectric FETs beyond memory
- Conclusions

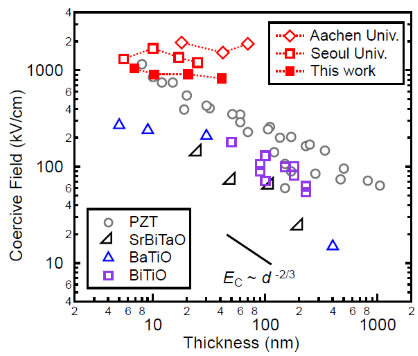
Memory window





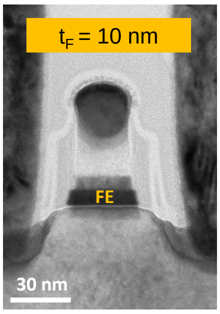
$$MW = 2 \cdot \alpha \cdot E_C \cdot t_F$$

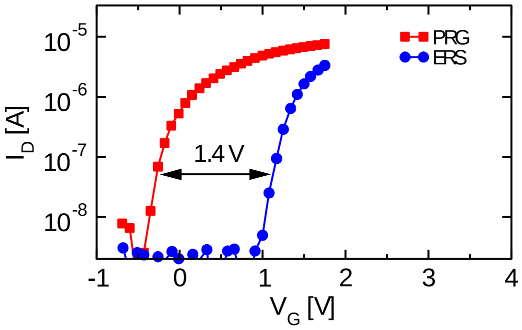
- Coercive field E_C relatively invariant
- Thickness of FE t_F can be increased

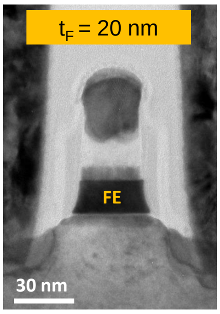


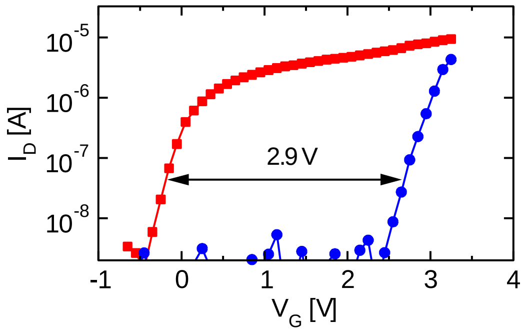
S. Migita et al., JJAP, 2018

Memory window









$$MW = 2 \cdot \alpha \cdot E_C \cdot t_F$$

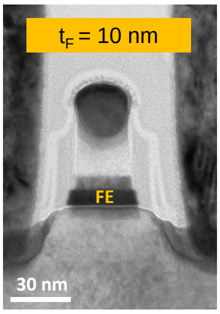
- Coercive field E_C relatively invariant
- Thickness of FE t_F can be increased

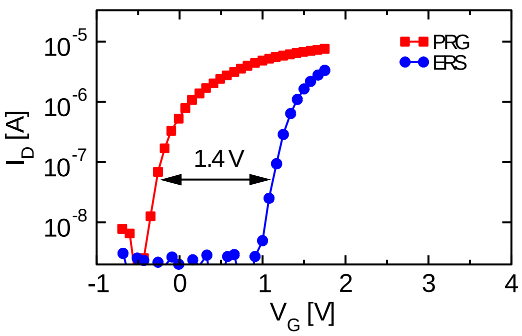
- MW up to 3 V is achieved
- Stable retention and endurance
- Possibility of multi-level storage

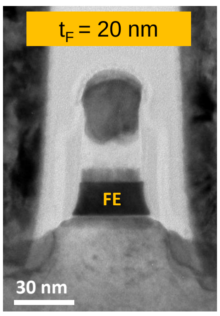
H. Mulaosmanovic, IEEE T-ED, 2019

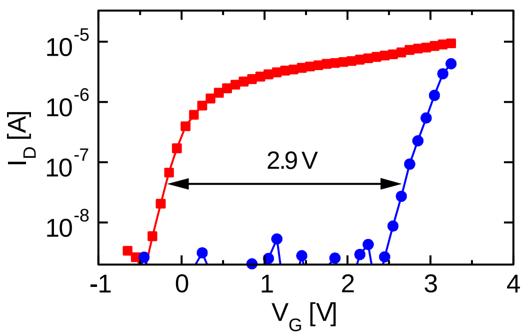
Memory window

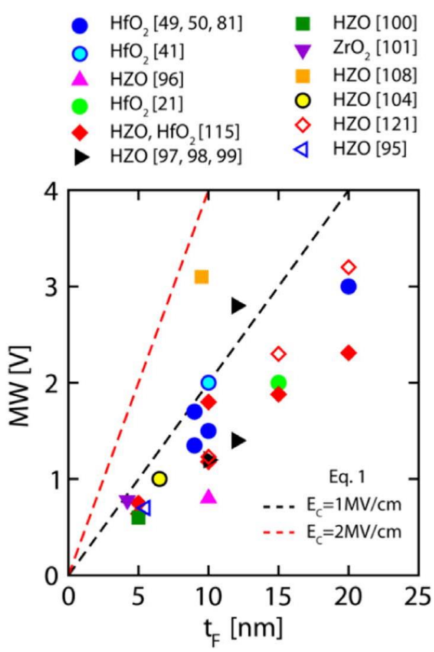
$$MW = 2 \cdot \alpha \cdot E_C \cdot t_F$$





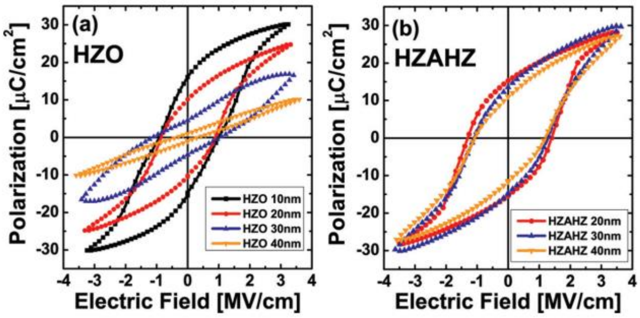




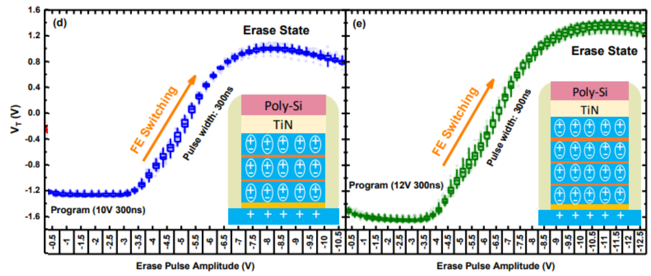


Mulaosmanovic et al., Nanotechnology, 2021

Memory window



Kim et al., APL, 2014

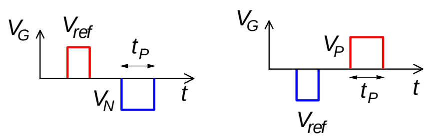


Ali et al., IEDM, 2019

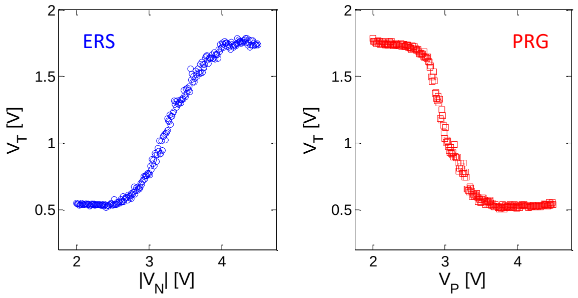
- Ferroelectric properties tend to rapidly degrade at higher film thicknesses
- Insertion of interlayers (e.g. AlO_x) may contrast this degradation
- Penalty: integration complexity and larger operation voltage (e.g. 12 V)

Switching kinetics

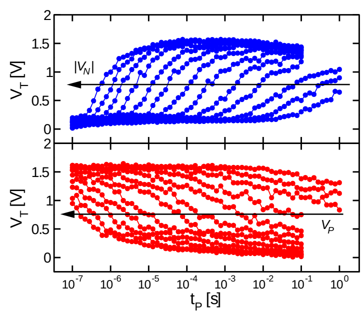
Mulaosmanovic et al, IEEE T-ED 67, 5804 (2020)



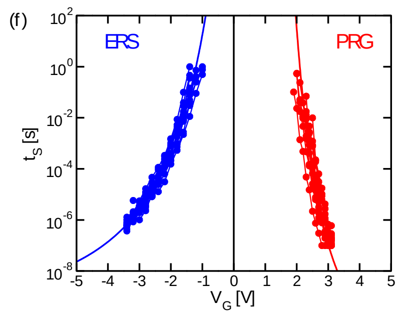
Amplitude variation



Pulse width variation

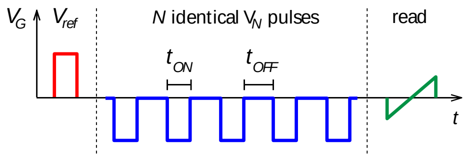


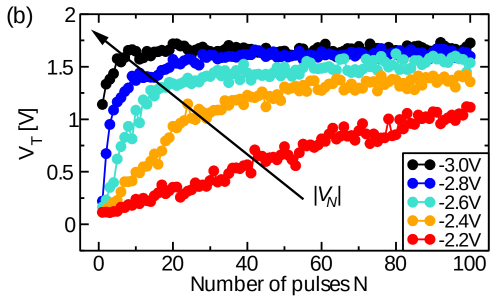
$$t_S = t_0 \exp\left[\frac{\alpha}{k_B T} \cdot \frac{1}{(V_G - V_0)^2}\right]$$

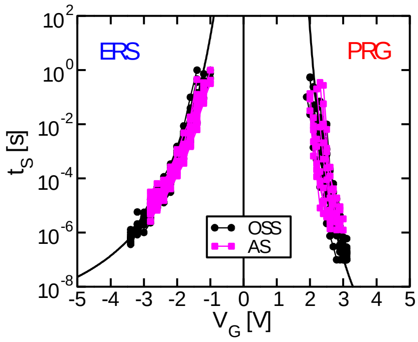


- Significant time-voltage switching dependency
- Trade-off: fast switching → larger amplitudes
- Sub-nanesecond switching demonstrated

Accumulative switching







Mulaosmanovic et al, ACS AMI 10, 23997 (2018)
Mulaosmanovic et al, IEEE T-ED 67, 5804 (2020)

- ➢ FeFETs undergo switching even upon sub-critical voltage pulses
- ➢ Accumulative effect demonstrated over a broad range of electrical conditions
- ➢ Same physical laws governing both one-shot and accumulative switching

Size Dependency

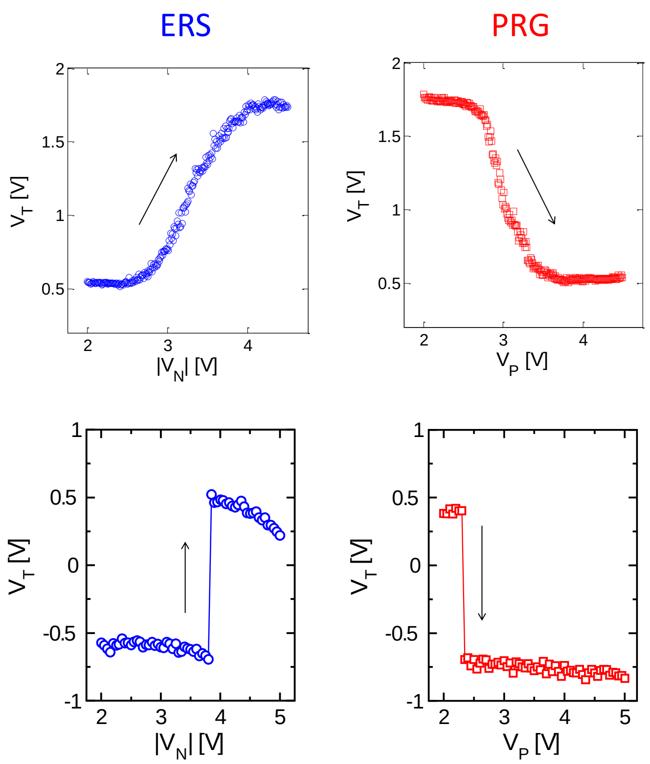
➢Large FeFETs:

- ***W* = 1 µm, *L* = 1 µm**
- Gradual switching between 2 states
- > 64 intermediate V_T states

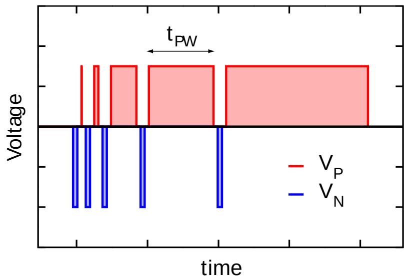
➢ Ultra-scaled FeFETs:

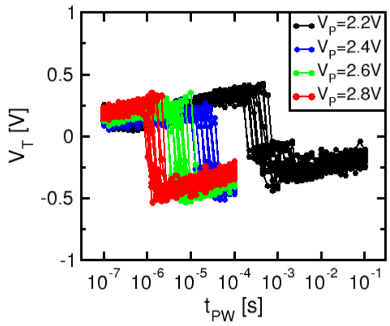
- ***W* = 80 nm, *L* = 30 nm**
- Abrupt switching between 2 states
- Apparently, no intermediate states

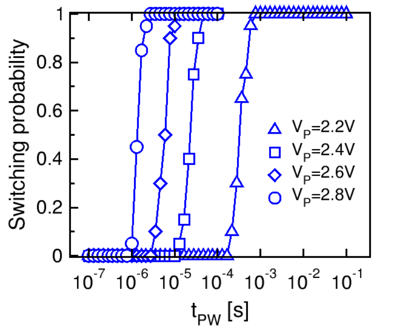
H. Mulaosmanovic et al., IEDM, 2015
H. Mulaosmanovic et al., EDTM, 2020



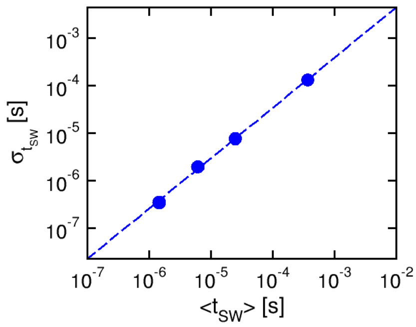
Stochastic switching







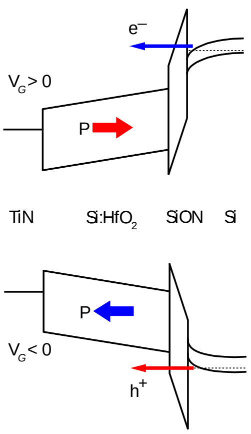
- ➢ Time-voltage trade-off for ferroelectric switching
- ➢ Switching is a stochastic process!
- ➢ Unity slope over several decades in the mean t_{SW} vs. standard deviation (σ_{tSW}) plot → Poisson process



H. Mulaosmanovic et al., ACS AMI, 2017

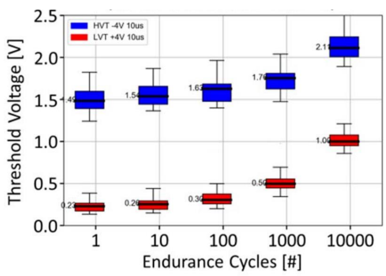
Reliability

Charge trapping

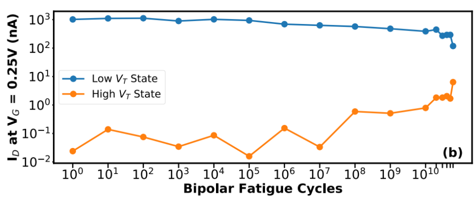


- ➢ Screening of MW
- ➢ Long read latency
- ➢ Degradation of IL
- ➢ Endurance walk-out

Cycling Endurance



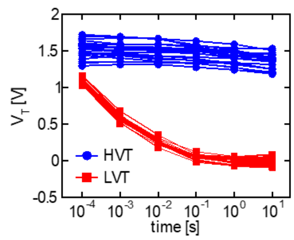
- ➢ Usually < 10^6 cycles



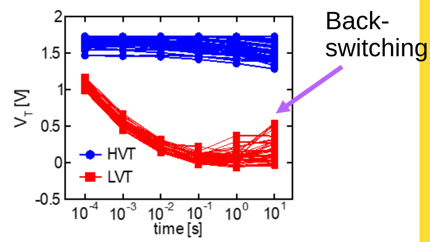
A. Tan, IEEE EDL, 2021

- ➢ But, some reports with > 10^{10} cycles avalable

Data Retention

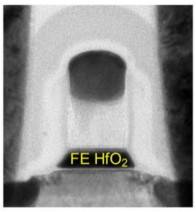


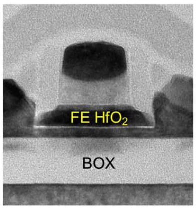
- ➢ Proper stack design → robust retention, even at T > 250°C
- ➢ Improper stack → depolarization



Integration

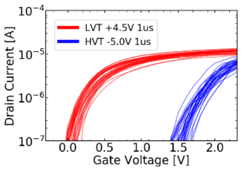
Planar

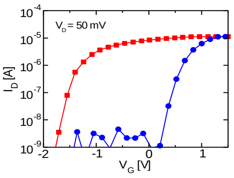




28 nm bulk

22 nm FDSOI

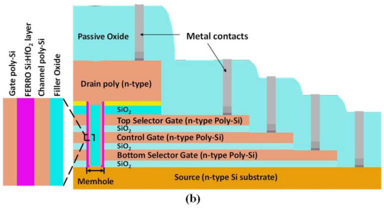


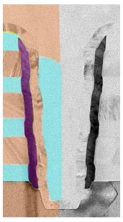


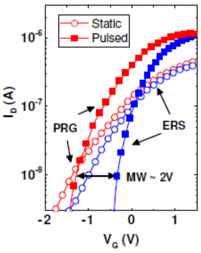
Trentzsch, IEDM, 2016

Dünkel, IEDM, 2017

Vertical 3D

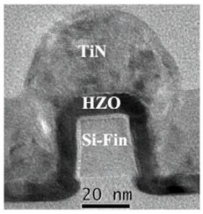


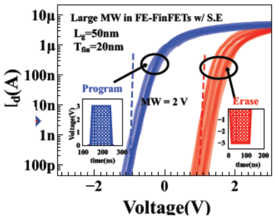




K. Florent, IEDM, 2018

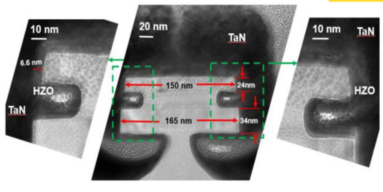
Fin-FeFET

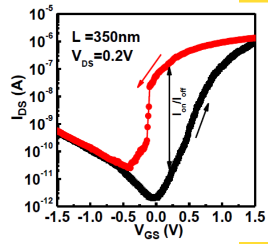




S. De, VLSI, 2021

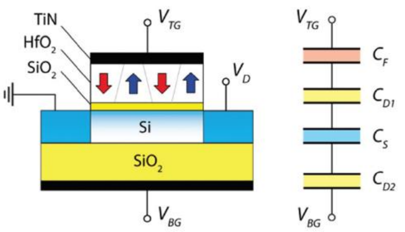
Omega-FeFET

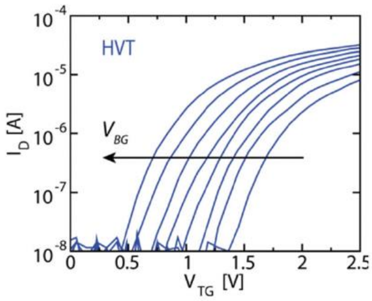


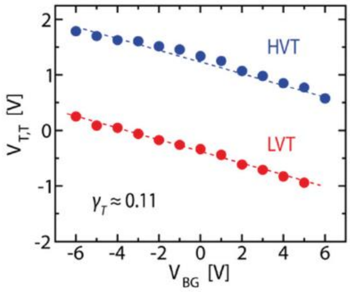


Chen, IRPS, 2020

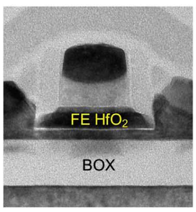
22 FDSOI FeFET: add-on functionality







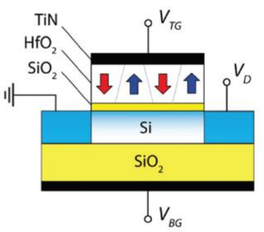
$$|\Delta V_{\mathrm{T,T}}| = \gamma_{\mathrm{T}} |\Delta V_{\mathrm{BG}}|$$

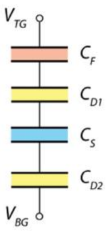


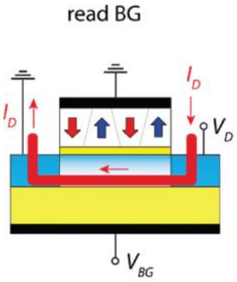
- Back bias enables independent tuning of V_T (HVT and LVT)
- No polarization disturb
- Targeting option for read out

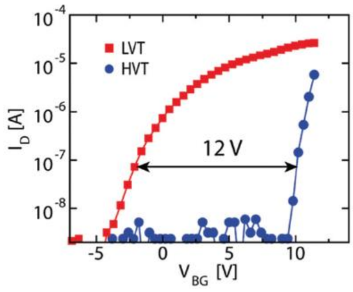
H. Mulaosmanovic, Nanoscale, 2021

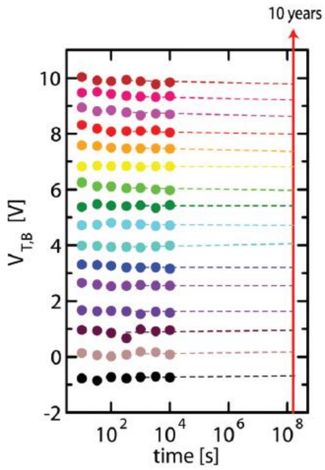
22 FDSOI FeFET: add-on functionality











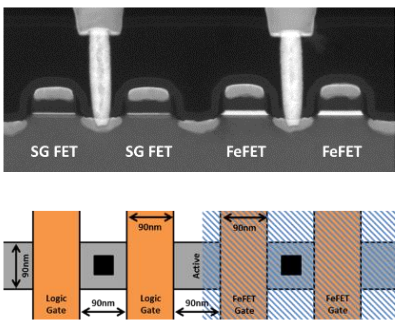
$$\gamma_{\mathrm{B}} \equiv \frac{|\Delta V_{\mathrm{T,B}}|}{|\Delta V_{\mathrm{TG}}|}$$

4 bits/cell

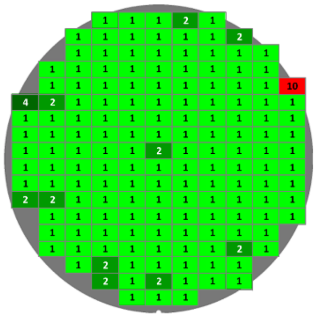
- ➢ Back-gate can act as an independent read-out terminal
- ➢ Artificial increase of MW up to 12 V
- ➢ Enables easy V_T distinction → 4 bit/cell storage
- ➢ Write and read paths are separated → no read disturb

H. Mulaosmanovic, Nanoscale, 2021

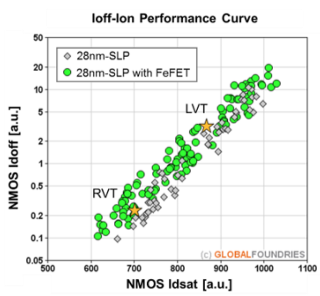
Co-Integration FeFET + CMOS

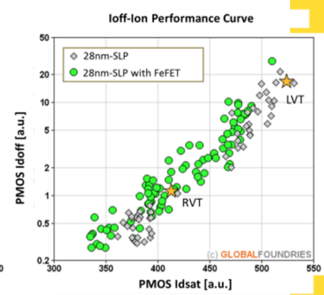


24Mb 0.120µm2 SRAM yield >90% & CMOS within 10% on 28SLPe with FeFET technology



24 Mb D120 QRAM yield





- FeFETs and logic FETs sharing the same active area
- 90 nm gate-to-gate distance

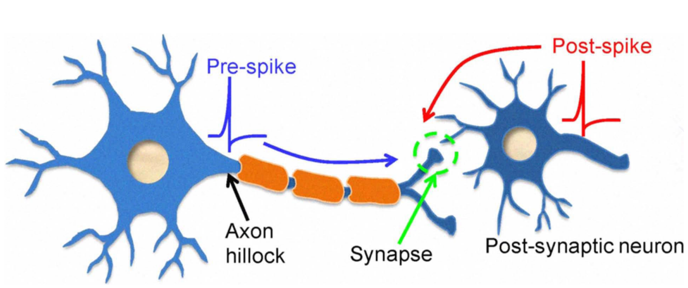
- SRAM yield comparable to the high-volume production CMOS base platform
- Device matching is within 10% of the base platform
- Can be further improved by target implants

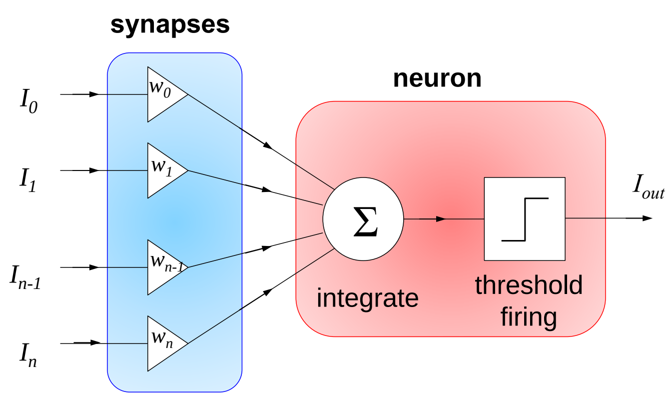
S. Beyer et al., IMW, 2020

- Introduction to Ferroelectric FETs
- Ferroelectric HfO_2
- Device characteristics
  - Memory window
  - Switching kinetics
  - Size dependence
  - Reliability
  - (Co)-Integration
- **Ferroelectric FETs beyond memory**
- Conclusions

Neuromorphic Computing

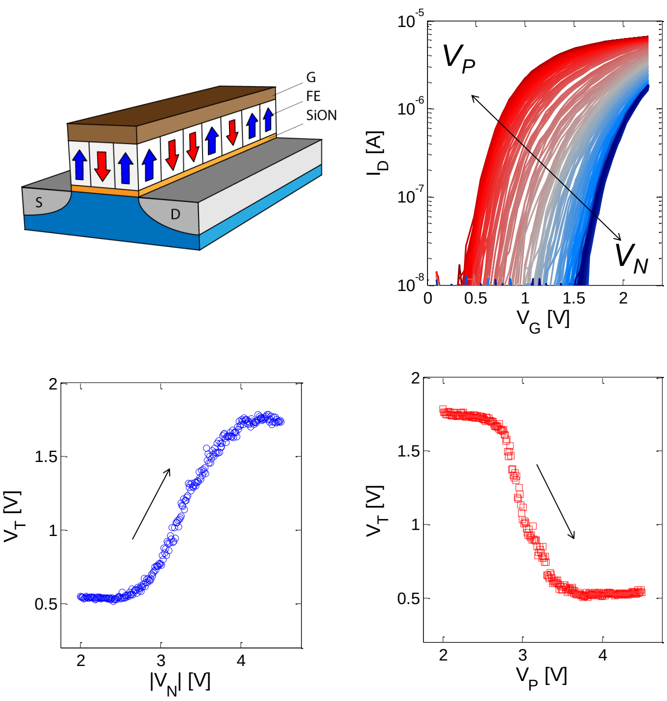
D. Kuzum et al., IEDM, 2011

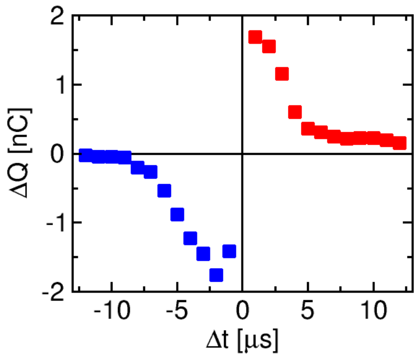




- Directly or remotely inspired by the computing in biological brains
- Main building blocks: synapses and neurons

Ferroelectric synapses



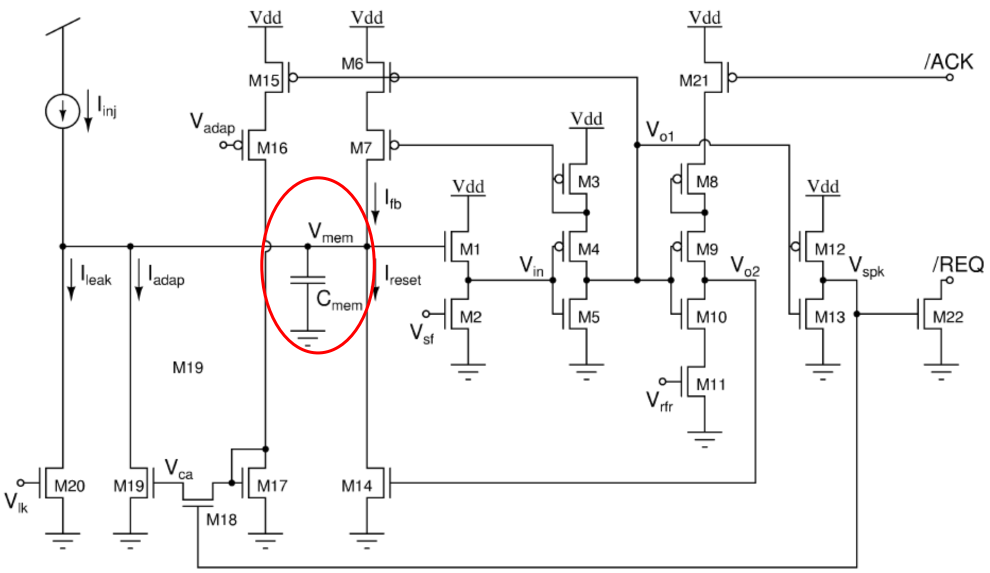


- Polycrystalline HfO_2 → multi-domain ferroelectric
- Gradual switching → analog conductivity tuning
- Spike-timing dependent plasticity (STDP) demonstrated
- Suitable for analog weights in DNNs as well

H. Mulaosmanovic et al., VLSI, 2017

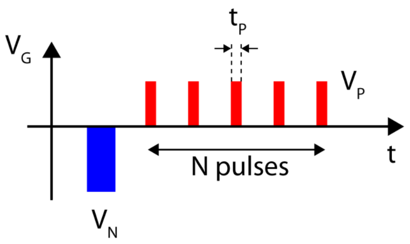
Artificial neurons with CMOS

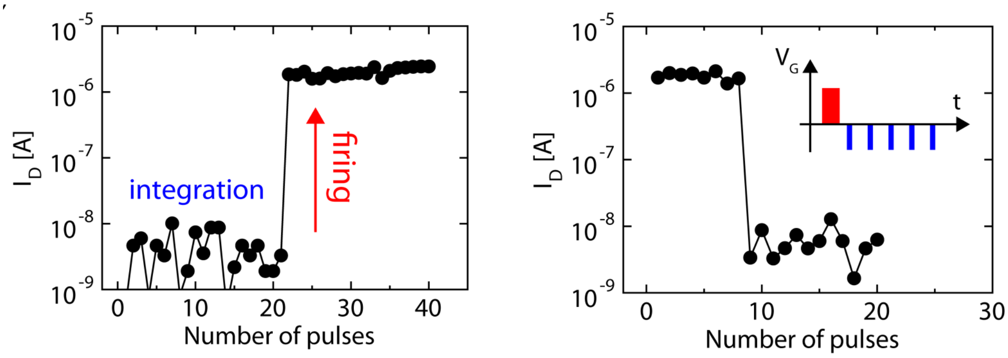
G. Indiveri et al.,
IEEE Trans. Neur. Net., 2006



- Large capacitor C_{mem} for the integration of spikes occupies a significant area
- Mimicking of additional neuronal dynamics → dramatic increase of n. of transistors

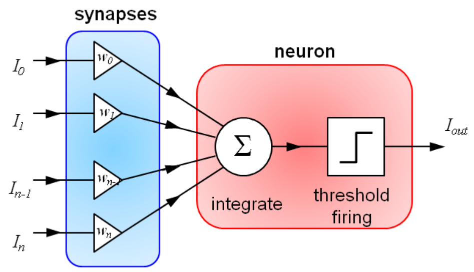
Ferroelectric neurons





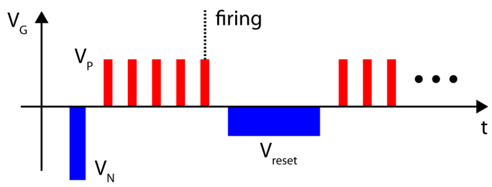
W = 80 nm; *L* = 30 nm

- ➢ Integration of gate pulses → integration of spikes coming from other neurons
- ➢ Abrupt switching → firing

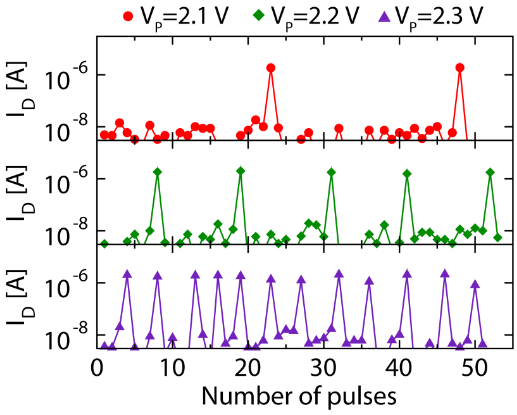


H. Mulaosmanovic et al., Nanoscale, 2018

Ferroelectric neurons



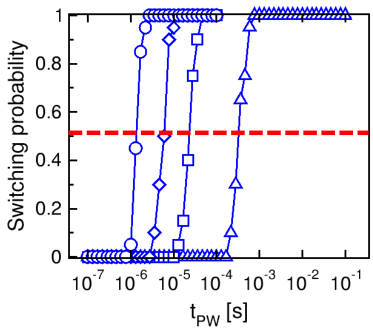
W = 80 nm; *L* = 30 nm



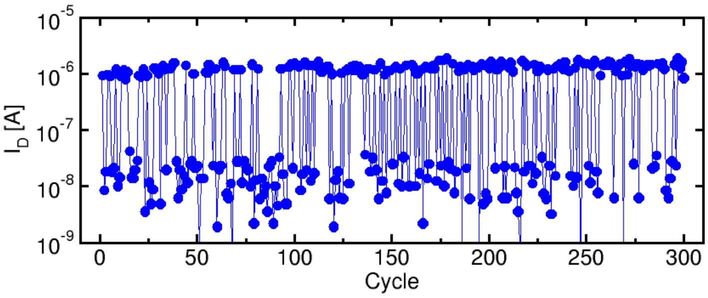
- ➢ Stronger neuronal input signals induce higher firing frequency → firing rate tuning
- ➢ Refractory period can be arbitrarily tuned over several orders of magnitude to satisfy circuital requirements (e.g. real-time as well as accelerated-time neuronal dynamics)

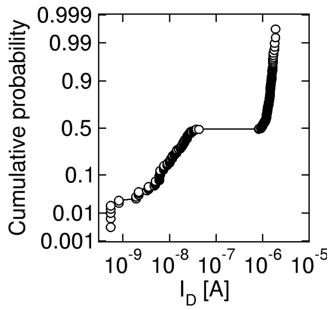
H. Mulaosmanovic et al., Nanoscale, 2018

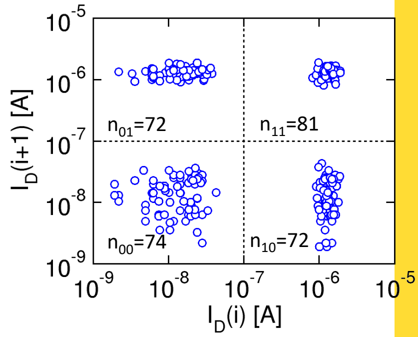
Random number generation



t_{PW} @ P = 0.5



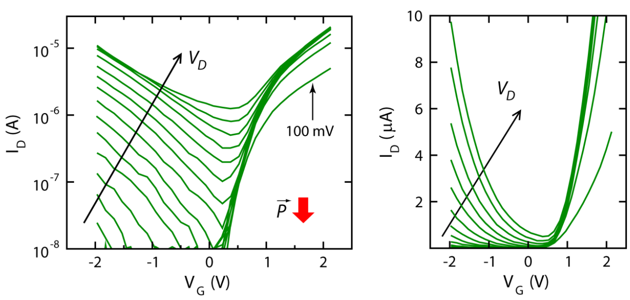


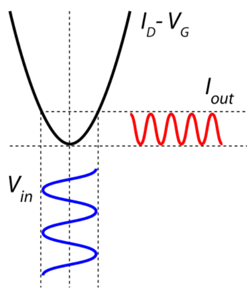


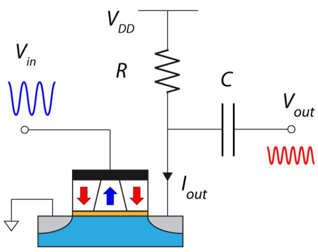
- Stream of equally probable "1" and "0"
- Populations of "00", "01", "10" and "11" is nearly matched

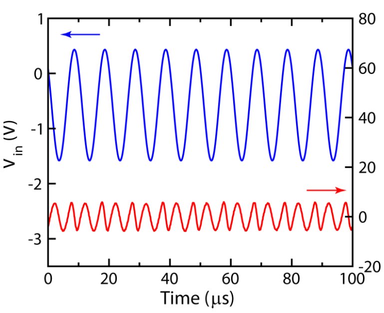
H. Mulaosmanovic et al., IEEE EDL, 2018

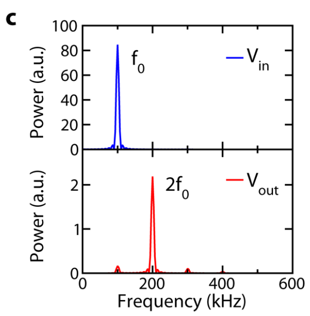
Frequency multiplication











- Tune the symmetry of FeFET's I_D-V_G by polarization switching and GIDL
- Frequency multiplication is achieved

Mulaosmanovic et al., Nature Electron. 3, 2020

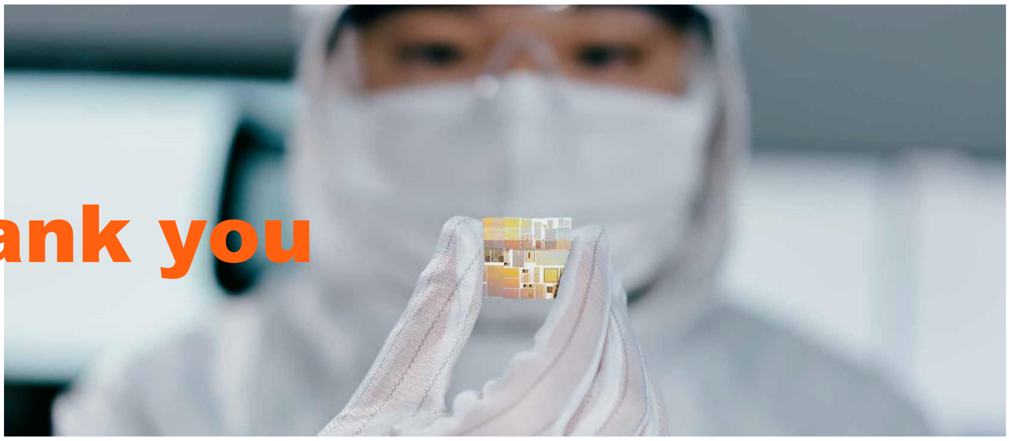
Conclusions

- Ferroelectricity – intrinsic memory functionality
- FeFETs – very attractive 1T memory solutions for embedded applications
- Ferroelectric HfO_2 opens new opportunities for FeFETs
- Significant progress in device physics, integration, reliability, and scaling over the years
- Main switching mechanisms revealed

  → important learning for retention, disturbs, endurance, array operation schemes
- Unconventional applications due to variety of switching patterns

  → Neuromorphic; Reconfigurable logic-in-memory; Frequency manipulation; Security

Thank you



GF FeFET Team

Dr. Sven Beyer
Dr. Stefan Dünkel
Dr. Johannes Müller
Dr. Martin Trentzsch

Partners

namlab nanoelectronic materials laboratory

FMC The Ferroelectric Memory Company

Fraunhofer IPMS

Funding

IPCEI on Microelectronics

This work is funded by the Federal Ministry for Economics and Energy (BMWi) and by the State of Saxony in the framework "Important Project of Common European Interest (IPCEI)."

Shosuke Fujii
Kioxia Corporation

Shosuke Fujii is a Chief Specialist at Kioxia Corporation, currently leading an emerging device technology team. He received the B.S. (2005) and M.S. (2007) degrees in materials science and engineering from Kyoto University, Japan. He joined Toshiba Corporation in 2007, where he was engaged in the research on reliability physics of MONOS memories. From 2009 to 2016, he was engaged in the research of emerging memory cell technology including resistive switching memory and ferroelectric memory. From 2016 to 2018, he was a visiting scholar in Stanford University, where he studied scaling effects of resistive switching memories. He is currently with Kioxia Corporation (renamed from Toshiba Memory Corporation), where he is engaged in research and development of emerging memory devices. He served as a technical committee member for memory reliability in IEEE IRPS in 2014 and 2015, and a tutorial lecturer in IEEE IRPS in 2016. He has been serving as a technical program committee member in VLSI Symposium since 2020.

Ferroelectric Tunnel Junction

Shosuke Fujii
KIOXIA

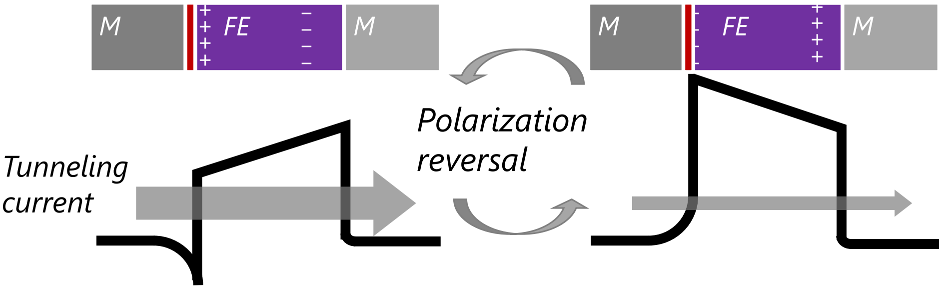
KIOXIA

Outline

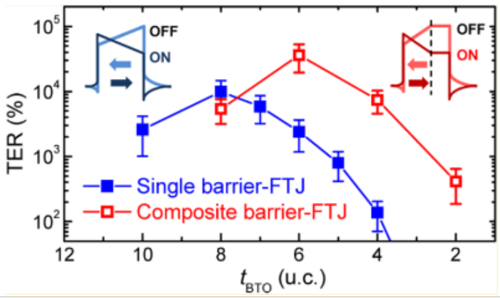
1. Ferroelectric tunnel junction
2. Ferroelectric HfO_2-based tunnel junction
3. FTJ for emerging application
4. Summary

Ferroelectric Tunnel Junction (FTJ)

- Emerging memory that utilizes polarization reversal



- Demonstration of $BaTiO_3$ FTJ

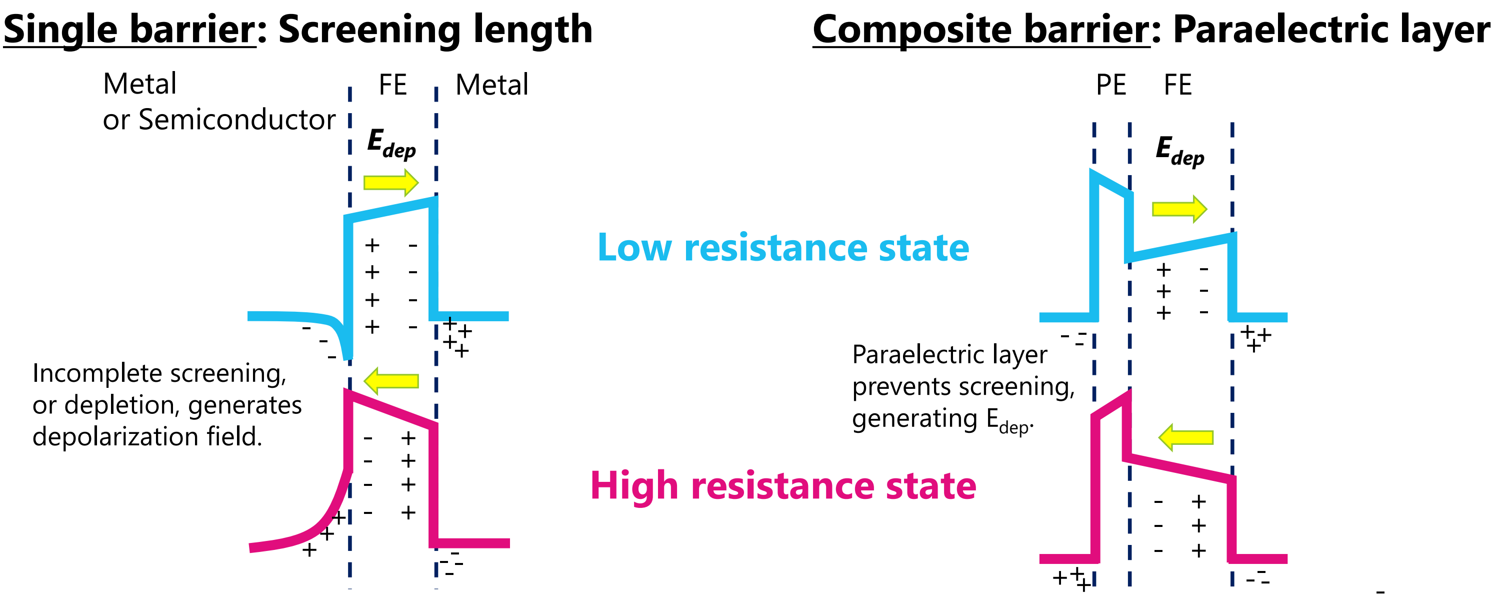


Mechanism of barrier modulation
-Single barrier: Screening length of electrodes
-Composite barrier: Paraelectric layer

L. Wang et al., Nano Letters, 16 (2016) p.3911

Switching Mechanism

Depolarization field, E_{dep}, modulates potential profile, inducing TER



Effect of E_{dep} on FTJ TER

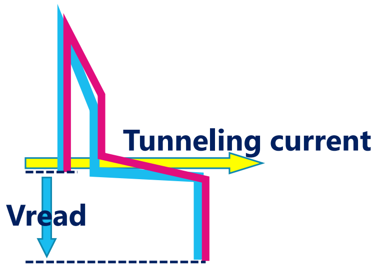
E_{dep} is the TER mechanism, but deceasing memory window

Ec > Edep required for stable polarization

$$E_c > E_{dep} = \frac{P}{\varepsilon_{FE}\varepsilon_0}\left(1+\frac{\varepsilon_{PE}t_{FE}}{\varepsilon_{FE}t_{PE}}\right)^{-1}$$

Data reading further destabilizes the polarization

$$E_c > E_{dep} = \frac{P}{\varepsilon_{FE}\varepsilon_0}\left(1+\frac{\varepsilon_{PE}t_{FE}}{\varepsilon_{FE}t_{PE}}\right)^{-1} + \underline{\frac{\varepsilon_{PE}}{\varepsilon_{FE}t_{PE}+\varepsilon_{PE}t_{FE}}V_{read}}$$



Voltage application is inevitable to obtain a sufficient amount of tunneling current, but it decreases the TER.

TER considering data reading

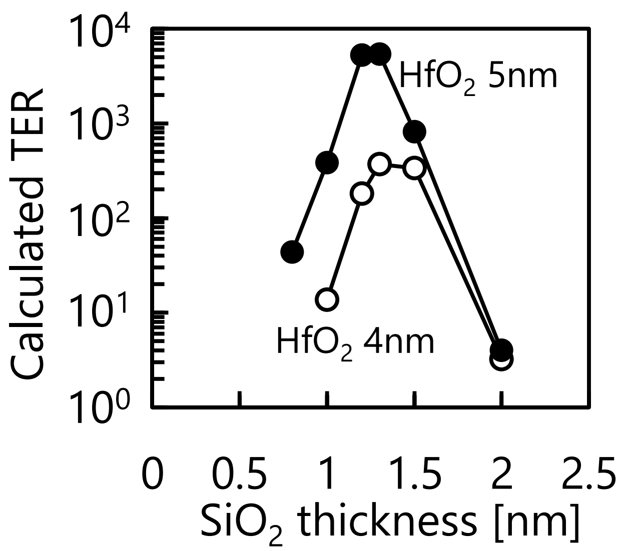
Precise stack design is needed to obtain a reasonable amount of TER

✓ FTJ with **THIN** PE

Weak modulation due to small E_{dep}.
→ Small TER

✓ FTJ with **THICK** PE

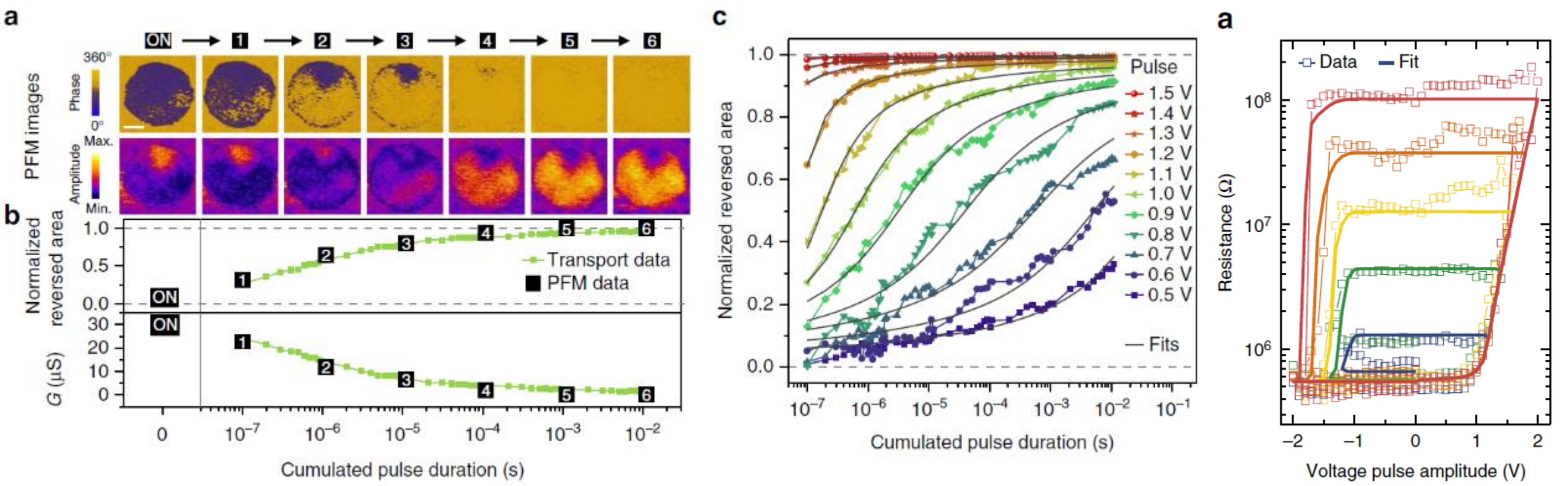
Larger voltage is needed for detectable read current, leading larger E_{dep}.
→ Small TER



S. Fujii et al., *SSDM* (2021) p.117-118.

Memristive switching

Continuous resistance change due to nucleation and expansion of domains
→ Opportunity for emerging in-memory computing application



Boyn, S., Grollier, J., Lecerf, G. et al. Learning through ferroelectric domain dynamics in solid-state synapses. Nature Communications 8, 14736 (2017). [3]
https://doi.org/10.1038/ncomms14736 https://creativecommons.org/licenses/by/4.0/

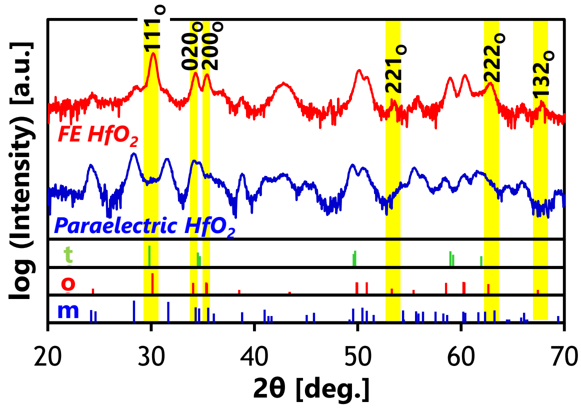
Outline

1. Ferroelectric tunnel junction
2. Ferroelectric HfO_2-based tunnel junction
3. FTJ for emerging application
4. Summary

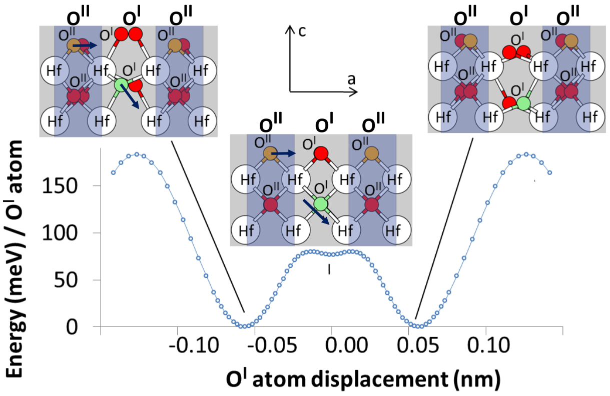
Ferroelectric HfO_2

- First reported in 2011
- Most common high-k material in CMOS technology
- HfO_2 with various dopant (Si, Y, etc), $Hf_{0.5}Zr_{0.5}O_2$
- 10nm or less (much thinner than conventional ferroelectric material such as PZT)

XRD of ferroelectric HfO_2: Orthorhombic crystalline structure



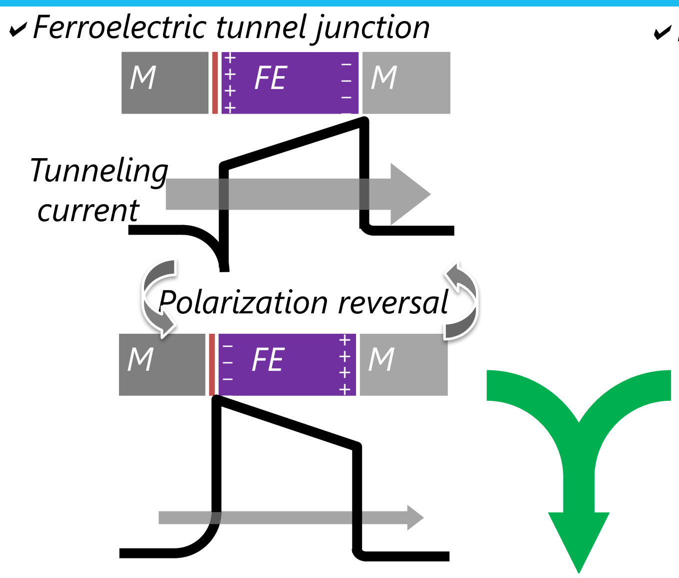
Displacement of O atoms contributes to ferroelectricity

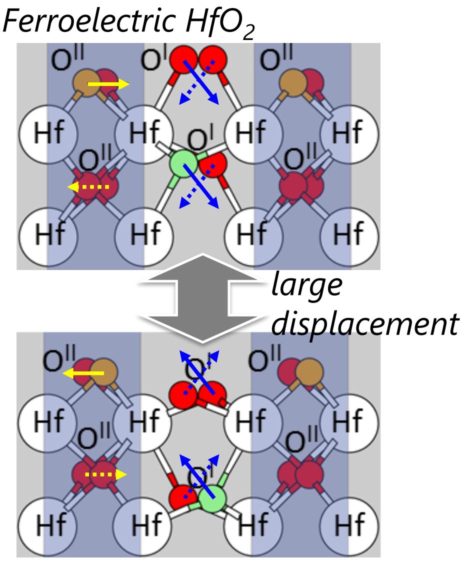


S. Fujii et al., *VLSI Tech.* (2016) p.148. © 2016 IEEE [4]

Combination of FTJ with HfO_2

HfO_2-based FTJ: A CMOS compatible emerging non-volatile memory





<u>HfO_2-based Ferroelectric Tunnel Junction</u>

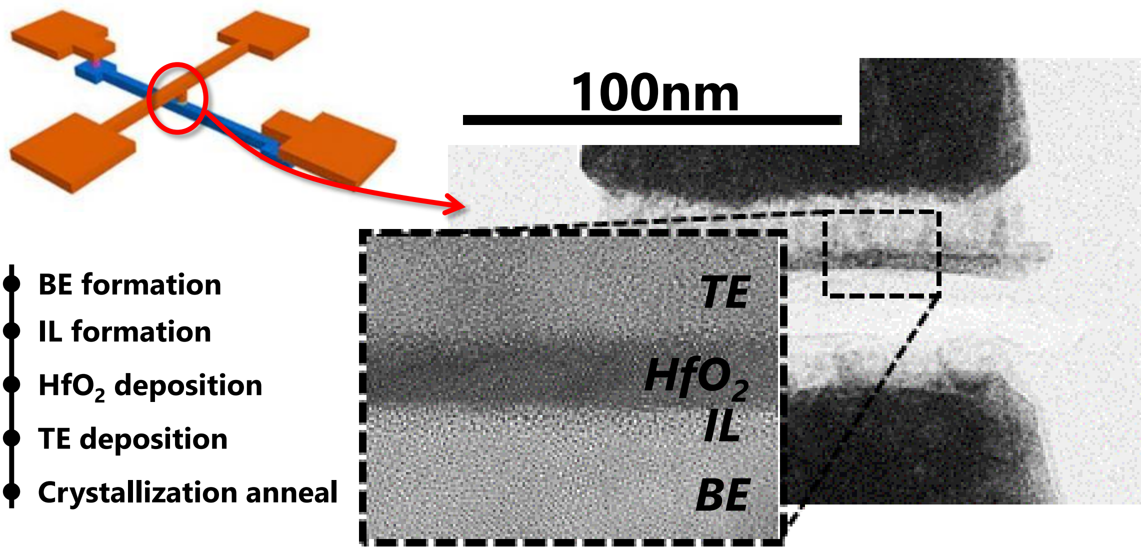
Recent Hafnia-FTJ research

Most of HfO_2-FTJ research is on composite barrier structure

	Fraunhofer	Namlab/Leti	UCB	KAIST	SNU/SKH	IBM	U Tokyo	WD/UCB
Top Electrode	TiN	TiN	W	TiN	TiN	TiN	Al	Pt
Ferro	HfZrO 4~8nm	HfZrO 10nm	HfZrO 1nm	HfZrO ~5nm	HfOx 6nm	HfZrO	HfZrO 4nm	HfZrO ~2nm
Interface layer (PE)	SiO_2 or Al_2O_3 1~2nm	Al_2O_3 2nm	SiO_2 1nm	Ta_2O_5	SiO_2 or Al_2O_3 1nm	WOx	SiO_2 1nm	Semicon ductor
Bottom electrode	Si-sub	TiN	Si-sub	TaN	Si-sub	TiN	Si-sub	LSMO
Reference	IEEE T-ED 2022 [5]	IEEE ISCAS 2021 [6]	Adv. Electron. Mater. 2021 [7]	IEDM 2021 [8]	Nanotech. 2021 [9]	EDTM 2021 [10]	IEEE JEDS 2018 [11]	Adv. Electron. Mater. 2021 [12]

Demonstration of HfO_2-based FTJ: Device structure

Simple cross point structure
No current limiter, such as series resistor or transistor , is required

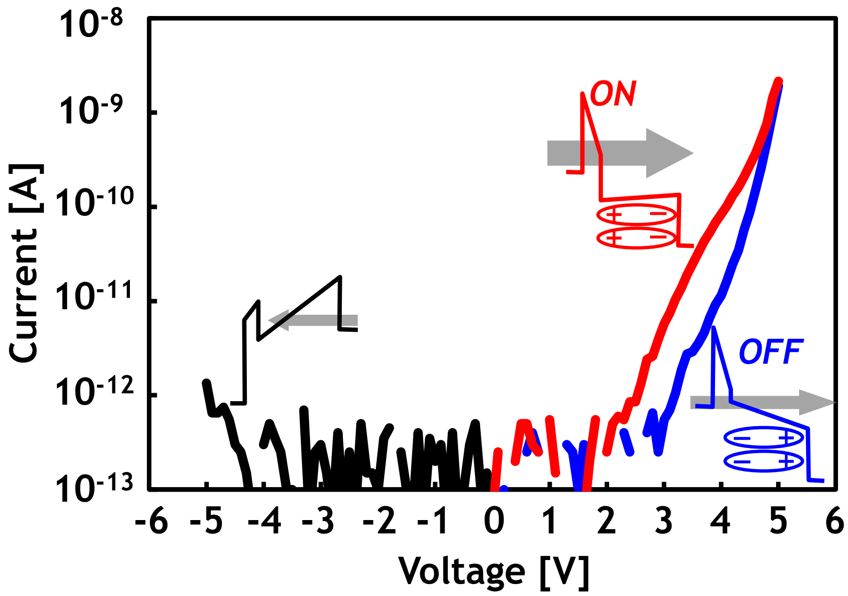


- **BE formation**
- **IL formation**
- **HfO_2 deposition**
- **TE deposition**
- **Crystallization anneal**

S. Fujii et al.,
VLSI Tech. (2016) p.148.
© 2016 IEEE [4]

Demonstration of HfO_2-based FTJ: Device performance

FTJ performance is suitable for cross-point architecture



Advantages

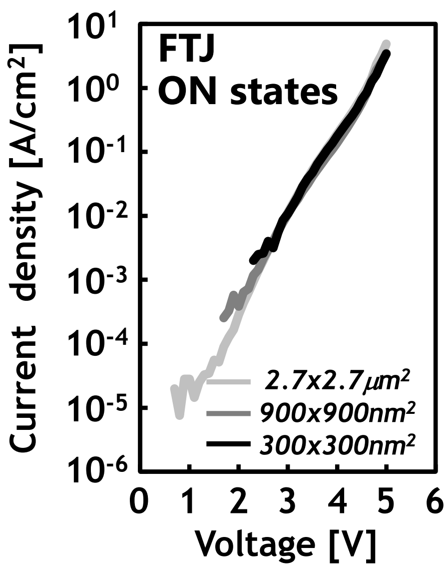
- Low current operation
- Self compliance
- Large non-linearity
- Intrinsic diode

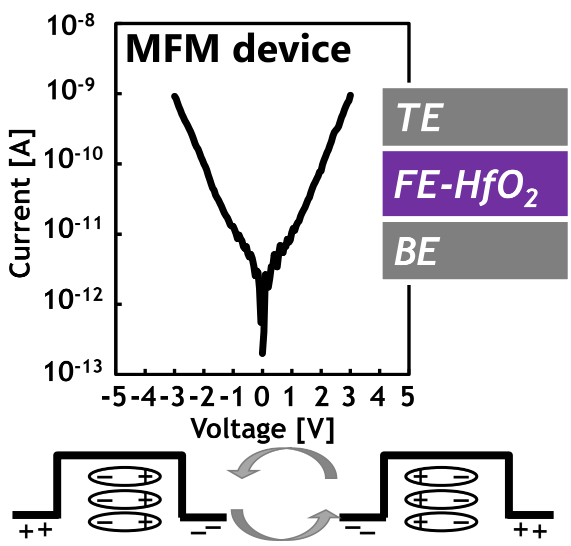
→ Suitable for cross-point architecture

S. Fujii et al., *VLSI Tech.* (2016) p.148. © 2016 IEEE [4]

Distinguish from Resistive RAM and Trap Asist Tunneling

Clear area scaling → Non-filamentary switching
No switching without IL → Consistent with FTJ mechanism



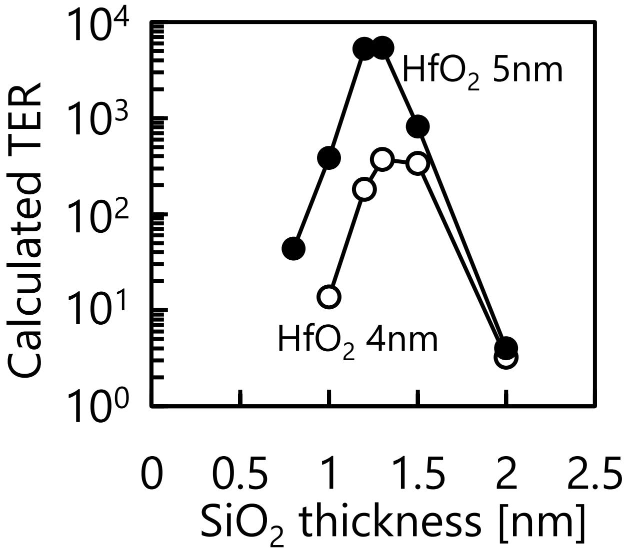


Barrier modulation does not occur

S. Fujii et al., *VLSI Tech.* (2016) p.148. © 2016 IEEE [4]

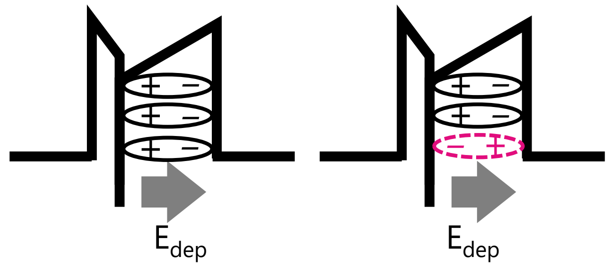
Device design for FTJ with composite barrier

Precise stack design is necessary for performance improvement



$$E_{dep} = \frac{P}{\varepsilon_{FE}}\left(1 + \frac{\varepsilon_{IL} t_{FE}}{\varepsilon_{FE} t_{IL}}\right)^{-1} < E_c$$

t_{FE}: FE thickness, t_{IL}: IL thickness

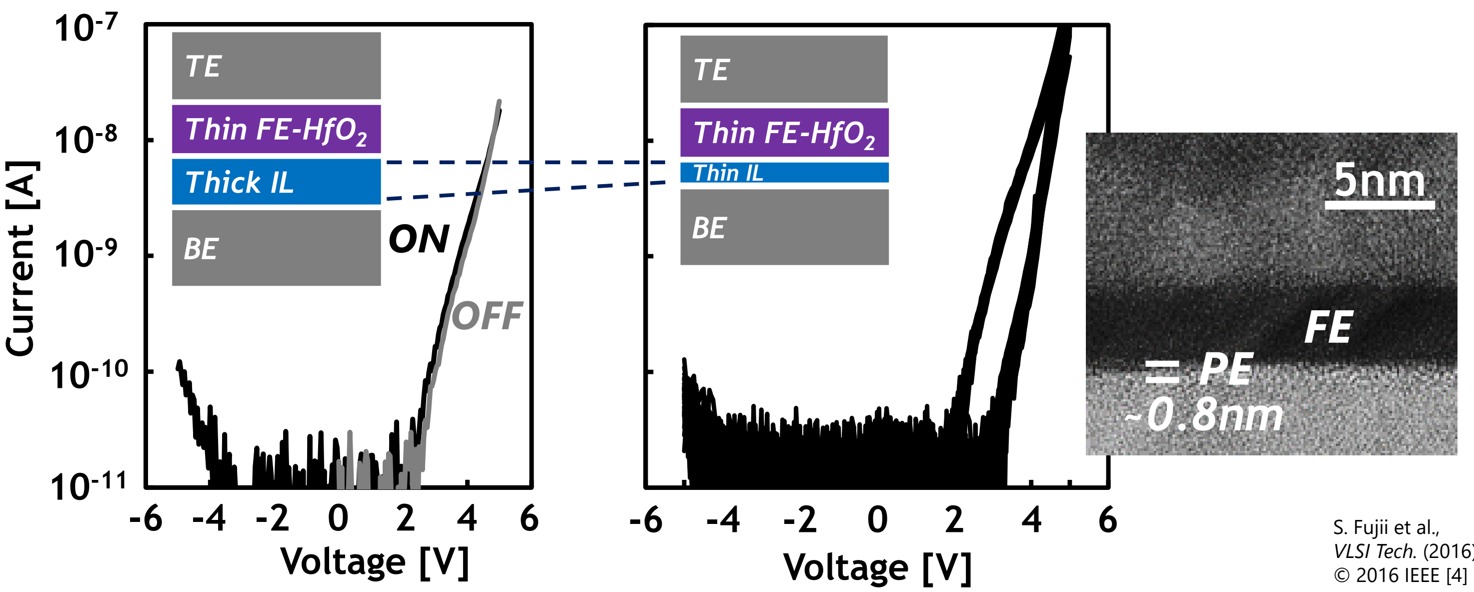


$E_{dep} > E_c$ destabilizes the polarization

S. Fujii et al., *SSDM* (2021) p.117-118.

Device design for FTJ with composite barrier

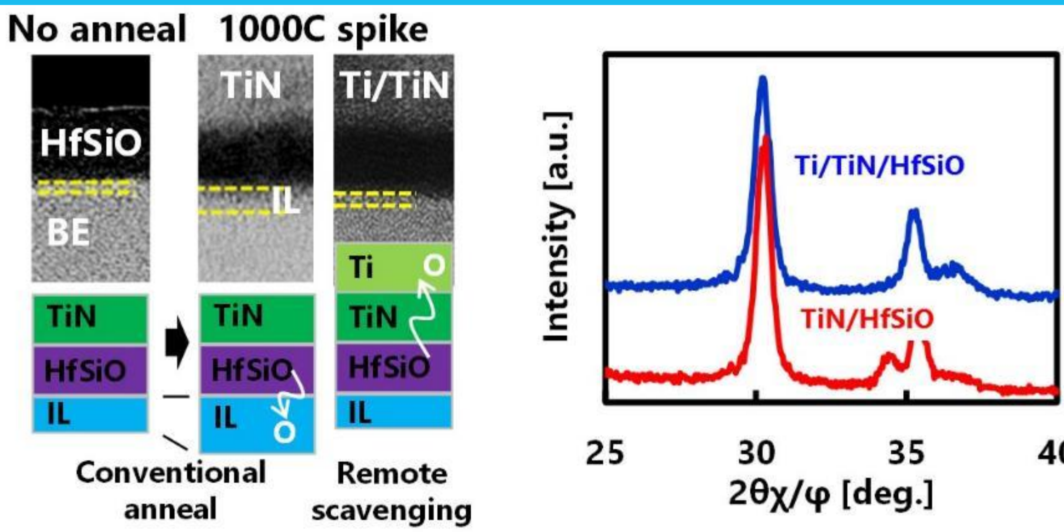
Thickness design is key for performance improvement



S. Fujii et al., *VLSI Tech.* (2016) p.148.

Thickness control technique: Remote scavenging

Remote scavenging process keeps IL thickness as designed

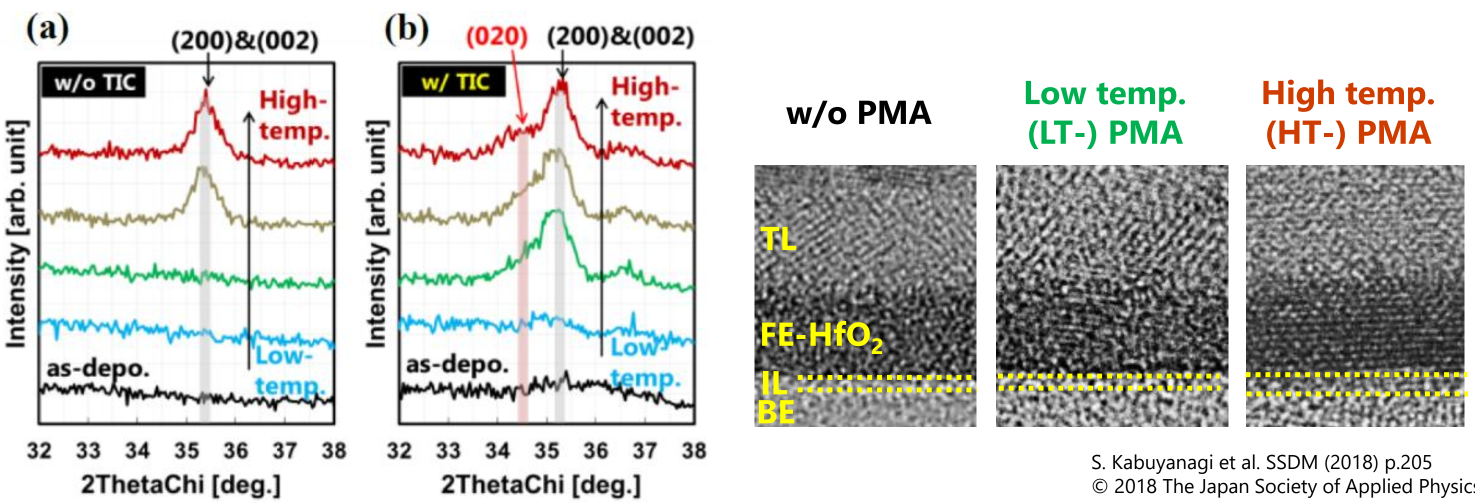


-Scavenger Ti traps O during crystallization
-IL thickness is kept as designed while suppressing monoclinic phase

S. Fujii et al., VLSI Technology 2020, p.1

Thickness control technique: Template-induced crystallization(TIC)

TIC reduces crystallization temperature, keeping IL as designed

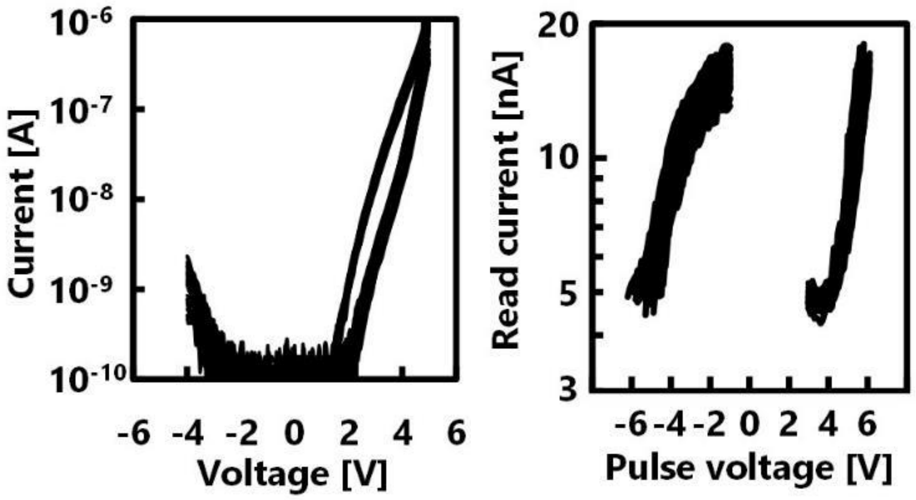


S. Kabuyanagi et al. SSDM (2018) p.205

-Crystallization temperature is lowered owing to the assist of template layer.
-Low temperature PMA keeps PE thickness as designed.

HfO_2-FTJ with precise thickness control

The improved FTJ shows large TER with low operation current



S. Fujii et al., VLSI Technology 2020, p.1

Device performance

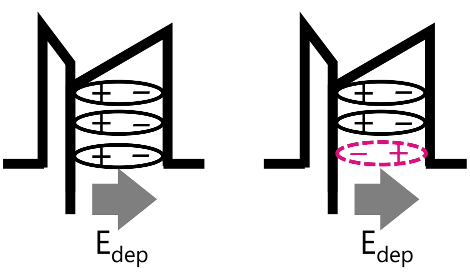
- **-Analog resistance change**
- **-Low operation current**
- **-Low variability**
- -State stability
- -Cycling endurance

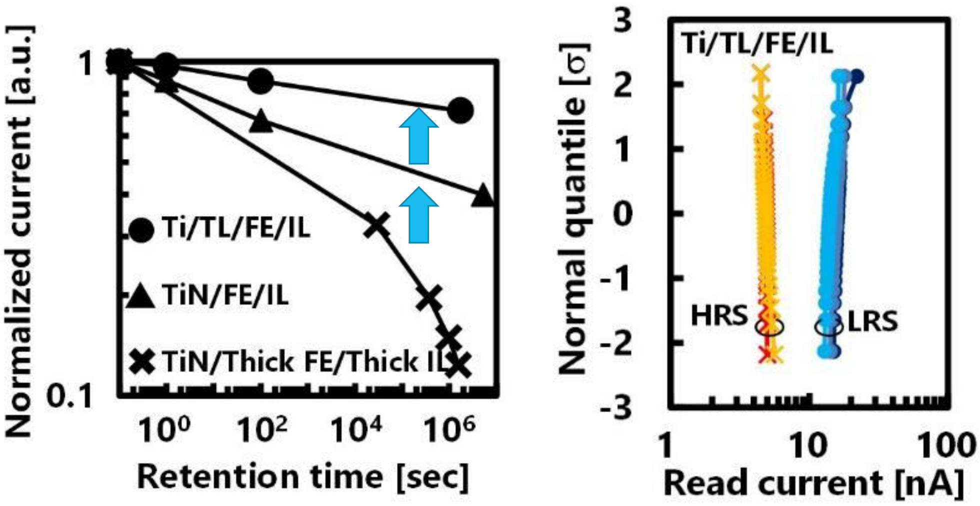
Although low operation current is mandatory for future low power application, too small read current could degrade operation speed

State stability : Long term data retention

Stable polarization is achieved by decreasing E_{dep}

$$E_{dep} = \frac{P}{\varepsilon_{FE}}\left(1 + \frac{\varepsilon_{IL} t_{FE}}{\varepsilon_{FE} \boxed{t_{IL}}}\right)^{-1}$$



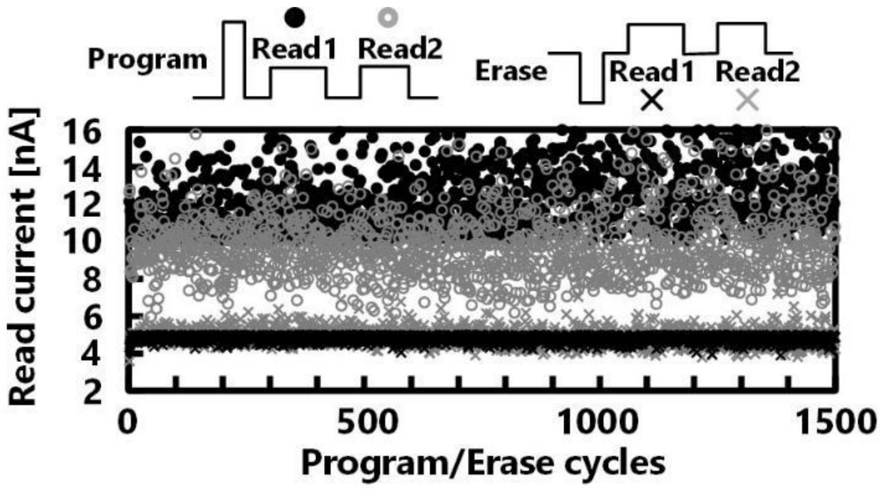


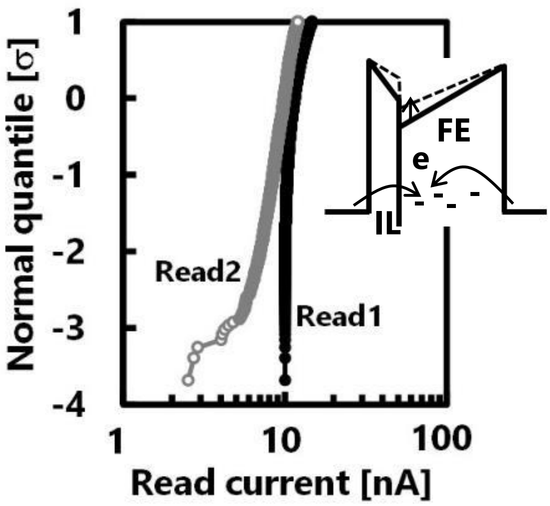
- E_{dep} is increased with IL thickness
- Thin IL with the assist of TL and remote scavenging achieves stable polarization

S. Fujii et al., VLSI Technology 2020, p.1

State stability : Short term relaxation

Quick electron trapping in LRS immediately after programming



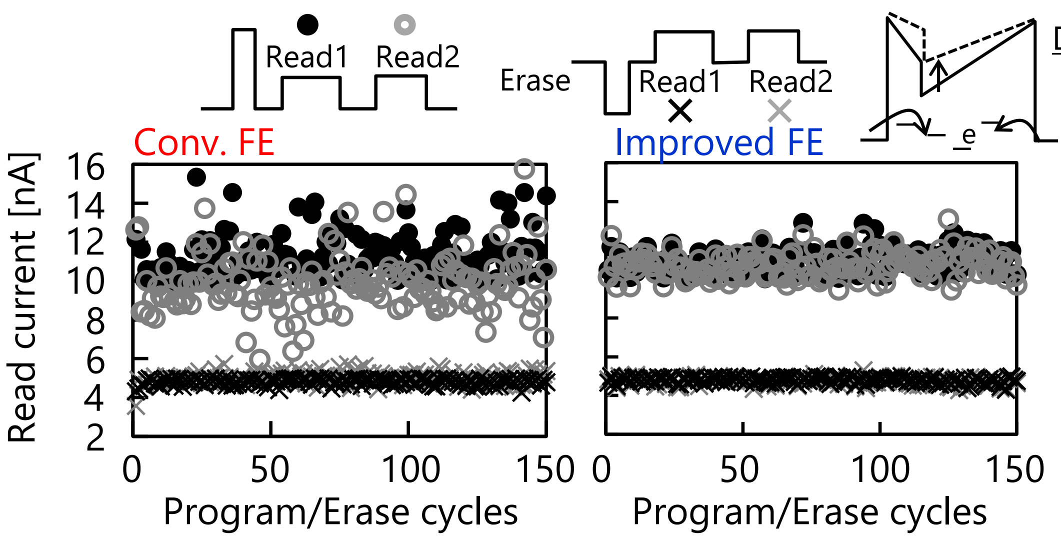


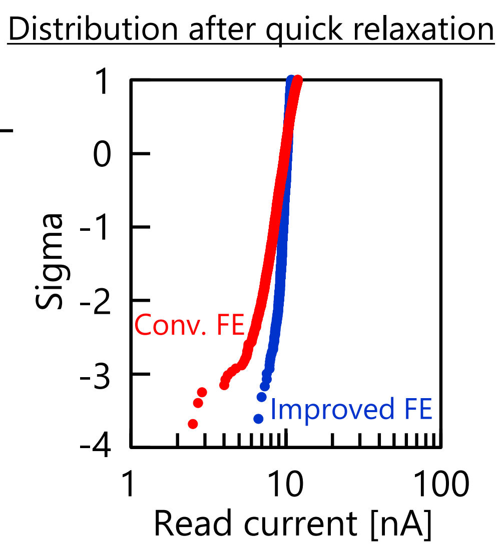
- A quick relaxation is observed after programming
- Electron trapping increases the effective barrier height

S. Fujii et al., VLSI Technology 2020, p.1

FE engineering for suppressing quick relaxation

Quick relaxation is suppressed by FE process optimization

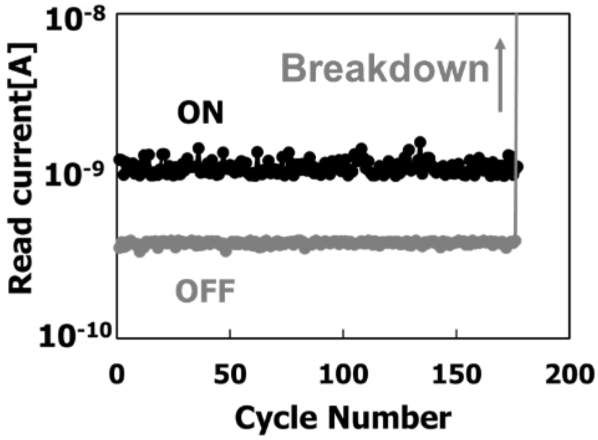




S. Fujii et al., *SSDM* (2021) p.117-118.

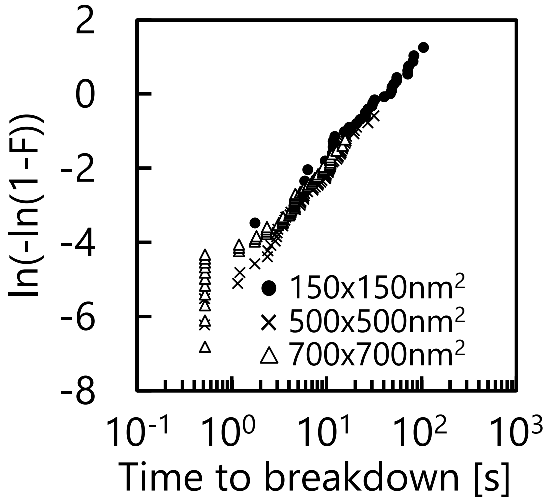
Cycling endurance of the FTJ

Endurance failure can be described as conventional breakdown model



- Endurance failure is caused by breakdown

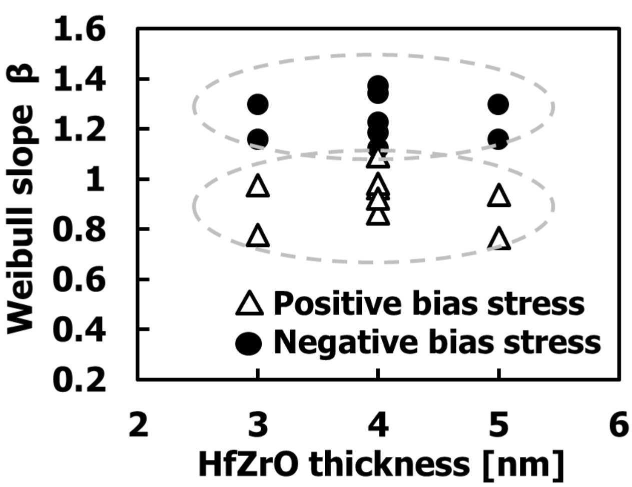
M. Yamaguchi et al., IRPS 2018, 6D.2

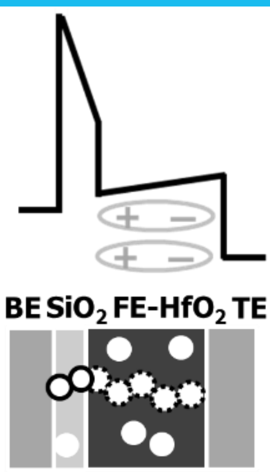


- FTJ breakdown is well normalized by Weibull distribution
- Conventional percolation model is applicable

Breakdown mechanism of the HfO_2 FTJ

Defects generation in IL determines breakdown





Device performance

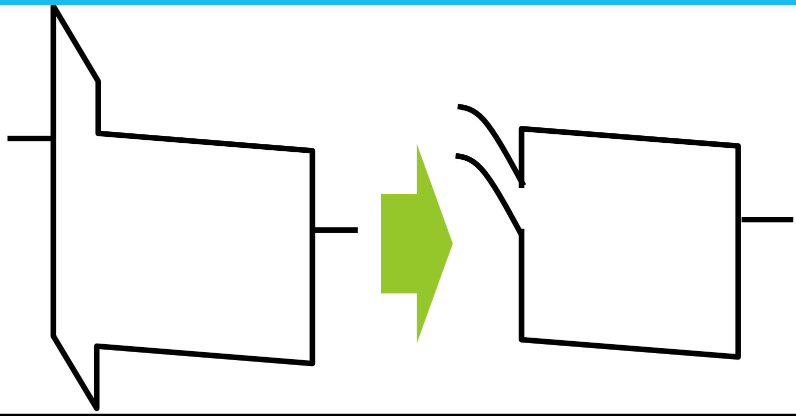
- Analog resistance change
- Low operation current
- Low variability
- State stability
- Cycling endurance

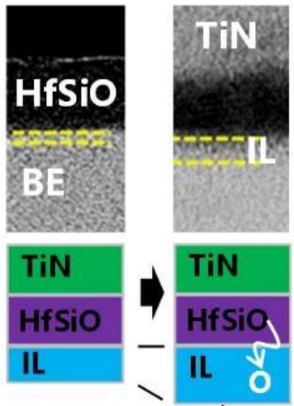
- Weibull slope is independent of the FE thickness → IL determines the breakdown

M. Yamaguchi et al., IRPS 2020

To improve the endurance

MFS structure with no IL is promising, but difficult to realize it



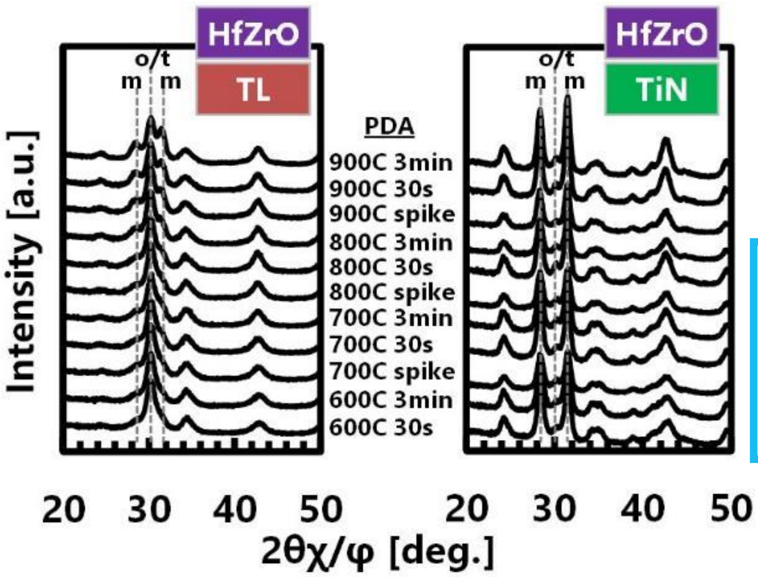


	MFIM	MFS
Operation current	Very low small current due to IL	No IL increases tunnel current
Cycling endurance	Low-k IL limits the lifetime	No IL improves the lifetime

Conventional process inevitably forms IL

Utilizing TIC for fabricating MFS structure

The TL improves controllability of stack structure



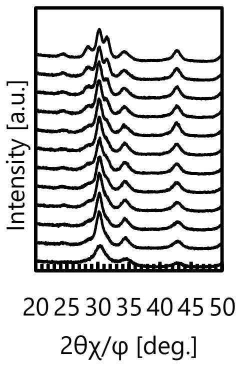
The TL does not need metal cap to form orthorhombic phase

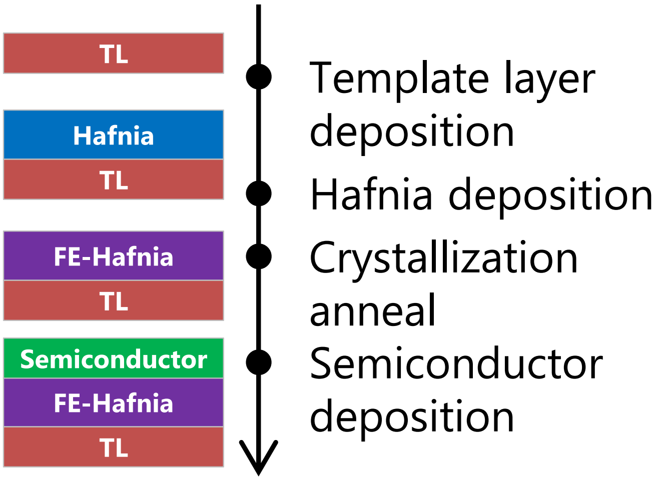
→ **Another route for performance improvement**

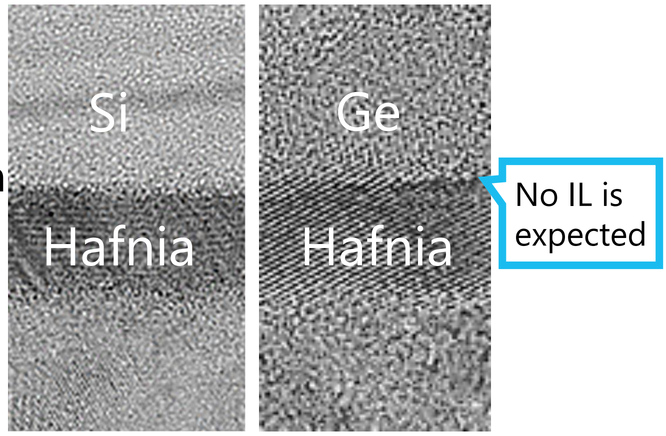
S. Fujii, VLSI Technology 2020, p.1

MFS FTJ structure

Novel process technologies could realize the MFS structure







The MFS FTJ using our novel process technologies could further improve operation current and endurance

S. Fujii et al., *SSDM* (2021) p.117-118.

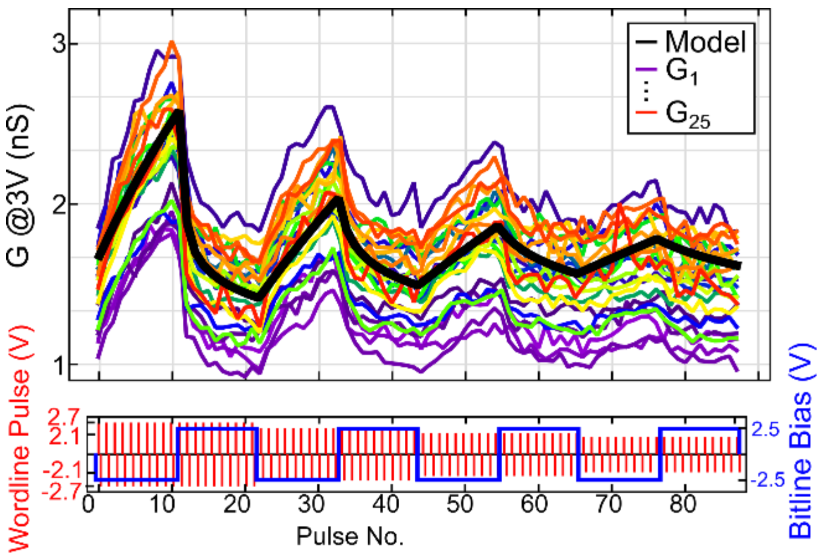
Outline

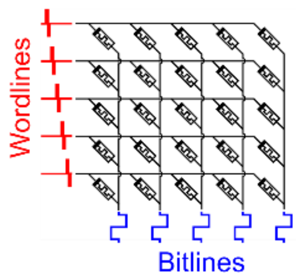
1. Ferroelectric tunnel junction
2. Ferroelectric HfO_2-based tunnel junction
3. FTJ for emerging applications
4. Summary

Memristive switching for emerging applications

Analog resistance change can be utilized for emerging applications

5x5 FTJ selectorless crossbar read and write.

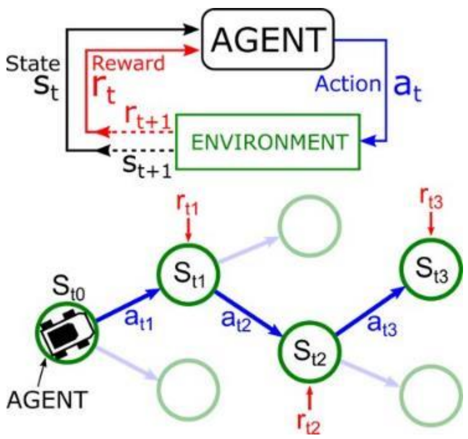




R. Berdan, VLSI Technology 2019, p.22

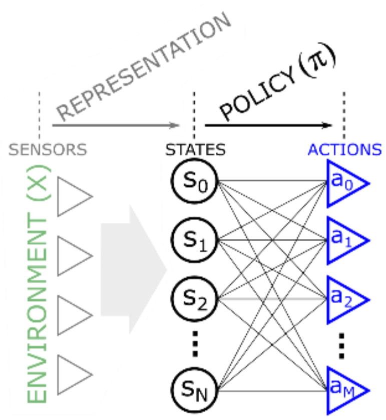
Memristor-based reinforcement learning

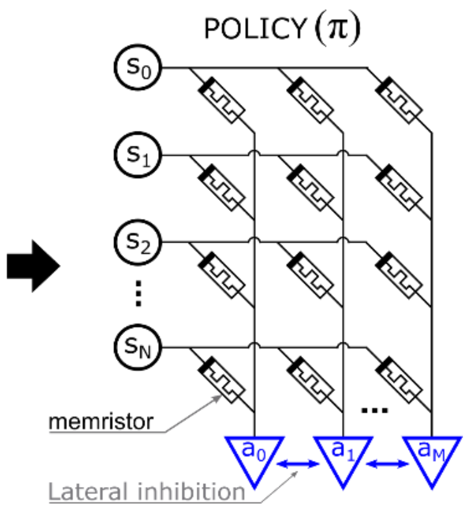
Reinforcement learning (RL)



- Evaluate state → execute action → get reward
- Maximize the reward by learning
- Requires massive computation

In-memory RL with memristive FTJ



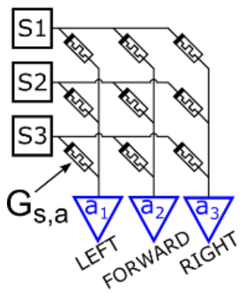


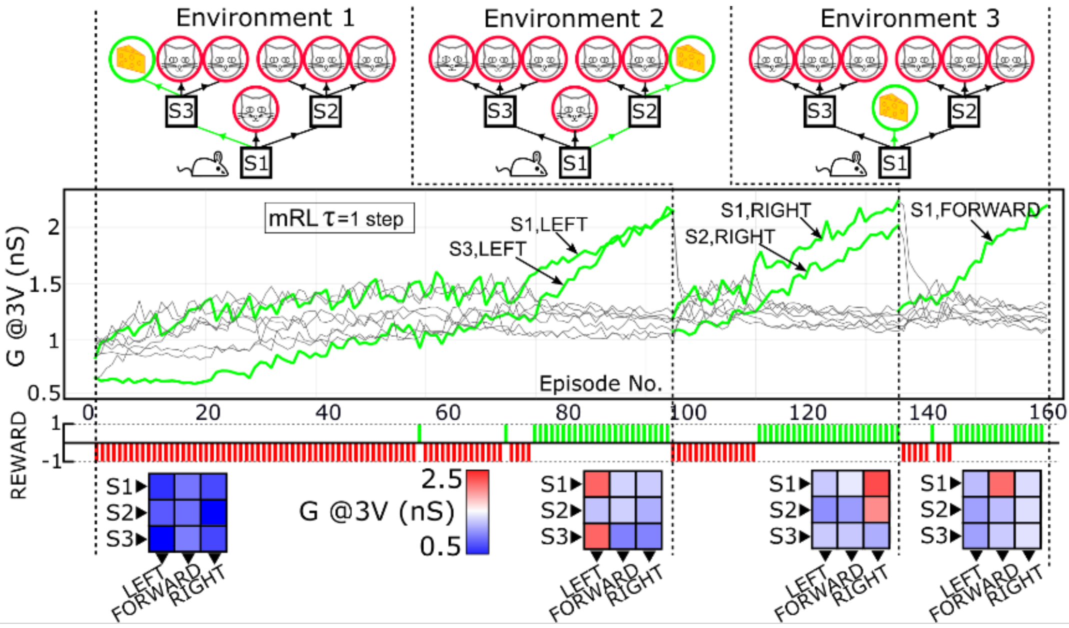
$$s_i:\ a_{t+1} = \max_j (G_{ij})$$

R. Berdan et al., *VLSI Tech. 2019*, p.22

In-memory reinforcement learning

Path-finding demonstration using FTJ memristor crossbar

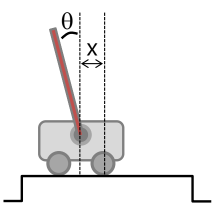


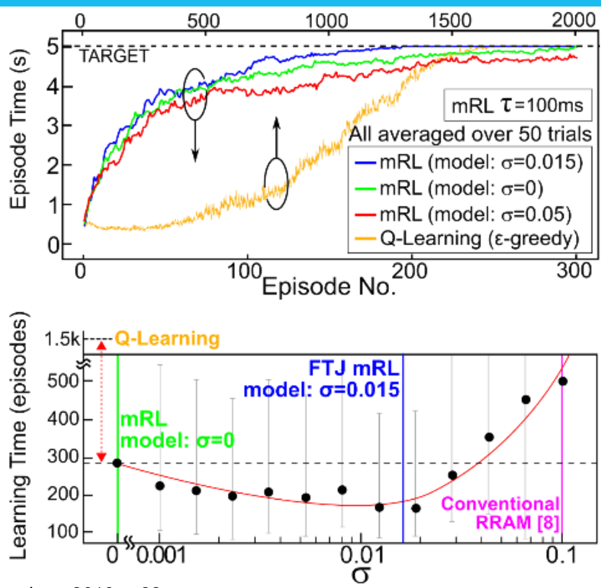


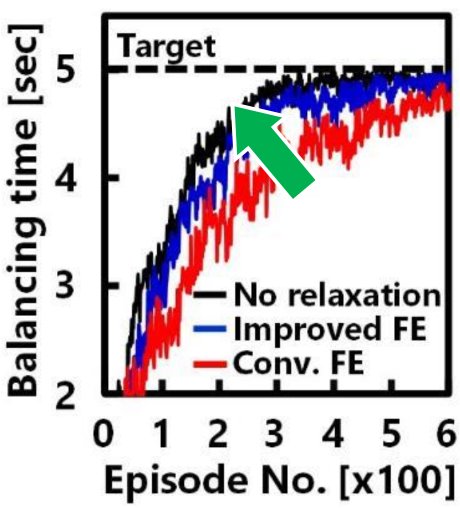
R. Berdan et al., *VLSI Tech. 2019*, p.22

Impact of variability and relaxation on learning performance

FTJ with moderate variability and reduced relaxation shows better performance







R. Berdan et al., VLSI technology 2019, p.22

S. Fujii et al., VLSI Technology 2020, p.1

Vector Matrix Multiplication (VMM)

Linear computation using FTJ tunneling current

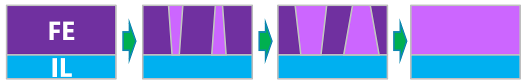
nature electronics
ARTICLES
https://doi.org/10.1038/s41928-020-0405-0
Check for updates

Low-power linear computation using nonlinear ferroelectric tunnel junction memristors

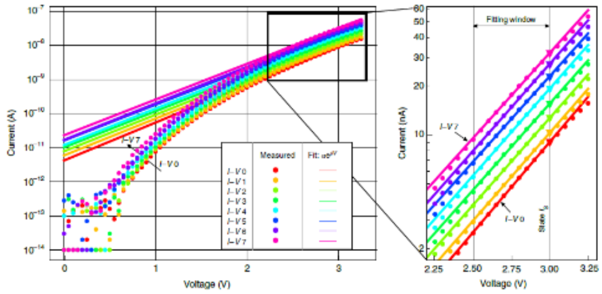
Radu Berdan[1,2], Takao Marukame[1], Kensuke Ota[3], Marina Yamaguchi[3], Masumi Saitoh[3], Shosuke Fujii[3], Jun Deguchi[2] and Yoshifumi Nishi[1]

Analogue in-memory computing using memristors could alleviate the performance constraints imposed by digital von Neumann systems in data-intensive tasks. Conventional linear memristors typically operate at high currents, potentially limiting power efficiency and scalability in practical applications. Here, we show that nonlinear ferroelectric tunnel junction memristors can perform linear computation at ultralow currents. Using logarithmic line drivers, we demonstrate that analogue-voltage-amplitude vector-matrix multiplication (VMM) can be performed in selectorless ferroelectric tunnel junction crossbars by exploiting a device nonlinearity factor that remains constant for multiple conductive states. We also show that our ferroelectric tunnel junction crossbars have the attributes required to scale analogue VMM-intensive applications, such as neural inference engines, towards energy efficiencies above 100 tera-operations per second per watt.

$$I = \underline{s}I_S e^{\bar{\beta}V}$$



Tunneling current is proportional to conductive domain area, *s*.



Each state has the same β with different α (effective conductive area).

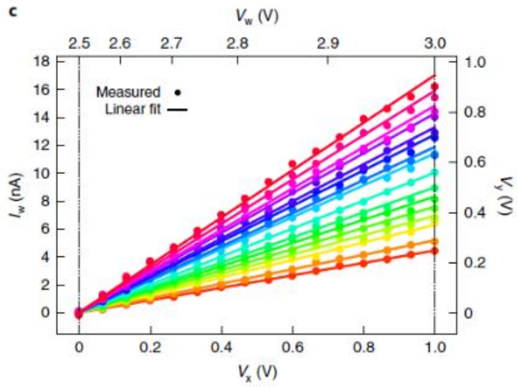
R. Berdan et al., Nature Electronics (2020) 259

Conversion to linear I-V

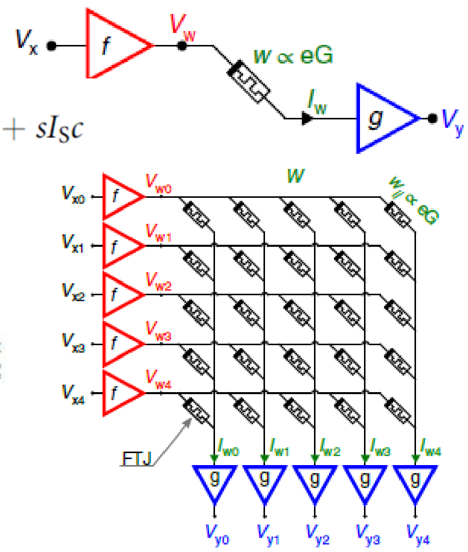
VMM operation owing to low variability of FTJ

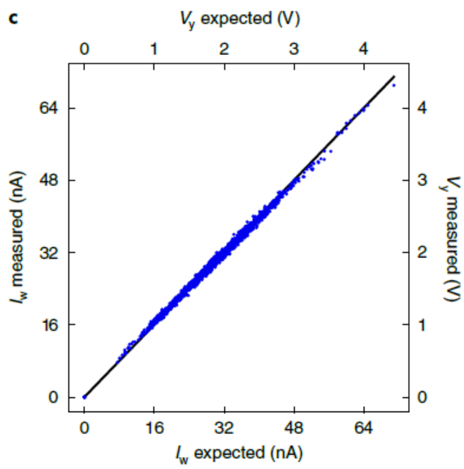
$$f(V_x) = \frac{1}{\bar{\beta}}\ln(kV_x + c)$$

$$\underline{I_w} = sI_S e^{\bar{\beta}f(V_x)} = sI_S e^{\bar{\beta}\frac{1}{\bar{\beta}}\ln(kV_x+c)} = sI_S k\underline{V_x} + sI_S c$$



Linear Iw-Vx curves obtained from non-linear FTJ I-V.



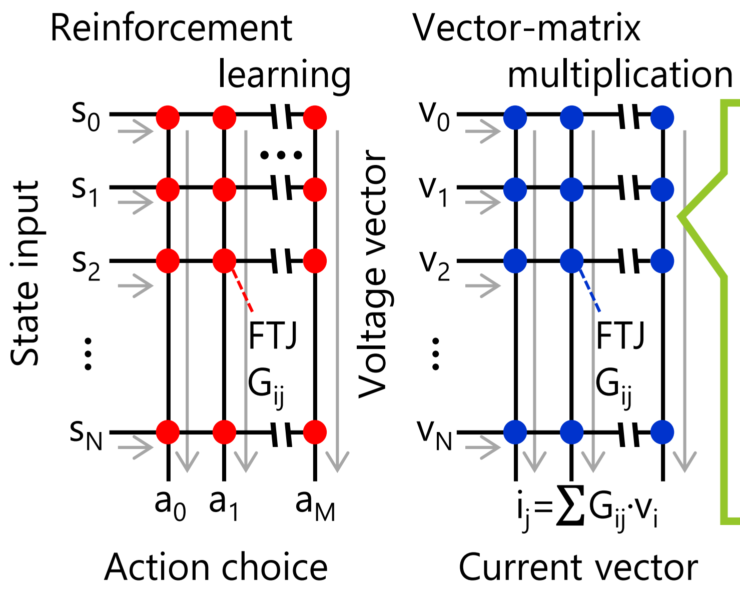


VMM error σ=0.77%

R. Berdan et al., Nature Electronics (2020) 259

FTJ for emerging applications

Device performance required for various emerging applications are almost the same
HfO_2 FTJ is suitable for those applications



Device requirements **HfO_2 FTJ**

- -Analog resistance change
- -Low operation current — MFS FTJ could be a solution.
- -Low variability
- -State stability
- -Cycling endurance — MFS FTJ could be a solution.

Outline

1. **Ferroelectric tunnel junction**
2. **Ferroelectric HfO_2-based tunnel junction**
3. **FTJ for emerging applications**
4. **Summary**

Summary

Ferroelectric tunnel junction

-Depolarization E_{dep} is the TER mechanism.
-Precise stack design is necessary for reasonable TER.
-Continuous resistance change owing to nucleation and expansion of the domain.

HfO_2 FTJ

-A CMOS compatible emerging memory, having low operation current, low variability, stable memory state, and analog resistance change.
-Precise stack design can be realized by sophisticated process technologies
-Cycling endurance and small read current could be improved using MFS structure

FTJ for emerging application

-Analog resistance change with low variability opened an opportunity for emerging computing application, reinforcement learning and vector-matrix-multiplication.
-Device performance required for various emerging applications are almost the same, and the HfO_2 FTJ is suitable for those applications.

References

[1] L. Wang, M. R. Cho, Y. J. Shin, J. R. Kim, S. Das, J-G. Yoon, J-S. Chung, and T. W. Noh, "Overcoming the Fundamental Barrier Thickness Limits of Ferroelectric Tunnel Junctions through BaTiO3/SrTiO3 Composite Barriers", Nano Letters, vol 6 (2016) p.3911-3918.
[2] S. Fujii, M. Yamaguchi, S. Kabuyanagi, K. Ota, and M. Saitoh, "Hafnia-based Ferroelectric Tunnel Junction for Emerging Applications", SSDM 2021, p117-118.
[3] S. Boyn, J. Grollier, G. Lecerf, et al. "Learning through ferroelectric domain dynamics in solid-state synapses", Nature Communications 8, 14736 (2017).
[4] S. Fujii, Y. Kamimuta, T. Ino, Y. Nakasaki, R. Takaishi, and M. Saitoh, "First demonstration and performance improvement of ferroelectric HfO_2-based resistive switch with low operation current and intrinsic diode property," 2016 IEEE Symposium on VLSI Technology, 2016, p.148-149.
[5] A. Sünbül et al., "Optimizing Ferroelectric and Interface Layers in HZO-Based FTJs for Neuromorphic Applications," in IEEE Transactions on Electron Devices, vol. 69, no. 2, pp. 808-815
[6] E. Covi et al., "Ferroelectric Tunneling Junctions for Edge Computing," 2021 IEEE International Symposium on Circuits and Systems (ISCAS), 2021, pp. 1-5
[7] Cheema, S. S., Shanker, N., Hsu, C.-H., Datar, A., Bae, J., Kwon, D., Salahuddin, S., One Nanometer HfO2-Based Ferroelectric Tunnel Junctions on Silicon. Adv. Electron. Mater. 2021, 2100499.
[8] Y. Goh, et al., "High Performance and Self-rectifying Hafnia-based Ferroelectric Tunnel Junction for Neuromorphic Computing and TCAM Applications", 2021 IEEE International Electron Devices Meeting (IEDM), 2021 p.378-381.
[9] Kyung Kyu Min et al, "Interlayer engineering for enhanced ferroelectric tunnel junction operations in HfO_x-based metal-ferroelectric-insulator-semiconductor stack", 2021 Nanotechnology 32 495203
[10] L. Bégon-Lours et al., "A BEOL Compatible, 2-Terminals, Ferroelectric Analog Non-Volatile Memory," 2021 5th IEEE Electron Devices Technology & Manufacturing Conference (EDTM), 2021, pp. 1-3
[11] M. Kobayashi, Y. Tagawa, F. Mo, T. Saraya and T. Hiramoto, "Ferroelectric HfO2 Tunnel Junction Memory With High TER and Multi-Level Operation Featuring Metal Replacement Process," in IEEE Journal of the Electron Devices Society, vol. 7, pp. 134-139
[12] Prasad, B., Thakare, V., Kalitsov, A., Zhang, Z., Terris, B., Ramesh, R., Large Tunnel Electroresistance with Ultrathin Hf0.5Zr0.5O2 Ferroelectric Tunnel Barriers. Adv. Electron. Mater. 2021, 7, 2001074.
[13] S. Fujii, M. Yamaguchi, S. Kabuyanagi, K. Ota, and M. Saitoh, "Improved state stability of HfO_2 ferroelectric tunnel junction by template-induced crystallization and remote scavenging for efficient in-memory reinforcement learning", 2020 IEEE Symposium on VLSI Technology, 2020, p.1-2
[14] S. Kabuyanagi, S. Fujii, K. Usuda, M. Yamaguchi, T. Ino, Y. Nakasaki, R. Takaishi, Y. Kamimuta, M. Saitoh, "Performance improvement by template-induced crystallization in ferroelectric HfO2 tunnel junction memory for cross-point high-density application", SSDM (2018) p.205.
[15] M. Yamaguchi, S. Fujii, Y. Kamimuta, S. Kabuyanagi, T. Ino, Y. Nakasaki, R. Takaishi, R. Ichihara, M. Saitoh, "Impact of specific failure mechanisms on endurance improvement for HfO2-based ferroelectric tunnel junction memory", IEEE International Reliability Physics Symposium (IRPS), 2018, p. 6D2.1-6D2.6.
[16] M. Yamaguchi, S. Fujii, K. Ota and M. Saitoh, "Breakdown Lifetime Analysis of HfO2-based Ferroelectric Tunnel Junction (FTJ) Memory for In-Memory Reinforcement Learning," 2020 IEEE International Reliability Physics Symposium (IRPS), 2020, pp. 1-6
[17] R. Berdan, T. Marukame, S. Kabuyanagi, K. Ota, M. Saitoh, S. Fujii, J. Deguchi, and Y. Nishi, "In-memory reinforcement learning with moderately-stochastic conductance switching of ferroelectric tunnel junction", 2019 Symposium on VLSI Technology, 2019, p.22-23
[18] R. Berdan, T. Marukame, K. Ota, M. Yamaguchi, M. Saitoh, S. Fujii, J. Deguchi, and Y. Nishi, "Low-power linear computation using nonlinear ferroelectric tunnel junction memristors", Nature Electronics, vol 3 (2020) p. 259-266.

KIOXIA

Onur Mutlu
ETH Zurich

Onur Mutlu is a Professor of Computer Science at ETH Zurich. He is also a faculty member at Carnegie Mellon University, where he previously held the Strecker Early Career Professorship. His current broader research interests are in computer architecture, systems, hardware security, and bioinformatics. A variety of techniques he, along with his group and collaborators, has invented over the years have influenced industry and have been employed in commercial microprocessors and memory/storage systems. He obtained his PhD and MS in ECE from the University of Texas at Austin and BS degrees in Computer Engineering and Psychology from the University of Michigan, Ann Arbor. He started the Computer Architecture Group at Microsoft Research (2006-2009) and held various product and research positions at Intel Corporation, Advanced Micro Devices, VMware, and Google. He received the Intel Outstanding Researcher Award, IEEE High Performance Computer Architecture Test of Time Award, the IEEE Computer Society Edward J. McCluskey Technical Achievement Award, ACM SIGARCH Maurice Wilkes Award, the inaugural IEEE Computer Society Young Computer Architect Award, the inaugural Intel Early Career Faculty Award, US National Science Foundation CAREER Award, Carnegie Mellon University Ladd Research Award, faculty partnership awards from various companies, and a healthy number of best paper or "Top Pick" paper recognitions at various computer systems, architecture, and security venues. He is an ACM Fellow "for contributions to computer architecture research, especially in memory systems", IEEE Fellow for "contributions to computer architecture research and practice", and an elected member of the Academy of Europe (Academia Europaea). His computer architecture and digital logic design course lectures and materials are freely available on YouTube (https://www.youtube.com/OnurMutluLectures), and his research group makes a wide variety of software and hardware artifacts freely available online (https://safari.ethz.ch/). For more information, please see his webpage at https://people.inf.ethz.ch/omutlu/.

Security Aspects of DRAM
The Story of RowHammer

Onur Mutlu
omutlu@gmail.com
https://people.inf.ethz.ch/omutlu
15 May 2022
IMW Tutorial

SAFARI ETH zürich Carnegie Mellon

How Reliable/Secure/Safe is This Bridge?

Source: http://www.technologystudent.com/struct1/tacom1.png

Collapse of the "Galloping Gertie"

Source: AP
http://www.wsdot.wa.gov/tnbhistory/connections/connections3.htm

How Secure Are These People?

Security is about preventing unforeseen consequences

Source: https://s-media-cache-ak0.pinimg.com/originals/48/09/54/4809543a9c7700246a0cf8acdae27abf.jpg

What Is RowHammer?

- One can predictably induce bit flips in commodity DRAM chips
  - >80% of the tested DRAM chips are vulnerable

- First example of how a simple hardware failure mechanism can create a widespread system security vulnerability

An "Early" Position Paper [IMW'13]

- Onur Mutlu,
  "Memory Scaling: A Systems Architecture Perspective"
  Proceedings of the 5th International Memory Workshop ***(IMW)***, Monterey, CA, May 2013. Slides (pptx) (pdf)
  EETimes Reprint

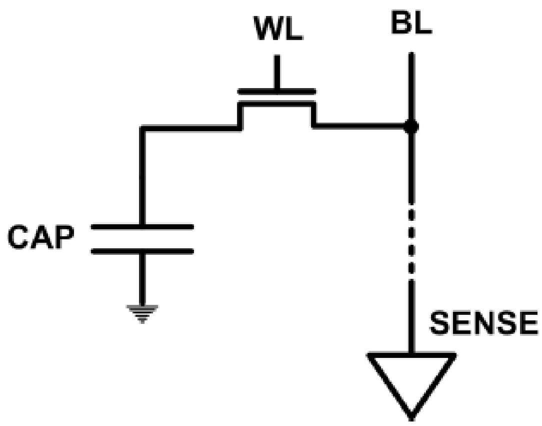
Memory Scaling: A Systems Architecture Perspective

Onur Mutlu
Carnegie Mellon University
onur@cmu.edu
http://users.ece.cmu.edu/~omutlu/

https://people.inf.ethz.ch/omutlu/pub/memory-scaling_memcon13.pdf

The DRAM Scaling Problem

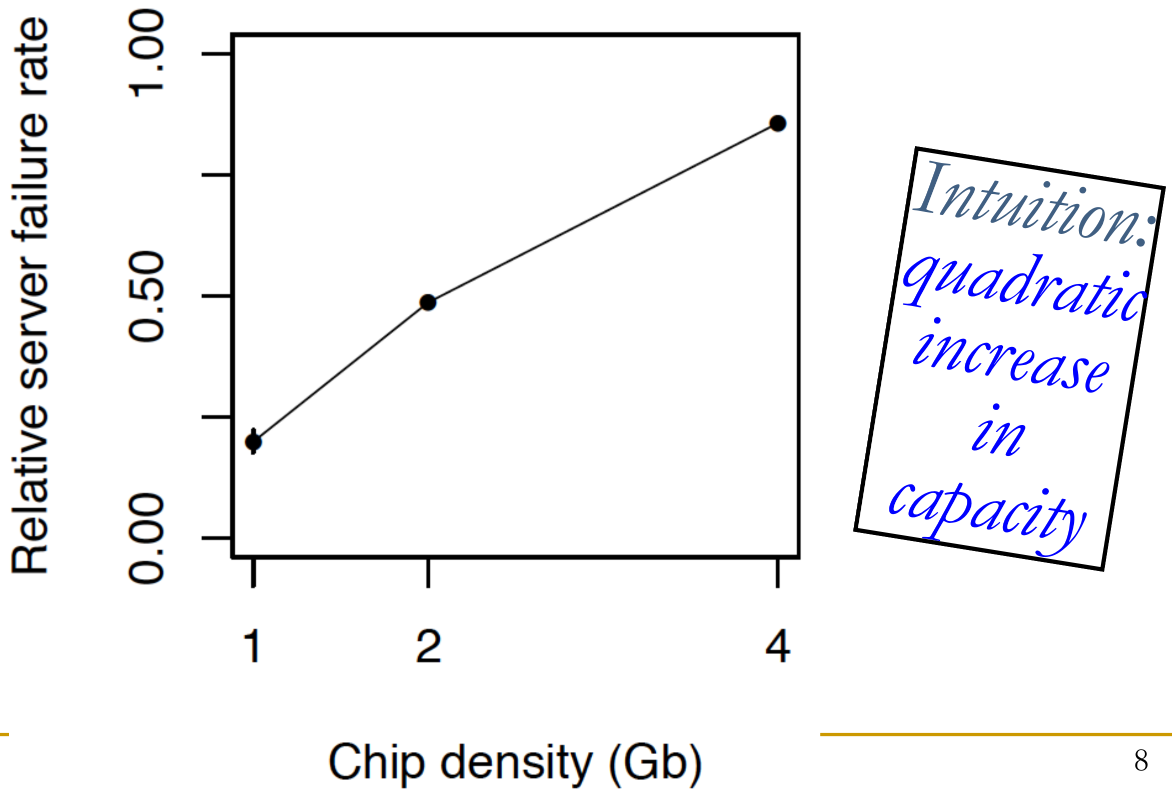
- DRAM stores charge in a capacitor (charge-based memory)
  - Capacitor must be large enough for reliable sensing
  - Access transistor should be large enough for low leakage and high retention time
  - Scaling beyond 40-35nm (2013) is challenging [ITRS, 2009]



- DRAM capacity, cost, and energy/power hard to scale

As Memory Scales, It Becomes Unreliable

- Data from all of Facebook's servers worldwide
- Meza+, "Revisiting Memory Errors in Large-Scale Production Data Centers," DSN'15.



Large-Scale Failure Analysis of DRAM Chips

- Analysis and modeling of memory errors found in all of Facebook's server fleet

- Justin Meza, Qiang Wu, Sanjeev Kumar, and Onur Mutlu,
"Revisiting Memory Errors in Large-Scale Production Data Centers: Analysis and Modeling of New Trends from the Field"
*Proceedings of the 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (**DSN**)*, Rio de Janeiro, Brazil, June 2015.
[Slides (pptx) (pdf)] [DRAM Error Model]

Revisiting Memory Errors in Large-Scale Production Data Centers: Analysis and Modeling of New Trends from the Field

Justin Meza Qiang Wu * Sanjeev Kumar * Onur Mutlu
Carnegie Mellon University * Facebook, Inc.

Infrastructures to Understand Such Issues

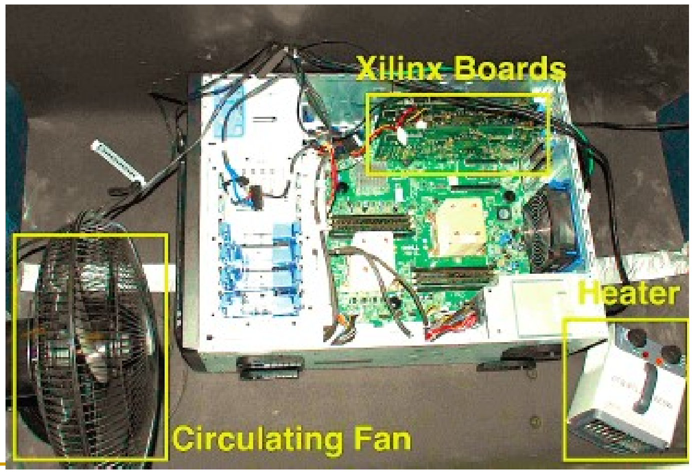
An Experimental Study of Data Retention Behavior in Modern DRAM Devices: Implications for Retention Time Profiling Mechanisms (Liu et al., ISCA 2013)

The Efficacy of Error Mitigation Techniques for DRAM Retention Failures: A Comparative Experimental Study (Khan et al., SIGMETRICS 2014)

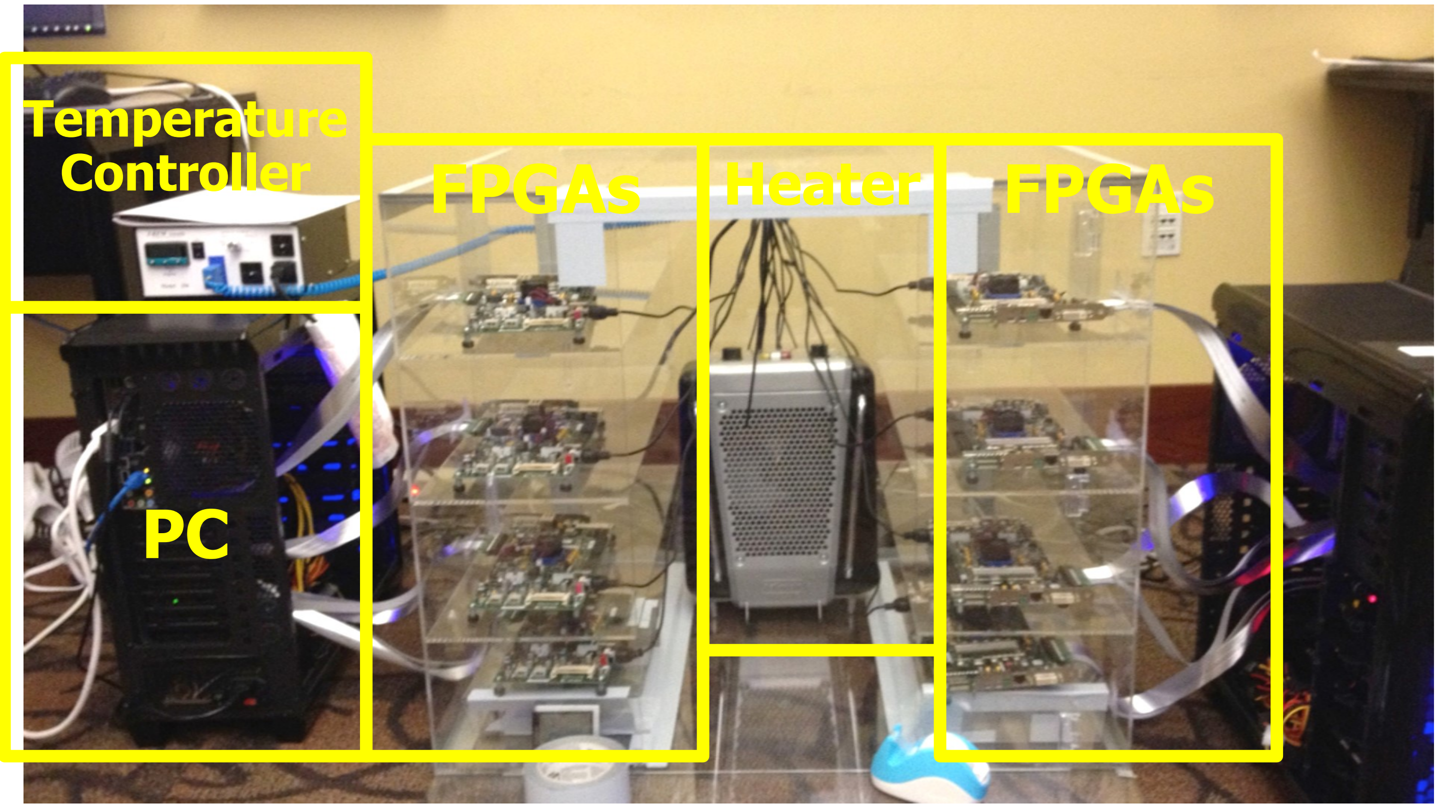
Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors (Kim et al., ISCA 2014)

Adaptive-Latency DRAM: Optimizing DRAM Timing for the Common-Case (Lee et al., HPCA 2015)

AVATAR: A Variable-Retention-Time (VRT) Aware Refresh for DRAM Systems (Qureshi et al., DSN 2015)



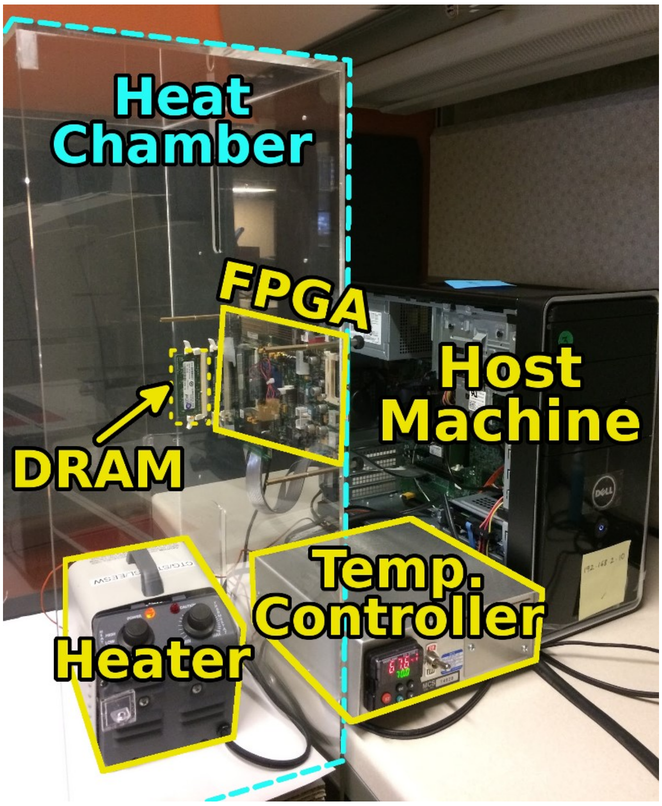
Infrastructures to Understand Such Issues



Kim+, "Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors," ISCA 2014.

SoftMC: Open Source DRAM Infrastructure

- Hasan Hassan et al., "**SoftMC: A Flexible and Practical Open-Source Infrastructure for Enabling Experimental DRAM Studies**," HPCA 2017.



- **Flexible**
- **Easy to Use (C++ API)**
- **Open-source**

  github.com/CMU-SAFARI/SoftMC

SoftMC: Open Source DRAM Infrastructure

- https://github.com/CMU-SAFARI/SoftMC

SoftMC: A Flexible and Practical Open-Source Infrastructure for Enabling Experimental DRAM Studies

Hasan Hassan[1,2,3] Nandita Vijaykumar[3] Samira Khan[4,3] Saugata Ghose[3] Kevin Chang[3]
Gennady Pekhimenko[5,3] Donghyuk Lee[6,3] Oguz Ergin[2] Onur Mutlu[1,3]

[1]*ETH Zürich* [2]*TOBB University of Economics & Technology* [3]*Carnegie Mellon University*
[4]*University of Virginia* [5]*Microsoft Research* [6]*NVIDIA Research*

SoftMC: Open Source DRAM Infrastructure

- https://github.com/CMU-SAFARI/SoftMC

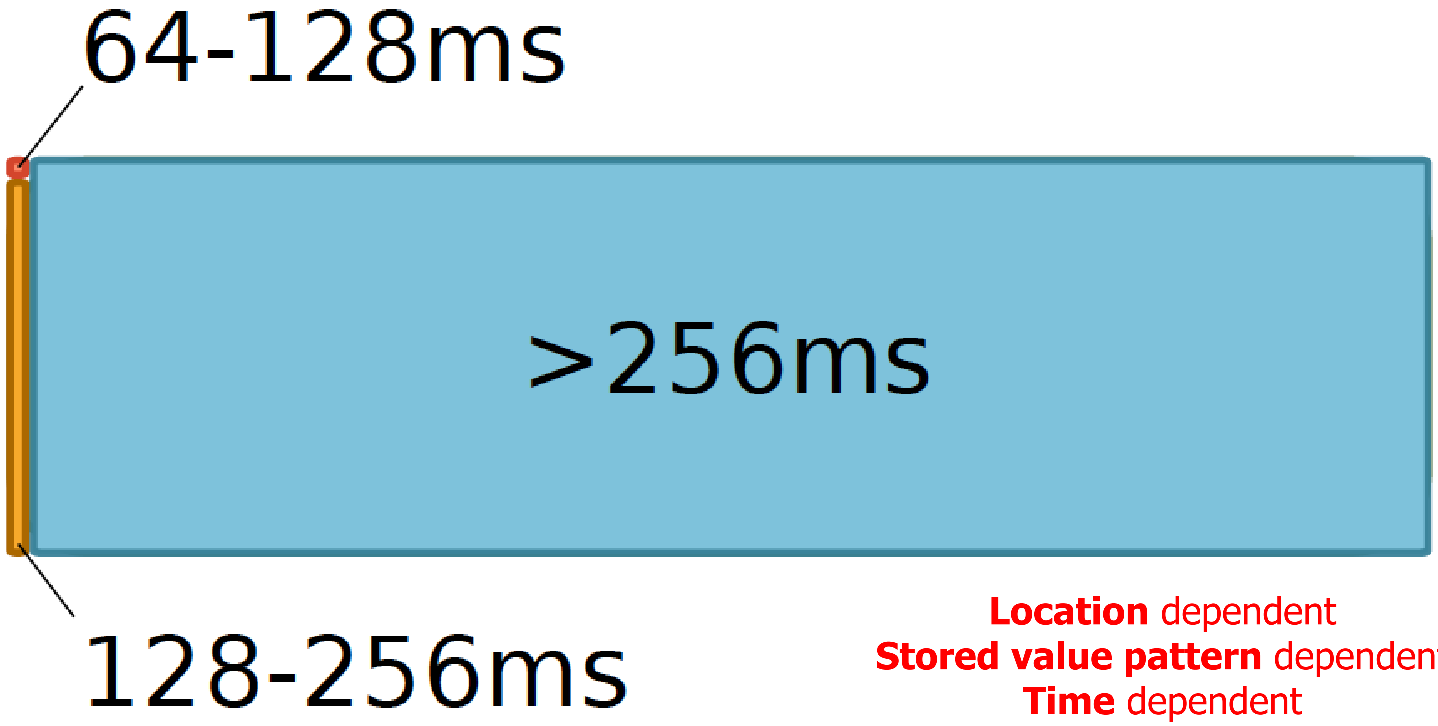
SoftMC: A Flexible and Practical Open-Source Infrastructure for Enabling Experimental DRAM Studies

Hasan Hassan[1,2,3] Nandita Vijaykumar[3] Samira Khan[4,3] Saugata Ghose[3] Kevin Chang[3]
Gennady Pekhimenko[5,3] Donghyuk Lee[6,3] Oguz Ergin[2] Onur Mutlu[1,3]

[1]*ETH Zürich* [2]*TOBB University of Economics & Technology* [3]*Carnegie Mellon University*
[4]*University of Virginia* [5]*Microsoft Research* [6]*NVIDIA Research*

Data Retention in Memory [Liu et al., ISCA 2013]

- Retention Time Profile of DRAM looks like this:



Location dependent
Stored value pattern dependent
Time dependent

Liu+, "RAIDR: Retention-Aware Intelligent DRAM Refresh," ISCA 2012.

RAIDR: Heterogeneous Refresh [ISCA'12]

- Jamie Liu, Ben Jaiyen, Richard Veras, and Onur Mutlu, **"RAIDR: Retention-Aware Intelligent DRAM Refresh"** *Proceedings of the 39th International Symposium on Computer Architecture* (***ISCA***), Portland, OR, June 2012. Slides (pdf)

RAIDR: Retention-Aware Intelligent DRAM Refresh

Jamie Liu Ben Jaiyen Richard Veras Onur Mutlu
Carnegie Mellon University

Analysis of Data Retention Failures [ISCA'13]

- Jamie Liu, Ben Jaiyen, Yoongu Kim, Chris Wilkerson, and Onur Mutlu,
  "An Experimental Study of Data Retention Behavior in Modern DRAM Devices: Implications for Retention Time Profiling Mechanisms"
  Proceedings of the 40th International Symposium on Computer Architecture (***ISCA***), Tel-Aviv, Israel, June 2013. Slides (ppt) Slides (pdf)

An Experimental Study of Data Retention Behavior in Modern DRAM Devices: Implications for Retention Time Profiling Mechanisms

Jamie Liu*
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213
jamiel@alumni.cmu.edu

Ben Jaiyen*
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213
bjaiyen@alumni.cmu.edu

Yoongu Kim
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213
yoonguk@ece.cmu.edu

Chris Wilkerson
Intel Corporation
2200 Mission College Blvd.
Santa Clara, CA 95054
chris.wilkerson@intel.com

Onur Mutlu
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213
onur@cmu.edu

Mitigation of Retention Issues [SIGMETRICS'14]

- Samira Khan, Donghyuk Lee, Yoongu Kim, Alaa Alameldeen, Chris Wilkerson, and Onur Mutlu,
  "The Efficacy of Error Mitigation Techniques for DRAM Retention Failures: A Comparative Experimental Study"
  Proceedings of the ACM International Conference on Measurement and Modeling of Computer Systems (***SIGMETRICS***), Austin, TX, June 2014. [Slides (pptx) (pdf)] [Poster (pptx) (pdf)] [Full data sets]

The Efficacy of Error Mitigation Techniques for DRAM Retention Failures: A Comparative Experimental Study

Samira Khan[†*]
samirakhan@cmu.edu

Donghyuk Lee[†]
donghyuk1@cmu.edu

Yoongu Kim[†]
yoongukim@cmu.edu

Alaa R. Alameldeen[*]
alaa.r.alameldeen@intel.com

Chris Wilkerson[*]
chris.wilkerson@intel.com

Onur Mutlu[†]
onur@cmu.edu

[†]Carnegie Mellon University [*]Intel Labs

Mitigation of Retention Issues [DSN'15]

- Moinuddin Qureshi, Dae Hyun Kim, Samira Khan, Prashant Nair, and Onur Mutlu,
  "AVATAR: A Variable-Retention-Time (VRT) Aware Refresh for DRAM Systems"
  Proceedings of the 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (***DSN***), Rio de Janeiro, Brazil, June 2015.
  [Slides (pptx) (pdf)]

AVATAR: A Variable-Retention-Time (VRT) Aware Refresh for DRAM Systems

Moinuddin K. Qureshi† Dae-Hyun Kim† Samira Khan‡ Prashant J. Nair† Onur Mutlu‡
†Georgia Institute of Technology
{moin, dhkim, pnair6}@ece.gatech.edu
‡Carnegie Mellon University
{samirakhan, onur}@cmu.edu

Mitigation of Retention Issues [DSN'16]

- Samira Khan, Donghyuk Lee, and Onur Mutlu,
  "PARBOR: An Efficient System-Level Technique to Detect Data-Dependent Failures in DRAM"
  Proceedings of the 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (***DSN***), Toulouse, France, June 2016.
  [Slides (pptx) (pdf)]

PARBOR: An Efficient System-Level Technique to Detect Data-Dependent Failures in DRAM

Samira Khan⋆ Donghyuk Lee†‡ Onur Mutlu*†
⋆University of Virginia †Carnegie Mellon University ‡Nvidia *ETH Zürich

Mitigation of Retention Issues [MICRO'17]

- Samira Khan, Chris Wilkerson, Zhe Wang, Alaa R. Alameldeen, Donghyuk Lee, and Onur Mutlu,
"Detecting and Mitigating Data-Dependent DRAM Failures by Exploiting Current Memory Content"
*Proceedings of the 50th International Symposium on Microarchitecture (**MICRO**)*, Boston, MA, USA, October 2017.
[Slides (pptx) (pdf)] [Lightning Session Slides (pptx) (pdf)] [Poster (pptx) (pdf)]

Detecting and Mitigating Data-Dependent DRAM Failures by Exploiting Current Memory Content

Samira Khan* Chris Wilkerson† Zhe Wang† Alaa R. Alameldeen† Donghyuk Lee‡ Onur Mutlu*
*University of Virginia †Intel Labs ‡Nvidia Research *ETH Zürich

Mitigation of Retention Issues [ISCA'17]

- Minesh Patel, Jeremie S. Kim, and Onur Mutlu,
"The Reach Profiler (REAPER): Enabling the Mitigation of DRAM Retention Failures via Profiling at Aggressive Conditions"
*Proceedings of the 44th International Symposium on Computer Architecture (**ISCA**)*, Toronto, Canada, June 2017.
[Slides (pptx) (pdf)]
[Lightning Session Slides (pptx) (pdf)]

- First experimental analysis of (mobile) LPDDR4 chips
- Analyzes the complex tradeoff space of retention time profiling
- Idea: enable fast and robust profiling at higher refresh intervals & temperatures

The Reach Profiler (REAPER): Enabling the Mitigation of DRAM Retention Failures via Profiling at Aggressive Conditions

Minesh Patel§‡ Jeremie S. Kim‡§ Onur Mutlu§‡
§ETH Zürich ‡Carnegie Mellon University

Mitigation of Retention Issues [DSN'19]

- Minesh Patel, Jeremie S. Kim, Hasan Hassan, and Onur Mutlu,
"Understanding and Modeling On-Die Error Correction in Modern DRAM: An Experimental Study Using Real Devices"
Proceedings of the 49th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (***DSN***), Portland, OR, USA, June 2019.
[Source Code for EINSim, the Error Inference Simulator]
Best paper award.

Understanding and Modeling On-Die Error Correction in Modern DRAM: An Experimental Study Using Real Devices

Minesh Patel[†] Jeremie S. Kim[‡†] Hasan Hassan[†] Onur Mutlu[†‡]
[†]*ETH Zürich* [‡]*Carnegie Mellon University*

Mitigation of Retention Issues [MICRO'20]

- Minesh Patel, Jeremie S. Kim, Taha Shahroodi, Hasan Hassan, and Onur Mutlu,
"Bit-Exact ECC Recovery (BEER): Determining DRAM On-Die ECC Functions by Exploiting DRAM Data Retention Characteristics"
Proceedings of the 53rd International Symposium on Microarchitecture (***MICRO***), Virtual, October 2020.
[Slides (pptx) (pdf)]
[Lightning Talk Slides (pptx) (pdf)]
[Talk Video (15 minutes)]
[Lightning Talk Video (1.5 minutes)]
Best paper award.

Bit-Exact ECC Recovery (BEER): Determining DRAM On-Die ECC Functions by Exploiting DRAM Data Retention Characteristics

Minesh Patel[†] Jeremie S. Kim[‡†] Taha Shahroodi[†] Hasan Hassan[†] Onur Mutlu[†‡]
[†]*ETH Zürich* [‡]*Carnegie Mellon University*

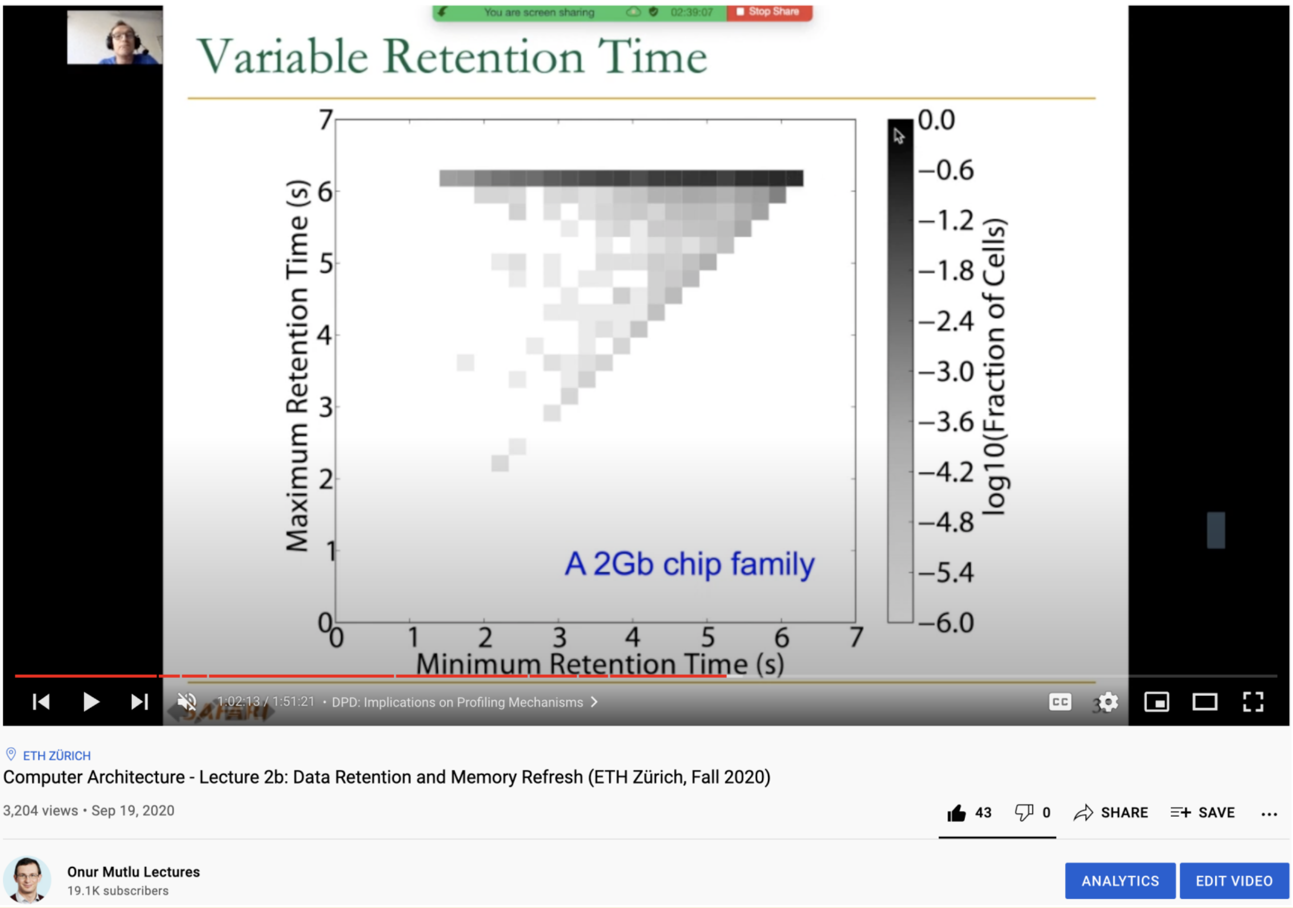
Mitigation of Retention Issues [MICRO'21]

- Minesh Patel, Geraldo F. de Oliveira Jr., and Onur Mutlu,
  "HARP: Practically and Effectively Identifying Uncorrectable Errors in Memory Chips That Use On-Die Error-Correcting Codes"
  Proceedings of the 54th International Symposium on Microarchitecture (***MICRO***), Virtual, October 2021.
  [Slides (pptx) (pdf)]
  [Short Talk Slides (pptx) (pdf)]
  [Lightning Talk Slides (pptx) (pdf)]
  [Talk Video (20 minutes)]
  [Lightning Talk Video (1.5 minutes)]
  [HARP Source Code (Officially Artifact Evaluated with All Badges)]

HARP: Practically and Effectively Identifying Uncorrectable Errors in Memory Chips That Use On-Die Error-Correcting Codes

Minesh Patel, ETH Zürich
Geraldo F. Oliveira, ETH Zürich
Onur Mutlu, ETH Zürich

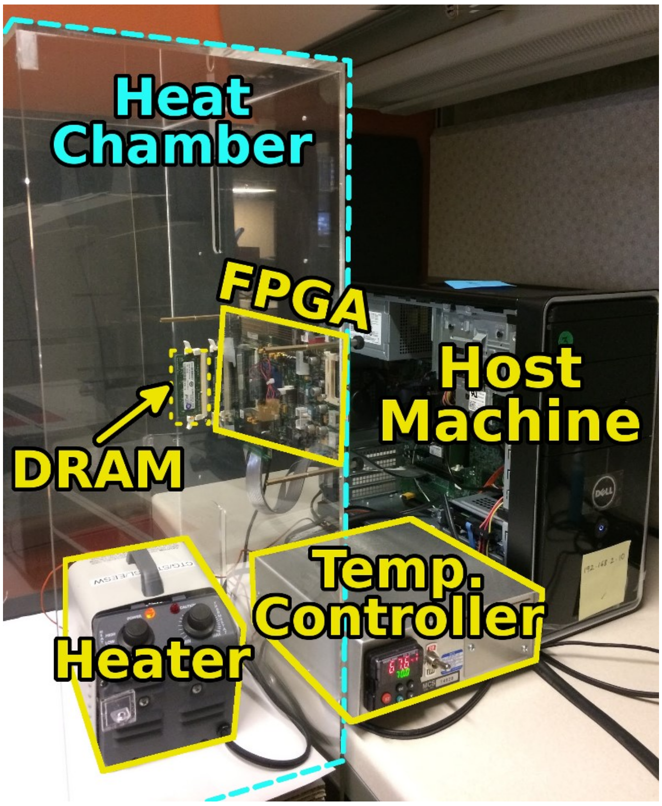
More on DRAM Refresh & Data Retention



https://www.youtube.com/watch?v=v702wUnaWGE&list=PL5Q2soXY2Zi9xidyIgBxUz7xRPS-wisBN&index=3

SoftMC: Enabling DRAM Infrastructure

- Hasan Hassan et al., "**SoftMC: A Flexible and Practical Open-Source Infrastructure for Enabling Experimental DRAM Studies**," HPCA 2017.



- **Flexible**
- **Easy to Use (C++ API)**
- **Open-source**
  github.com/CMU-SAFARI/SoftMC

A Curious Phenomenon

A Curious Discovery [Kim et al., ISCA 2014]

One can

predictably induce errors

in most DRAM memory chips

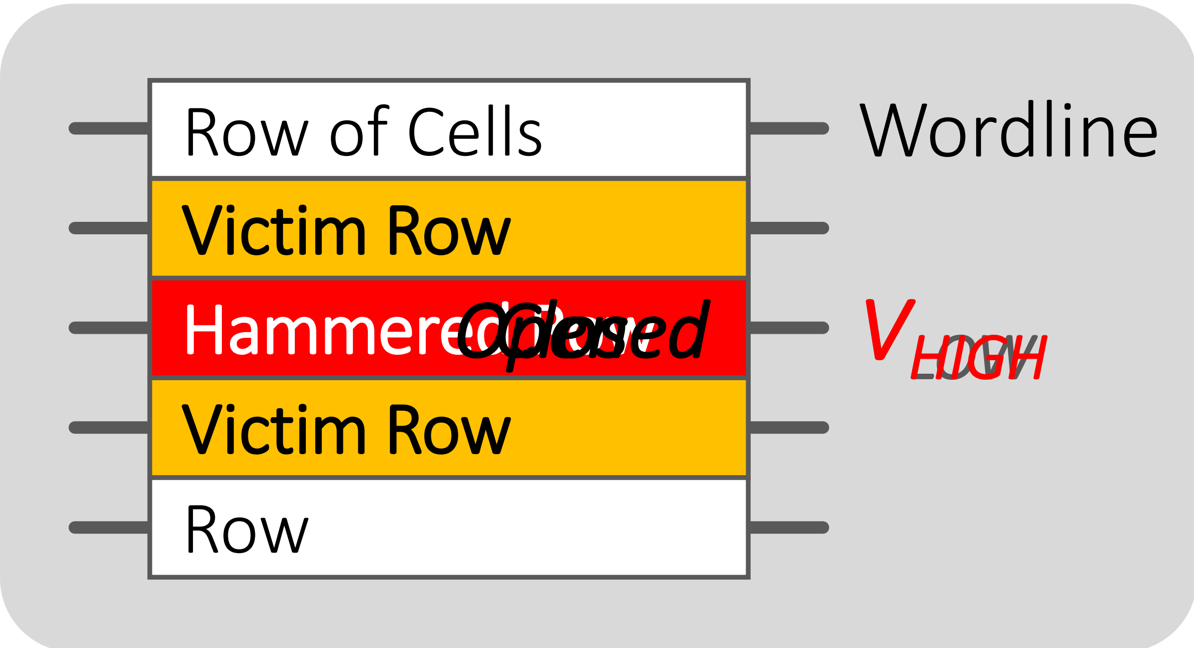
DRAM RowHammer

A simple hardware failure mechanism

can create a widespread

system security vulnerability

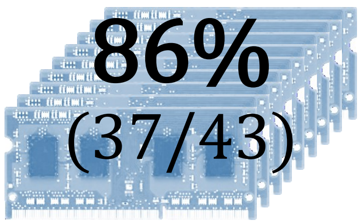
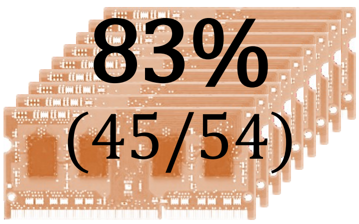
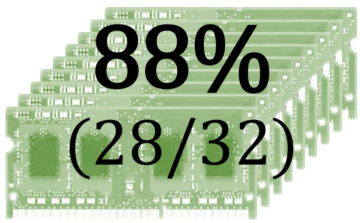
Modern DRAM is Prone to Disturbance Errors



Repeatedly reading a row enough times (before memory gets refreshed) induces **disturbance errors** in **adjacent rows** in **most real DRAM chips you can buy today**

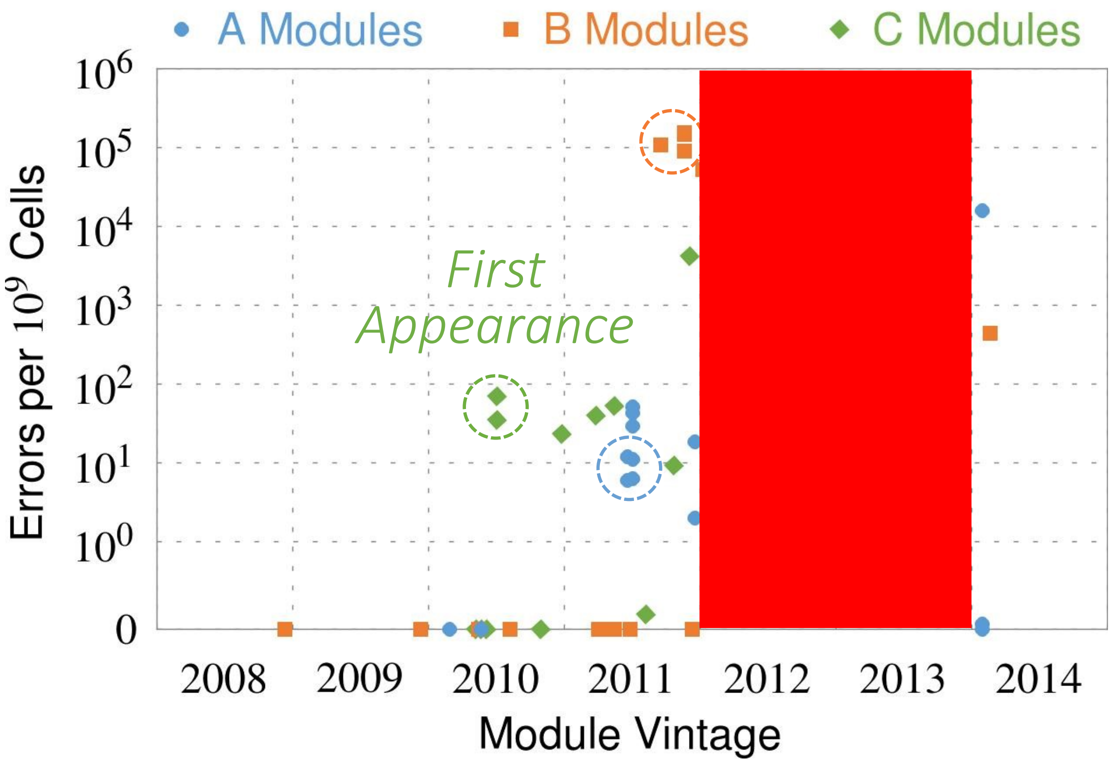
[Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors](), (Kim et al., ISCA 2014)

Most DRAM Modules Are Vulnerable

A company	B company	C company
		
86% (37/43)	**83%** (45/54)	**88%** (28/32)
Up to **1.0×10^{7}** errors	Up to **2.7×10^{6}** errors	Up to **3.3×10^{5}** errors

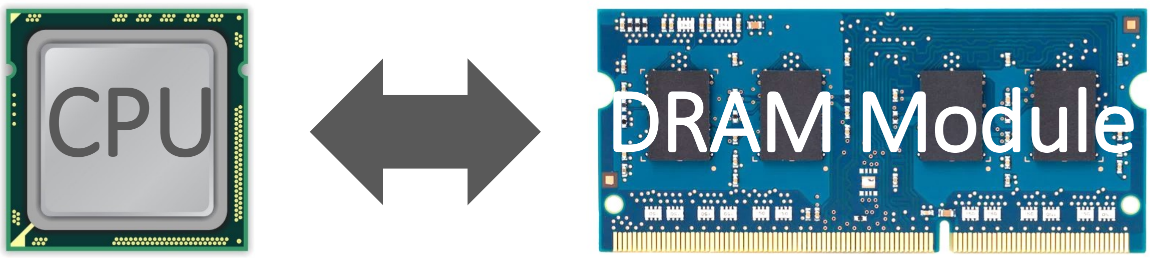
[Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors](), (Kim et al., ISCA 2014)

Recent DRAM Is More Vulnerable

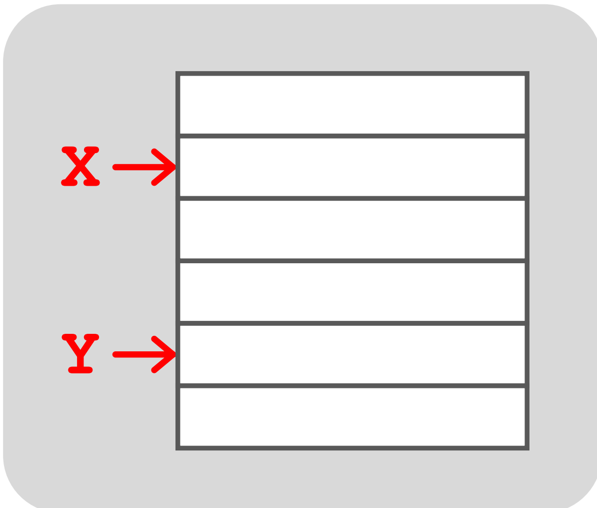


All modules from 2012–2013 are vulnerable

A Simple Program Can Induce Many Errors

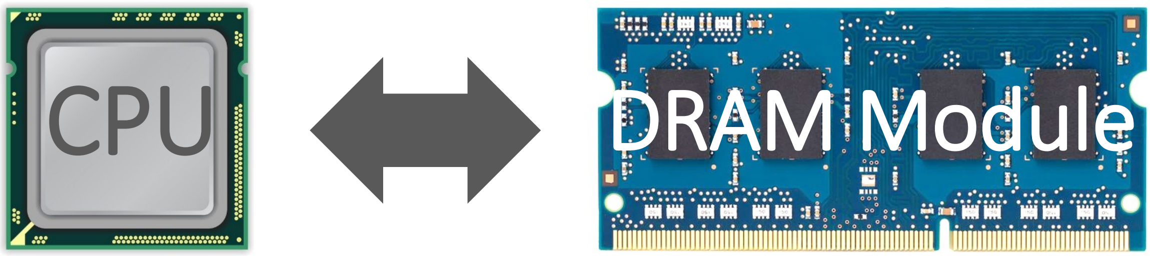



```
loop:
  mov (X), %eax
  mov (Y), %ebx
  clflush (X)
  clflush (Y)
  mfence
  jmp loop
```

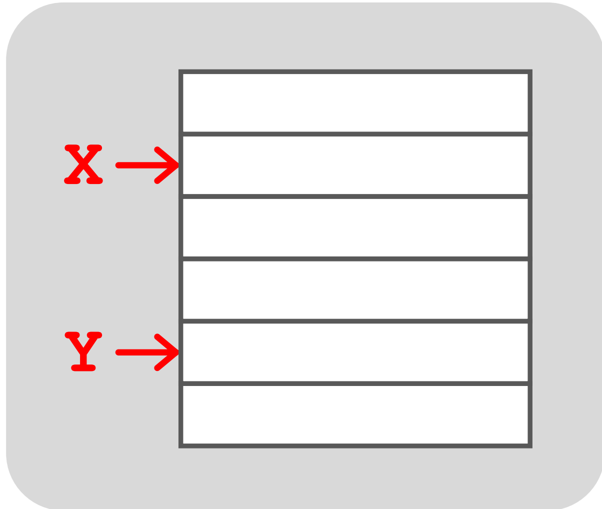


Download from: https://github.com/CMU-SAFARI/rowhammer

A Simple Program Can Induce Many Errors

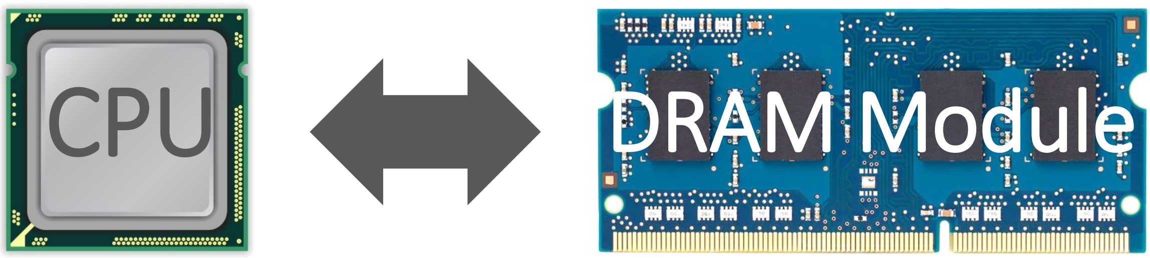


1. Avoid *cache hits*
   – Flush **X** from cache
2. Avoid *row hits* to **X**
   – Read **Y** in another row

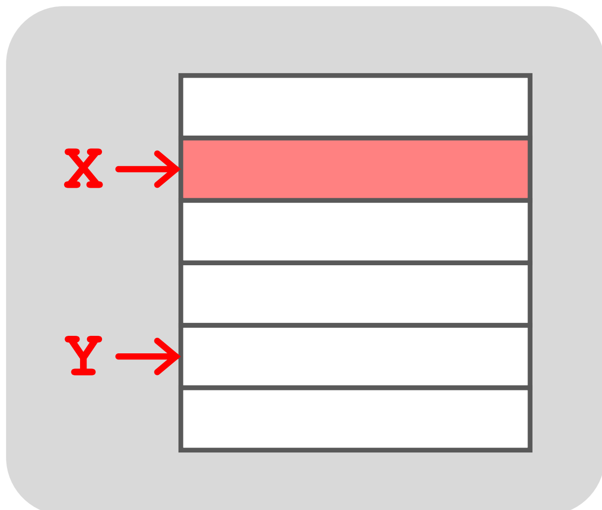


Download from: https://github.com/CMU-SAFARI/rowhammer

A Simple Program Can Induce Many Errors

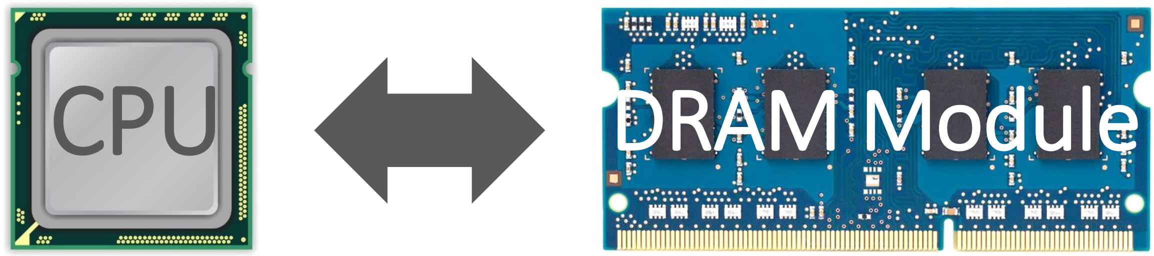



```
loop:
  mov (X), %eax
  mov (Y), %ebx
  clflush (X)
  clflush (Y)
  mfence
  jmp loop
```

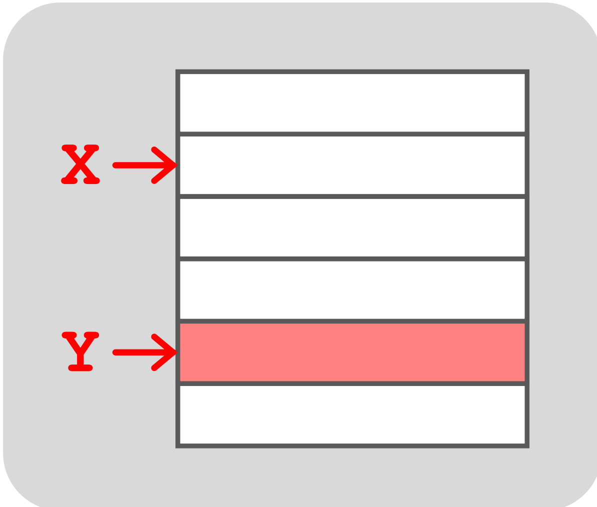


Download from: https://github.com/CMU-SAFARI/rowhammer

A Simple Program Can Induce Many Errors

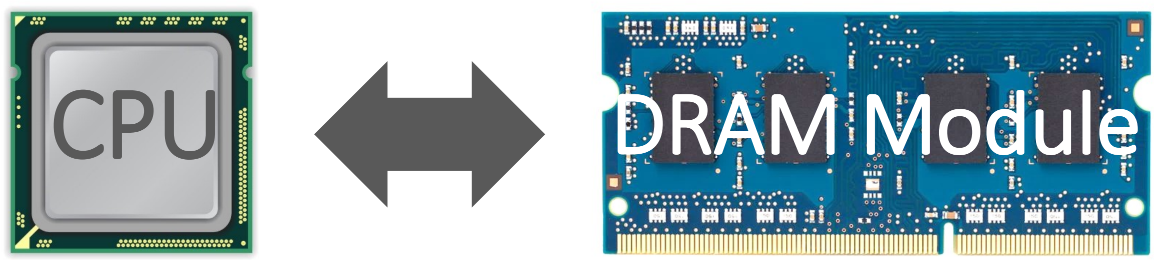



```
loop:
  mov  (X), %eax
  mov  (Y), %ebx
  clflush  (X)
  clflush  (Y)
  mfence
  jmp loop
```

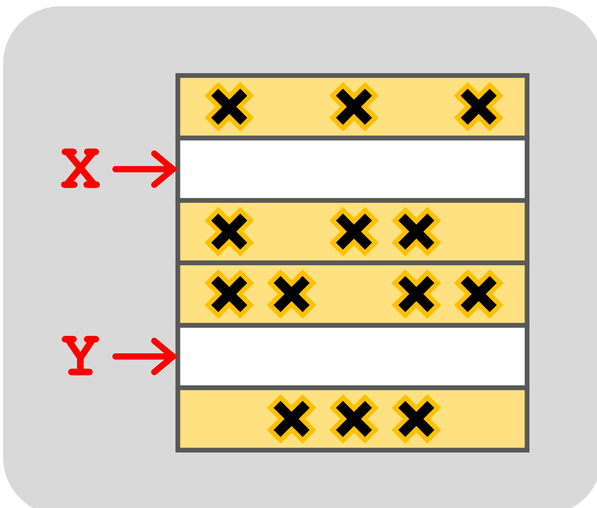


Download from: https://github.com/CMU-SAFARI/rowhammer

A Simple Program Can Induce Many Errors




```
loop:
  mov  (X), %eax
  mov  (Y), %ebx
  clflush  (X)
  clflush  (Y)
  mfence
  jmp loop
```



Download from: https://github.com/CMU-SAFARI/rowhammer

Observed Errors in Real Systems

CPU Architecture	Errors	Access-Rate
Intel Haswell (2013)	22.9K	12.3M/sec
Intel Ivy Bridge (2012)	20.7K	11.7M/sec
Intel Sandy Bridge (2011)	16.1K	11.6M/sec
AMD Piledriver (2012)	59	6.1M/sec

A real reliability & security issue

Kim+, "Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors," ISCA 2014.

One Can Take Over an Otherwise-Secure System

Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors

Abstract. Memory isolation is a key property of a reliable and secure computing system — an access to one memory address should not have unintended side effects on data stored in other addresses. However, as DRAM process technology

Project Zero

Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors (Kim et al., ISCA 2014)

News and updates from the Project Zero team at Google

Exploiting the DRAM rowhammer bug to gain kernel privileges (Seaborn, 2015)

Monday, March 9, 2015

Exploiting the DRAM rowhammer bug to gain kernel privileges

RowHammer Security Attack Example

- "Rowhammer" is a problem with some recent DRAM devices in which repeatedly accessing a row of memory can cause bit flips in adjacent rows (Kim et al., ISCA 2014).
  - Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors (Kim et al., ISCA 2014)

- We tested a selection of laptops and found that a subset of them exhibited the problem.
- We built two working privilege escalation exploits that use this effect.
  - Exploiting the DRAM rowhammer bug to gain kernel privileges (Seaborn+, 2015)

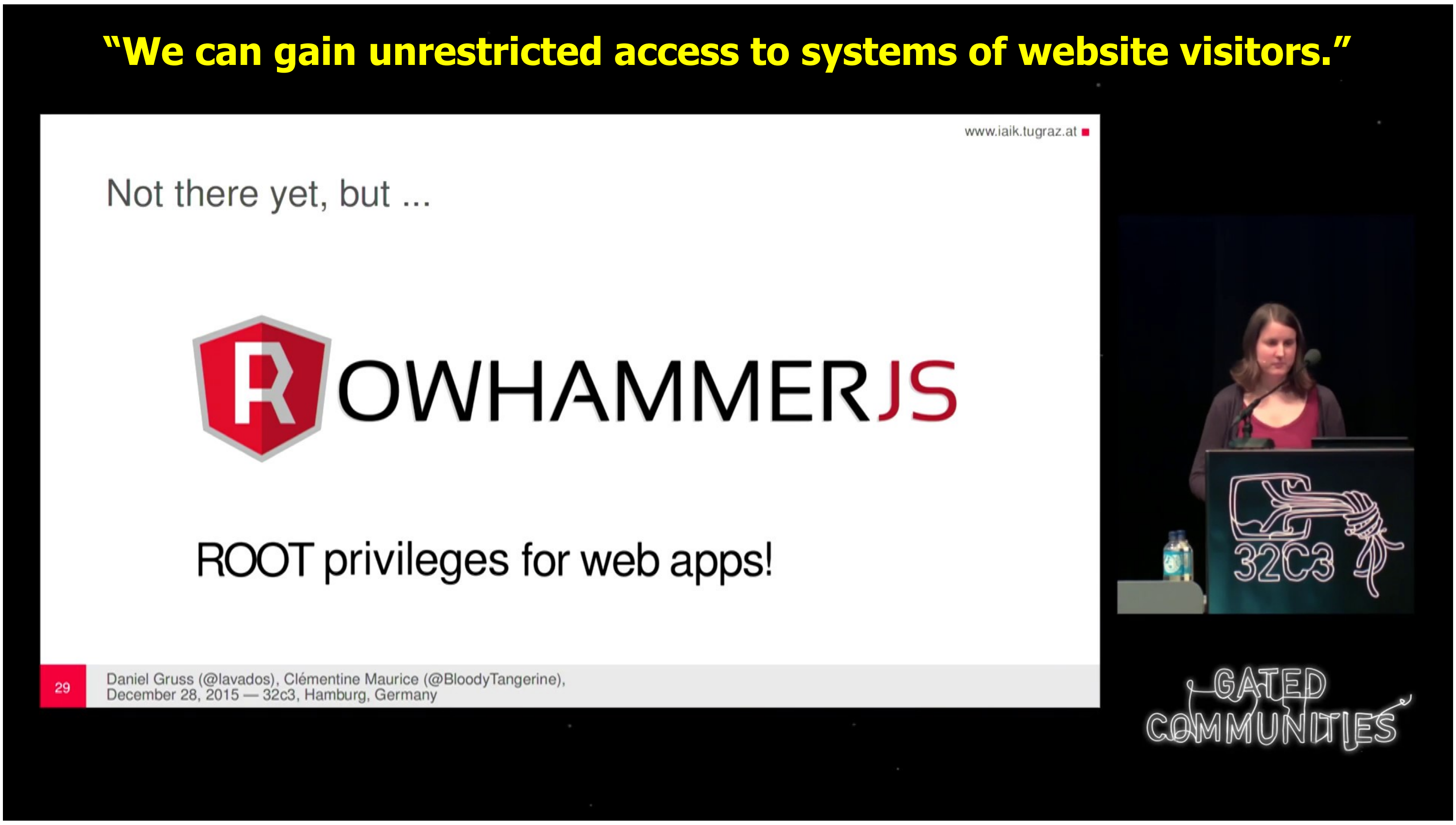
- One exploit uses rowhammer-induced bit flips to gain kernel privileges on x86-64 Linux when run as an unprivileged userland process.
- When run on a machine vulnerable to the rowhammer problem, the process was able to induce bit flips in page table entries (PTEs).
- It was able to use this to gain write access to its own page table, and hence gain read-write access to all of physical memory.

Exploiting the DRAM rowhammer bug to gain kernel privileges (Seaborn & Dullien, 2015)

Security Implications

Security Implications

More Security Implications (I)



Rowhammer.js: A Remote Software-Induced Fault Attack in JavaScript (DIMVA'16)

Source: https://lab.dsst.io/32c3-slides/7197.html

More Security Implications (II)

"Can gain control of a smart phone deterministically"

Drammer: Deterministic Rowhammer Attacks on Mobile Platforms, CCS'16

Source: https://fossbytes.com/drammer-rowhammer-attack-android-root-devices/

More Security Implications (III)

- Using an integrated GPU in a mobile system to remotely escalate privilege via the WebGL interface. IEEE S&P 2018

ars TECHNICA   BIZ & IT   TECH   SCIENCE   POLICY   CARS   GAMING & CULTURE

"GRAND PWNING UNIT" —

Drive-by Rowhammer attack uses GPU to compromise an Android phone

JavaScript based GLitch pwns browsers by flipping bits inside memory chips.

DAN GOODIN - 5/3/2018, 12:00 PM

Grand Pwning Unit: Accelerating Microarchitectural Attacks with the GPU

Pietro Frigo, Vrije Universiteit Amsterdam, p.frigo@vu.nl

Cristiano Giuffrida, Vrije Universiteit Amsterdam, giuffrida@cs.vu.nl

Herbert Bos, Vrije Universiteit Amsterdam, herbertb@cs.vu.nl

Kaveh Razavi, Vrije Universiteit Amsterdam, kaveh@cs.vu.nl

More Security Implications (IV)

- Rowhammer over RDMA (I) USENIX ATC 2018

ars TECHNICA

BIZ & IT TECH SCIENCE POLICY CARS GAMING & CULTURE

THROWHAMMER —

Packets over a LAN are all it takes to trigger serious Rowhammer bit flips

The bar for exploiting potentially serious DDR weakness keeps getting lower.

DAN GOODIN - 5/10/2018, 5:26 PM

Throwhammer: Rowhammer Attacks over the Network and Defenses

Andrei Tatar
VU Amsterdam

Radhesh Krishnan
VU Amsterdam

Elias Athanasopoulos
University of Cyprus

Cristiano Giuffrida
VU Amsterdam

Herbert Bos
VU Amsterdam

Kaveh Razavi
VU Amsterdam

More Security Implications (V)

- Rowhammer over RDMA (II)

The Hacker News™
Security in a serious way

Nethammer—Exploiting DRAM Rowhammer Bug Through Network Requests

Nethammer:
Inducing Rowhammer Faults through Network Requests

Moritz Lipp
Graz University of Technology

Misiker Tadesse Aga
University of Michigan

Michael Schwarz
Graz University of Technology

Daniel Gruss
Graz University of Technology

Clémentine Maurice
Univ Rennes, CNRS, IRISA

Lukas Raab
Graz University of Technology

Lukas Lamster
Graz University of Technology

More Security Implications (VI)

- IEEE S&P 2020

RAMBleed

RAMBleed: Reading Bits in Memory Without Accessing Them

Andrew Kwong
University of Michigan
ankwong@umich.edu

Daniel Genkin
University of Michigan
genkin@umich.edu

Daniel Gruss
Graz University of Technology
daniel.gruss@iaik.tugraz.at

Yuval Yarom
University of Adelaide and Data61
yval@cs.adelaide.edu.au

More Security Implications (VII)

- USENIX Security 2019

Terminal Brain Damage: Exposing the Graceless Degradation in Deep Neural Networks Under Hardware Fault Attacks

Sanghyun Hong, Pietro Frigo[†], Yiğitcan Kaya, Cristiano Giuffrida[†], Tudor Dumitraş

University of Maryland, College Park
[†]*Vrije Universiteit Amsterdam*

A Single Bit-flip Can Cause Terminal Brain Damage to DNNs

One specific bit-flip in a DNN's representation leads to accuracy drop over 90%

Our research found that a specific bit-flip in a DNN's bitwise representation can cause the accuracy loss up to 90%, and the DNN has 40-50% parameters, on average, that can lead to the accuracy drop over 10% when individually subjected to such single bitwise corruptions...

Read More

More Security Implications (VIII)

- USENIX Security 2020

DeepHammer: Depleting the Intelligence of Deep Neural Networks through Targeted Chain of Bit Flips

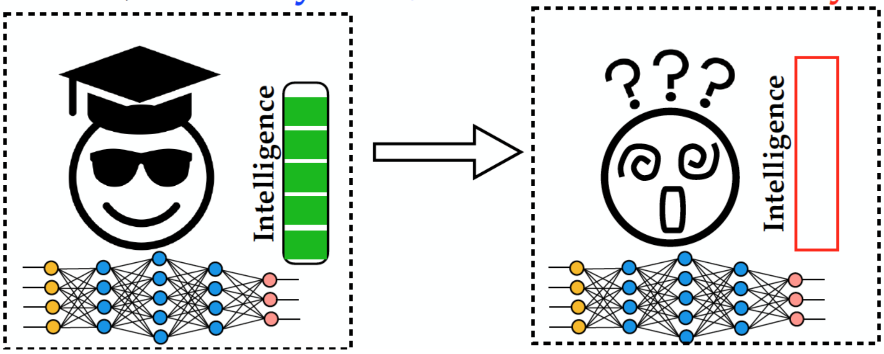
Fan Yao
University of Central Florida
fan.yao@ucf.edu

Adnan Siraj Rakin
Deliang Fan
Arizona State University
asrakin@asu.edu
dfan@asu.edu

Degrade the **inference accuracy** to the level of **Random Guess**

Example: ResNet-20 for CIFAR-10, **10** output classes

Before attack, **Accuracy: 90.2%** After attack, **Accuracy: ~10% (1/10)**



More Security Implications (IX)

- Rowhammer on MLC NAND Flash (based on [Cai+, HPCA 2017])

The Register®
Biting the hand that feeds IT

Security

Rowhammer RAM attack adapted to hit flash storage

Project Zero's two-year-old dog learns a new trick

By Richard Chirgwin 17 Aug 2017 at 04:27 17 SHARE ▼

From random block corruption to privilege escalation: A filesystem attack vector for rowhammer-like attacks

Anil Kurmus Nikolas Ioannou Matthias Neugschwandtner Nikolaos Papandreou
Thomas Parnell
IBM Research – Zurich

More Security Implications?

A RowHammer Survey Across the Stack

- Onur Mutlu and Jeremie Kim,
  "RowHammer: A Retrospective"
  IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (***TCAD***) *Special Issue on Top Picks in Hardware and Embedded Security*, 2019.
  [Preliminary arXiv version]
  [Slides from COSADE 2019 (pptx)]
  [Slides from VLSI-SOC 2020 (pptx) (pdf)]
  [Talk Video (1 hr 15 minutes, with Q&A)]

RowHammer: A Retrospective

Onur Mutlu[§‡] Jeremie S. Kim[‡§]
[§]ETH Zürich [‡]Carnegie Mellon University

Understanding RowHammer

First RowHammer Analysis

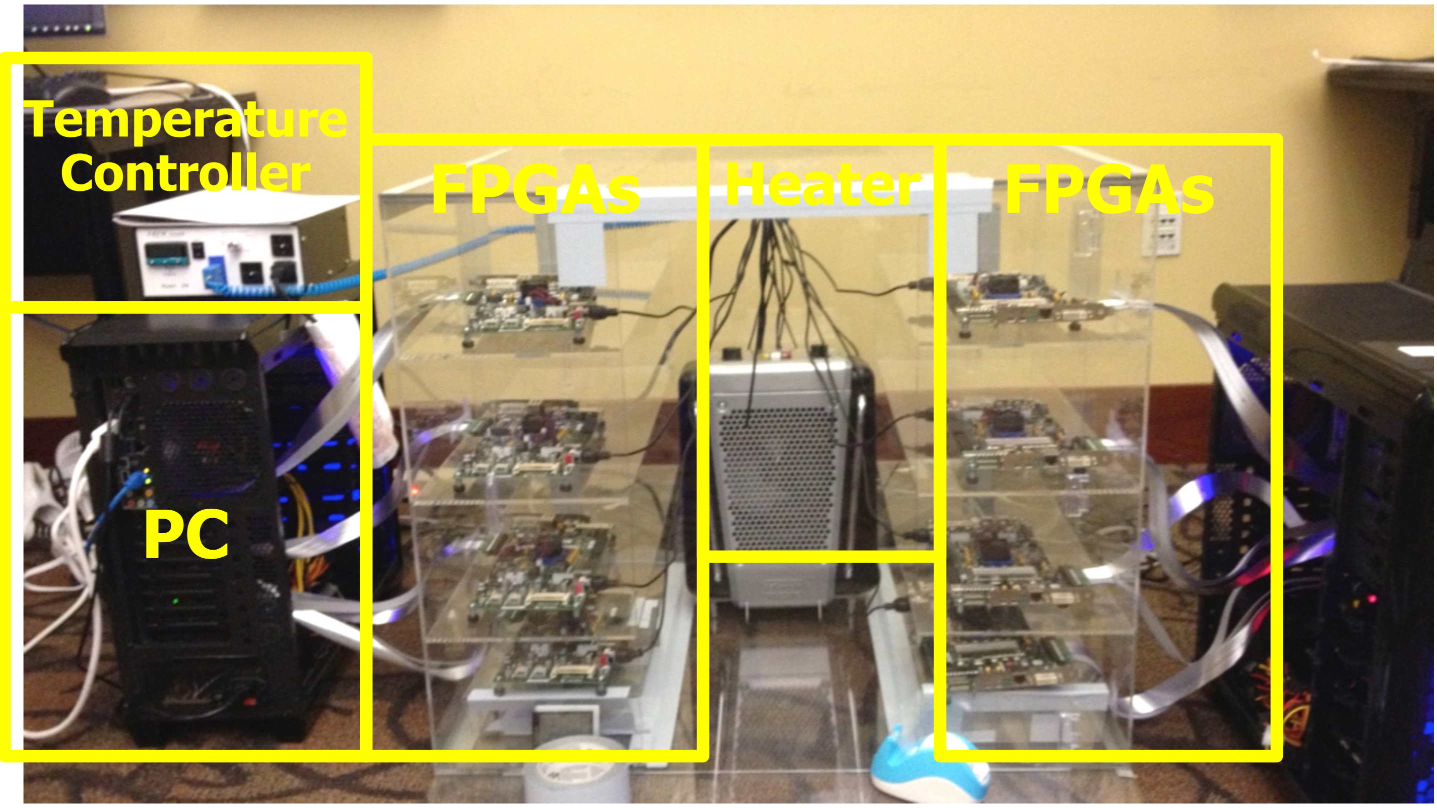
- Yoongu Kim, Ross Daly, Jeremie Kim, Chris Fallin, Ji Hye Lee, Donghyuk Lee, Chris Wilkerson, Konrad Lai, and Onur Mutlu,
"Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors"
Proceedings of the 41st International Symposium on Computer Architecture (***ISCA***), Minneapolis, MN, June 2014.
[Slides (pptx) (pdf)] [Lightning Session Slides (pptx) (pdf)] [Source Code and Data] [Lecture Video (1 hr 49 mins), 25 September 2020]
One of the 7 papers of 2012-2017 selected as Top Picks in Hardware and Embedded Security for IEEE TCAD (link).

Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors

Yoongu Kim[1] Ross Daly* Jeremie Kim[1] Chris Fallin* Ji Hye Lee[1]
Donghyuk Lee[1] Chris Wilkerson[2] Konrad Lai Onur Mutlu[1]

[1]Carnegie Mellon University [2]Intel Labs

RowHammer Infrastructure (2012-2014)



Tested DRAM Modules from 2008-2014

(129 total)

Manufacturer	Module	Date* (yy-ww)	Timing† Freq (MT/s)	Timing† t_{RC} (ns)	Organization Size (GB)	Organization Chips	Chip Size (Gb)‡	Chip Pins	Chip DieVersion§	Victims-per-Module Average	Victims-per-Module Minimum	Victims-per-Module Maximum	RI_{th} (ms) Min
A (Total of 43 Modules)	A_1	10-08	1066	50.625	0.5	4	1	×16	$\mathcal{B}$	0	0	0	–
	A_2	10-20	1066	50.625	1	8	1	×8	$\mathcal{F}$	0	0	0	–
	A_{3-5}	10-20	1066	50.625	0.5	4	1	×16	$\mathcal{B}$	0	0	0	–
	A_{6-7}	11-24	1066	49.125	1	4	2	×16	$\mathcal{D}$	7.8×10^1	5.2×10^1	1.0×10^2	21.3
	A_{8-12}	11-26	1066	49.125	1	4	2	×16	$\mathcal{D}$	2.4×10^2	5.4×10^1	4.4×10^2	16.4
	A_{13-14}	11-50	1066	49.125	1	4	2	×16	$\mathcal{D}$	8.8×10^1	1.7×10^1	1.6×10^2	26.2
	A_{15-16}	12-22	1600	50.625	1	4	2	×16	$\mathcal{D}$	9.5	9	1.0×10^1	34.4
	A_{17-18}	12-26	1600	49.125	2	8	2	×8	$\mathcal{M}$	1.2×10^2	3.7×10^1	2.0×10^2	21.3
	A_{19-30}	12-40	1600	48.125	2	8	2	×8	$\mathcal{K}$	8.6×10^6	7.0×10^6	$\mathbf{1.0 \times 10^7}$	**8.2**
	A_{31-34}	13-02	1600	48.125	2	8	2	×8	–	1.8×10^6	1.0×10^6	3.5×10^6	11.5
	A_{35-36}	13-14	1600	48.125	2	8	2	×8	–	4.0×10^1	1.9×10^1	6.1×10^1	21.3
	A_{37-38}	13-20	1600	48.125	2	8	2	×8	$\mathcal{K}$	1.7×10^6	1.4×10^6	2.0×10^6	9.8
	A_{39-40}	13-28	1600	48.125	2	8	2	×8	$\mathcal{K}$	5.7×10^4	5.4×10^4	6.0×10^4	16.4
	A_{41}	14-04	1600	49.125	2	8	2	×8	–	2.7×10^5	2.7×10^5	2.7×10^5	18.0
	A_{42-43}	14-04	1600	48.125	2	8	2	×8	$\mathcal{K}$	0.5	0	1	62.3
B (Total of 54 Modules)	B_1	08-49	1066	50.625	1	8	1	×8	$\mathcal{D}$	0	0	0	–
	B_2	09-49	1066	50.625	1	8	1	×8	$\mathcal{E}$	0	0	0	–
	B_3	10-19	1066	50.625	1	8	1	×8	$\mathcal{F}$	0	0	0	–
	B_4	10-31	1333	49.125	2	8	2	×8	$\mathcal{C}$	0	0	0	–
	B_5	11-13	1333	49.125	2	8	2	×8	$\mathcal{C}$	0	0	0	–
	B_6	11-16	1066	50.625	1	8	1	×8	$\mathcal{F}$	0	0	0	–
	B_7	11-19	1066	50.625	1	8	1	×8	$\mathcal{F}$	0	0	0	–
	B_8	11-25	1333	49.125	2	8	2	×8	$\mathcal{C}$	0	0	0	–
	B_9	11-37	1333	49.125	2	8	2	×8	$\mathcal{D}$	1.9×10^6	1.9×10^6	1.9×10^6	11.5
	B_{10-12}	11-46	1333	49.125	2	8	2	×8	$\mathcal{D}$	2.2×10^6	1.5×10^6	$\mathbf{2.7 \times 10^6}$	11.5
	B_{13}	11-49	1333	49.125	2	8	2	×8	$\mathcal{C}$	0	0	0	–
	B_{14}	12-01	1866	47.125	2	8	2	×8	$\mathcal{D}$	9.1×10^5	9.1×10^5	9.1×10^5	**9.8**
	B_{15-31}	12-10	1866	47.125	2	8	2	×8	$\mathcal{D}$	9.8×10^5	7.8×10^5	1.2×10^6	11.5
	B_{32}	12-25	1600	48.125	2	8	2	×8	$\mathcal{E}$	7.4×10^5	7.4×10^5	7.4×10^5	11.5
	B_{33-42}	12-28	1600	48.125	2	8	2	×8	$\mathcal{E}$	5.2×10^5	1.9×10^5	7.3×10^5	11.5
	B_{43-47}	12-31	1600	48.125	2	8	2	×8	$\mathcal{E}$	4.0×10^5	2.9×10^5	5.5×10^5	13.1
	B_{48-51}	13-19	1600	48.125	2	8	2	×8	$\mathcal{E}$	1.1×10^5	7.4×10^4	1.4×10^5	14.7
	B_{52-53}	13-40	1333	49.125	2	8	2	×8	$\mathcal{D}$	2.6×10^4	2.3×10^4	2.9×10^4	21.3
	B_{54}	14-07	1333	49.125	2	8	2	×8	$\mathcal{D}$	7.5×10^3	7.5×10^3	7.5×10^3	26.2
C (Total of 32 Modules)	C_1	10-18	1333	49.125	2	8	2	×8	$\mathcal{A}$	0	0	0	–
	C_2	10-20	1066	50.625	2	8	2	×8	$\mathcal{A}$	0	0	0	–
	C_3	10-22	1066	50.625	2	8	2	×8	$\mathcal{A}$	0	0	0	–
	C_{4-5}	10-26	1333	49.125	2	8	2	×8	$\mathcal{B}$	8.9×10^2	6.0×10^2	1.2×10^3	29.5
	C_6	10-43	1333	49.125	1	8	1	×8	$\mathcal{T}$	0	0	0	–
	C_7	10-51	1333	49.125	2	8	2	×8	$\mathcal{B}$	4.0×10^2	4.0×10^2	4.0×10^2	29.5
	C_8	11-12	1333	46.25	2	8	2	×8	$\mathcal{B}$	6.9×10^2	6.9×10^2	6.9×10^2	21.3
	C_9	11-19	1333	46.25	2	8	2	×8	$\mathcal{B}$	9.2×10^2	9.2×10^2	9.2×10^2	27.9
	C_{10}	11-31	1333	49.125	2	8	2	×8	$\mathcal{B}$	3	3	3	39.3
	C_{11}	11-42	1333	49.125	2	8	2	×8	$\mathcal{B}$	1.6×10^2	1.6×10^2	1.6×10^2	39.3
	C_{12}	11-48	1600	48.125	2	8	2	×8	$\mathcal{C}$	7.1×10^4	7.1×10^4	7.1×10^4	19.7
	C_{13}	12-08	1333	49.125	2	8	2	×8	$\mathcal{C}$	3.9×10^4	3.9×10^4	3.9×10^4	21.3
	C_{14-15}	12-12	1333	49.125	2	8	2	×8	$\mathcal{C}$	3.7×10^4	2.1×10^4	5.4×10^4	21.3
	C_{16-18}	12-20	1600	48.125	2	8	2	×8	$\mathcal{C}$	3.5×10^3	1.2×10^3	7.0×10^3	27.9
	C_{19}	12-23	1600	48.125	2	8	2	×8	$\mathcal{E}$	1.4×10^5	1.4×10^5	1.4×10^5	18.0
	C_{20}	12-24	1600	48.125	2	8	2	×8	$\mathcal{C}$	6.5×10^4	6.5×10^4	6.5×10^4	21.3
	C_{21}	12-26	1600	48.125	2	8	2	×8	$\mathcal{C}$	2.3×10^4	2.3×10^4	2.3×10^4	24.6
	C_{22}	12-32	1600	48.125	2	8	2	×8	$\mathcal{C}$	1.7×10^4	1.7×10^4	1.7×10^4	22.9
	C_{23-24}	12-37	1600	48.125	2	8	2	×8	$\mathcal{C}$	2.3×10^4	1.1×10^4	3.4×10^4	18.0
	C_{25-30}	12-41	1600	48.125	2	8	2	×8	$\mathcal{C}$	2.0×10^4	1.1×10^4	3.2×10^4	19.7
	C_{31}	13-11	1600	48.125	2	8	2	×8	$\mathcal{C}$	3.3×10^5	3.3×10^5	$\mathbf{3.3 \times 10^5}$	**14.7**
	C_{32}	13-35	1600	48.125	2	8	2	×8	$\mathcal{C}$	3.7×10^4	3.7×10^4	3.7×10^4	21.3

∗ We report the manufacture date marked on the chip packages, which is more accurate than other dates that can be gleaned from a module.
† We report timing constraints stored in the module's on-board ROM [33], which is read by the system BIOS to calibrate the memory controller.
‡ The maximum DRAM chip size supported by our testing platform is 2Gb.
§ We report DRAM die versions marked on the chip packages, which typically progress in the following manner: $\mathcal{M} \rightarrow \mathcal{A} \rightarrow \mathcal{B} \rightarrow \mathcal{C} \rightarrow \cdots$.

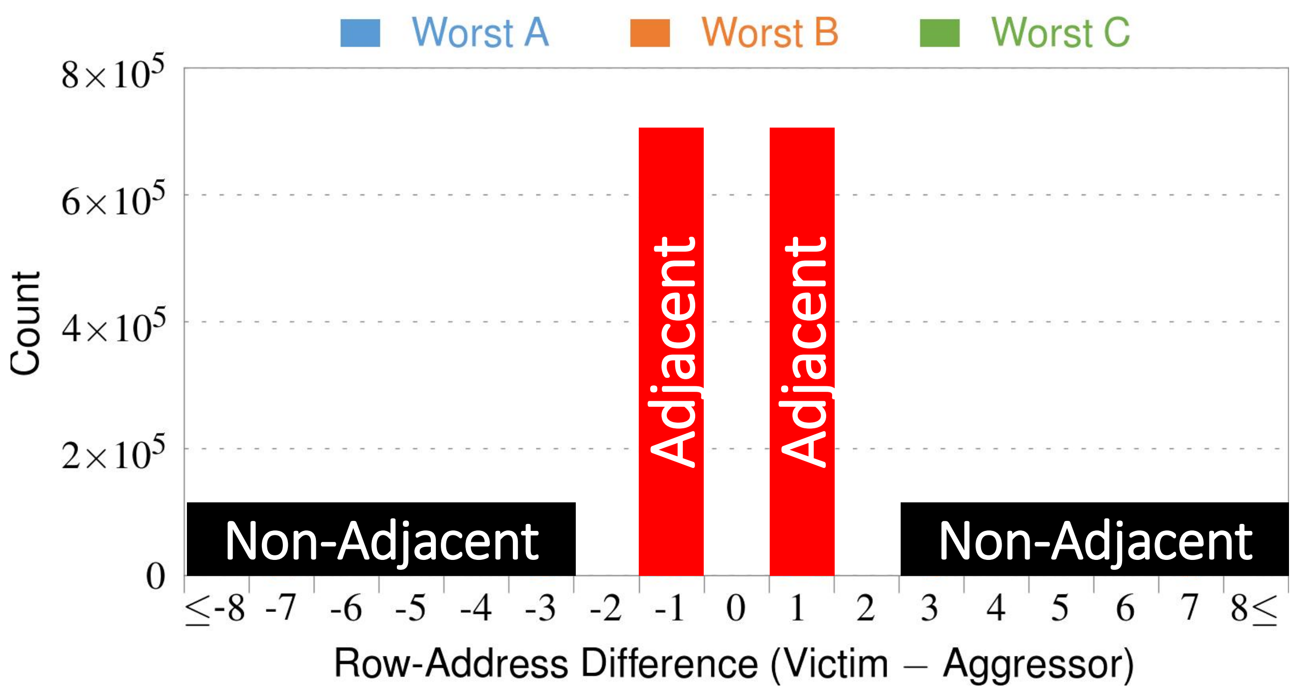
Table 3. Sample population of 129 DDR3 DRAM modules, categorized by manufacturer and sorted by manufacture date

RowHammer Characterization Results

1. Most Modules Are at Risk
2. Errors vs. Vintage
3. Error = Charge Loss
4. Adjacency: Aggressor & Victim
5. Sensitivity Studies
6. Other Results in Paper
7. Solution Space

Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors, (Kim et al., ISCA 2014)

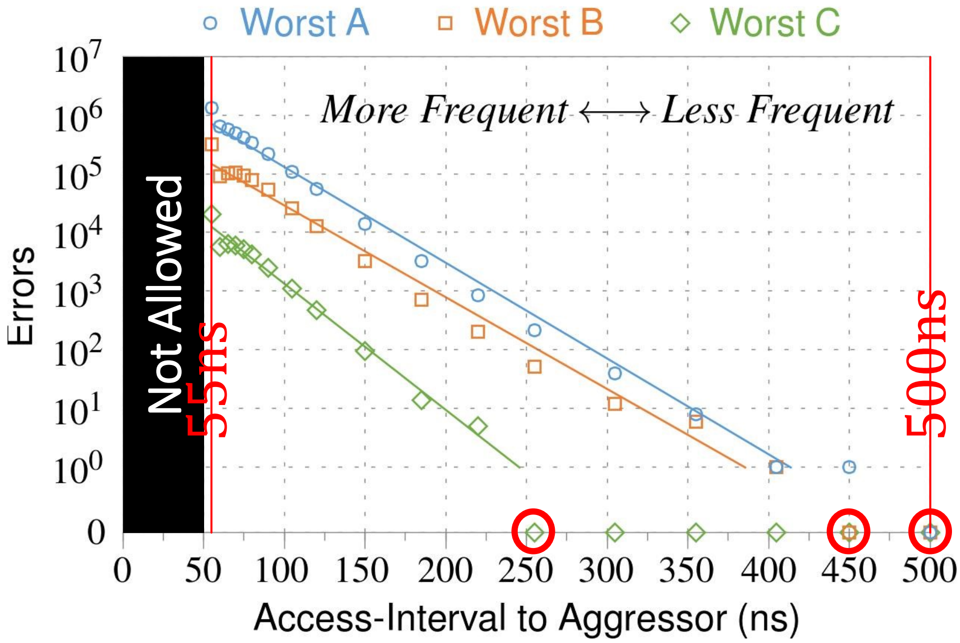
4. Adjacency: Aggressor & Victim



Note: For three modules with the most errors (only first bank)

Most aggressors & victims are adjacent

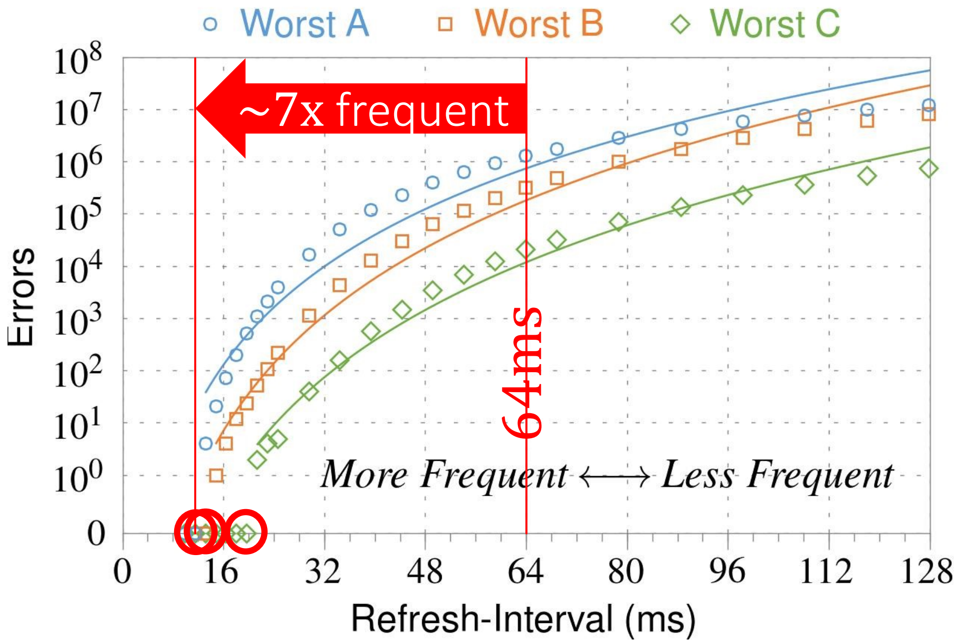
❶ Access Interval (Aggressor)



Note: For three modules with the most errors (only first bank)

Less frequent accesses → Fewer errors

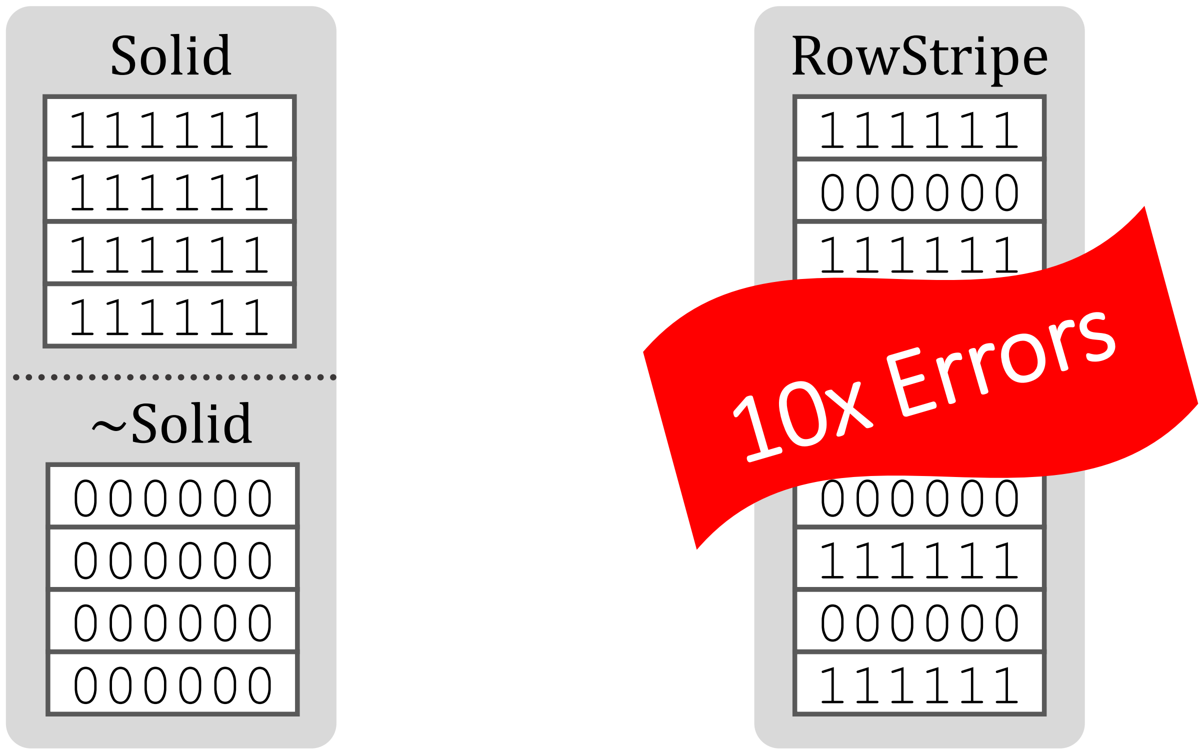
❷ Refresh Interval



Note: Using three modules with the most errors (only first bank)

More frequent refreshes → Fewer errors

❸ Data Pattern



Errors affected by data stored in other cells

6. Other Key Observations [ISCA'14]

- *Victim Cells ≠ Retention-Weak Cells*
  - Almost no overlap between them
- *Errors are repeatable*
  - Across ten iterations of testing, >70% of victim cells had errors in every iteration
- *As many as 4 errors per cache-line*
  - Simple ECC (e.g., SECDED) cannot prevent all errors
- *Cells affected by two aggressors on either side*
  - Double sided hammering

Major RowHammer Characteristics (2014)

- Yoongu Kim, Ross Daly, Jeremie Kim, Chris Fallin, Ji Hye Lee, Donghyuk Lee, Chris Wilkerson, Konrad Lai, and Onur Mutlu,
  "Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors"
  Proceedings of the 41st International Symposium on Computer Architecture (***ISCA***), Minneapolis, MN, June 2014.
  [Slides (pptx) (pdf)] [Lightning Session Slides (pptx) (pdf)] [Source Code and Data] [Lecture Video (1 hr 49 mins), 25 September 2020]
  One of the 7 papers of 2012-2017 selected as Top Picks in Hardware and Embedded Security for IEEE TCAD (link).

Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors

Yoongu Kim[1] Ross Daly* Jeremie Kim[1] Chris Fallin* Ji Hye Lee[1]
Donghyuk Lee[1] Chris Wilkerson[2] Konrad Lai Onur Mutlu[1]

[1]Carnegie Mellon University [2]Intel Labs

RowHammer is Getting Much Worse (2020)

- Jeremie S. Kim, Minesh Patel, A. Giray Yaglikci, Hasan Hassan, Roknoddin Azizi, Lois Orosa, and Onur Mutlu,
  "Revisiting RowHammer: An Experimental Analysis of Modern Devices and Mitigation Techniques"
  Proceedings of the 47th International Symposium on Computer Architecture (***ISCA***), Valencia, Spain, June 2020.
  [Slides (pptx) (pdf)]
  [Lightning Talk Slides (pptx) (pdf)]
  [Talk Video (20 minutes)]
  [Lightning Talk Video (3 minutes)]

Revisiting RowHammer: An Experimental Analysis of Modern DRAM Devices and Mitigation Techniques

Jeremie S. Kim[§†] Minesh Patel[§] A. Giray Yağlıkçı[§]
Hasan Hassan[§] Roknoddin Azizi[§] Lois Orosa[§] Onur Mutlu[§†]

[§]*ETH Zürich* [†]*Carnegie Mellon University*

New RowHammer Dimensions (2021)

- Lois Orosa, Abdullah Giray Yaglikci, Haocong Luo, Ataberk Olgun, Jisung Park, Hasan Hassan, Minesh Patel, Jeremie S. Kim, and Onur Mutlu,
  "A Deeper Look into RowHammer's Sensitivities: Experimental Analysis of Real DRAM Chips and Implications on Future Attacks and Defenses"
  Proceedings of the 54th International Symposium on Microarchitecture (***MICRO***), Virtual, October 2021.
  [Slides (pptx) (pdf)]
  [Short Talk Slides (pptx) (pdf)]
  [Lightning Talk Slides (pptx) (pdf)]
  [Talk Video (21 minutes)]
  [Lightning Talk Video (1.5 minutes)]
  [arXiv version]

A Deeper Look into RowHammer's Sensitivities: Experimental Analysis of Real DRAM Chips and Implications on Future Attacks and Defenses

Lois Orosa* ETH Zürich
A. Giray Yağlıkçı* ETH Zürich
Haocong Luo ETH Zürich
Ataberk Olgun ETH Zürich, TOBB ETÜ
Jisung Park ETH Zürich
Hasan Hassan ETH Zürich
Minesh Patel ETH Zürich
Jeremie S. Kim ETH Zürich
Onur Mutlu ETH Zürich

RowHammer vs. Wordline Voltage (2022)

- To appear in DSN 2022

Understanding the RowHammer Vulnerability Under Reduced Wordline Voltage: An Experimental Study Using Real DRAM Devices

A. Giray Yağlıkçı[1] Haocong Luo[1] Geraldo F. de Oliviera[1] Ataberk Olgun[1] Jisung Park[1]
Minesh Patel[1] Hasan Hassan[1] Jeremie S. Kim[1] Lois Orosa[1,2] Onur Mutlu[1]
[1] *ETH Zürich* [2] *Galicia Supercomputing Center (CESGA)*

RowHammer Solutions

Two Types of RowHammer Solutions

- Immediate
  - To protect the vulnerable DRAM chips in the field
  - Limited possibilities
- Longer-term
  - To protect future DRAM chips
  - Wider range of protection mechanisms
- Our ISCA 2014 paper proposes both types of solutions
  - Seven solutions in total
  - PARA proposed as best solution → already employed in the field

Some Potential Solutions (ISCA 2014)

- Make better DRAM chips — Cost
- Refresh frequently — Power, Performance
- Sophisticated ECC — Cost, Power
- Access counters — Cost, Power, Complexity

Apple's Security Patch for RowHammer

- https://support.apple.com/en-gb/HT204934

Available for: OS X Mountain Lion v10.8.5, OS X Mavericks v10.9.5

Impact: A malicious application may induce memory corruption to escalate privileges

Description: A disturbance error, also known as Rowhammer, exists with some DDR3 RAM that could have led to memory corruption. This issue was mitigated by increasing memory refresh rates.

CVE-ID

CVE-2015-3693 : Mark Seaborn and Thomas Dullien of Google, working from original research by Yoongu Kim et al (2014)

HP, Lenovo, and many other vendors released similar patches

Our Solution to RowHammer

- **PARA:** *Probabilistic Adjacent Row Activation*

- Key Idea
  - After closing a row, we activate (i.e., refresh) one of its neighbors with a low probability: $p = 0.005$

- Reliability Guarantee
  - When $p=0.005$, errors in one year: 9.4×10^{-14}
  - By adjusting the value of p, we can vary the strength of protection against errors

Advantages of PARA

- *PARA refreshes rows infrequently*
  - Low power
  - Low performance-overhead
    - Average slowdown: 0.20% (for 29 benchmarks)
    - Maximum slowdown: 0.75%
- *PARA is stateless*
  - Low cost
  - Low complexity

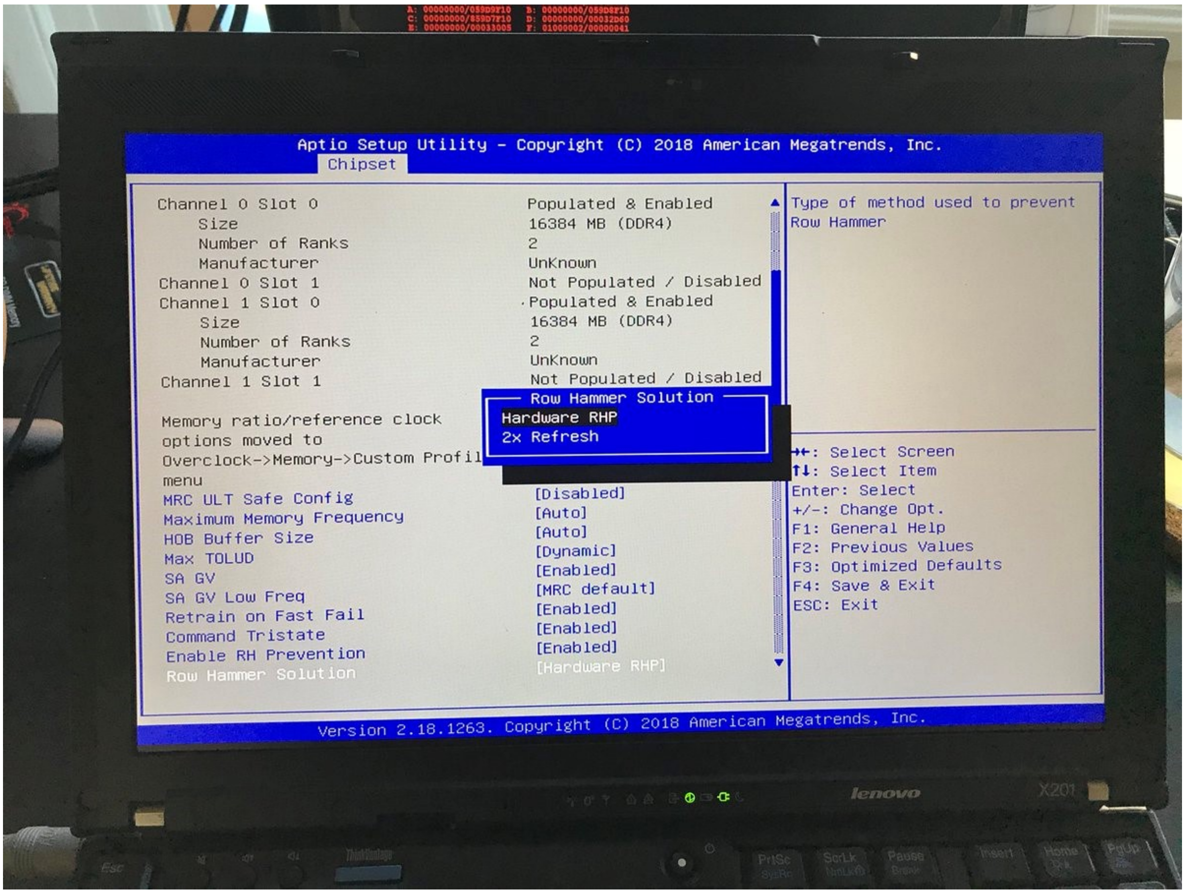
- *PARA is an effective and low-overhead solution to prevent disturbance errors*

Requirements for PARA

- If implemented in DRAM chip (done today)
  - Enough slack in timing and refresh parameters
  - Plenty of slack today:
    - Lee et al., "Adaptive-Latency DRAM: Optimizing DRAM Timing for the Common Case," HPCA 2015.
    - Chang et al., "Understanding Latency Variation in Modern DRAM Chips," SIGMETRICS 2016.
    - Lee et al., "Design-Induced Latency Variation in Modern DRAM Chips," SIGMETRICS 2017.
    - Chang et al., "Understanding Reduced-Voltage Operation in Modern DRAM Devices," SIGMETRICS 2017.
    - Ghose et al., "What Your DRAM Power Models Are Not Telling You: Lessons from a Detailed Experimental Study," SIGMETRICS 2018.
    - Kim et al., "Solar-DRAM: Reducing DRAM Access Latency by Exploiting the Variation in Local Bitlines," ICCD 2018.

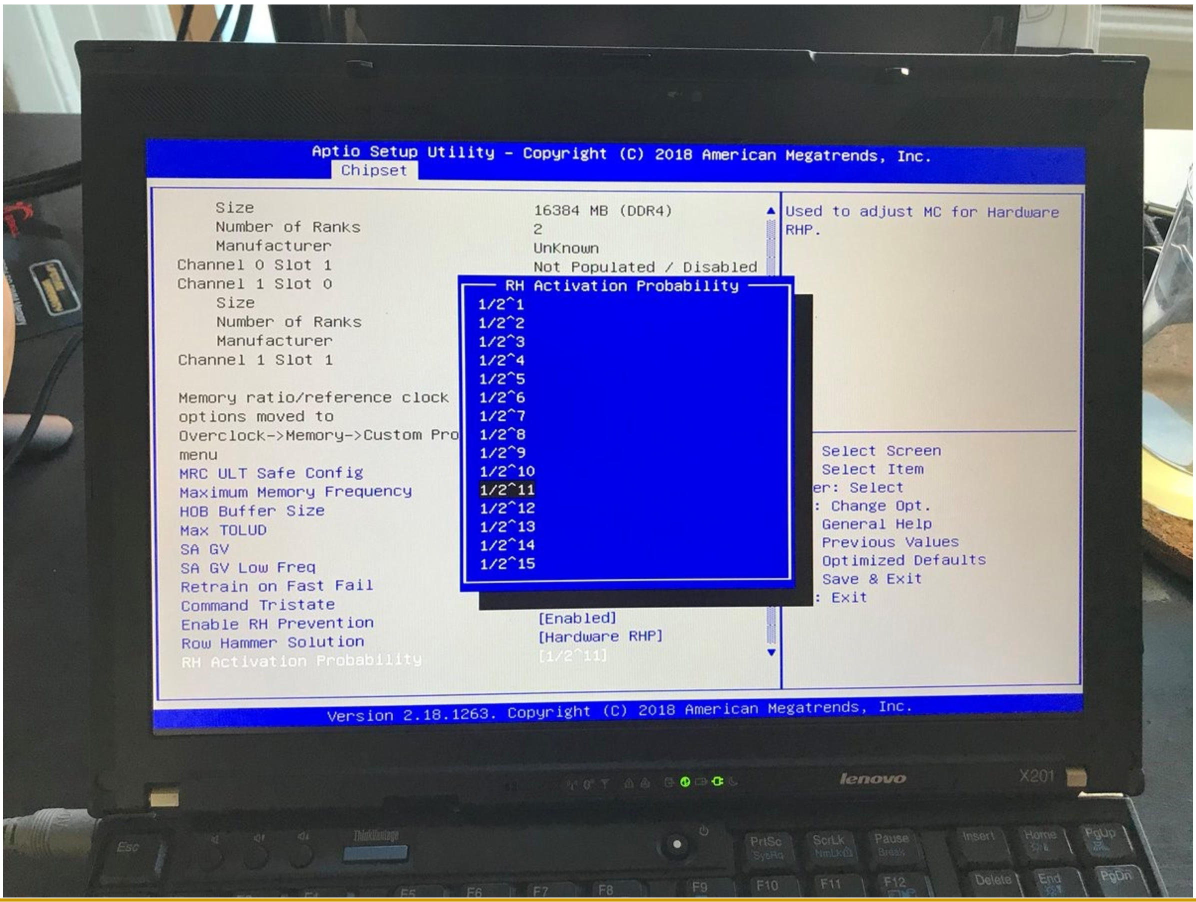
- If implemented in memory controller
  - Better coordination between memory controller and DRAM
  - Memory controller should know which rows are physically adjacent

Probabilistic Activation in Real Life (I)



https://twitter.com/isislovecruft/status/1021939922754723841

Probabilistic Activation in Real Life (II)



https://twitter.com/isislovecruft/status/1021939922754723841

Seven RowHammer Solutions Proposed

- Yoongu Kim, Ross Daly, Jeremie Kim, Chris Fallin, Ji Hye Lee, Donghyuk Lee, Chris Wilkerson, Konrad Lai, and Onur Mutlu,
  "Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors"
  Proceedings of the 41st International Symposium on Computer Architecture (***ISCA***), Minneapolis, MN, June 2014.
  [Slides (pptx) (pdf)] [Lightning Session Slides (pptx) (pdf)] [Source Code and Data] [Lecture Video (1 hr 49 mins), 25 September 2020]
  One of the 7 papers of 2012-2017 selected as Top Picks in Hardware and Embedded Security for IEEE TCAD (link).

Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors

Yoongu Kim[1] Ross Daly* Jeremie Kim[1] Chris Fallin* Ji Hye Lee[1]
Donghyuk Lee[1] Chris Wilkerson[2] Konrad Lai Onur Mutlu[1]

[1]Carnegie Mellon University [2]Intel Labs

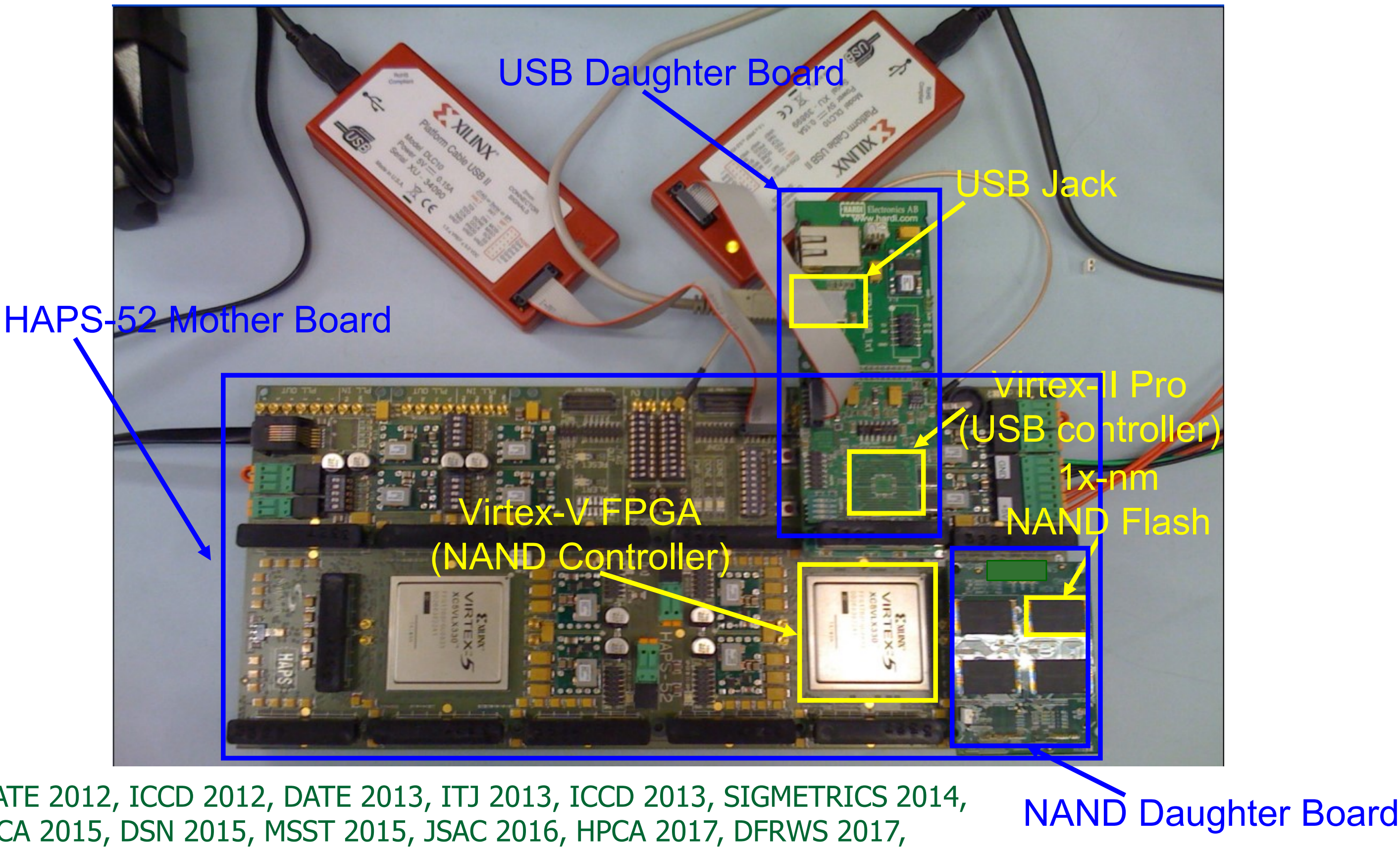
A Takeaway

Main Memory Needs

Intelligent Controllers

for Security, Safety, Reliability, Scaling

Aside: Intelligent Controller for NAND Flash



[DATE 2012, ICCD 2012, DATE 2013, ITJ 2013, ICCD 2013, SIGMETRICS 2014, HPCA 2015, DSN 2015, MSST 2015, JSAC 2016, HPCA 2017, DFRWS 2017, PIEEE 2017, HPCA 2018, SIGMETRICS 2018]

Cai+, "Error Characterization, Mitigation, and Recovery in Flash Memory Based Solid State Drives," Proc. IEEE 2017.

Intelligent Flash Controllers [PIEEE'17]

INVITED PAPER

Proceedings of the IEEE, Sept. 2017

Error Characterization, Mitigation, and Recovery in Flash-Memory-Based Solid-State Drives

This paper reviews the most recent advances in solid-state drive (SSD) error characterization, mitigation, and data recovery techniques to improve both SSD's reliability and lifetime.

By Yu Cai, Saugata Ghose, Erich F. Haratsch, Yixin Luo, and Onur Mutlu

https://arxiv.org/pdf/1706.08642

Detailed Lectures on RowHammer

- Computer Architecture, Fall 2021, Lecture 5
  - RowHammer (ETH Zürich, Fall 2021)
  - https://www.youtube.com/watch?v=7wVKnPj3NVw&list=PL5Q2soXY2Zi-Mnk1PxjEIG32HAGILkTOF&index=5
- Computer Architecture, Fall 2021, Lecture 6
  - RowHammer and Secure & Reliable Memory (ETH Zürich, Fall 2021)
  - https://www.youtube.com/watch?v=HNd4skQrt6I&list=PL5Q2soXY2Zi-Mnk1PxjEIG32HAGILkTOF&index=6

https://www.youtube.com/onurmutlulectures

First RowHammer Analysis

- Yoongu Kim, Ross Daly, Jeremie Kim, Chris Fallin, Ji Hye Lee, Donghyuk Lee, Chris Wilkerson, Konrad Lai, and Onur Mutlu,
  "Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors"
  Proceedings of the 41st International Symposium on Computer Architecture (***ISCA***), Minneapolis, MN, June 2014.
  [Slides (pptx) (pdf)] [Lightning Session Slides (pptx) (pdf)] [Source Code and Data] [Lecture Video (1 hr 49 mins), 25 September 2020]
  One of the 7 papers of 2012-2017 selected as Top Picks in Hardware and Embedded Security for IEEE TCAD (link).

Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors

Yoongu Kim[1] Ross Daly* Jeremie Kim[1] Chris Fallin* Ji Hye Lee[1]
Donghyuk Lee[1] Chris Wilkerson[2] Konrad Lai Onur Mutlu[1]

[1]Carnegie Mellon University [2]Intel Labs

Retrospective on RowHammer & Future

- Onur Mutlu,
  "The RowHammer Problem and Other Issues We May Face as Memory Becomes Denser"
  Invited Paper in Proceedings of the Design, Automation, and Test in Europe Conference (***DATE***), Lausanne, Switzerland, March 2017.
  [Slides (pptx) (pdf)]

The RowHammer Problem and Other Issues We May Face as Memory Becomes Denser

Onur Mutlu
ETH Zürich
onur.mutlu@inf.ethz.ch
https://people.inf.ethz.ch/omutlu

A More Recent RowHammer Retrospective

- Onur Mutlu and Jeremie Kim,
  "RowHammer: A Retrospective"
  IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (***TCAD***) *Special Issue on Top Picks in Hardware and Embedded Security*, 2019.
  [Preliminary arXiv version]
  [Slides from COSADE 2019 (pptx)]
  [Slides from VLSI-SOC 2020 (pptx) (pdf)]
  [Talk Video (1 hr 15 minutes, with Q&A)]

RowHammer: A Retrospective

Onur Mutlu[§‡] Jeremie S. Kim[‡§]
[§]ETH Zürich [‡]Carnegie Mellon University

RowHammer in 2020-2022

Revisiting RowHammer

RowHammer is Getting Much Worse

- Jeremie S. Kim, Minesh Patel, A. Giray Yaglikci, Hasan Hassan, Roknoddin Azizi, Lois Orosa, and Onur Mutlu,
  "Revisiting RowHammer: An Experimental Analysis of Modern Devices and Mitigation Techniques"
  Proceedings of the 47th International Symposium on Computer Architecture (***ISCA***), Valencia, Spain, June 2020.
  [Slides (pptx) (pdf)]
  [Lightning Talk Slides (pptx) (pdf)]
  [Talk Video (20 minutes)]
  [Lightning Talk Video (3 minutes)]

Revisiting RowHammer: An Experimental Analysis of Modern DRAM Devices and Mitigation Techniques

Jeremie S. Kim[§†] Minesh Patel[§] A. Giray Yağlıkçı[§]
Hasan Hassan[§] Roknoddin Azizi[§] Lois Orosa[§] Onur Mutlu[§†]

[§]*ETH Zürich* [†]*Carnegie Mellon University*

Key Takeaways from 1580 Chips

- **Newer DRAM chips are much more vulnerable to RowHammer (more bit flips, happening earlier)**

- There are new chips whose weakest cells fail after **only 4800 hammers**

- Chips of newer DRAM technology nodes can exhibit RowHammer bit flips 1) in **more rows** and 2) **farther away** from the victim row.

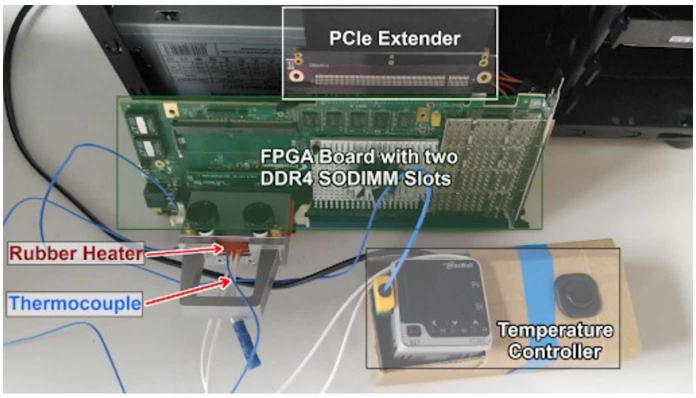
- **Existing mitigation mechanisms are NOT effective at future technology nodes**

DRAM Testing Infrastructures

Three separate testing infrastructures

1. **DDR3:** FPGA-based SoftMC [Hassan+, HPCA'17] (Xilinx ML605)
2. **DDR4:** FPGA-based SoftMC [Hassan+, HPCA'17] (Xilinx Virtex UltraScale 95)
3. **LPDDR4:** In-house testing hardware for LPDDR4 chips

All provide fine-grained control over DRAM commands, timing parameters and temperature



DDR4 DRAM testing infrastructure

1580 DRAM Chips Tested

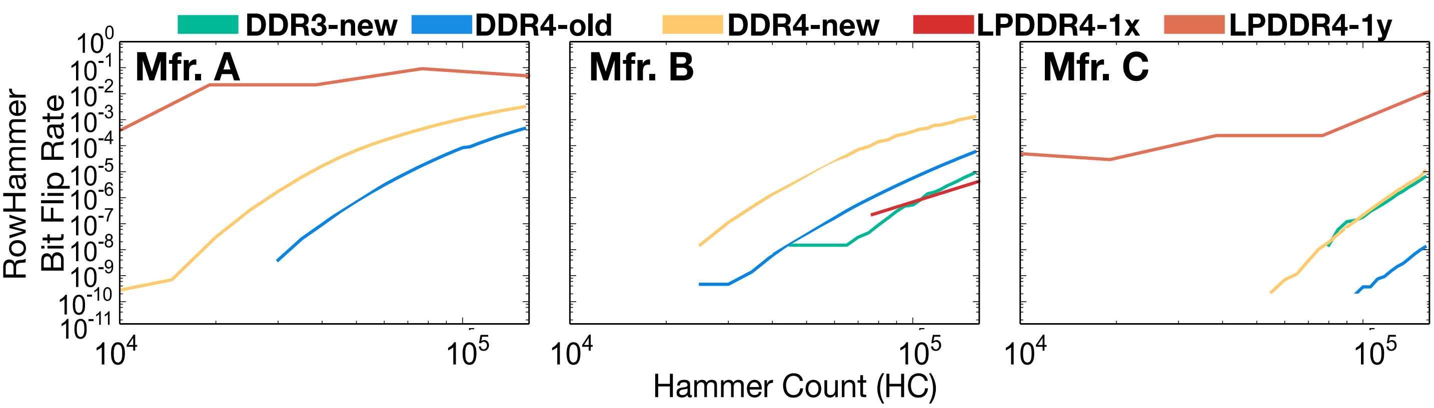
	Number of Chips (Modules) Tested			
				Total
	56 (10)	88 (11)	28 (7)	**172 (28)**
	80 (10)	52 (9)	104 (13)	**236 (32)**
	112 (16)	24 (3)	128 (18)	**264 (37)**
	264 (43)	16 (2)	108 (28)	**388 (73)**
	12 (3)	180 (45)	N/A	**192 (48)**
	184 (46)	N/A	144 (36)	**328 (82)**

1580 total DRAM chips tested from **300** DRAM modules

- **Three** major DRAM manufacturers {A, B, C}
- **Three** DRAM *types* or *standards* {DDR3, DDR4, LPDDR4}
  - LPDDR4 chips we test implement on-die ECC
- **Two** technology nodes per DRAM type {old/new, 1x/1y}
  - Categorized based on manufacturing date, datasheet publication date, purchase date, and characterization results

Type-node: configuration describing a chip's type and technology node generation: **DDR3-old/new, DDR4-old/new, LPDDR4-1x/1y**

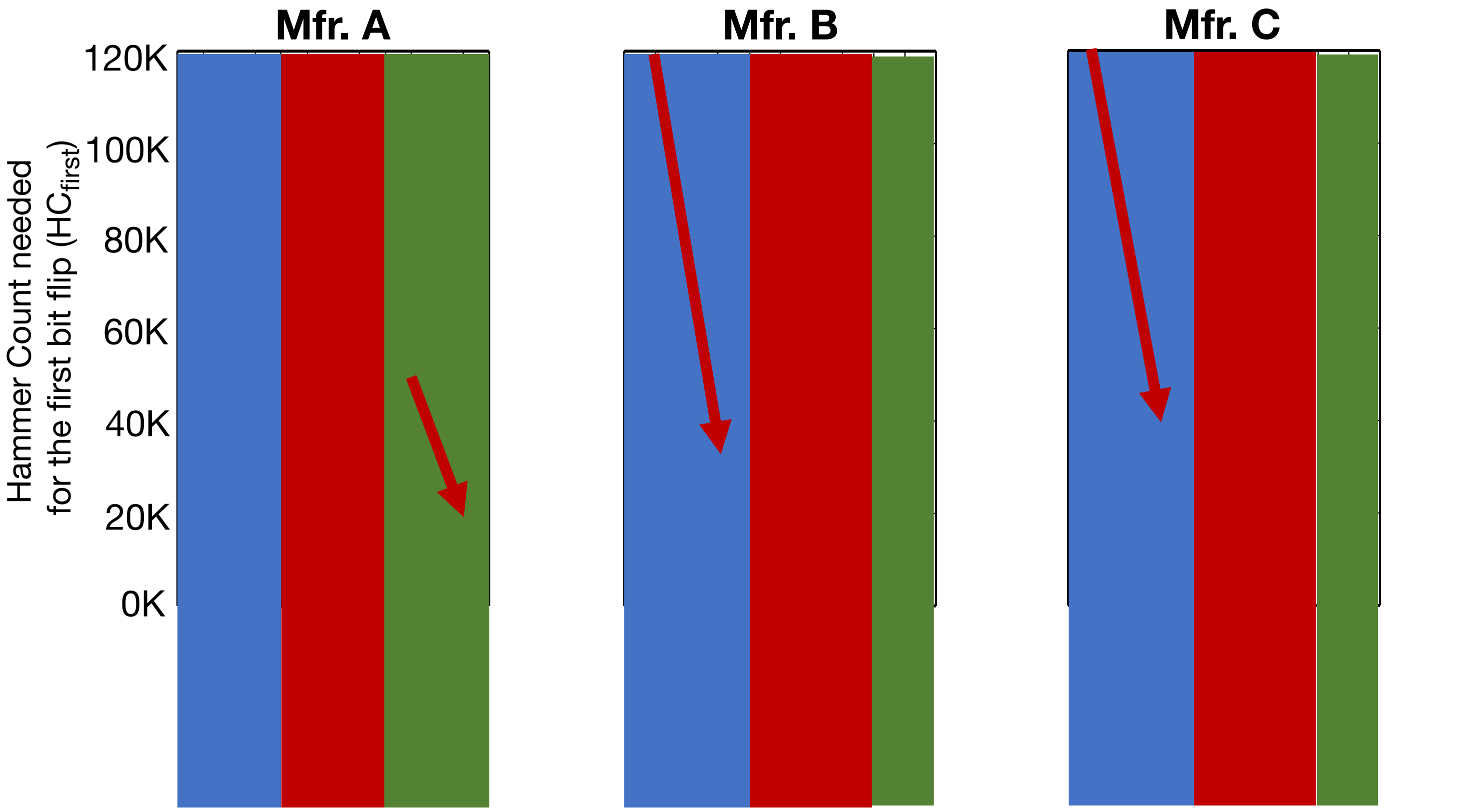
3. Hammer Count (HC) Effects



RowHammer bit flip rates **increase**
when going **from old to new** DDR4 technology node generations

RowHammer bit flip rates (i.e., RowHammer vulnerability) increase with technology node generation

5. First RowHammer Bit Flips per Chip



Newer chips from each DRAM manufacturer are more vulnerable to RowHammer

5. First RowHammer Bit Flips per Chip

In a DRAM type, HC_{first} reduces significantly from old to new chips, i.e., DDR3: 69.2k to 22.4k, DDR4: 17.5k to 10k, LPDDR4: 16.8k to 4.8k

There are chips whose weakest cells fail after only 4800 hammers

RowHammer is Getting Much Worse

- Jeremie S. Kim, Minesh Patel, A. Giray Yaglikci, Hasan Hassan, Roknoddin Azizi, Lois Orosa, and Onur Mutlu,
  "Revisiting RowHammer: An Experimental Analysis of Modern Devices and Mitigation Techniques"
  *Proceedings of the 47th International Symposium on Computer Architecture (**ISCA**)*, Valencia, Spain, June 2020.
  [Slides (pptx) (pdf)]
  [Lightning Talk Slides (pptx) (pdf)]
  [Talk Video (20 minutes)]
  [Lightning Talk Video (3 minutes)]

Revisiting RowHammer: An Experimental Analysis of Modern DRAM Devices and Mitigation Techniques

Jeremie S. Kim[§†] Minesh Patel[§] A. Giray Yağlıkçı[§]
Hasan Hassan[§] Roknoddin Azizi[§] Lois Orosa[§] Onur Mutlu[§†]

[§]*ETH Zürich* [†]*Carnegie Mellon University*

Detailed Lecture on Revisiting RowHammer

- Computer Architecture, Fall 2020, Lecture 5b
  - RowHammer in 2020: Revisiting RowHammer (ETH Zürich, Fall 2020)
  - https://www.youtube.com/watch?v=gR7XR-Eepcg&list=PL5Q2soXY2Zi9xidyIgBxUz7xRPS-wisBN&index=10

https://www.youtube.com/onurmutlulectures

TRRespass

Industry-Adopted Solutions Do Not Work

- Pietro Frigo, Emanuele Vannacci, Hasan Hassan, Victor van der Veen, Onur Mutlu, Cristiano Giuffrida, Herbert Bos, and Kaveh Razavi,
  "TRRespass: Exploiting the Many Sides of Target Row Refresh"
  *Proceedings of the 41st IEEE Symposium on Security and Privacy (**S&P**)*, San Francisco, CA, USA, May 2020.
  [Slides (pptx) (pdf)]
  [Lecture Slides (pptx) (pdf)]
  [Talk Video (17 minutes)]
  [Lecture Video (59 minutes)]
  [Source Code]
  [Web Article]
  Best paper award.
  Pwnie Award 2020 for Most Innovative Research. Pwnie Awards 2020

TRRespass: Exploiting the Many Sides of Target Row Refresh

Pietro Frigo[*†] Emanuele Vannacci[*†] Hasan Hassan[§] Victor van der Veen[¶]
Onur Mutlu[§] Cristiano Giuffrida[*] Herbert Bos[*] Kaveh Razavi[*]

[*]Vrije Universiteit Amsterdam [§]ETH Zürich [¶]Qualcomm Technologies Inc.

TRRespass

- First work to show that TRR-protected DRAM chips are vulnerable to RowHammer in the field
  - Mitigations advertised as secure are not secure

- Introduces the Many-sided RowHammer attack
  - Idea: Hammer many rows to bypass TRR mitigations (e.g., by overflowing proprietary TRR tables that detect aggressor rows)

- (Partially) reverse-engineers the TRR and pTRR mitigation mechanisms implemented in DRAM chips and memory controllers

- Provides an automatic tool that can effectively create many-sided RowHammer attacks in DDR4 and LPDDR4(X) chips

Example Many-Sided Hammering Patterns

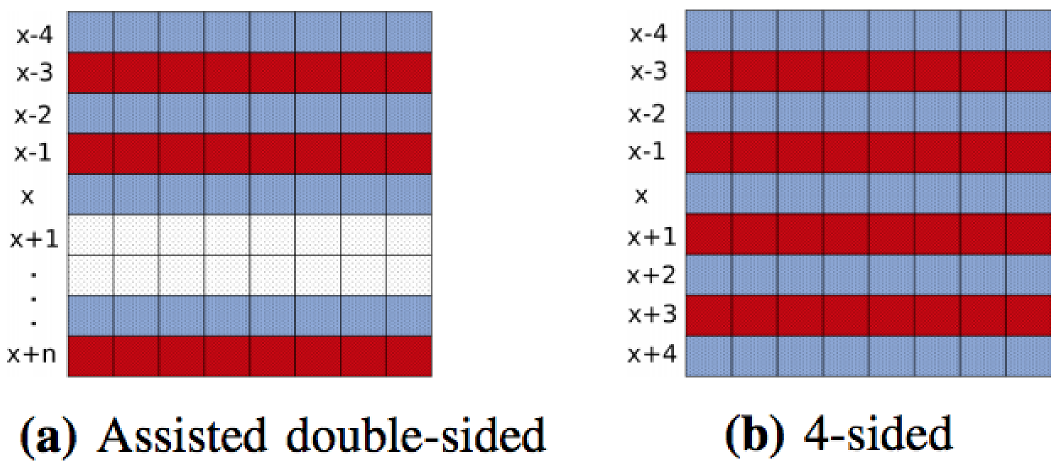


Fig. 12: Hammering patterns discovered by *TRRespass*. Aggressor rows are in red (■) and victim rows are in blue (■).

BitFlips vs. Number of Aggressor Rows

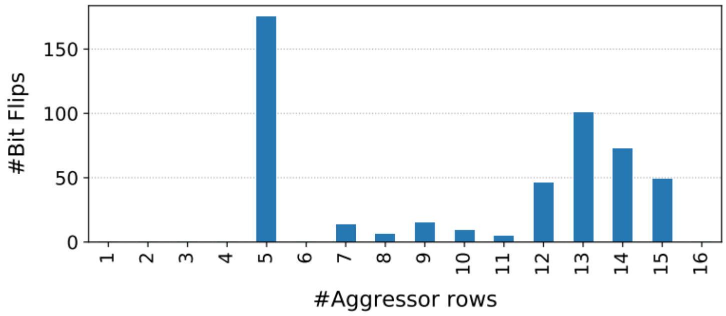


Fig. 10: Bit flips vs. number of aggressor rows. Module $\mathcal{C}_{12}$: Number of bit flips in bank 0 as we vary the number of aggressor rows. Using SoftMC, we refresh DRAM with standard `tREFI` and run the tests until each aggressor rows is hammered 500K times.

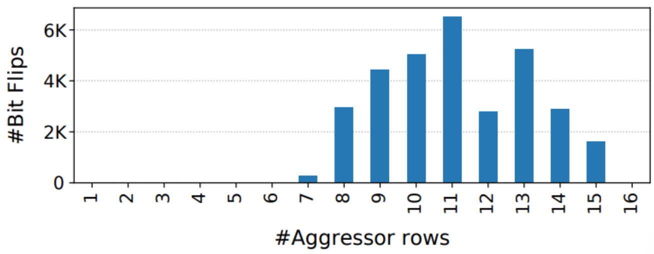


Fig. 11: Bit flips vs. number of aggressor rows. Module $\mathcal{A}_{15}$: Number of bit flips in bank 0 as we vary the number of aggressor rows. Using SoftMC, we refresh DRAM with standard `tREFI` and run the tests until each aggressor rows is hammered 500K times.

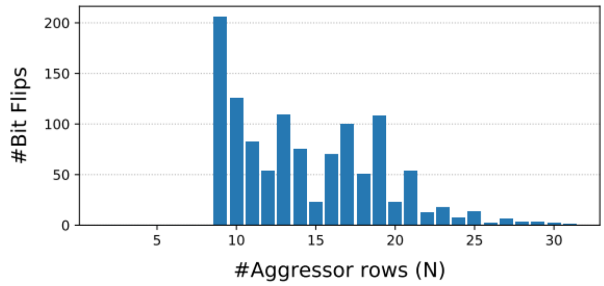


Fig. 13: Bit flips vs. number of aggressor rows. Module $\mathcal{A}_{10}$: Number of bit flips triggered with *N-sided* RowHammer for varying number of N on Intel Core i7-7700K. Each aggressor row is one row away from the closest aggressor row (i.e., `VAVAVA...` configuration) and aggressor rows are hammered in a round-robin fashion.

TRRespass Vulnerable DRAM Modules

TABLE II: *TRRespass* results. We report the number of patterns found and bit flips detected for the 42 DRAM modules in our set.

Module	*Date (yy-ww)*	*Freq. (MHz)*	*Size (GB)*	*Organization*			*MAC*	*Found Patterns*	*Best Pattern*	*Corruptions*			*Double Refresh*
				Ranks	*Banks*	*Pins*				*Total*	$1 \to 0$	$0 \to 1$	
$\mathcal{A}_{0,1,2,3}$	16-37	2132	4	1	16	×8	UL	—	—	—	—	—	—
$\mathcal{A}_{4}$	16-51	2132	4	1	16	×8	UL	4	9-sided	7956	4008	3948	—
$\mathcal{A}_{5}$	18-51	2400	4	1	8	×16	UL	—	—	—	—	—	—
$\mathcal{A}_{6,7}$	18-15	2666	4	1	8	×16	UL	—	—	—	—	—	—
$\mathcal{A}_{8}$	17-09	2400	8	1	16	×8	UL	33	19-sided	20808	10289	10519	—
$\mathcal{A}_{9}$	17-31	2400	8	1	16	×8	UL	33	19-sided	24854	12580	12274	—
$\mathcal{A}_{10}$	19-02	2400	16	2	16	×8	UL	488	10-sided	11342	1809	11533	✓
$\mathcal{A}_{11}$	19-02	2400	16	2	16	×8	UL	523	10-sided	12830	1682	11148	✓
$\mathcal{A}_{12,13}$	18-50	2666	8	1	16	×8	UL	—	—	—	—	—	—
$\mathcal{A}_{14}$	19-08†	3200	16	2	16	×8	UL	120	14-sided	32723	16490	16233	—
$\mathcal{A}_{15}$‡	17-08	2132	4	1	16	×8	UL	2	9-sided	22397	12351	10046	—
$\mathcal{B}_{0}$	18-11	2666	16	2	16	×8	UL	2	3-sided	17	10	7	—
$\mathcal{B}_{1}$	18-11	2666	16	2	16	×8	UL	2	3-sided	22	16	6	—
$\mathcal{B}_{2}$	18-49	3000	16	2	16	×8	UL	2	3-sided	5	2	3	—
$\mathcal{B}_{3}$	19-08†	3000	8	1	16	×8	UL	—	—	—	—	—	—
$\mathcal{B}_{4,5}$	19-08†	2666	8	2	16	×8	UL	—	—	—	—	—	—
$\mathcal{B}_{6,7}$	19-08†	2400	4	1	16	×8	UL	—	—	—	—	—	—
$\mathcal{B}_{8}$◇	19-08†	2400	8	1	16	×8	UL	—	—	—	—	—	—
$\mathcal{B}_{9}$◇	19-08†	2400	8	1	16	×8	UL	2	3-sided	12	—	12	✓
$\mathcal{B}_{10,11}$	16-13†	2132	8	2	16	×8	UL	—	—	—	—	—	—
$\mathcal{C}_{0,1}$	18-46	2666	16	2	16	×8	UL	—	—	—	—	—	—
$\mathcal{C}_{2,3}$	19-08†	2800	4	1	16	×8	UL	—	—	—	—	—	—
$\mathcal{C}_{4,5}$	19-08†	3000	8	1	16	×8	UL	—	—	—	—	—	—
$\mathcal{C}_{6,7}$	19-08†	3000	16	2	16	×8	UL	—	—	—	—	—	—
$\mathcal{C}_{8}$	19-08†	3200	16	2	16	×8	UL	—	—	—	—	—	—
$\mathcal{C}_{9}$	18-47	2666	16	2	16	×8	UL	—	—	—	—	—	—
$\mathcal{C}_{10,11}$	19-04	2933	8	1	16	×8	UL	—	—	—	—	—	—
$\mathcal{C}_{12}$‡	15-01†	2132	4	1	16	×8	UT	25	10-sided	190037	63904	126133	✓
$\mathcal{C}_{13}$‡	18-49	2132	4	1	16	×8	UT	3	9-sided	694	239	455	—

† The module does not report manufacturing date. Therefore, we report purchase date as an approximation.
‡ Analyzed using the FPGA-based SoftMC.
◇ The system runs with double refresh frequency in standard conditions. We configured the refresh interval to be 64 ms in the BIOS settings.

UL = Unlimited
UT = Untested

TRRespass Vulnerable Mobile Phones

TABLE III: LPDDR4(X) results. Mobile phones tested against *TRRespass* on ARMv8 sorted by production date. We found bit flip inducing RowHammer patterns on 5 out of 13mobile phones.

Mobile Phone	*Year*	*SoC*	*Memory (GB)*	*Found Patterns*
Google Pixel	2016	MSM8996	4†	✓
Google Pixel 2	2017	MSM8998	4	—
Samsung G960F/DS	2018	Exynos 9810	4	—
Huawei P20 DS	2018	Kirin 970	4	—
Sony XZ3	2018	SDM845	4	—
HTC U12+	2018	SDM845	6	—
LG G7 ThinQ	2018	SDM845	4†	✓
Google Pixel 3	2018	SDM845	4	✓
Google Pixel 4	2019	SM8150	6	—
OnePlus 7	2019	SM8150	8	✓
Samsung G970F/DS	2019	Exynos 9820	6	✓
Huawei P30 DS	2019	Kirin 980	6	—
Xiaomi Redmi Note 8 Pro	2019	Helio G90T	6	—

† LPDDR4 (not LPDDR4X)

TRRespass Based RowHammer Attack

TABLE IV: Time to exploit. Time to find the first exploitable template on two sample modules from each DRAM vendor.

Module	τ *(ms)*	*PTE* [81]	*RSA-2048* [79]	*sudo* [27]
$\mathcal{A}_{14}$	188.7	4.9s	6m 27s	—
$\mathcal{A}_{4}$	180.8	38.8s	39m 28s	—
$\mathcal{B}_{1}$	360.7	—	—	—
$\mathcal{B}_{2}$	331.2	—	—	—
$\mathcal{C}_{12}$	300.0	2.3s	74.6s	54m16s
$\mathcal{C}_{13}$	180.9	3h 15m	—	—

τ: Time to template a single row: time to fill the victim and aggressor rows + hammer time + time to scan the row.

TRRespass Key Results

- 13 out of 42 tested DDR4 DRAM modules are vulnerable
  - From all 3 major manufacturers
  - 3-, 9-, 10-, 14-, 19-sided hammer attacks needed
- 5 out of 13 mobile phones tested vulnerable
  - From 4 major manufacturers
  - With LPDDR4(X) DRAM chips
- These results are scratching the surface
  - TRRespass tool is not exhaustive
  - There is a lot of room for uncovering more vulnerable chips and phones

TRRespass Key Takeaways

RowHammer is still an open problem

Security by obscurity is likely not a good solution

Detailed Lecture on TRRespass

- Computer Architecture, Fall 2020, Lecture 5a
  - RowHammer in 2020: TRRespass (ETH Zürich, Fall 2020)
  - https://www.youtube.com/watch?v=pwRw7QqK_qA&list=PL5Q2soXY2Zi9xidyIgBxUz7xRPS-wisBN&index=9

https://www.youtube.com/onurmutlulectures

Industry-Adopted Solutions Do Not Work

- Pietro Frigo, Emanuele Vannacci, Hasan Hassan, Victor van der Veen, Onur Mutlu, Cristiano Giuffrida, Herbert Bos, and Kaveh Razavi,
  "TRRespass: Exploiting the Many Sides of Target Row Refresh"
  Proceedings of the 41st IEEE Symposium on Security and Privacy (***S&P***), San Francisco, CA, USA, May 2020.
  [Slides (pptx) (pdf)]
  [Lecture Slides (pptx) (pdf)]
  [Talk Video (17 minutes)]
  [Lecture Video (59 minutes)]
  [Source Code]
  [Web Article]
  Best paper award.
  Pwnie Award 2020 for Most Innovative Research. Pwnie Awards 2020

TRRespass: Exploiting the Many Sides of Target Row Refresh

Pietro Frigo*† Emanuele Vannacci*† Hasan Hassan§ Victor van der Veen¶
Onur Mutlu§ Cristiano Giuffrida* Herbert Bos* Kaveh Razavi*

*Vrije Universiteit Amsterdam §ETH Zürich ¶Qualcomm Technologies Inc.

How to Guarantee That a Chip is RowHammer-Free?

Hard to Guarantee RowHammer-Free Chips

- Lucian Cojocar, Jeremie Kim, Minesh Patel, Lillian Tsai, Stefan Saroiu, Alec Wolman, and Onur Mutlu,
  "Are We Susceptible to Rowhammer? An End-to-End Methodology for Cloud Providers"
  Proceedings of the 41st IEEE Symposium on Security and Privacy (***S&P***), San Francisco, CA, USA, May 2020.
  [Slides (pptx) (pdf)]
  [Talk Video (17 minutes)]

Are We Susceptible to Rowhammer? An End-to-End Methodology for Cloud Providers

Lucian Cojocar, Jeremie Kim[§†], Minesh Patel[§], Lillian Tsai[‡],
Stefan Saroiu, Alec Wolman, and Onur Mutlu[§†]
Microsoft Research, [§]ETH Zürich, [†]CMU, [‡]MIT

Uncovering TRR Almost Completely

Industry-Adopted Solutions Are Very Poor

- Hasan Hassan, Yahya Can Tugrul, Jeremie S. Kim, Victor van der Veen, Kaveh Razavi, and Onur Mutlu,
  "Uncovering In-DRAM RowHammer Protection Mechanisms: A New Methodology, Custom RowHammer Patterns, and Implications"
  Proceedings of the 54th International Symposium on Microarchitecture (***MICRO***), Virtual, October 2021.
  [Slides (pptx) (pdf)]
  [Short Talk Slides (pptx) (pdf)]
  [Lightning Talk Slides (pptx) (pdf)]
  [Talk Video (25 minutes)]
  [Lightning Talk Video (100 seconds)]
  [arXiv version]

Uncovering In-DRAM RowHammer Protection Mechanisms: A New Methodology, Custom RowHammer Patterns, and Implications

Hasan Hassan[†] Yahya Can Tuğrul[†‡] Jeremie S. Kim[†] Victor van der Veen[σ]
Kaveh Razavi[†] Onur Mutlu[†]
[†]*ETH Zürich* [‡]*TOBB University of Economics & Technology* [σ]*Qualcomm Technologies Inc.*

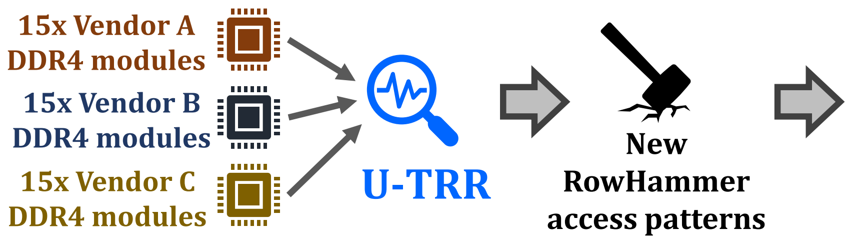
U-TRR Summary & Key Results

Target Row Refresh (TRR):
a set of obscure, undocumented, and proprietary RowHammer mitigation techniques

We cannot easily study the *security properties* of TRR

Is TRR fully secure? How can we validate its security guarantees?

U-TRR | A new methodology that leverages *data retention failures* to uncover the inner workings of TRR and study its security



All 45 modules we test are **vulnerable**

99.9% of rows in a DRAM bank experience **at least one RowHammer bit flip**

Up to **7** RowHammer **bit flips** in an 8-byte dataword, **making ECC ineffective**

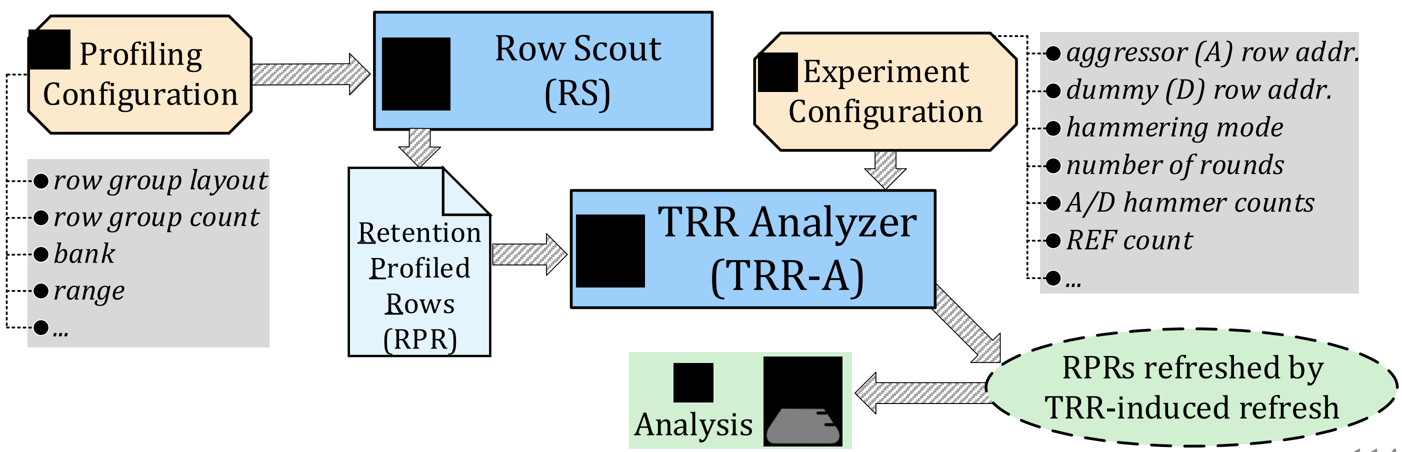
TRR **does not provide security** against RowHammer

U-TRR can facilitate the development of **new RowHammer attacks** and **more secure RowHammer protection** mechanisms

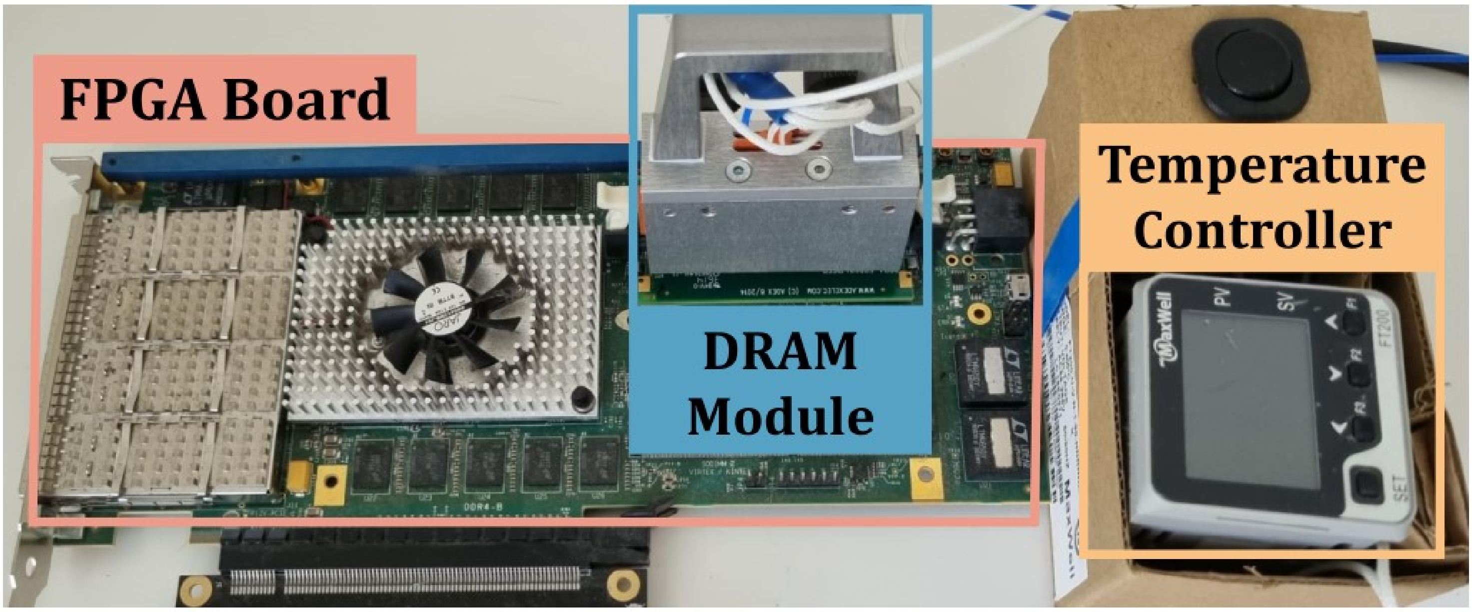
Overview of U-TRR

U-TRR: A new methodology to *uncover* the inner workings of TRR

Key idea: Use data retention failures as a side channel to detect when a row is refreshed by TRR

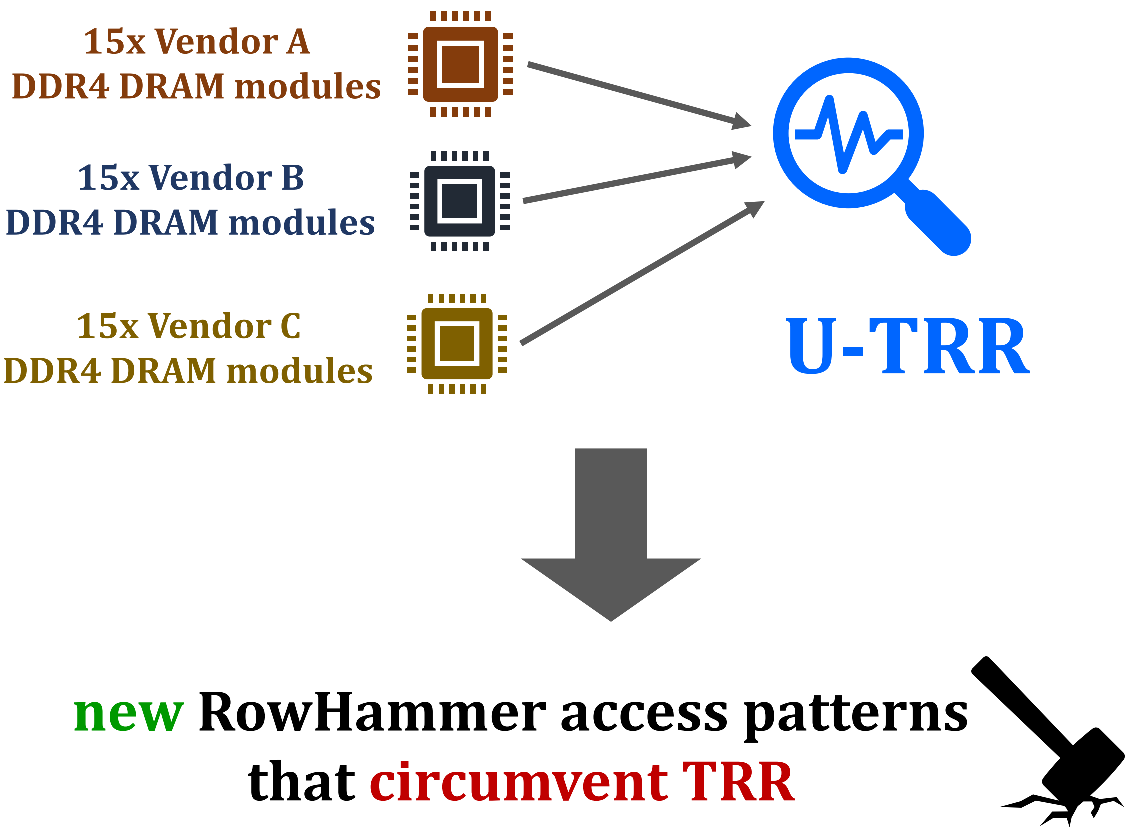


Analyzing TRR-Protected DDR4 Chips



** SoftMC [Hassan+, HPCA'17] enhanced for DDR4*

U-TRR Analysis Summary



Key Takeaways

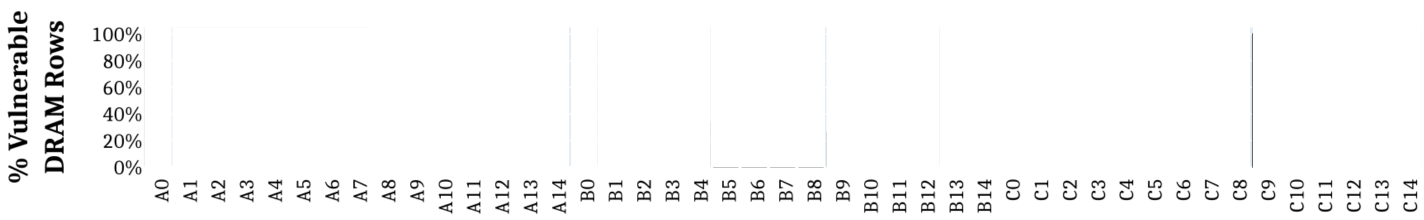
All 45 modules we test are vulnerable

99.9% of rows in a DRAM bank
experience **at least one RowHammer bit flip**

ECC is ineffective: up to 7 RowHammer bit flips
in an 8-byte dataword

Module	Date (yy-ww)	Chip Density (Gbit)	Organization			HC_{first}†	Our Key TRR Observations and Results							
			Ranks	Banks	Pins		Version	Aggressor Detection	Aggressor Capacity	Per-Bank TRR	TRR-to-REF Ratio	Neighbors Refreshed	% Vulnerable DRAM Rows†	Max. Bit Flips per Row per Hammer†
A0	19-50	8	1	16	8	$16K$	A_{TRR1}	Counter-based	16	✓	1/9	4	73.3%	1.16
A1-5	19-36	8	1	8	16	$13K$-$15K$	A_{TRR1}	Counter-based	16	✓	1/9	4	99.2% - 99.4%	2.32 - 4.73
A6-7	19-45	8	1	8	16	$13K$-$15K$	A_{TRR1}	Counter-based	16	✓	1/9	4	99.3% - 99.4%	2.12 - 3.86
A8-9	20-07	8	1	16	8	$12K$-$14K$	A_{TRR1}	Counter-based	16	✓	1/9	4	74.6% - 75.0%	1.96 - 2.96
A10-12	19-51	8	1	16	8	$12K$-$13K$	A_{TRR1}	Counter-based	16	✓	1/9	4	74.6% - 75.0%	1.48 - 2.86
A13-14	20-31	8	1	8	16	$11K$-$14K$	A_{TRR2}	Counter-based	16	✓	1/9	2	94.3% - 98.6%	1.53 - 2.78
B0	18-22	4	1	16	8	$44K$	B_{TRR1}	Sampling-based	1	✗	1/4	2	99.9%	2.13
B1-4	20-17	4	1	16	8	$159K$-$192K$	B_{TRR1}	Sampling-based	1	✗	1/4	2	23.3% - 51.2%	0.06 - 0.11
B5-6	16-48	4	1	16	8	$44K$-$50K$	B_{TRR1}	Sampling-based	1	✗	1/4	2	99.9%	1.85 - 2.03
B7	19-06	8	2	16	8	$20K$	B_{TRR1}	Sampling-based	1	✗	1/4	2	99.9%	31.14
B8	18-03	4	1	16	8	$43K$	B_{TRR1}	Sampling-based	1	✗	1/4	2	99.9%	2.57
B9-12	19-48	8	1	16	8	$42K$-$65K$	B_{TRR2}	Sampling-based	1	✗	1/9	2	36.3% - 38.9%	16.83 - 24.26
B13-14	20-08	4	1	16	8	$11K$-$14K$	B_{TRR3}	Sampling-based	1	✓	1/2	4	99.9%	16.20 - 18.12
C0-3	16-48	4	1	16	x8	$137K$-$194K$	C_{TRR1}	Mix	Unknown	✓	1/17	2	1.0% - 23.2%	0.05 - 0.15
C4-6	17-12	8	1	16	x8	$130K$-$150K$	C_{TRR1}	Mix	Unknown	✓	1/17	2	7.8% - 12.0%	0.06 - 0.08
C7-8	20-31	8	1	8	x16	$40K$-$44K$	C_{TRR1}	Mix	Unknown	✓	1/17	2	39.8% - 41.8%	9.66 - 14.56
C9-11	20-31	8	1	8	x16	$42K$-$53K$	C_{TRR2}	Mix	Unknown	✓	1/9	2	99.7%	9.30 - 32.04
C12-14	20-46	16	1	8	x16	$6K$-$7K$	C_{TRR3}	Mix	Unknown	✓	1/8	2	99.9%	4.91 - 12.64

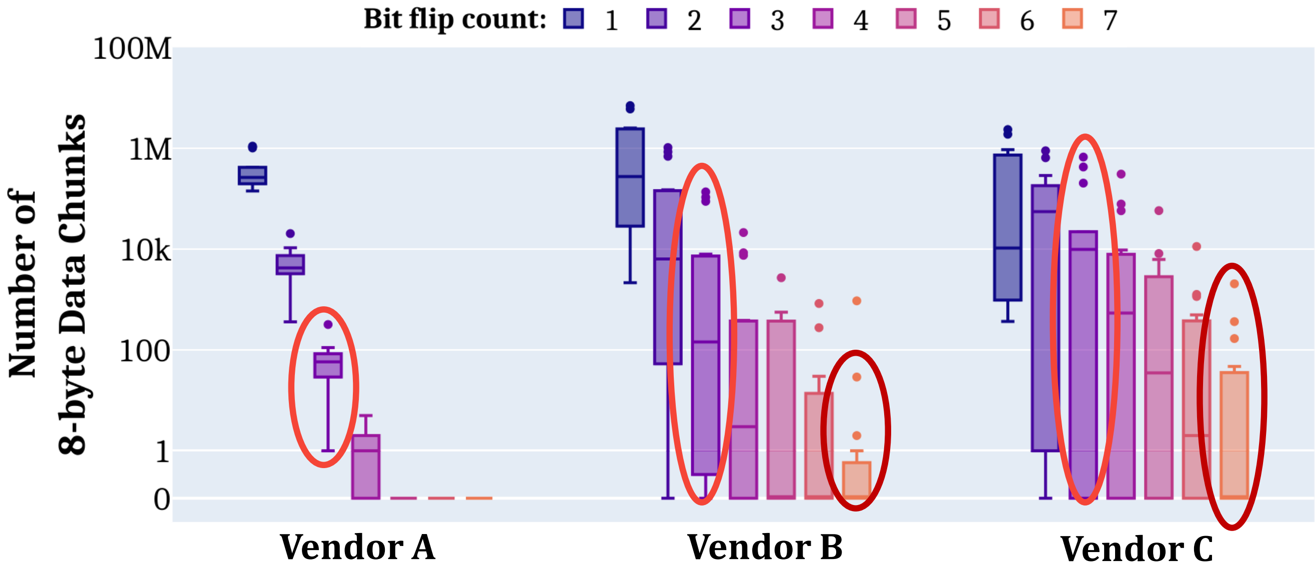
Effect on Individual Rows



All 45 modules we tested are vulnerable
to our new RowHammer access patterns

Our RowHammer access patterns
cause bit flips in more than 99.9% of the rows

Bypassing ECC with New RowHammer Patterns



Modules from all three vendors have many **8-byte data chunks** with 3 and more (up to 7) RowHammer bit flips

Conventional DRAM ECC cannot protect against our new RowHammer access patterns

SAFARI

Many Observations & Results in the Paper

- More observations on the TRRs of the three vendors
- Detailed description of the crafted access patterns
- Hammers per aggressor row sensitivity analysis
- Observations and results for individual modules
- ...

Module	Date (yy-ww)	Chip Density (Gbit)	Organization: Ranks	Organization: Banks	Organization: Pins	HC_{first}†	Our Key TRR Observations and Results: Version	Aggressor Detection	Aggressor Capacity	Per-Bank TRR	TRR-to-REF Ratio	Neighbors Refreshed	% Vulnerable DRAM Rows†	Max. Bit Flips per Row per Hammer†
A0	19-50	8	1	16	8	16K	A_{TRR1}	Counter-based	16	✓	1/9	4	73.3%	1.16
A1-5	19-36	8	1	8	16	13K-15K	A_{TRR1}	Counter-based	16	✓	1/9	4	99.2% - 99.4%	2.32 - 4.73
A6-7	19-45	8	1	8	16	13K-15K	A_{TRR1}	Counter-based	16	✓	1/9	4	99.3% - 99.4%	2.12 - 3.86
A8-9	20-07	8	1	16	8	12K-14K	A_{TRR1}	Counter-based	16	✓	1/9	4	74.6% - 75.0%	1.96 - 2.96
A10-12	19-51	8	1	16	8	12K-13K	A_{TRR1}	Counter-based	16	✓	1/9	4	74.6% - 75.0%	1.48 - 2.86
A13-14	20-31	8	1	8	16	11K-14K	A_{TRR2}	Counter-based	16	✓	1/9	2	94.3% - 98.6%	1.53 - 2.78
B0	18-22	4	1	16	8	44K	B_{TRR1}	Sampling-based	1	✗	1/4	2	99.9%	2.13
B1-4	20-17	4	1	16	8	159K-192K	B_{TRR1}	Sampling-based	1	✗	1/4	2	23.3% - 51.2%	0.06 - 0.11
B5-6	16-48	4	1	16	8	44K-50K	B_{TRR1}	Sampling-based	1	✗	1/4	2	99.9%	1.85 - 2.03
B7	19-06	8	2	16	8	20K	B_{TRR1}	Sampling-based	1	✗	1/4	2	99.9%	31.14
B8	18-03	4	1	16	8	43K	B_{TRR1}	Sampling-based	1	✗	1/4	2	99.9%	2.57
B9-12	19-48	8	1	16	8	42K-65K	B_{TRR2}	Sampling-based	1	✗	1/9	2	36.3% - 38.9%	16.83 - 24.26
B13-14	20-08	4	1	16	8	11K-14K	B_{TRR3}	Sampling-based	1	✓	1/2	4	99.9%	16.20 - 18.12
C0-3	16-48	4	1	16	x8	137K-194K	C_{TRR1}	Mix	Unknown	✓	1/17	2	1.0% - 23.2%	0.05 - 0.15
C4-6	17-12	8	1	16	x8	130K-150K	C_{TRR1}	Mix	Unknown	✓	1/17	2	7.8% - 12.0%	0.06 - 0.08
C7-8	20-31	8	1	8	x16	40K-44K	C_{TRR1}	Mix	Unknown	✓	1/17	2	39.8% - 41.8%	9.66 - 14.56
C9-11	20-31	8	1	8	x16	42K-53K	C_{TRR2}	Mix	Unknown	✓	1/9	2	99.7%	9.30 - 32.04
C12-14	20-46	16	1	8	x16	6K-7K	C_{TRR3}	Mix	Unknown	✓	1/8	2	99.9%	4.91 - 12.64

SAFARI

Uncovering TRR Can Help Future Solutions

- Hasan Hassan, Yahya Can Tugrul, Jeremie S. Kim, Victor van der Veen, Kaveh Razavi, and Onur Mutlu,
  "Uncovering In-DRAM RowHammer Protection Mechanisms: A New Methodology, Custom RowHammer Patterns, and Implications"
  Proceedings of the 54th International Symposium on Microarchitecture (***MICRO***), Virtual, October 2021.
  [Slides (pptx) (pdf)]
  [Short Talk Slides (pptx) (pdf)]
  [Lightning Talk Slides (pptx) (pdf)]
  [Talk Video (25 minutes)]
  [Lightning Talk Video (100 seconds)]
  [arXiv version]

Uncovering In-DRAM RowHammer Protection Mechanisms:
A New Methodology, Custom RowHammer Patterns, and Implications

Hasan Hassan[†] Yahya Can Tuğrul[†‡] Jeremie S. Kim[†] Victor van der Veen[σ]
Kaveh Razavi[†] Onur Mutlu[†]
[†]*ETH Zürich* [‡]*TOBB University of Economics & Technology* [σ]*Qualcomm Technologies Inc.*

New RowHammer Characteristics

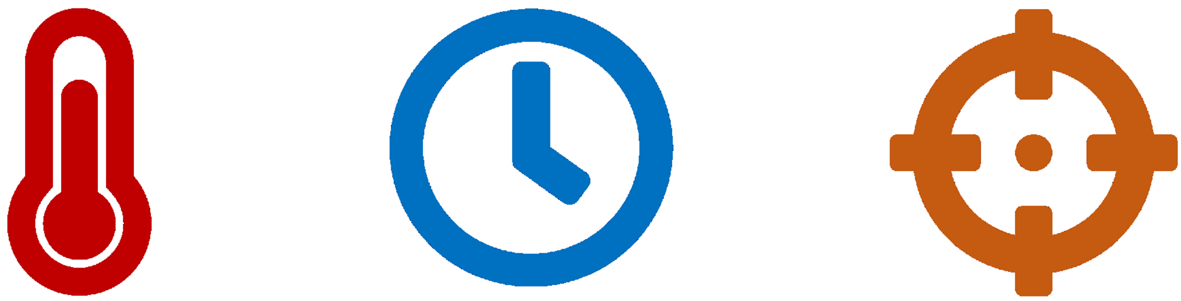
RowHammer Has Many Dimensions

- Lois Orosa, Abdullah Giray Yaglikci, Haocong Luo, Ataberk Olgun, Jisung Park, Hasan Hassan, Minesh Patel, Jeremie S. Kim, and Onur Mutlu,
  "A Deeper Look into RowHammer's Sensitivities: Experimental Analysis of Real DRAM Chips and Implications on Future Attacks and Defenses"
  Proceedings of the 54th International Symposium on Microarchitecture (***MICRO***), Virtual, October 2021.
  [Slides (pptx) (pdf)]
  [Short Talk Slides (pptx) (pdf)]
  [Lightning Talk Slides (pptx) (pdf)]
  [Talk Video (21 minutes)]
  [Lightning Talk Video (1.5 minutes)]
  [arXiv version]

A Deeper Look into RowHammer's Sensitivities: Experimental Analysis of Real DRAM Chips and Implications on Future Attacks and Defenses

Lois Orosa* ETH Zürich
A. Giray Yağlıkçı* ETH Zürich
Haocong Luo ETH Zürich
Ataberk Olgun ETH Zürich, TOBB ETÜ
Jisung Park ETH Zürich
Hasan Hassan ETH Zürich
Minesh Patel ETH Zürich
Jeremie S. Kim ETH Zürich
Onur Mutlu ETH Zürich

Our Goal



Temperature

Aggressor Row Active Time

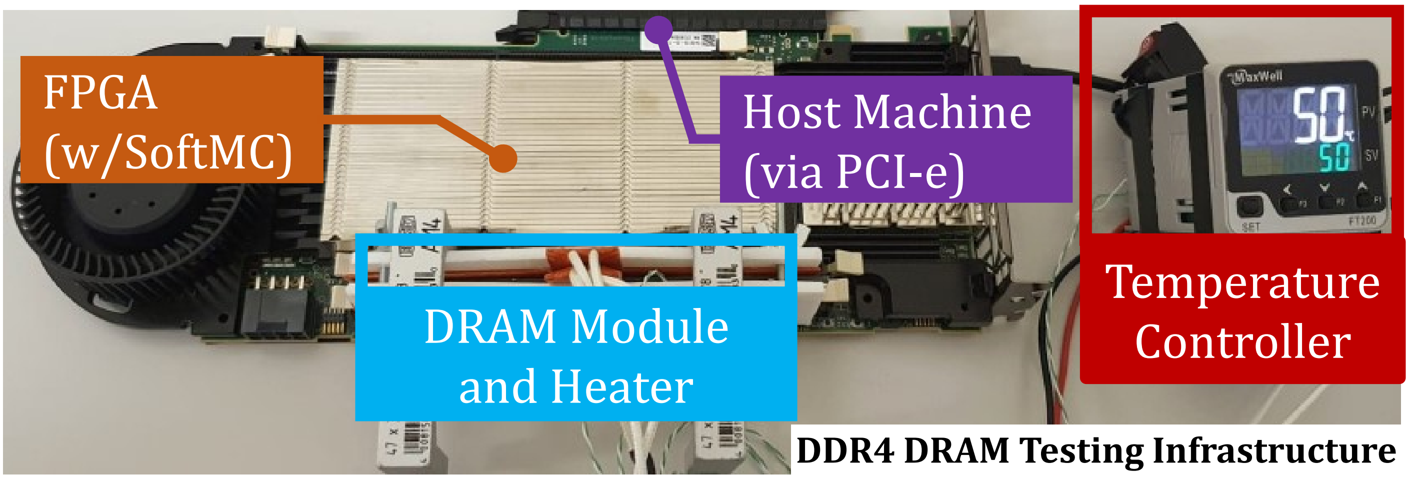
'ictim DRAM Cell's Physical Location

To find **effective and efficient** attacks and defenses

DRAM Testing Infrastructures

Two separate testing infrastructures

1. **DDR3:** FPGA-based SoftMC (Xilinx ML605)
2. **DDR4:** FPGA-based SoftMC (Xilinx Virtex UltraScale+ XCU200)



DDR4 DRAM Testing Infrastructure

Fine-grained control over **DRAM commands**, **timing parameters** and **temperature (±0.1°C)**

DRAM Chips Tested

Two DRAM standards

Mfr.	DDR4 DIMMs	DDR3 SODIMMs	# Chips	Density	Die	Org.
A (Micron)	9	1	144 (8)	8Gb (4Gb)	B (P)	x4 (x8)
B (Samsung)	4	1	32 (8)	4Gb (4Gb)	F (Q)	x8 (x8)
C (SK Hynix)	5	1	40 (8)	4Gb (4Gb)	B (B)	x8 (x8)
D (Nanya)	4	-	32 (-)	8Gb (-)	C (-)	x8 (-)

4 Major Manufacturers

272 DRAM Chips in total

Summary of The Study & Key Results

- **272 DRAM chips** from **four major manufacturers**

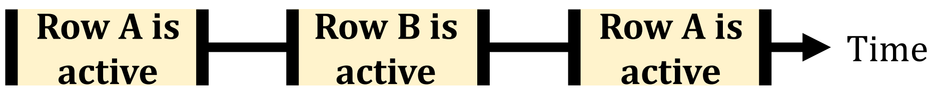
- **6 major takeaways** from **16 novel observations**

- A RowHammer bit flip is **more likely to occur**
  1) in **a bounded range of temperature**
  2) if the aggressor row is **active for longer time**
  3) in **certain physical regions** of the DRAM module under attack

- Our novel observations can inspire and aid future work
  - Craft **more effective attacks**
  - Design **more effective and efficient defenses**

Example Attack Improvement 3:
Bypassing Defenses with Aggressor Row Active Time

Activating aggressor rows as frequently as possible:



Keeping aggressor rows active for a longer time:

36% reduction in HC_{first}

Reduces the minimum activation count to induce a bit flip **by 36%**

Bypasses defenses that do not account for this reduction

Key Takeaways from Spatial Variation Analysis

Key Takeaway 5

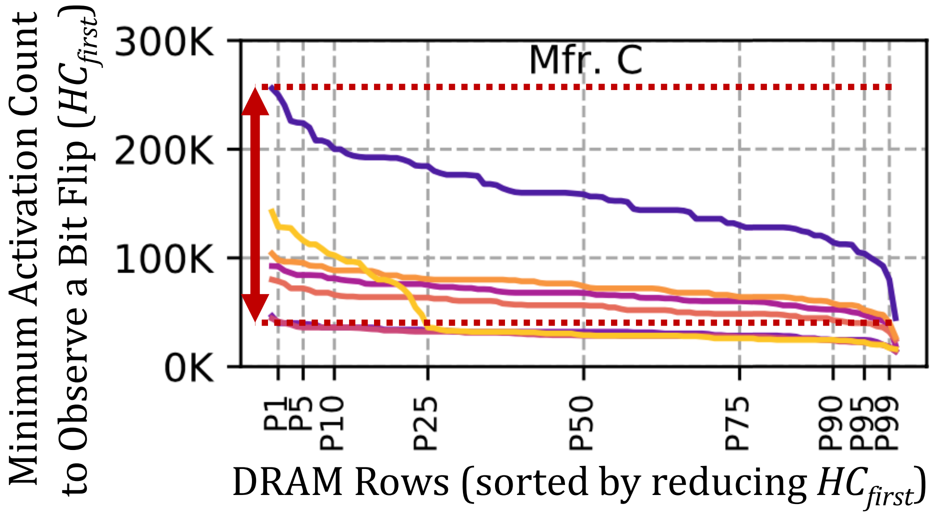
RowHammer vulnerability **significantly varies** across DRAM rows and columns due to **design-induced** and **manufacturing-process-induced** variation

Key Takeaway 6

The distribution of **the minimum activation count to observe bit flips (HC_{first})** exhibits **a diverse set of values in a subarray** but **similar values across subarrays** in the same DRAM module

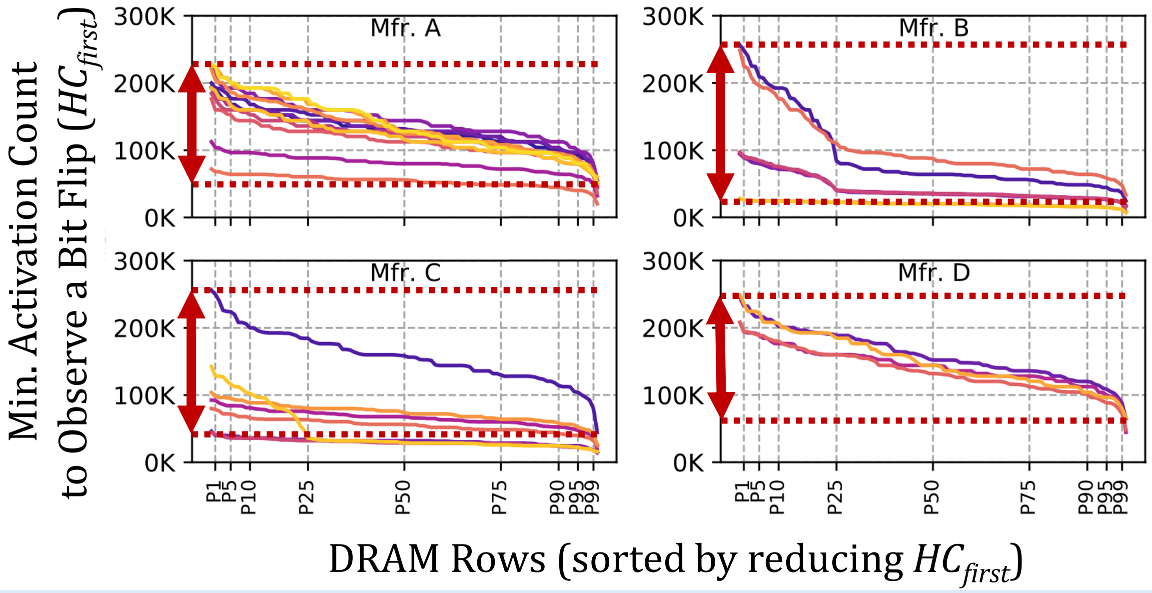
Spatial Variation across Rows

The **minimum activation count** to observe bit flips **(HC_{first})** across **DRAM rows:**



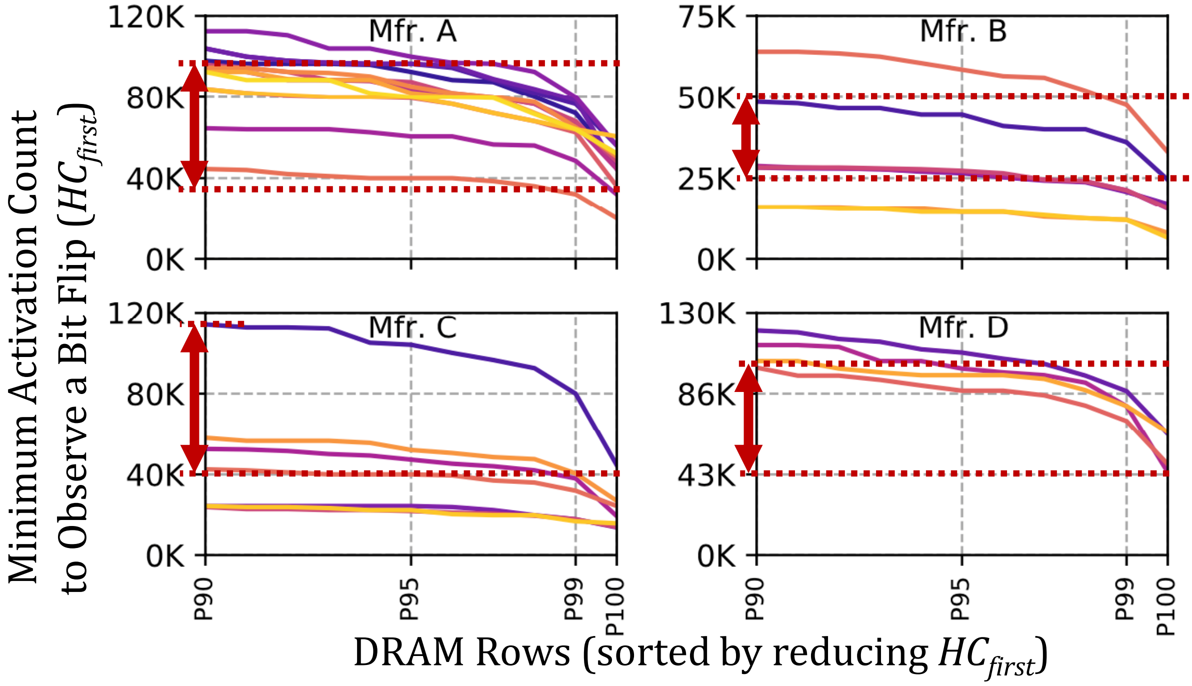
The RowHammer vulnerability **significantly varies** across DRAM rows

Spatial Variation across Rows



The RowHammer vulnerability
significantly varies across DRAM rows

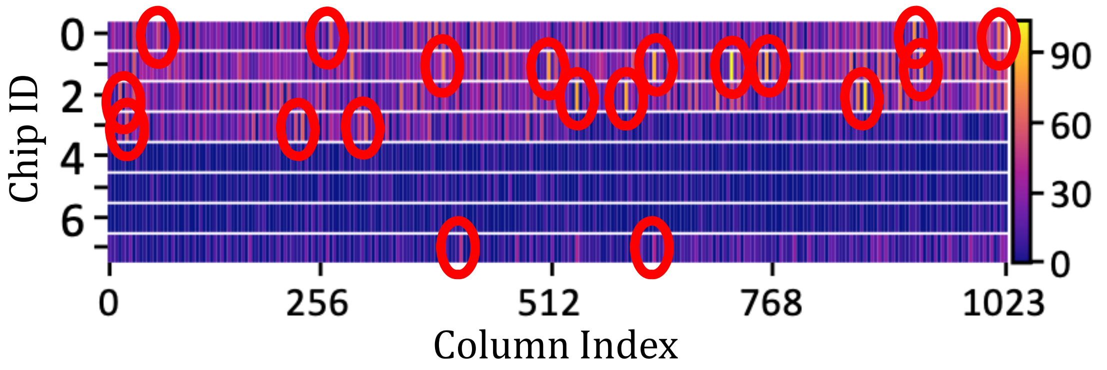
Spatial Variation across Rows



OBSERVATION 12

A small fraction of DRAM rows are **significantly more vulnerable** to RowHammer than **the vast majority** of the rows

Spatial Variation across Columns

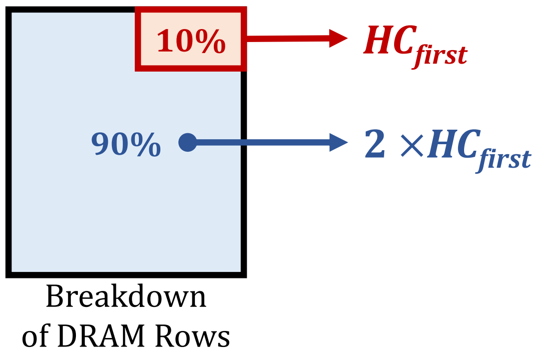


OBSERVATION 13

Certain columns are **significantly more vulnerable** to RowHammer than other columns

Example Defense Improvements

- **Example 1: Leveraging variation across DRAM rows**



Aggressiveness can be reduced:

33% area reduction
for BlockHammer [Yağlıkçı+, HPCA'21]

80% area reduction
for Graphene [Park+, MICRO'20]

- **Example 2: Leveraging variation with temperature**
  - A DRAM cell experiences **bit flips** within **a bounded temperature range**

no bit flips | **Vulnerable Temperature Range** | no bit flips → Temperature

  - A row can be **disabled** within the row's **vulnerable temperature range**

Disable RowA | **Disable RowB** → Temperature

Many More Analyses In The Paper

- Lois Orosa, Abdullah Giray Yaglikci, Haocong Luo, Ataberk Olgun, Jisung Park, Hasan Hassan, Minesh Patel, Jeremie S. Kim, and Onur Mutlu,
  "A Deeper Look into RowHammer's Sensitivities: Experimental Analysis of Real DRAM Chips and Implications on Future Attacks and Defenses"
  Proceedings of the 54th International Symposium on Microarchitecture (***MICRO***), Virtual, October 2021.
  [Slides (pptx) (pdf)]
  [Short Talk Slides (pptx) (pdf)]
  [Lightning Talk Slides (pptx) (pdf)]
  [Talk Video (21 minutes)]
  [Lightning Talk Video (1.5 minutes)]
  [arXiv version]

A Deeper Look into RowHammer's Sensitivities: Experimental Analysis of Real DRAM Chips and Implications on Future Attacks and Defenses

Lois Orosa* ETH Zürich
A. Giray Yağlıkçı* ETH Zürich
Haocong Luo ETH Zürich
Ataberk Olgun ETH Zürich, TOBB ETÜ
Jisung Park ETH Zürich
Hasan Hassan ETH Zürich
Minesh Patel ETH Zürich
Jeremie S. Kim ETH Zürich
Onur Mutlu ETH Zürich

More RowHammer Analysis

RowHammer vs. Wordline Voltage (2022)

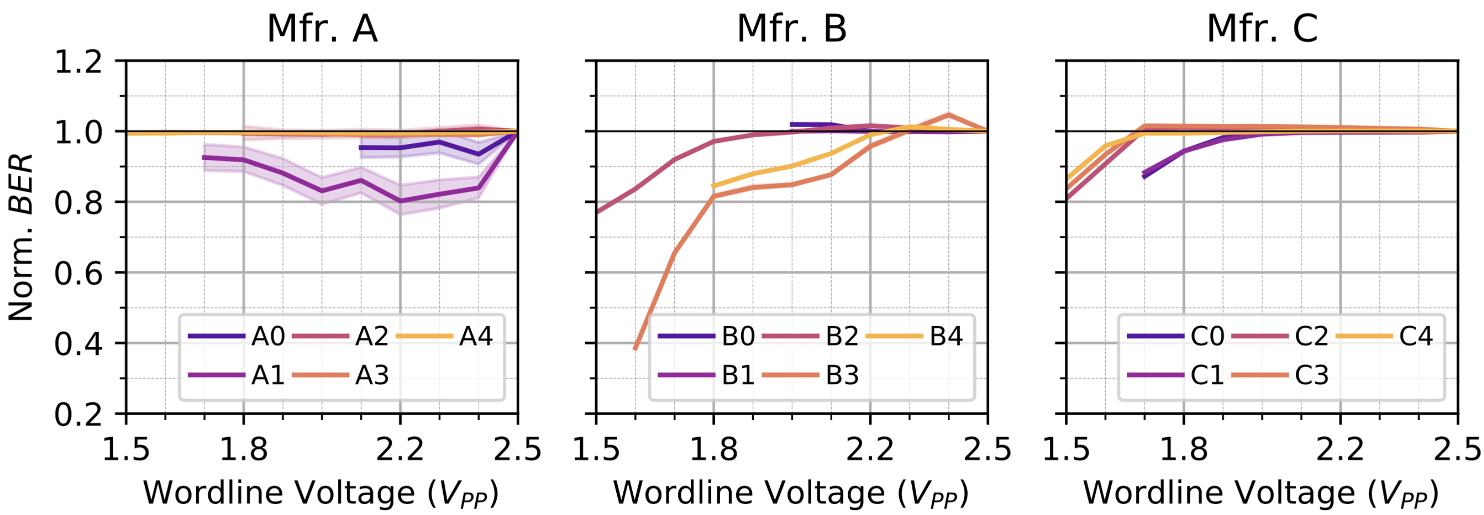
- To appear in DSN 2022

Understanding the RowHammer Vulnerability Under Reduced Wordline Voltage: An Experimental Study Using Real DRAM Devices

A. Giray Yağlıkçı[1] Haocong Luo[1] Geraldo F. de Oliviera[1] Ataberk Olgun[1] Jisung Park[1]
Minesh Patel[1] Hasan Hassan[1] Jeremie S. Kim[1] Lois Orosa[1,2] Onur Mutlu[1]
[1]ETH Zürich [2]Galicia Supercomputing Center (CESGA)

Sneak Peak: RowHammer vs. Voltage [DSN'22]

- Voltage swing on a DRAM row's wordline causes RowHammer
- No prior study on the impact of voltage on RowHammer



RowHammer vulnerability can be reduced via voltage scaling

New RowHammer Solutions

BlockHammer Solution in 2021

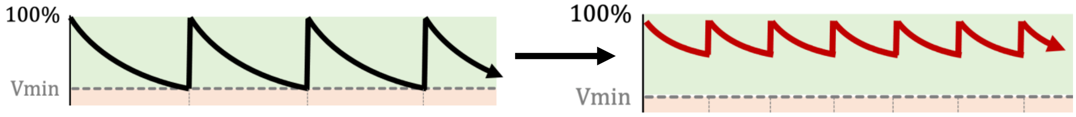
- A. Giray Yaglikci, Minesh Patel, Jeremie S. Kim, Roknoddin Azizi, Ataberk Olgun, Lois Orosa, Hasan Hassan, Jisung Park, Konstantinos Kanellopoulos, Taha Shahroodi, Saugata Ghose, and Onur Mutlu,
"BlockHammer: Preventing RowHammer at Low Cost by Blacklisting Rapidly-Accessed DRAM Rows"
*Proceedings of the 27th International Symposium on High-Performance Computer Architecture (**HPCA**)*, Virtual, February-March 2021.
[Slides (pptx) (pdf)]
[Short Talk Slides (pptx) (pdf)]
[Talk Video (22 minutes)]
[Short Talk Video (7 minutes)]

BlockHammer: Preventing RowHammer at Low Cost by Blacklisting Rapidly-Accessed DRAM Rows

A. Giray Yağlıkçı[1] Minesh Patel[1] Jeremie S. Kim[1] Roknoddin Azizi[1] Ataberk Olgun[1] Lois Orosa[1] Hasan Hassan[1] Jisung Park[1] Konstantinos Kanellopoulos[1] Taha Shahroodi[1] Saugata Ghose[2] Onur Mutlu[1]
[1]*ETH Zürich* [2]*University of Illinois at Urbana–Champaign*

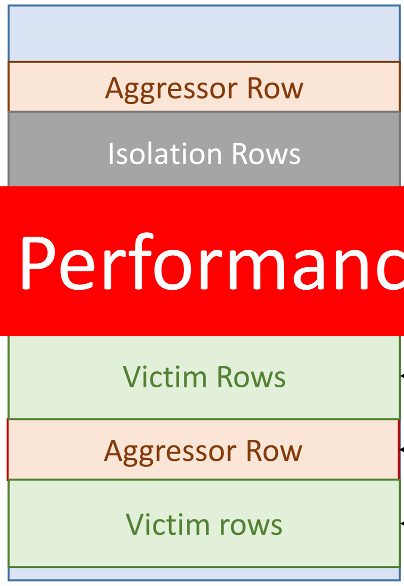
RowHammer Solution Approaches

- More robust DRAM chips **and/or** error-correcting codes
- Increased refresh rate



Fewer activations possible in a refresh interval

- Physical isolation



Large-enough distance

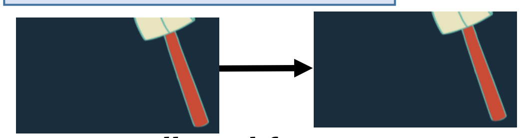
Refresh

Rapidly activated (hammered)

Refresh

Cost, Power, Performance, Complexity

- Reactive refresh
- Proactive throttling



Fewer activations allowed for aggressive applications

Two Key Challenges

1. **Scalability** with worsening RowHammer vulnerability
2. **Compatibility** with commodity DRAM chips

Our Goal

To prevent RowHammer efficiently and scalably *without* knowledge of or modifications to DRAM internals

BlockHammer
Key Idea

Selectively throttle memory accesses that may cause RowHammer bit-flips

BlockHammer
Overview of Approach

RowBlocker

Tracks row activation rates using area-efficient Bloom filters

Blacklists rows that are activated at a high rate

Throttles activations targeting a blacklisted row

No row can be activated at a high enough rate to induce bit-flips

AttackThrottler

Identifies threads that perform a RowHammer attack

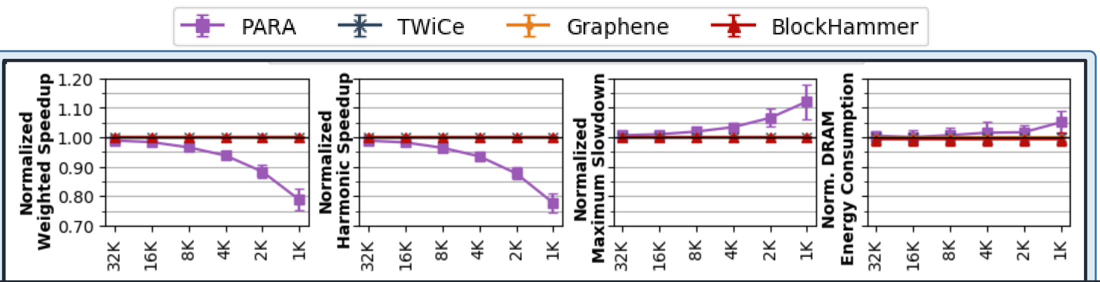
Reduces memory bandwidth usage of identified threads

Greatly reduces the **performance degradation**
and **energy wastage** a RowHammer attack inflicts on a system

Evaluation
Scaling with RowHammer Vulnerability

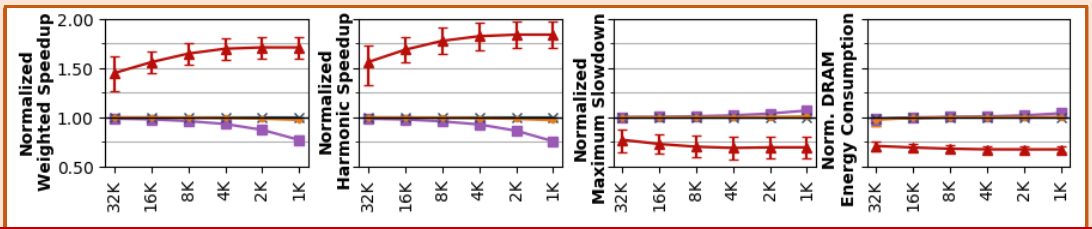
- System throughput (weighted speedup)
- Job turnaround time (harmonic speedup)
- Unfairness (maximum slowdown)
- DRAM energy consumption

No RowHammer Attack



BlockHammer's performance and energy overheads remain **negligible (<0.6%)**

RowHammer Attack Present



BlockHammer scalably provides **much higher performance** (71% on average)
and **lower energy consumption** (32% on average) than state-of-the-art mechanisms

Key Results: BlockHammer

- **Competitive** with state-of-the-art mechanisms **when there is no attack**
- **Superior** performance and DRAM energy **when RowHammer attack present**
- **Better hardware area scaling with RowHammer vulnerability**
- **Security Proof**
- Addresses **Many-Sided Attacks**
- Evaluation of **14 mechanisms** representing **four mitigation approaches**
  - Comprehensive Protection
  - Compatibility with Commodity DRA
  - Scalability with RowHammer Vulne
  - Deterministic Protection

Approach	Mechanism	Comprehensive Protection	Compatible w/ Commodity DRAM Chips	Scaling with RowHammer Vulnerability	Deterministic Protection
Increased Refresh Rate [2, 73]		✓	✓	✗	✓
Physical Isolation	CATT [14]	✗	✗	✗	✓
	GuardION [148]	✗	✗	✗	✓
	ZebRAM [78]	✗	✗	✗	✓
Reactive Refresh	ANVIL [5]	✗	✗	✗	✓
	PARA [73]	✓	✗	✗	✗
	PRoHIT [137]	✓	✗	✗	✗
	MRLoc [161]	✓	✗	✗	✗
	CBT [132]	✓	✗	✗	✓
	TWiCe [84]	✓	✗	✗	✓
	Graphene [113]	✓	✗	✓	✓
Proactive Throttling	Naive Thrott. [102]	✓	✓	✗	✓
	Thrott. Supp. [40]	✓	✗	✗	✓
	BlockHammer	✓	✓	✓	✓

A Takeaway

Main Memory Needs

Intelligent Controllers

for Security, Safety, Reliability, Scaling

More RowHammer in 2020-2022

RowHammer in 2020 (I)

MICRO 2020 | Submit Work | Program | Atte

Session 1A: Security & Privacy I

5:00 PM CEST – 5:15 PM CEST
Graphene: Strong yet Lightweight Row Hammer Protection
Yeonhong Park, Woosuk Kwon, Eojin Lee, Tae Jun Ham, Jung Ho Ahn, Jae W. Lee (Seoul National University)

5:15 PM CEST – 5:30 PM CEST
Persist Level Parallelism: Streamlining Integrity Tree Updates for Secure Persistent Memory
Alexander Freij, Shougang Yuan, Huiyang Zhou (NC State University); Yan Solihin (University of Central Florida)

5:30 PM CEST – 5:45 PM CEST
PThammer: Cross-User-Kernel-Boundary Rowhammer through Implicit Accesses
Zhi Zhang (University of New South Wales and Data61, CSIRO, Australia); Yueqiang Cheng (Baidu Security); Dongxi Liu, Surya Nepal (Data61, CSIRO, Australia); Zhi Wang (Florida State University); Yuval Yarom (University of Adelaide and Data61, CSIRO, Australia)

RowHammer in 2020 (II)

S & P | Home | Program | Call For... | Attend | Workshops

Session #5: Rowhammer — Room 2

Session chair: Michael Franz (UC Irvine)

RAMBleed: Reading Bits in Memory Without Accessing Them
Andrew Kwong (University of Michigan), Daniel Genkin (University of Michigan), Daniel Gruss
Data61)

Are We Susceptible to Rowhammer? An End-to-End Methodology for Cloud Providers
Lucian Cojocar (Microsoft Research), Jeremie Kim (ETH Zurich, CMU), Minesh Patel (ETH Zu
(Microsoft Research), Onur Mutlu (ETH Zurich, CMU)

Leveraging EM Side-Channel Information to Detect Rowhammer Attacks
Zhenkai Zhang (Texas Tech University), Zihao Zhan (Vanderbilt University), Daniel Balasubrar
Peter Volgyesi (Vanderbilt University), Xenofon Koutsoukos (Vanderbilt University)

TRRespass: Exploiting the Many Sides of Target Row Refresh
Pietro Frigo (Vrije Universiteit Amsterdam, The Netherlands), Emanuele Vannacci (Vrije Univer
Veen (Qualcomm Technologies, Inc.), Onur Mutlu (ETH Zürich), Cristiano Giuffrida (Vrije Unive
The Netherlands), Kaveh Razavi (Vrije Universiteit Amsterdam, The Netherlands)

RowHammer in 2020 (III)

29TH USENIX SECURITY SYMPOSIUM | ATTEND | PROGRAM | PARTICIPATE | SPONSORS | ABOUT

DeepHammer: Depleting the Intelligence of Deep Neural Networks through Targeted Chain of Bit Flips
Fan Yao, *University of Central Florida;* Adnan Siraj Rakin and Deliang Fan, *Arizona State University*

AVAILABLE MEDIA

Show details

RowHammer in 2021 (I)

HotOS XVIII

The 18th Workshop on Hot Topics in Operating Systems

~~31 May~~ 1 June–3 June 2021, Cyberspace, People's Couches, and Zoom

Stop! Hammer Time: Rethinking Our Approach to Rowhammer Mitigations

RowHammer in 2021 (II)

30TH USENIX SECURITY SYMPOSIUM | ATTEND | PROGRAM | PARTICIPATE | SPONSORS | ABOUT

SMASH: Synchronized Many-sided Rowhammer Attacks from JavaScript

RowHammer in 2021 (III)

Session 10A: Security & Privacy III

Session Chair: Hoda Naghibijouybari (Binghamton)

9:00 PM CEST – 9:15 PM CEST

A Deeper Look into RowHammer's Sensitivities: Experimental Analysis of Real DRAM Chips and Implications on Future Attacks and Defenses

Lois Orosa, Abdullah Giray Yaglikci, Haocong Luo (ETH Zurich); Ataberk Olgun (TOBB University of Economics and Technology); Jisung Park, Hasan Hassan, Minesh Patel, Jeremie S. Kim, Onur Mutlu (ETH Zurich)

Paper

9:15 PM CEST – 9:30 PM CEST

Uncovering In-DRAM RowHammer Protection Mechanisms: A New Methodology, Custom RowHammer Patterns, and Implications

Hasan Hassan (ETH Zurich); Yahya Can Tugrul (TOBB University of Economics and Technology); Jeremie S. Kim (ETH Zurich); Victor van der Veen (Qualcomm); Kaveh Razavi, Onur Mutlu (ETH Zurich)

Paper

RowHammer in 2022 (I)

MAY 22-26, 2022 AT THE HYATT REGENCY, SAN FRANCISCO, CA

43rd IEEE Symposium on Security and Privacy

BLACKSMITH: Scalable Rowhammering in the Frequency Domain

SpecHammer: Combining Spectre and Rowhammer for New Speculative Attacks

PROTRR: Principled yet Optimal In-DRAM Target Row Refresh

RowHammer in 2022 (II)

Randomized Row-Swap: Mitigating Row Hammer by Breaking Spatial Correlation between Aggressor and Victim Rows

RowHammer in 2022 (III)

HPCA 2022

The 28th IEEE International Symposium on High-Performance Computer Architecture (HPCA-28), Seoul, South Korea

SafeGuard: Reducing the Security Risk from Row-Hammer via Low-Cost Integrity Protection

Mithril: Cooperative Row Hammer Protection on Commodity DRAM Leveraging Managed Refresh

RowHammer in 2022 (IV)

IRPS
International Reliability Physics Symposium

IRPS 2022

The Price of Secrecy: How Hiding Internal DRAM Topologies Hurts Rowhammer Defenses

Stefan Saroiu, Alec Wolman, Lucian Cojocar
Microsoft

More to Come…

Future Memory Reliability/Security Challenges

Future of Main Memory Security

- DRAM is becoming less reliable → more vulnerable
- Due to difficulties in DRAM scaling, other problems may also appear (or they may be going unnoticed)
- Some errors may already be slipping into the field
  - Read disturb errors (Rowhammer)
  - Retention errors
  - Read errors, write errors
  - ...
- These errors can also pose security vulnerabilities

Future of Main Memory Security

- DRAM
- Flash memory
- Emerging Technologies
  - Phase Change Memory
  - STT-MRAM
  - RRAM, memristors
  - ...

A Takeaway

Main Memory Needs

Intelligent Controllers

for Security, Safety, Reliability, Scaling

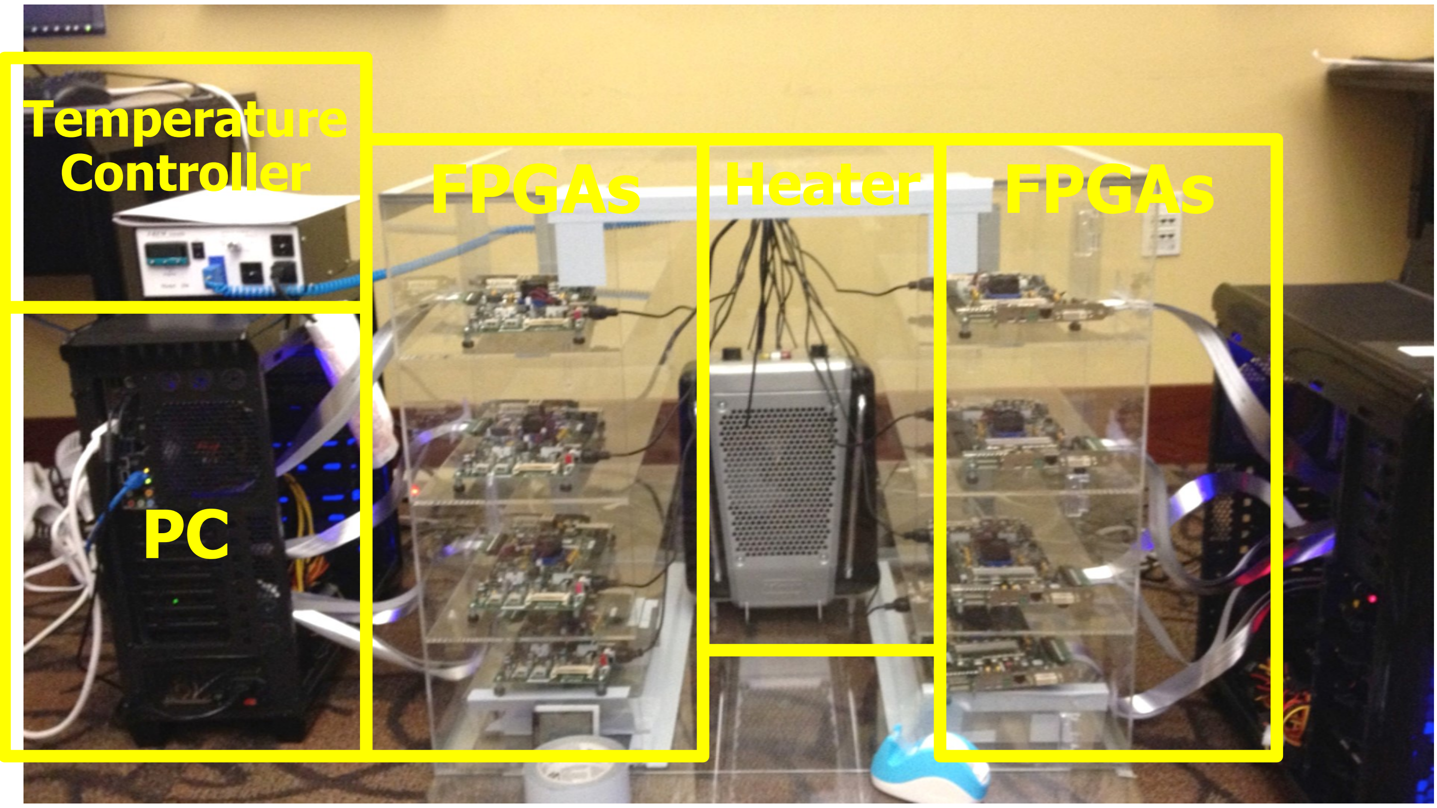
The Takeaway

Intelligent
Memory Controllers
Can Avoid Many Failures
& Enable Better Scaling

Architecting Future Memory for Security

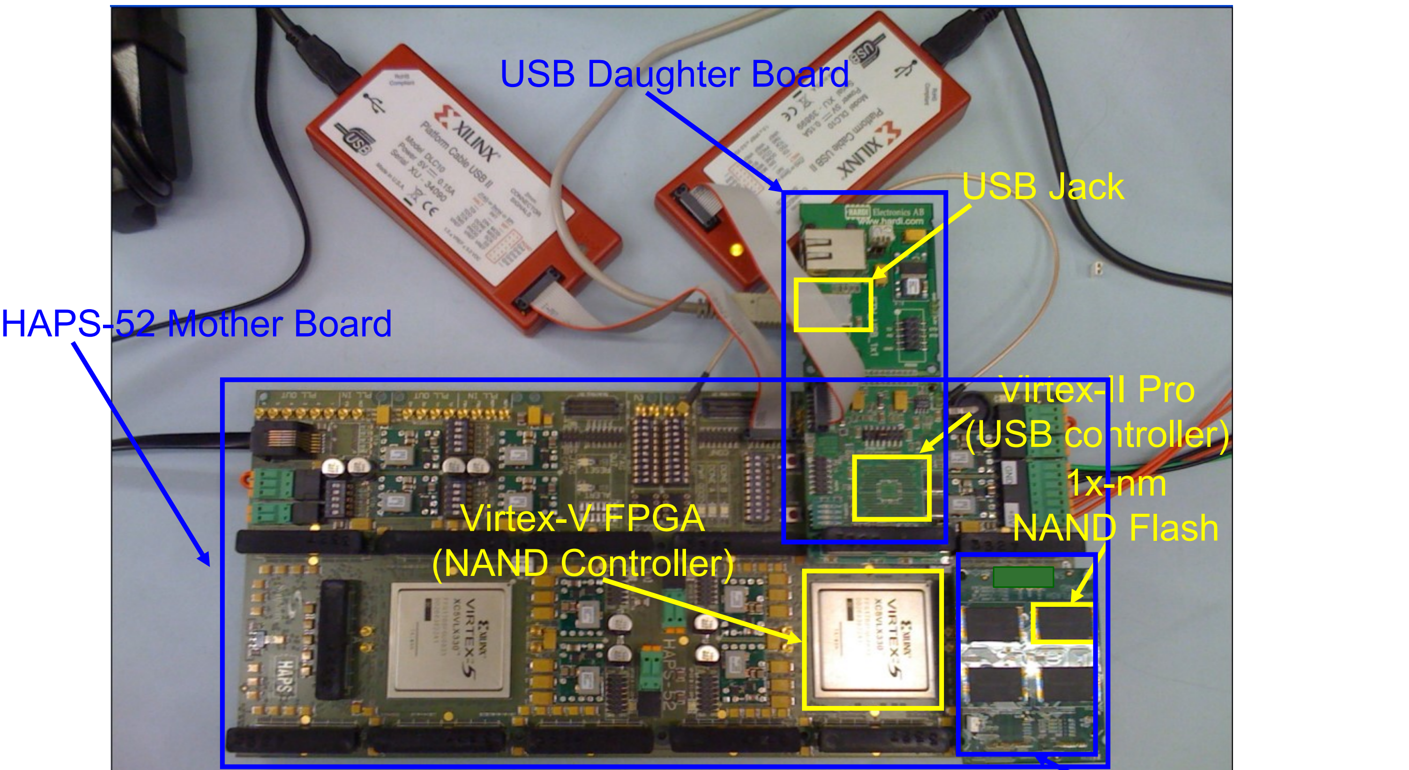
- **Understand**: Methods for vulnerability modeling & discovery
  - Modeling and prediction based on real (device) data and analysis
  - Understanding vulnerabilities
  - Developing reliable metrics
- **Architect**: Principled architectures with security as key concern
  - Good partitioning of duties across the stack
  - Cannot give up performance and efficiency
  - Patch-ability in the field
- **Design & Test**: Principled design, automation, (online) testing
  - Design for security
  - High coverage and good interaction with system reliability methods

Understand and Model with Experiments (DRAM)



Kim+, "Flipping Bits in Memory Without Accessing Them: An Experimental Study of DRAM Disturbance Errors," ISCA 2014.

Understand and Model with Experiments (Flash)



[DATE 2012, ICCD 2012, DATE 2013, ITJ 2013, ICCD 2013, SIGMETRICS 2014, HPCA 2015, DSN 2015, MSST 2015, JSAC 2016, HPCA 2017, DFRWS 2017, PIEEE 2017, HPCA 2018, SIGMETRICS 2018]

Cai+, "Error Characterization, Mitigation, and Recovery in Flash Memory Based Solid State Drives," Proc. IEEE 2017.

An Example Intelligent Controller

INVITED PAPER

Proceedings of the IEEE, Sept. 2017

Error Characterization, Mitigation, and Recovery in Flash-Memory-Based Solid-State Drives

This paper reviews the most recent advances in solid-state drive (SSD) error characterization, mitigation, and data recovery techniques to improve both SSD's reliability and lifetime.

By Yu Cai, Saugata Ghose, Erich F. Haratsch, Yixin Luo, and Onur Mutlu

https://arxiv.org/pdf/1706.08642

Collapse of the "Galloping Gertie" (1940)

Source: AP
http://www.wsdot.wa.gov/tnbhistory/connections/connections3.htm

Another Example (1994)

Source: By 최광모 - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=35197984

Yet Another Example (2007)

Source: Morry Gash/AP,
https://www.npr.org/2017/08/01/540669701/10-years-after-bridge-collapse-america-is-still-crumbling?t=1535427165809

A More Recent Example (2018)

Source: AFP / Valery HACHE, https://www.capitalfm.co.ke/news/2018/08/genoa-bridge-collapse-what-we-know/

The Takeaway, Again

In-Field Patch-ability

(Intelligent Memory)

Can Avoid Such Failures

An Early Proposal for Intelligent Controllers [IMW'13]

- Onur Mutlu,
  "Memory Scaling: A Systems Architecture Perspective"
  Proceedings of the 5th International Memory Workshop (***IMW***), Monterey, CA, May 2013. Slides (pptx) (pdf)
  EETimes Reprint

Memory Scaling: A Systems Architecture Perspective

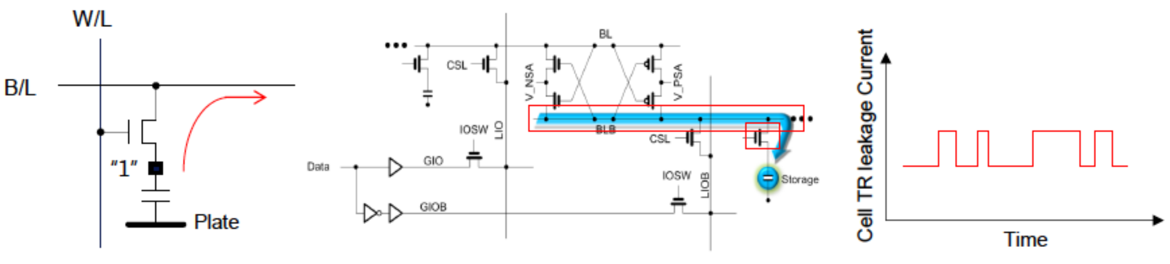
Onur Mutlu
Carnegie Mellon University
onur@cmu.edu
http://users.ece.cmu.edu/~omutlu/

https://people.inf.ethz.ch/omutlu/pub/memory-scaling_memcon13.pdf

Industry Is Writing Papers About It, Too

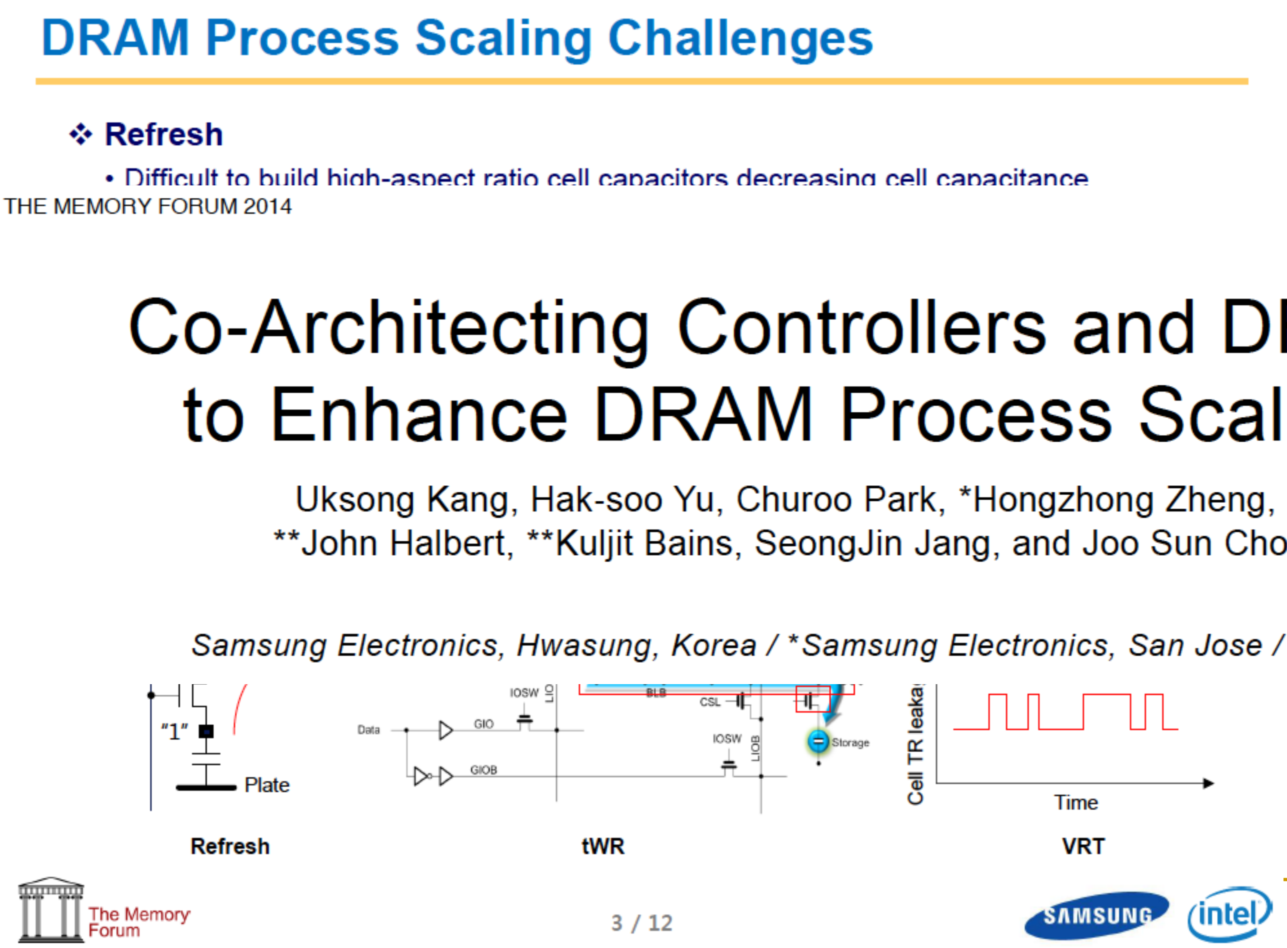
DRAM Process Scaling Challenges

- **Refresh**
  - Difficult to build high-aspect ratio cell capacitors decreasing cell capacitance
  - Leakage current of cell access transistors increasing
- **tWR**
  - Contact resistance between the cell capacitor and access transistor increasing
  - On-current of the cell access transistor decreasing
  - Bit-line resistance increasing
- **VRT**
  - Occurring more frequently with cell capacitance decreasing



Refresh **tWR** **VRT**

Industry Is Writing Papers About It, Too



Final Thoughts on RowHammer

Before RowHammer (I)

Using Memory Errors to Attack a Virtual Machine

Sudhakar Govindavajhala * Andrew W. Appel
Princeton University
{sudhakar,appel}@cs.princeton.edu

We present an experimental study showing that soft memory errors can lead to serious security vulnerabilities in Java and .NET virtual machines, or in any system that relies on type-checking of untrusted programs as a protection mechanism. Our attack works by sending to the JVM a Java program that is designed so that almost any memory error in its address space will allow it to take control of the JVM. All conventional Java and .NET virtual machines are vulnerable to this attack. The technique of the attack is broadly applicable against other language-based security schemes such as proof-carrying code.

We measured the attack on two commercial Java Virtual Machines: Sun's and IBM's. We show that a single-bit error in the Java program's data space can be exploited to execute arbitrary code with a probability of about 70%, and multiple-bit errors with a lower probability.

Our attack is particularly relevant against smart cards or tamper-resistant computers, where the user has physical access (to the outside of the computer) and can use various means to induce faults; we have successfully used heat. Fortunately, there are some straightforward defenses against this attack.

7 Physical fault injection

If the attacker has physical access to the outside of the machine, as in the case of a smart card or other tamper-resistant computer, the attacker can induce memory errors. We considered attacks on boxes in form factors ranging from a credit card to a palmtop to a desktop PC.

We considered several ways in which the attacker could induce errors.[4]

IEEE S&P 2003

https://www.cs.princeton.edu/~appel/papers/memerr.pdf

Before RowHammer (II)

Using Memory Errors to Attack a Virtual Machine

Sudhakar Govindavajhala * Andrew W. Appel
Princeton University
{sudhakar,appel}@cs.princeton.edu

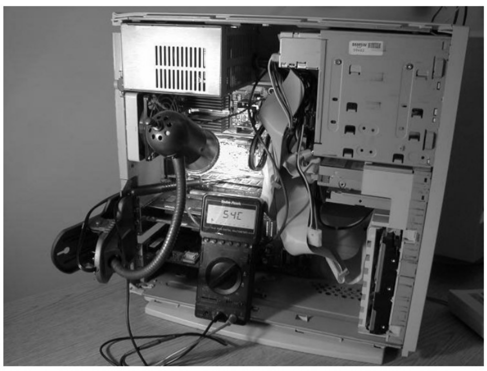


Figure 3. Experimental setup to induce memory errors, showing a PC built from surplus components, clip-on gooseneck lamp, 50-watt spotlight bulb, and digital thermometer. Not shown is the variable AC power supply for the lamp.

IEEE S&P 2003

https://www.cs.princeton.edu/~appel/papers/memerr.pdf

After RowHammer

A simple memory error

can be induced by software

RowHammer: Retrospective

- New mindset that has enabled a renewed interest in HW security attack research:
  - Real (memory) chips are vulnerable, in a simple and widespread manner → this causes real security problems
  - Hardware reliability → security connection is now mainstream discourse

- Many new RowHammer attacks…
  - Tens of papers in top security & architecture venues
  - **More to come** as RowHammer is getting worse (DDR4 & beyond)

- Many new RowHammer solutions…
  - Apple security release; Memtest86 updated
  - Many solution proposals in top venues (latest in ASPLOS 2022)
  - Principled system-DRAM co-design (in original RowHammer paper)
  - **More to come…**

Perhaps Most Importantly…

- RowHammer enabled a shift of mindset in mainstream security researchers
  - General-purpose hardware is fallible, in a widespread manner
  - Its problems are exploitable

- This mindset has enabled many systems security researchers to examine hardware in more depth
  - And understand HW's inner workings and vulnerabilities

- It is no coincidence that two of the groups that discovered Meltdown and Spectre heavily worked on RowHammer attacks before
  - **More to come…**

Conclusion

Summary: RowHammer

- Memory reliability is reducing
- Reliability issues open up security vulnerabilities
  - Very hard to defend against
- **Rowhammer is a prime example**
  - First example of how a simple hardware failure mechanism can create a widespread system security vulnerability
  - Its implications on system security research are tremendous & exciting

- Bad news: RowHammer is getting worse

- **Good news: We have a lot more to do**
  - We are now fully aware hardware is easily fallible
  - We are developing both attacks and solutions
  - We are developing principled models, methodologies, solutions

A RowHammer Survey Across the Stack

- Onur Mutlu and Jeremie Kim,
  "RowHammer: A Retrospective"
  IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (***TCAD***) *Special Issue on Top Picks in Hardware and Embedded Security*, 2019.
  [Preliminary arXiv version]
  [Slides from COSADE 2019 (pptx)]
  [Slides from VLSI-SOC 2020 (pptx) (pdf)]
  [Talk Video (1 hr 15 minutes, with Q&A)]

RowHammer: A Retrospective

Onur Mutlu[§‡] Jeremie S. Kim[‡§]
[§]ETH Zürich [‡]Carnegie Mellon University

Detailed Lectures on RowHammer

- Computer Architecture, Fall 2021, Lecture 5
  - RowHammer (ETH Zürich, Fall 2021)
  - https://www.youtube.com/watch?v=7wVKnPj3NVw&list=PL5Q2soXY2Zi-Mnk1PxjEIG32HAGILkTOF&index=5

- Computer Architecture, Fall 2021, Lecture 6
  - RowHammer and Secure & Reliable Memory (ETH Zürich, Fall 2021)
  - https://www.youtube.com/watch?v=HNd4skQrt6I&list=PL5Q2soXY2Zi-Mnk1PxjEIG32HAGILkTOF&index=6

https://www.youtube.com/onurmutlulectures

Funding Acknowledgments

- Alibaba, AMD, ASML, Google, Facebook, Hi-Silicon, HP Labs, Huawei, IBM, Intel, Microsoft, Nvidia, Oracle, Qualcomm, Rambus, Samsung, Seagate, VMware, Xilinx
- NSF
- NIH
- GSRC
- SRC
- CyLab
- EFCL

Thank you!

Acknowledgments

SAFARI

SAFARI Research Group

safari.ethz.ch

Think BIG, Aim HIGH!

https://safari.ethz.ch

SAFARI Research Group

- https://safari.ethz.ch/safari-newsletter-december-2021/

Security Aspects of DRAM
The Story of RowHammer

Onur Mutlu
omutlu@gmail.com
https://people.inf.ethz.ch/omutlu

15 May 2022
IMW Tutorial

SAFARI ETH zürich Carnegie Mellon

Swaroop Ghosh
Pennsylvania State University

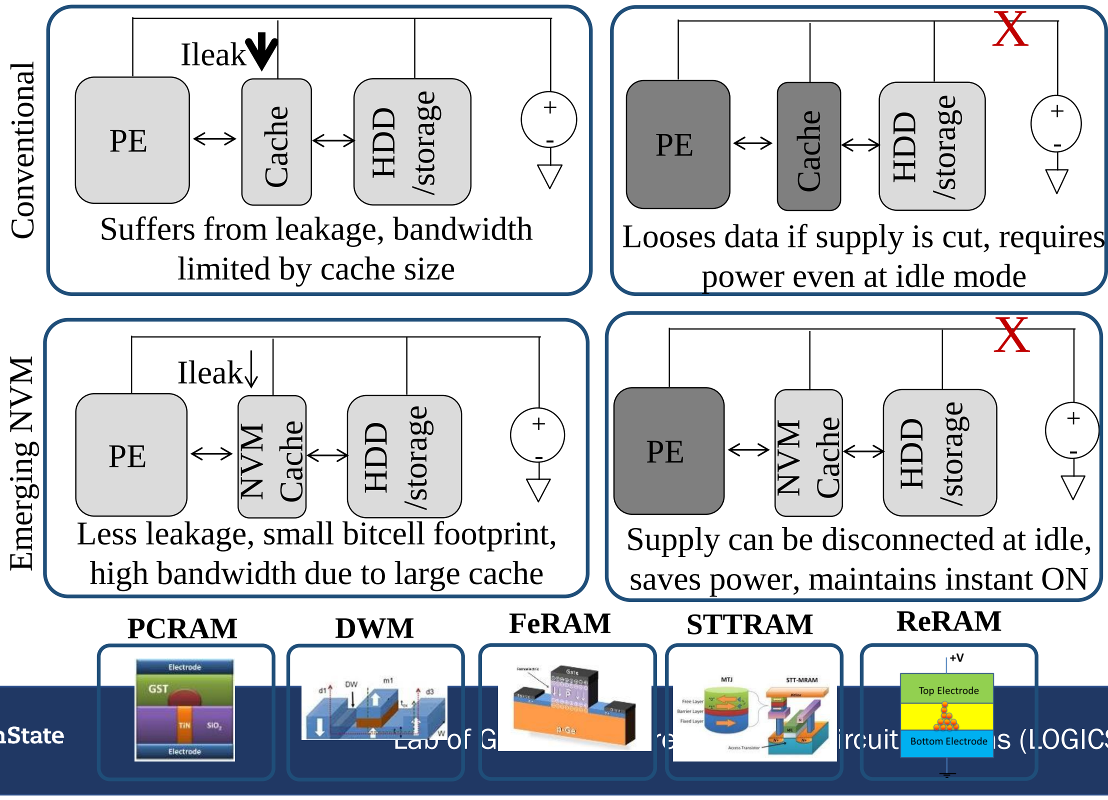
Swaroop Ghosh received the B.E. from IIT, Roorkee and Ph.D. from Purdue. He is an Associate Professor at Penn State. Prior to that, he was a Senior Research and Development Engineer at Intel. His research interests include low-power circuits, hardware security, quantum computing and digital testing for nanometer technologies.

Security Aspects of Non-Volatile Memories

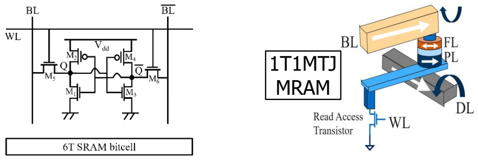
Dr. Swaroop Ghosh

School of Electrical Engineering and Computer Science, The Pennsylvania State University

Why Emerging NVM?



Why Emerging NVMs?

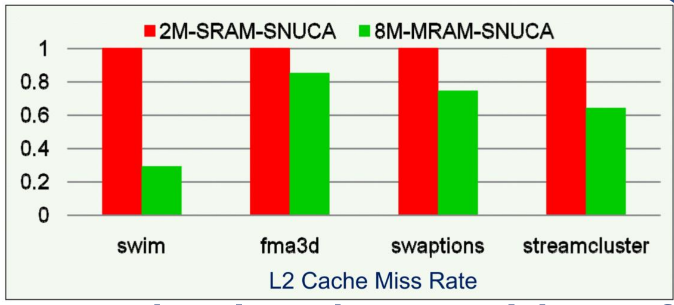


	SRAM	MRAM
Area	3.66mm²	3.30mm²
Capacity	128KB	512KB
Read latency	2.25ns	2.32ns
Write latency	2.26ns	11.02ns
Read energy	0.9nJ	0.86nJ
Write energy	0.8nJ	5.0nJ

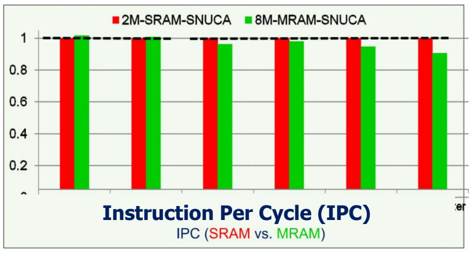
Cache	Leakage
2MB (16x128KB) SRAM	2.09W
8MB (16x512KB) MRAM	0.26W

- Pro: High Density, Low Leakage Power
- Con: Long write latency, high write energy

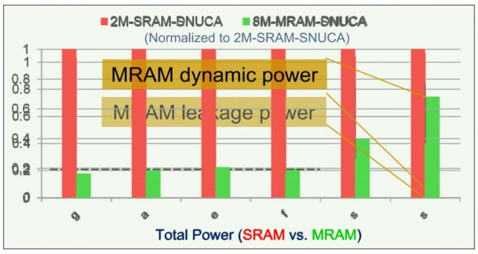
- Same area
- Same #of banks
- L2 cache size 4X (MRAM)



MRAM Cache miss reduces! → Higher Perf.



Last 4 benchmark: write extensive



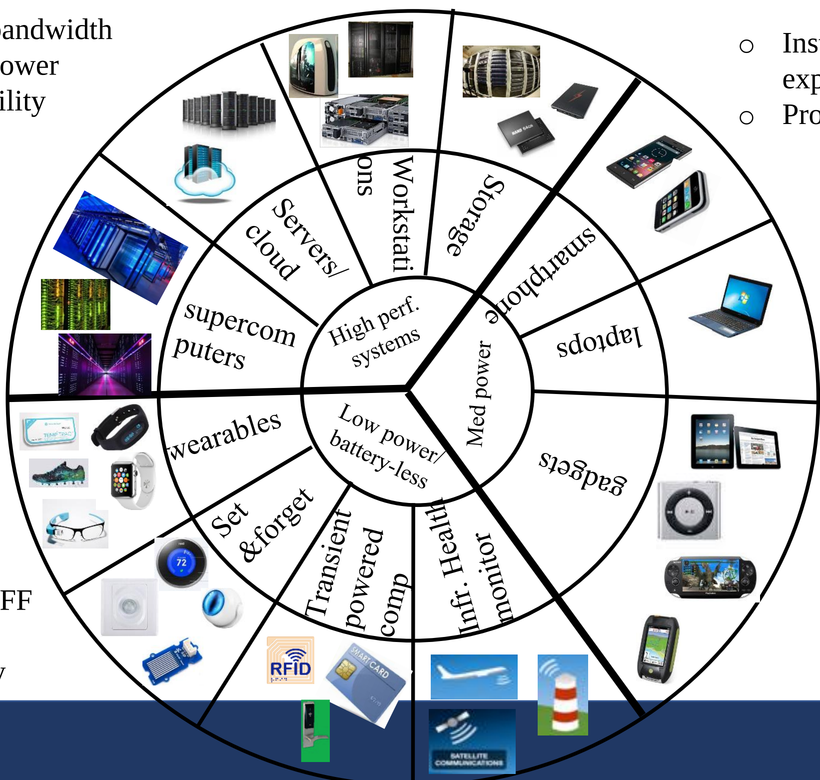
Significant lower power (MRAM)

[Guangyu Sun et al, HPCA 2009]

Non-Volatile Memory Design Space

- High bandwidth
- Low-power
- Reliability

- Instant ON experience
- Prolongs battery life



- Mostly OFF
- Power-efficiency

Emerging NVM Application

Industry & Transportation	Medical Sector	Programmable Logic Controller	Storage Accelerator	Storage

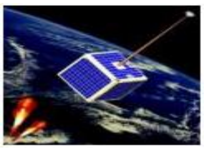
Automotive
BMW 1000RR Motorrad

Hemodialysis Machine by Nikkiso

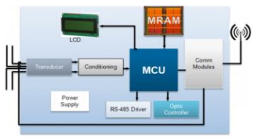
Direct Logic 205 PLC

Card
ES1GB-N03 nvNITRO

Optane by Intel



Aerospace
TAMU Magnetometer

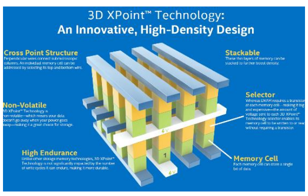


Smart Meter
RIM789 Smart Meter

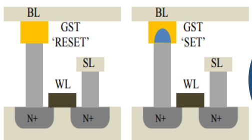
Accelerator
ES1GB-U201 NVMe

DDR3
EMD3D 256Mb by Everspin

Emerging NVM Technologies



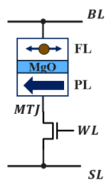
Intel: 750GB, SSD



Emerging NVMs

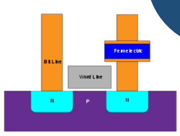
PCRAM — Main Memory

STTRAM — Cache

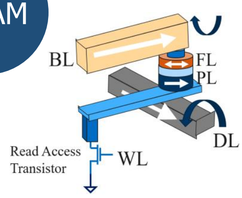


Everspin: 1Gb, DDR4

FeRAM/ FRAM — Main Memory

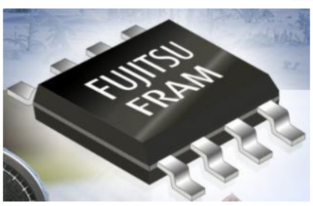


MRAM — Cache

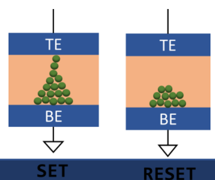


Samsung: 8Mb, Standalone

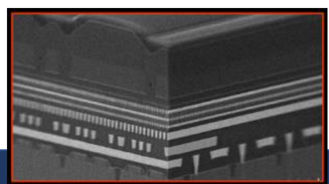
RRAM — Main Memory /*Cache



Fujitsu: 2Mb, Standalone



W. J. Lee et al. Nature Material, 2011

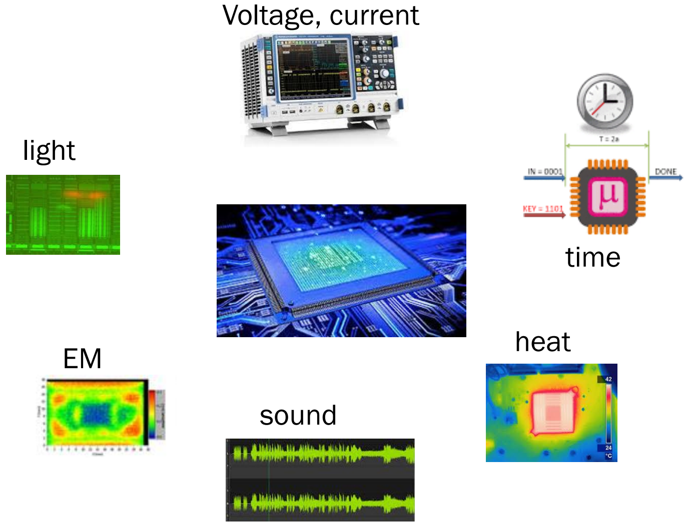


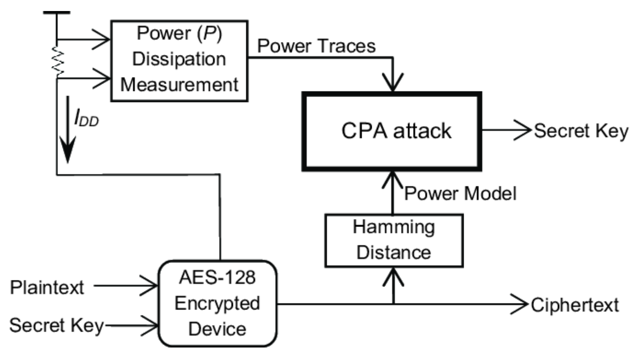
WD: 32Gb, SCM

Outline

- Basics of hardware security
- CMOS security primitives
- Emerging technologies
- Application in security
- Security vulnerabilities and defenses
- Conclusions

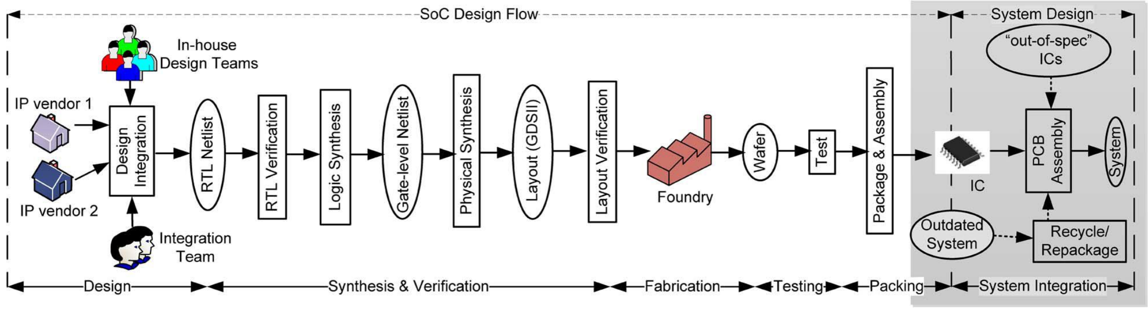
Side Channel Attacks





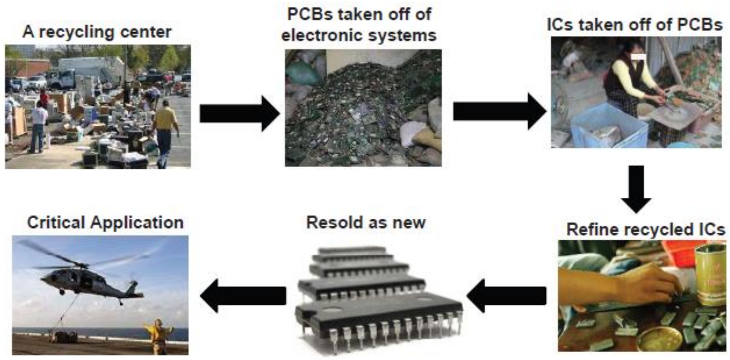
- Possible sources: electrical, ambient, acoustical, temporal…
- Objective: extract valuable information

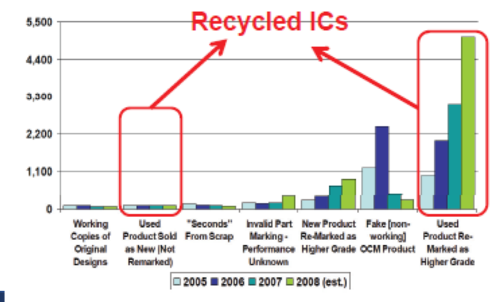
Semiconductor Supply Chain



- Profit driven business model that relies on outsourcing
  - Security vulnerabilities present at many stages of design and manufacturing process
- Attacks
  - Counterfeiting
  - Hardware Trojan Horses
  - Cloning
  - Overproduction
  - Reverse engineering
  - Non-invasive tampering

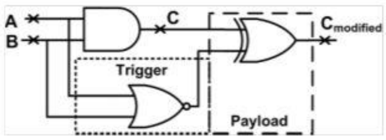
IC Recycling/Counterfeiting



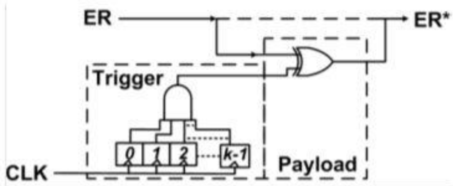


Hardware Trojans

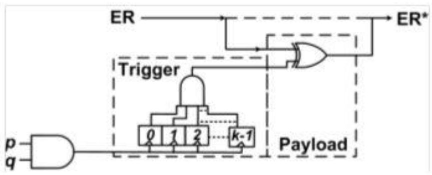
- Undesirable/ unintended design features to
  - Bypass security features
  - Bypasses convention test methods
  - Triggers in-field failures



(a) Combinationally triggered Trojan

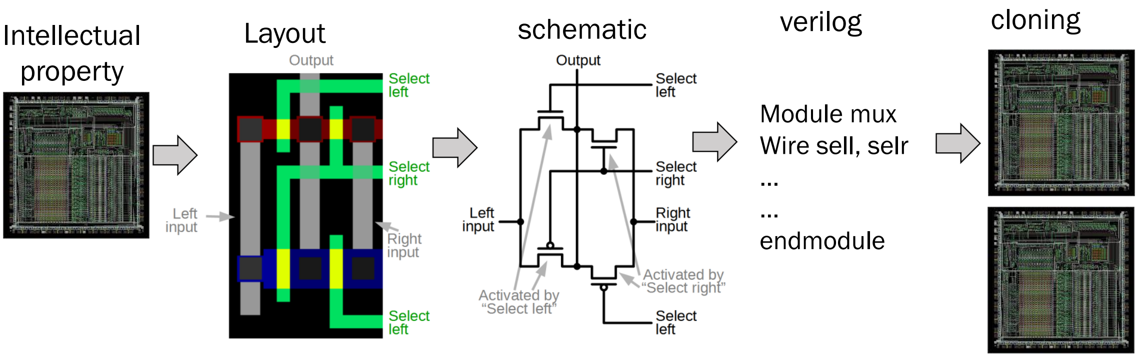


(b) Synchronous counter ("time-bomb") Trojan



(c) Asynchronous counter Trojan

Reverse Engineering and Cloning



- Delayering of chip, identification of gates and their connectivity information, and, reconstruction of netlist
  - Goals: competitive analysis, cloning and overproduction, siphoning profit

Non-Invasive Tampering

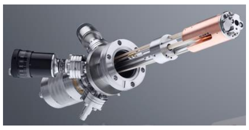
X-ray gun

magnet

laser

UV light

heat



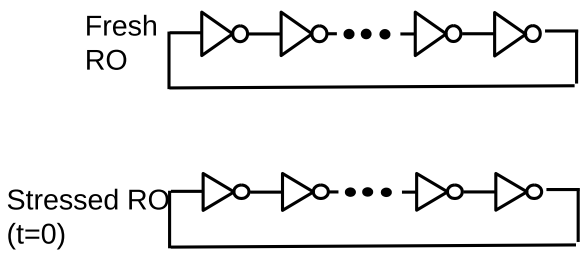
Ion gun

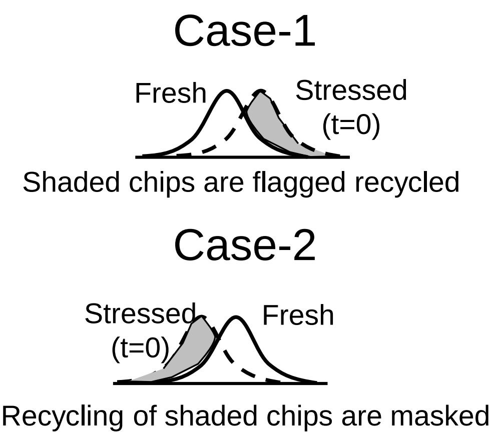
- Objective is to
  - Bypass security features
  - Launch denial-of-service attack
  - Extract valuable information

Outline

- Basics of hardware security
- CMOS security primitives
- Emerging technologies
- Application in security
- Security vulnerabilities and defenses
- Conclusions

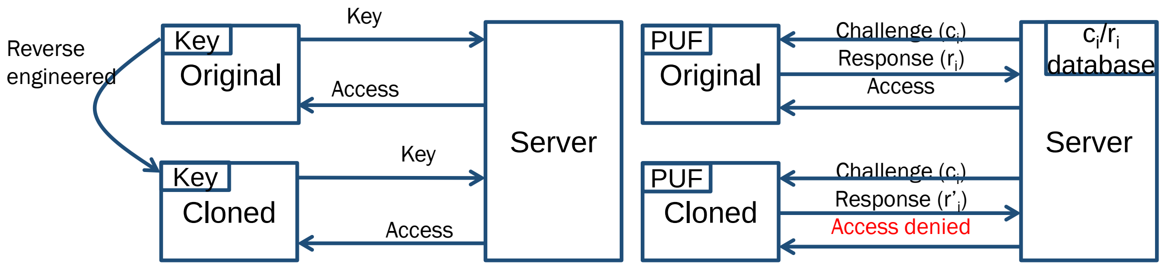
Recycling Sensor

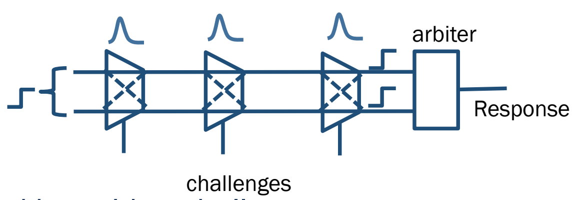




- Aged RO is compared with fresh RO
- Challenges
  - Process variation results in wrong decision or masking
  - Limited by aging of RO and delay sensitivity of RO on aging
- Prevents recycling/counterfeit ICs

Physically Unclonable Functions

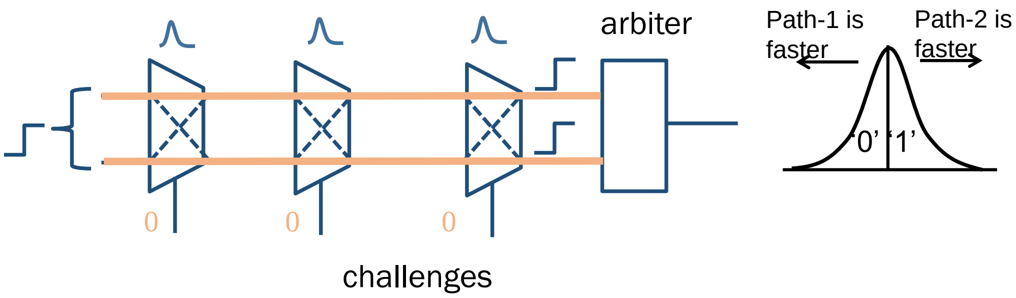




Process variation results in unique die-to-die response

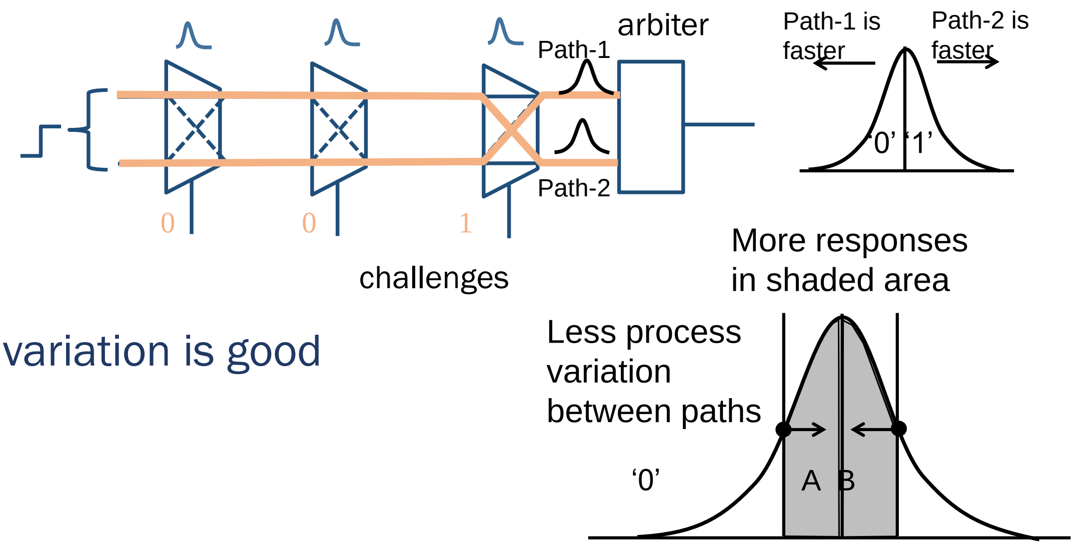
- PUF replaces the hardcoded key with a challenge response system
  - Response is generated from physical properties of the chip
  - Cannot be cloned
- Prevents cloning, counterfeit IC

Physically Unclonable Functions



- Different chips produce different responses for same challenge

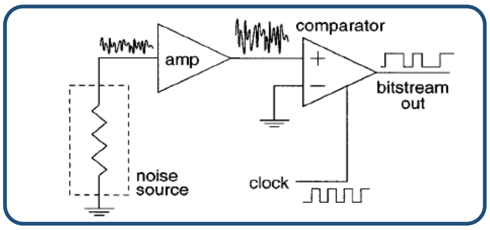
Physically Unclonable Functions



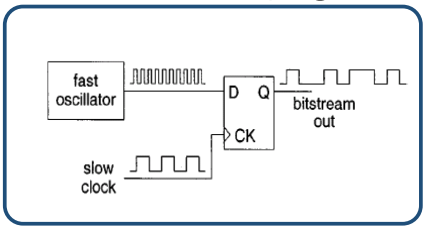
- Process variation is good

True Random Number Generator

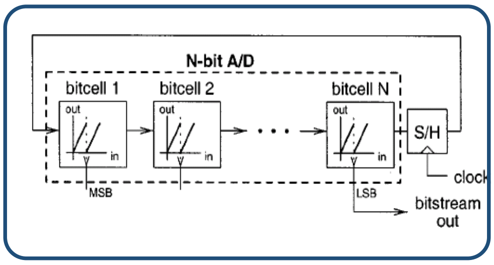
Direct amplification



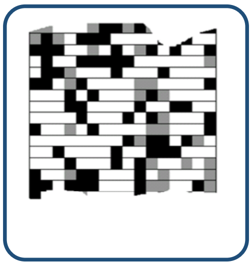
Oscillator sampling



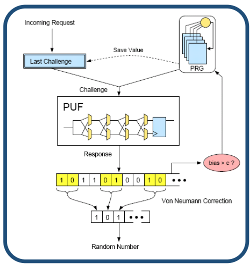
Discrete time chaos



SRAM PUF



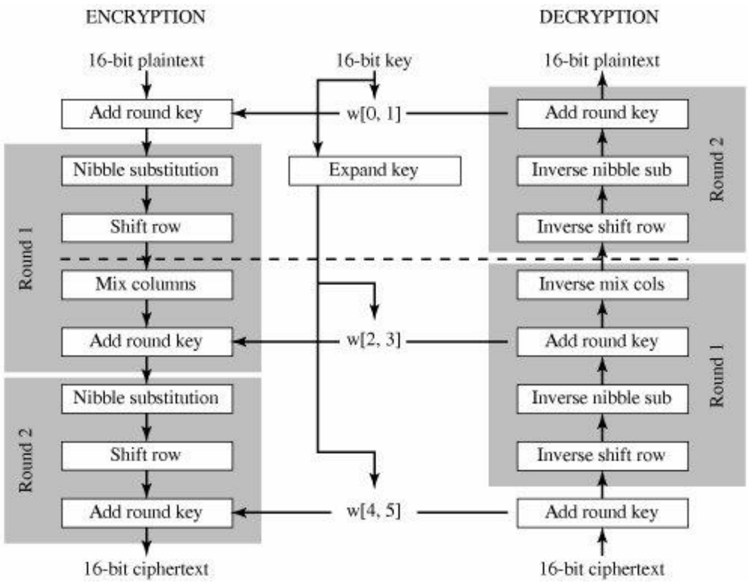
PUF



- Key generation and seed generation for authentication, secure communication

Encryption Engines

- Ensures privacy in communication
- Requires extensive shift, XOR and addition operation
  - Prone to side channel attack



Hardware Security Primitives- Key Requirements

Security primitive	Key Requirements
Recycling sensor	Low process variation, high sensitivity to usage
PUF	High process variation, nonlinearity
TRNG	High entropy
Encryption	Recursive shift, multiplication, addition
Miscellaneous	Sensitivity to ambient parameters

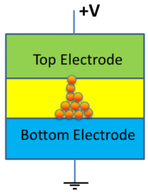
Outline

- Basics of hardware security
- CMOS security primitives
- Emerging technologies
- Application in security
- Security vulnerabilities and defenses
- Conclusions

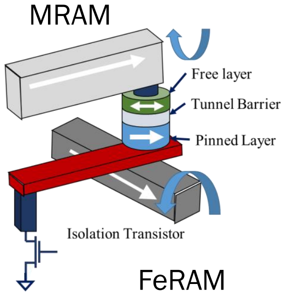
Emerging Technologies

- Opportunities
  - Non-volatility, electroforming, asymmetric read/write, retention, magnetization noise, stochastic resistance, non-linearity, random DW dynamics…
- Challenges
  - Vulnerabilities
- Need deeper understanding for right application

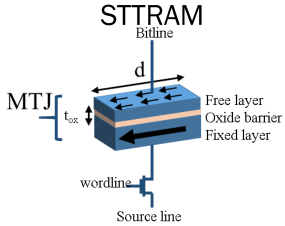
ReRAM



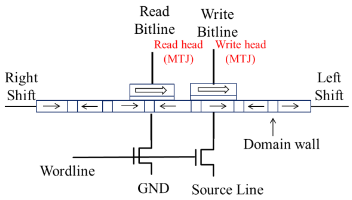
MRAM



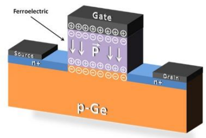
STTRAM



DWM



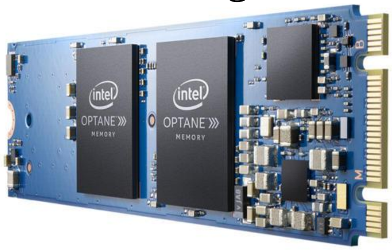
FeRAM



PCRAM

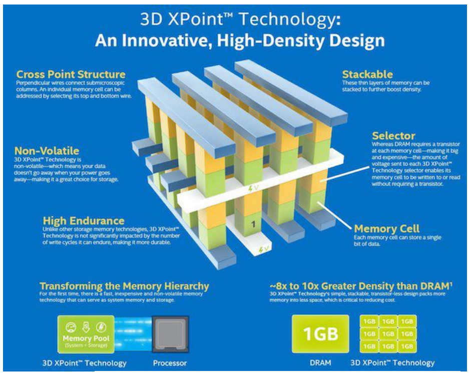
Recent Commercialization of Emerging NVMs

Phase Change RAM*

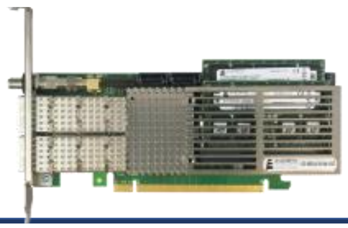


Intel unveils its Optane hyperfast memory

Intel released few key details around its new non-volatile memory



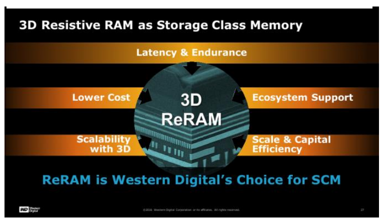
STT- MRAM



Published: *March 9, 2017*

Everspin unveils a new low latency, PCIe NVMe card based on Spin Torque MRAM

ReRAM

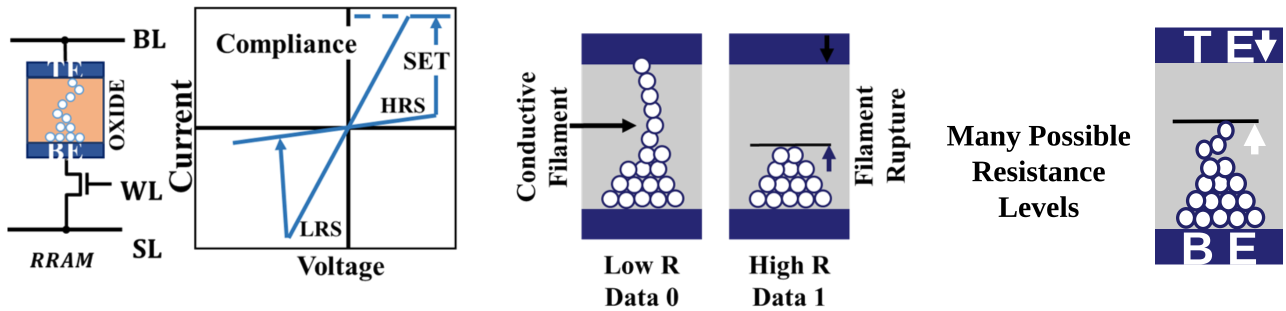


Western Digital to Use 3D ReRAM as Storage Class Memory for Special-Purpose SSDs

by Anton Shilov on August 12, 2016 8:00 AM EST

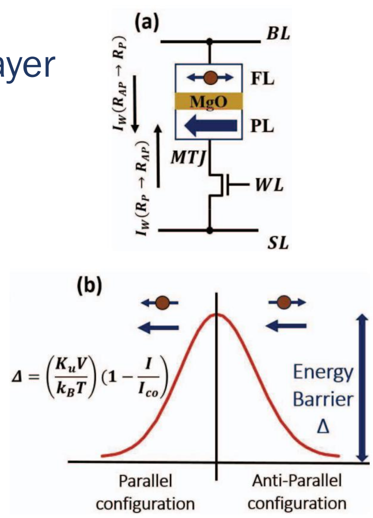
NVM: Resistive RAM (ReRAM)

- ReRAM Features
  - Bits stored as resistance state
  - Low R ⟶ Data "0", High R ⟶ Data "1"
  - Possible Oxides: HfO_2, TiO_2, TaO_x,WO_x
- Offers lowest footprint ($4F^2$ for xpoint)



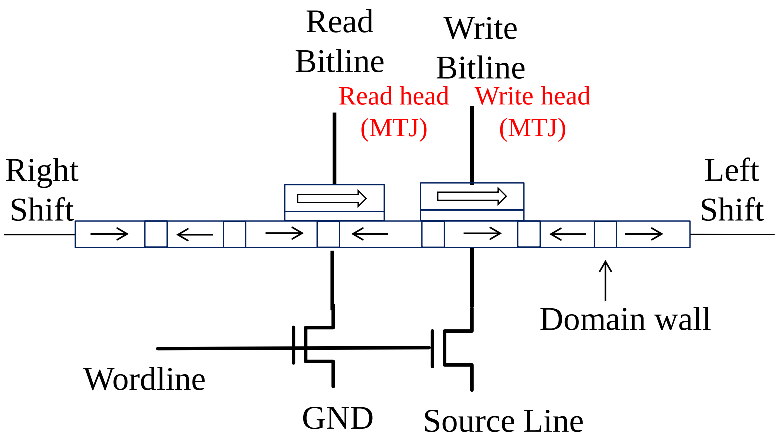
NVM: Spin Torque Transfer RAM (STTRAM)

- STTRAM Features
  - Magnetic Tunnel Junction (MTJ) as Storage element
  - MTJ consists of free (FL) and pinned (PL) magnetic layer
  - Bits stored as resistance state
  - Magnetic Orientation
    - **Data "0"**: Parallel (Low resistance)
    - **Data "1"**: Anti-parallel (High resistance)



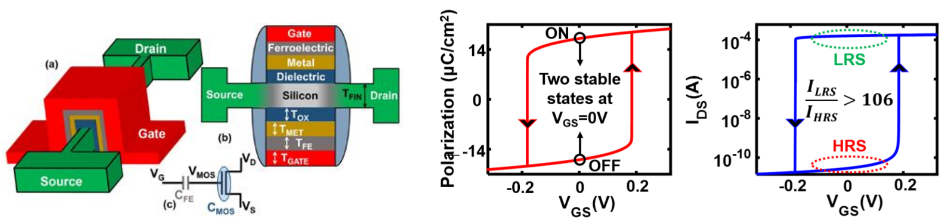
Domain Wall Memory

- DWM Features
  - Three components: Read MTJ, Write MTJ, Nanowire
  - Bits are stored in nanowire that acts like a shift register
  - Access mechanism is serial



NVM: Ferroelectric FET (FeFET)

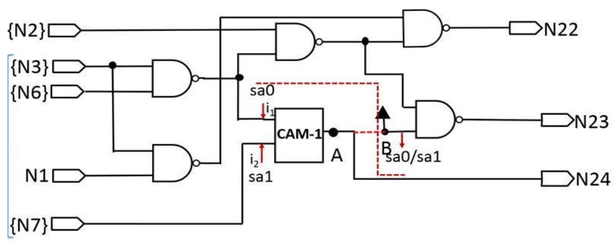
- FeFET features
  - Ferroelectric (FE) layer between metal gate and dielectric layer
  - Stores data as polarization state (+ve or -ve) of FE layer
  - Inherent 3-terminal structure allows isolation of read and write ports
  - If +ve V_{GS} > gate critical voltage ⟶ polarization switches to positive

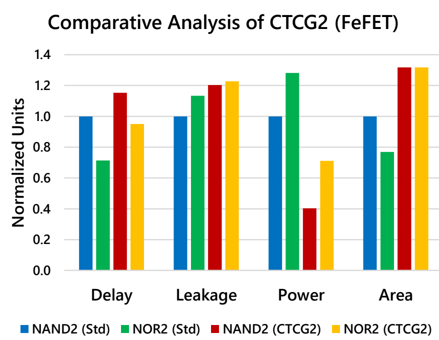


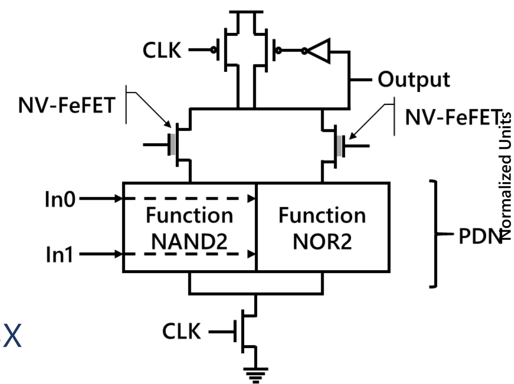
Outline

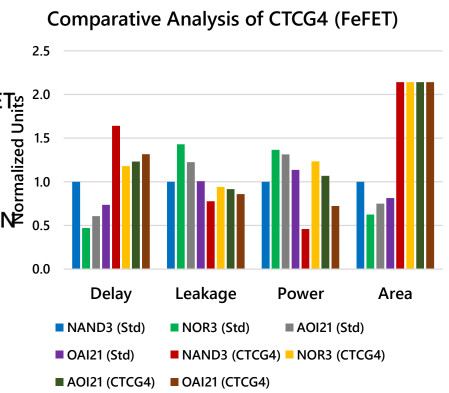
- Basics of hardware security
- CMOS security primitives
- Emerging technologies
- Application in security
- Security vulnerabilities and defenses
- Conclusions

Exploiting Persistence-Obfuscation







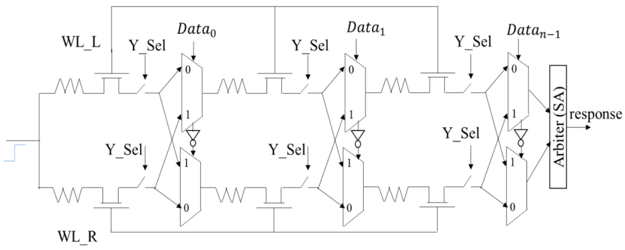


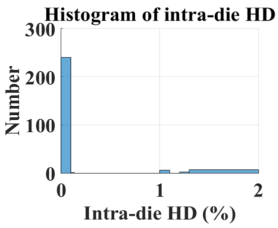
- Average delay overhead: 1.7X
- Average leakage overhead: 0.9X
- Average total power overhead: 0.6X
- Average area overhead: 2.3X

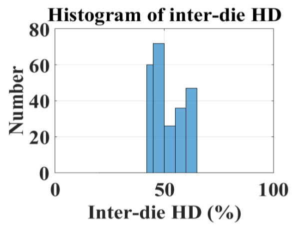
A. De, HOST'18

Exploiting Variations- Physically Unclonable Functions

- Design
  - PV in emerging NVM
  - RRAM based design using RC signal delay to generate PUF response

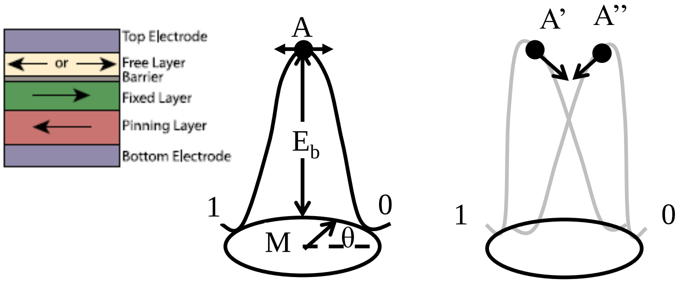






MRAM PUF

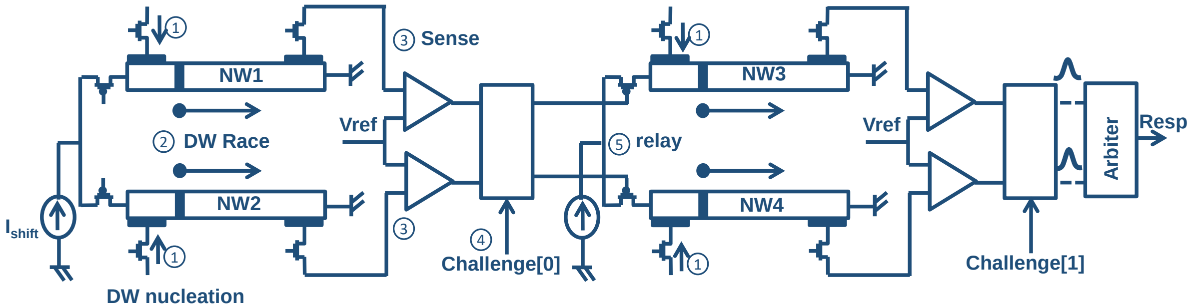
- Employs random initialization of the MTJ due to physical variations in the MTJ
- Variations create random tilt of energy barrier
- MTJ free layer is prone to prefer certain initial orientation much similar to SRAM PUF
- Intra-die HD of 0.0225 and an entropy of 0.99
- Decreasing the aspect ratio at constant volume and increasing the volume at constant aspect ratio is proposed to increase the tilt angle variation and enhance the stability of the PUF



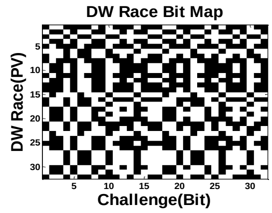
PUF Types	$\mathcal{D}_{intra}$	$\mathcal{D}_{inter}$	$\rho(Y^n) \leq$	area (μm^2)
SRAM	0.078	0.49	0.94	51.99
Latch	0.26	0.3	0.71	531.25
D flip-flop	0.19	0.39	0.81	765.63
Arbiter	0.07	0.46	0.5-0.9	690.56
Ring Oscillator	0.099	0.46	0.86	7774.2
Memristor *	-	$\simeq 0.5$	-	-
STT-PUF *	$\sim$ 10e-6	$\simeq 0.5$	0.985	6.79
MRAM	0.0225	0.47	0.99	6.74

Das, Jayita, Kevin Scott, Srinath Rajaram, Drew Burgett, and Sanjukta Bhanja. "MRAM PUF: A Novel Geometry Based Magnetic PUF With Integrated CMOS." (2015).

DWM-Relay PUF



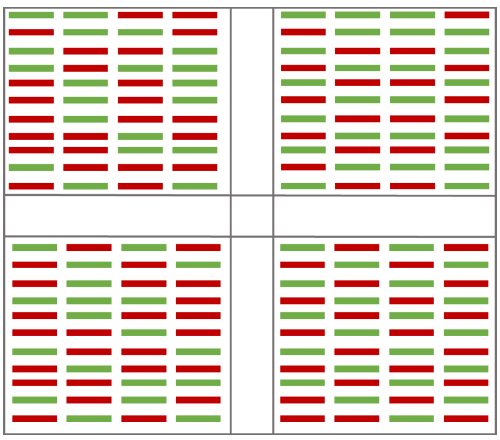
- Operation of relay-PUF
    - DW nucleation
    - Race
    - Sense
    - Relay
    - Race…
- Variation in DW velocity due to variation is exploited
- Hamming distance=50%

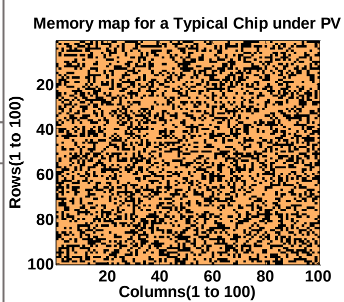


Anirudh Iyengar, Swaroop Ghosh and Kenneth Ramclam, "Domain wall magnet for embedded memory and hardware security", Special Issue of Journal of Emerging Topics on Circuits and Systems (JETCAS Special Issue), 2015

Memory PUF

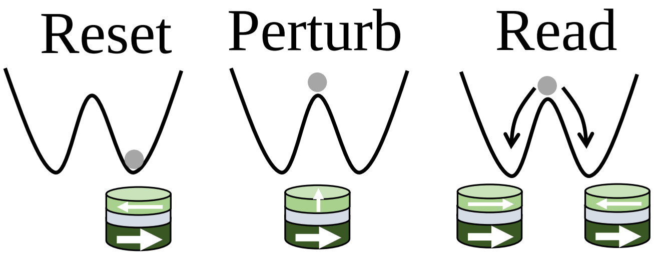
- DW is raced in memory array (pinned: '0' (red), remaining: '1' (green))
- Uneven number of '0's and '1's at high voltage
- Temperature variation changes response
- Hamming distance is ~44%





Spintronic TRNG

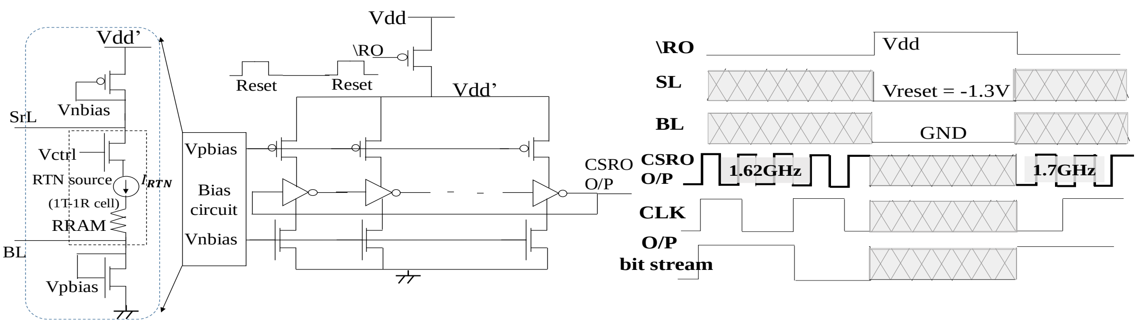
- Key ideas:
  - Reset the MTJ to AP state
  - Excite the free layer of the MTJ to the bifurcation point by applying a current pulse
  - Magnetization settle in random state due to thermal noise
  - To improve randomness and kill correlation bits are XOR'ed with each other
- Reset pulse is detrimental to MTJ reliability
- Sharing of reset and sense circuit makes sense MTJ susceptible to read disturb



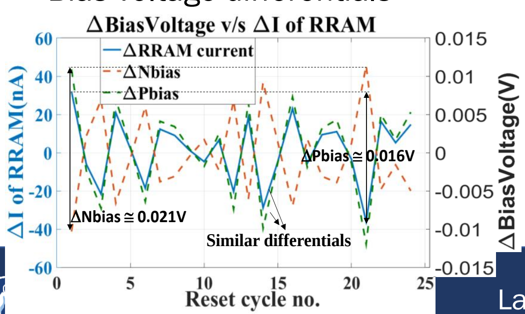
Choi, Won Ho, L. V. Yang, Jongyeon Kim, Abhishek Deshpande, Gyuseong Kang, Jian-ping Wang, and Chris H. Kim. "A Magnetic Tunnel Junction based True Random Number Generator with conditional perturb and real-time output probability tracking." In Electron Devices Meeting (IEDM), 2014 IEEE International, pp. 12-5. IEEE, 2014.

RRAM TRNG

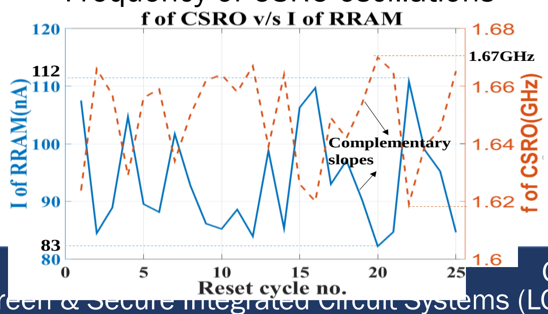
- RTN to generate random bit strings



Bias voltage differentials



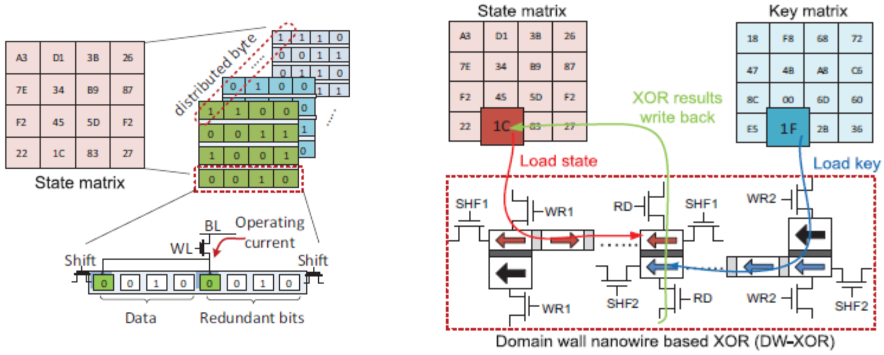
Frequency of CSRO oscillations



Govindaraj, TVLSI'18

Spintronic Encryption Engine

- SubByte: The DW-based Look-Up Table (LUT) is used to save leakage power
- ShiftRows: To mimic cyclic rotation in nanowire, redundant bits are employed in DW nanowire
- MixColumns: multiplication by shift and addition. For addition domain wall XOR gate is employed
- AddRoundKey: This step XORs the SM with the round key



] Wang, Yuhao, Hao Yu, Dennis Sylvester, and Pingfan Kong. "Energy efficient in-memory aes encryption based on nonvolatile domain-wall nanowire." In Design, Automation and Test in Europe Conference and Exhibition (DATE), 2014, pp. 1-4. IEEE, 2014.

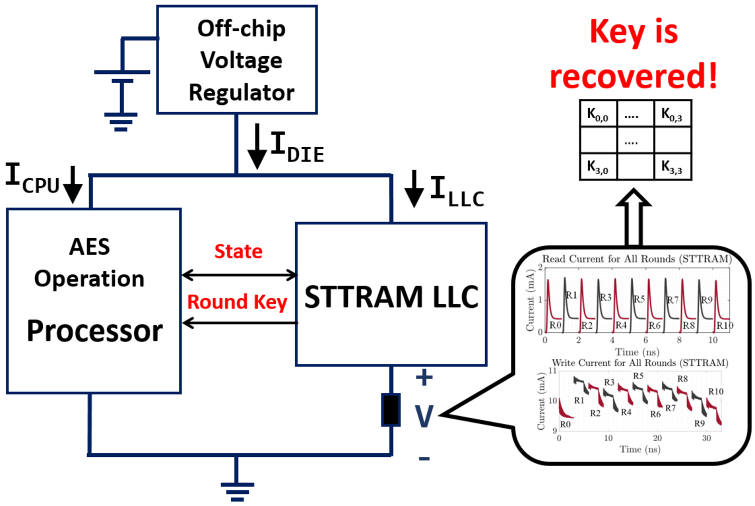
Outline

- Basics of hardware security
- CMOS security primitives
- Emerging technologies
- Application in security
- Security vulnerabilities and defenses
- Conclusions

Side Channel Attack & System Level Overview

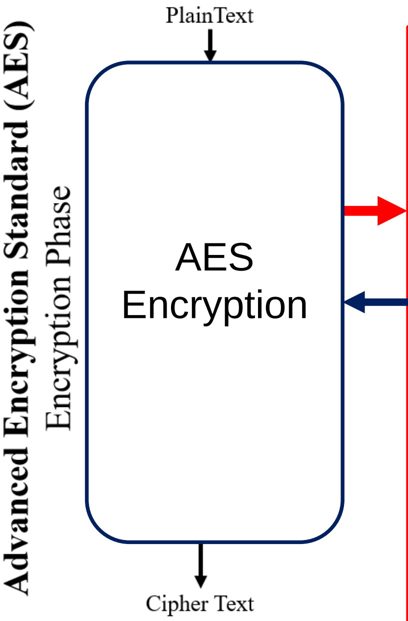
Side Channel Attack (SCA)

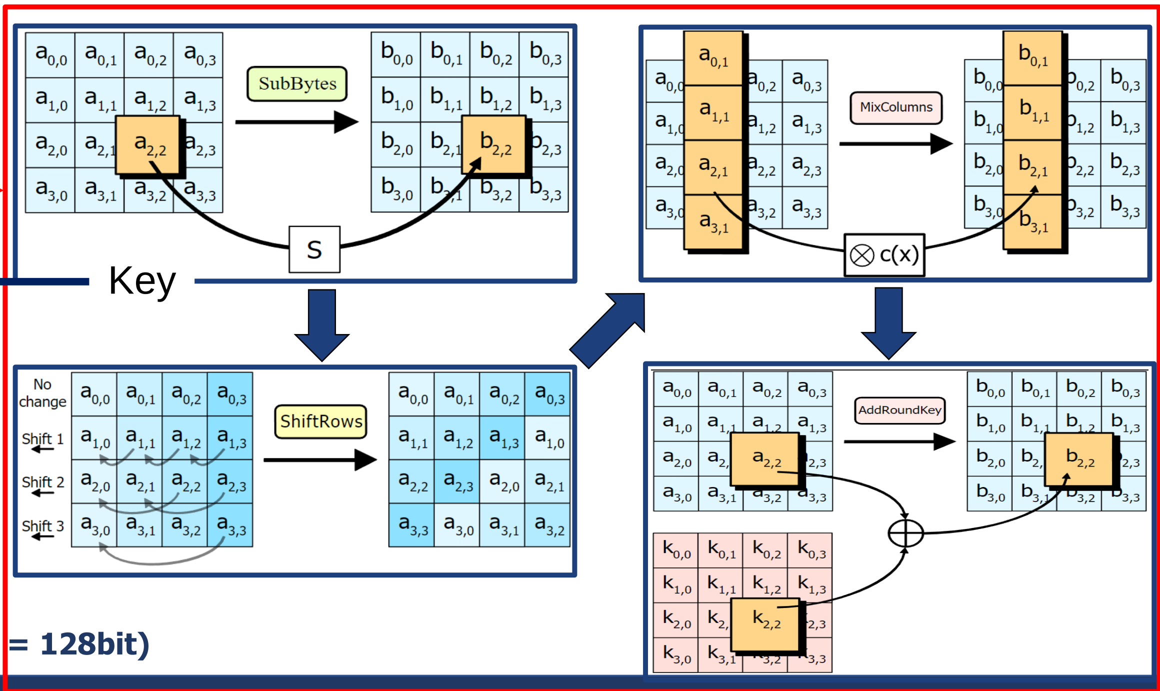
- Powerful physical attack
- Attacks the weakness of the physical implementation of a crypto-algorithm
- **Implements divide and conquer approach**
  - Let's say, the key is 128bit.
  - The possible cases are 2^{128}
  - SCA attacks one byte at a time
  - **Complexity reduces to $16 \times 2^8 = 2^{12}$**



- Attack Model:
  - Attacker has physical access to the system and can measure the power drawn by the system

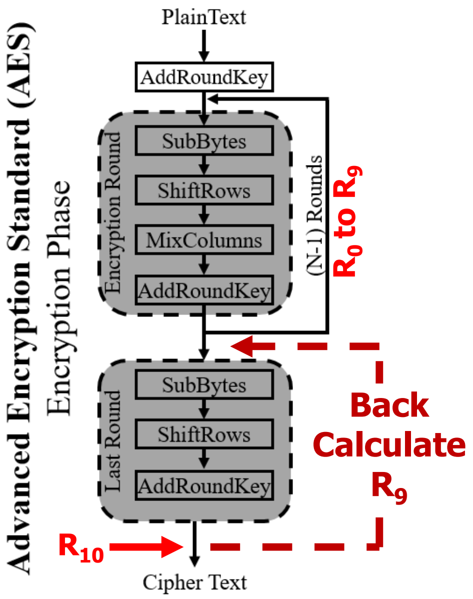
Advanced Encryption Standard (AES)

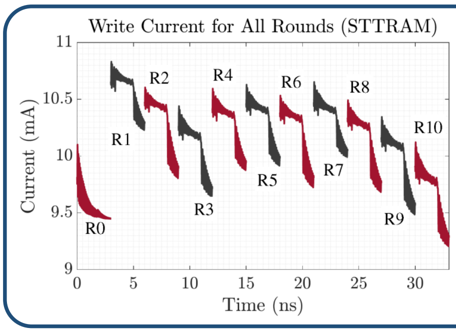


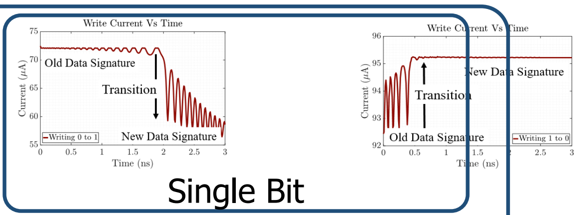


- **AES-128 (Key = 128bit)**
  - → 10 Rounds

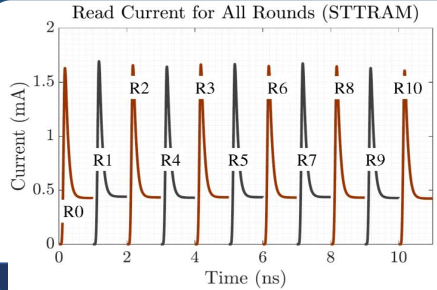
Write/Read Current of All Rounds of AES







- Falling transition prevails



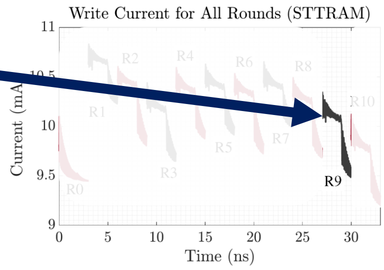
Peak value of read current depends on $\frac{Number\ of\ 0's}{Length\ of\ the\ data}$

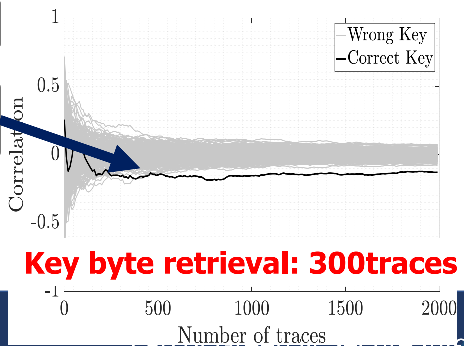
- Naming convention:
  - R9 means Round 9

Attack Steps and Result

Attack Steps (Retrieve one byte at a time)

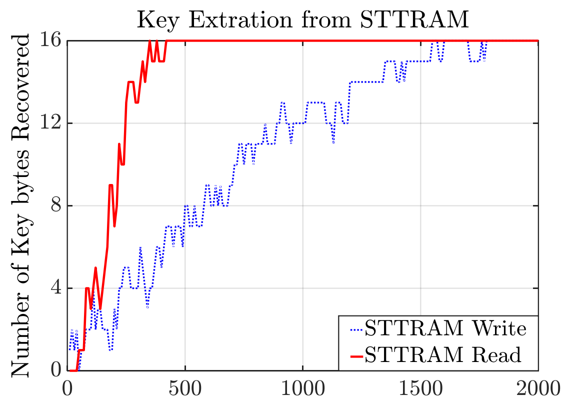
1. Capture Actual R_9 Wave
2. Get Ciphertext (C) (Public)
3. Back Calc. **256** R_9s from C (Attack 1 byte = 2^8 = 256)
4. Correlate Step 1 with Step 3
5. Attack successful: If one correlation stands out





Key byte retrieval: 300traces

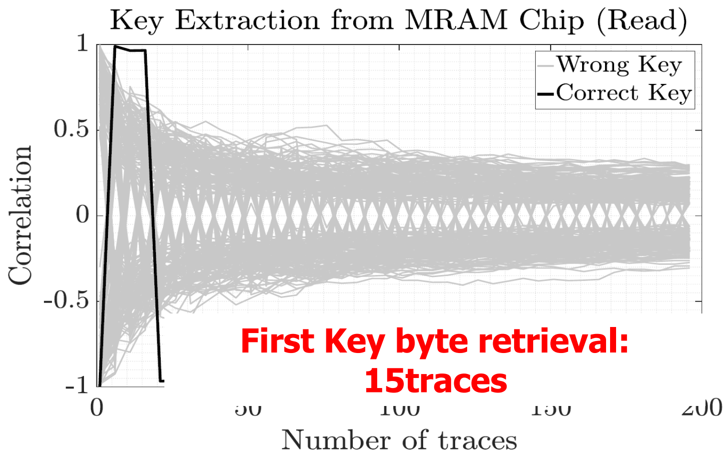
Attack Result



Key = 128bits = 16bytes

STTRAM read is more susceptible

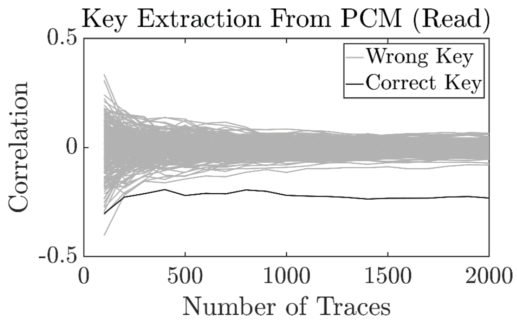
Correlation Analysis

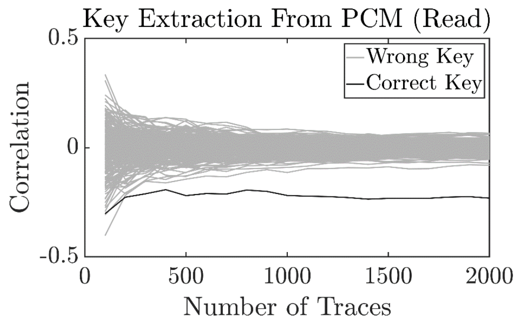


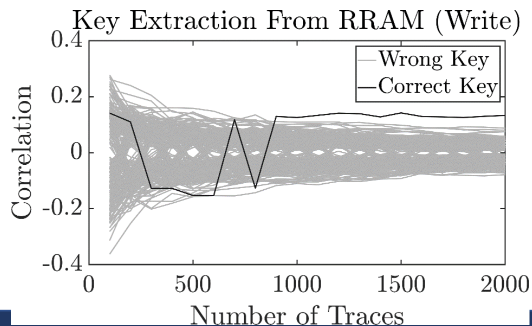
Leaks the first byte of the key in just 15traces!

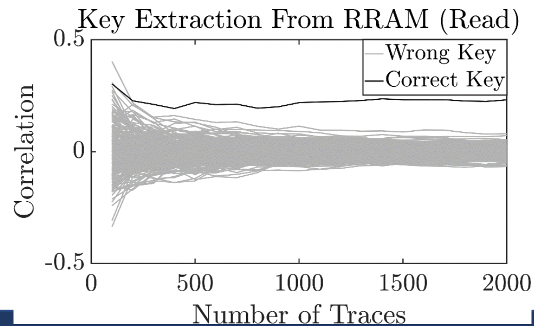
*Published in ICCD, 2017

Side Channel Vulnerability of Other NVMs

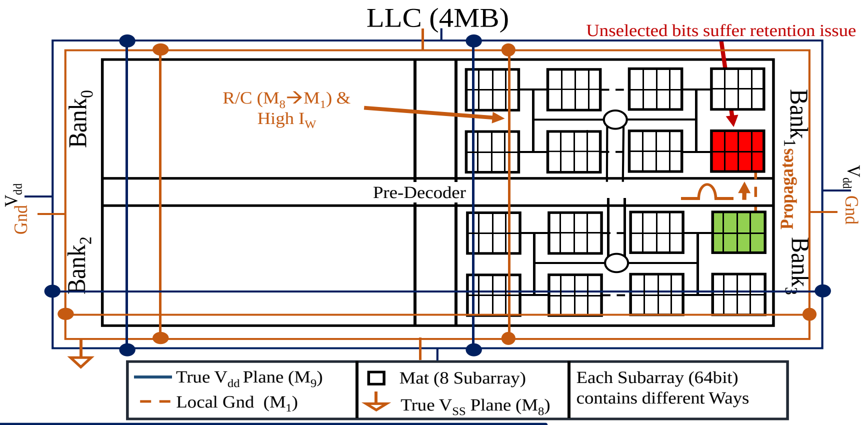


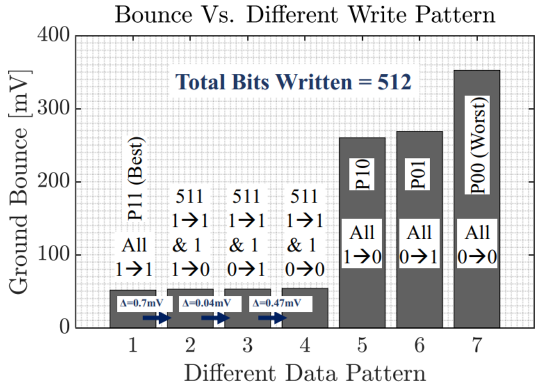






NVM Issues- Supply Noise





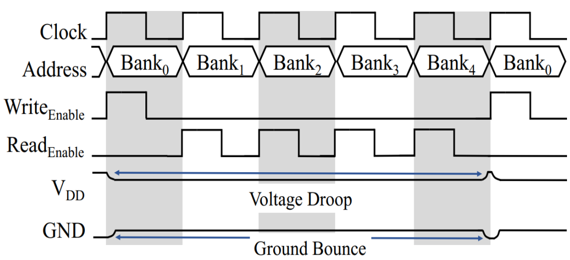
Attack Model:

- Server kind of setting
- Two users: Adversary and Victim

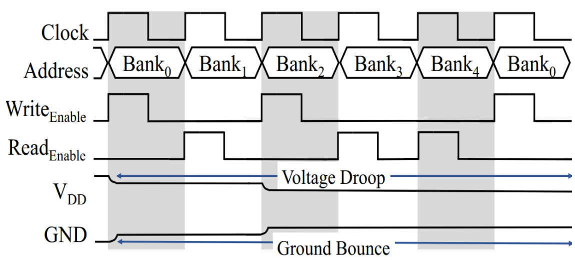
- Noise depends on the data pattern
- Noise can propagate to other memory bank and affecting parallel operation

Supply Noise can be leveraged to launch fault injection/information leakage attack (HASP'18, ISLPED'18)

Parallel Accesses



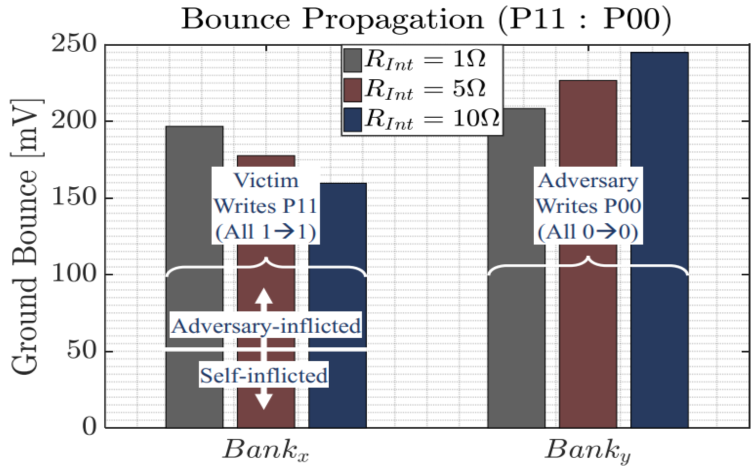
1X Write

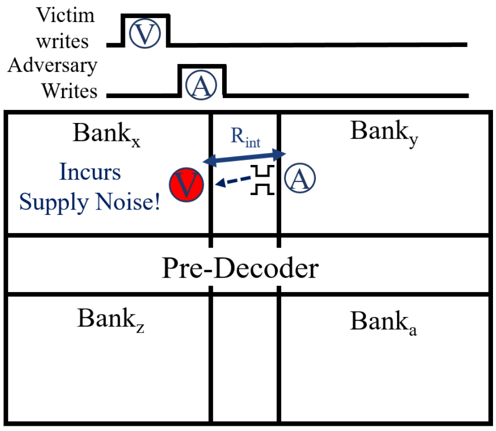


2X Write

- Read/write takes multiple clock cycles
- Parallel operations on independent banks
  - Increases throughput
- Worsen supply noise
- Operations can affect each other

Supply Noise Induced Fault Injection

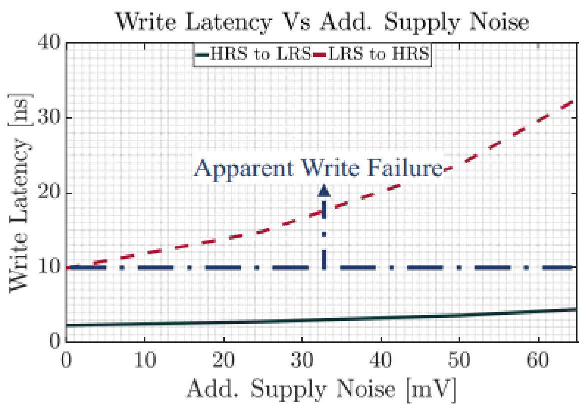


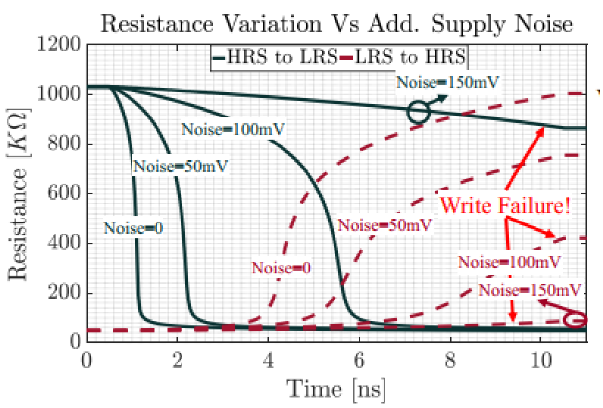


- Victim/adversary writes P11/P00 in $Bank_x$/$Bank_y$ simultaneously
- Victim incurs both
  - Self inflicted bounce
  - Adversary inflicted bounce

Impact of Supply Noise on Write Operation

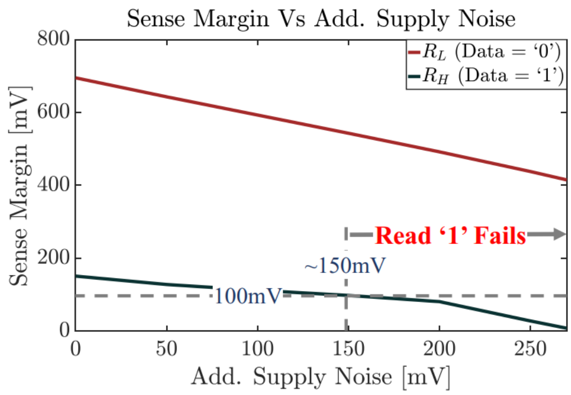
- Supply noise:
  - 0 to 50mV: No failure
  - 50 to 120mV: 0→1 write fails
  - >120mV: both write polarity fails

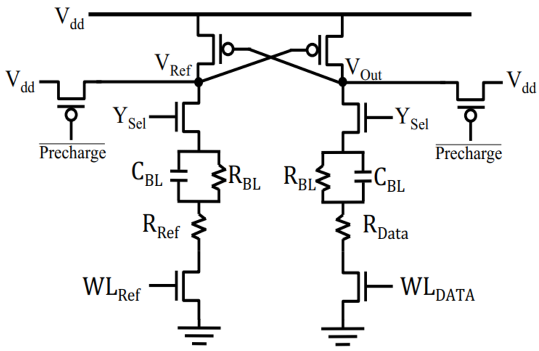




Impact of Supply Noise on Read Operation

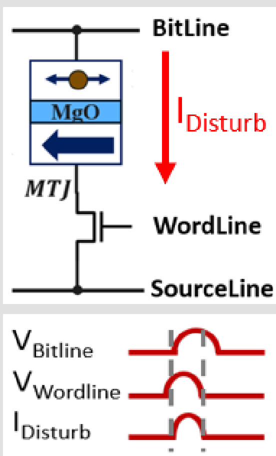
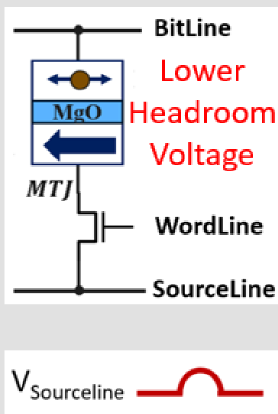
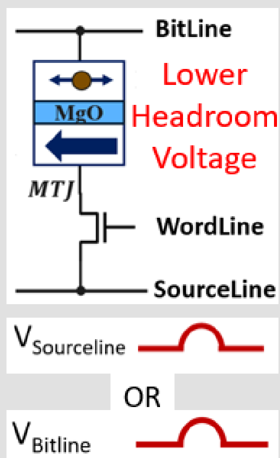
- Supply noise:
  - 0 to 150mV : No failure
  - >150mV : Read ‘1’ Fails





Supply Noise-Induced Row Hammer Attack

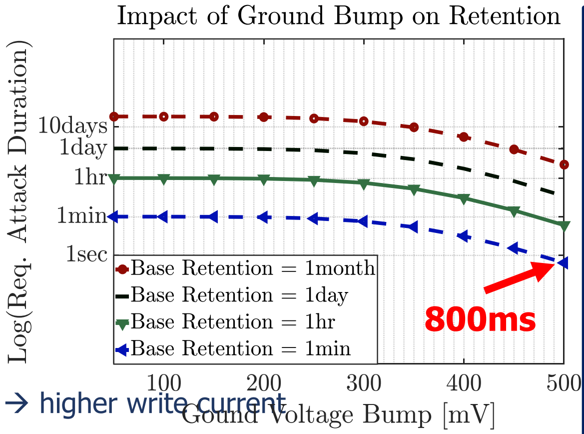
- Retention Failure
  - Data→ Not reliable anymore
- Read Failure
  - 0 is read as 1 or vice versa
- Write failure
  - 1→0 or 0→1 flipping fails

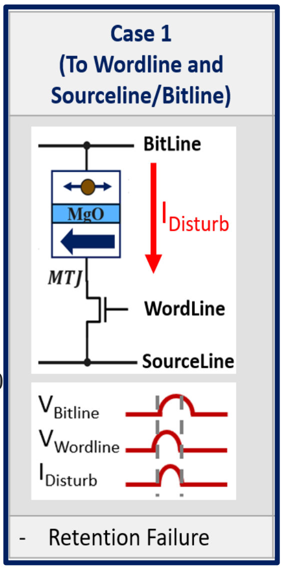
Case 1 (To Wordline and Sourceline/Bitline)	Case 2 (To Sourceline only)	Case 3 (To Sourceline/Bitline)
		
- Retention Failure	- Read Failure	- Write Failure

ISLPED’18, HASP’18, ICCD’18.

Retention Failure/Rowhammer

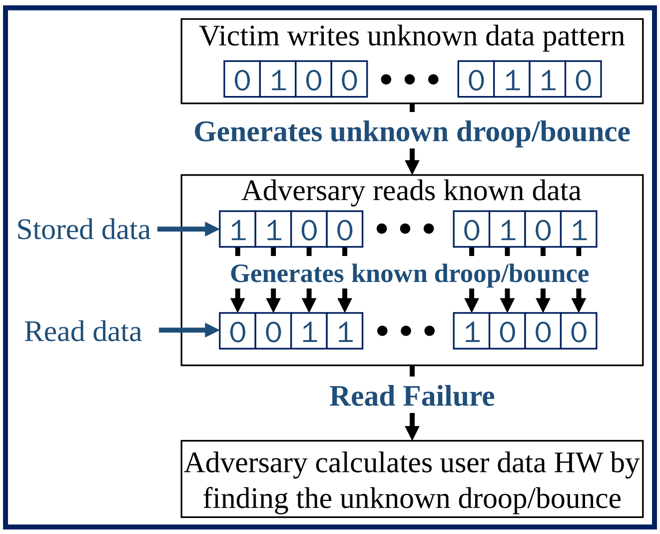
Vol. of FL (cm^3)	Thermal Stability	Base Ret. Time
1.041×10^{-17}	37.99	~1year
0.973×10^{-17}	35.50	~1month
0.845×10^{-17}	32.10	~1day
0.758×10^{-17}	28.95	~1hr
0.681×10^{-17}	24.85	~1min





- Higher Volume → Higher base retention → higher write current
- For LLC, low base retention desired!
- Retention time reduces as ground bounce increases
- Lower Base Retention → lower attack duration

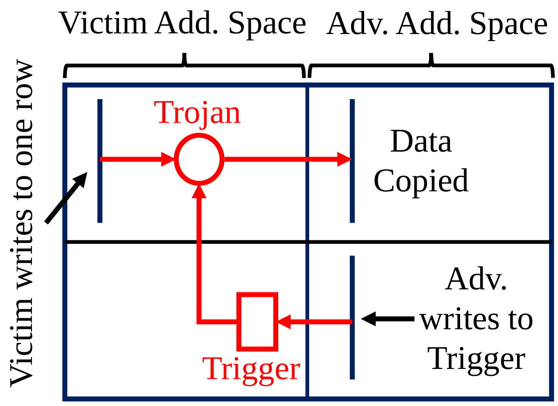
Information Leakage Attacks



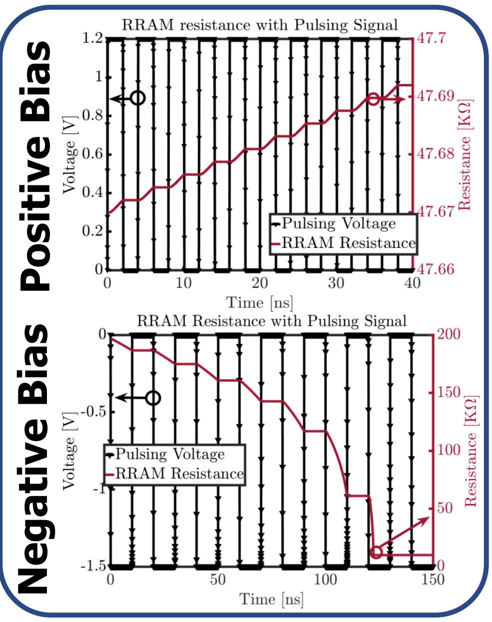
Information Leakage Attack

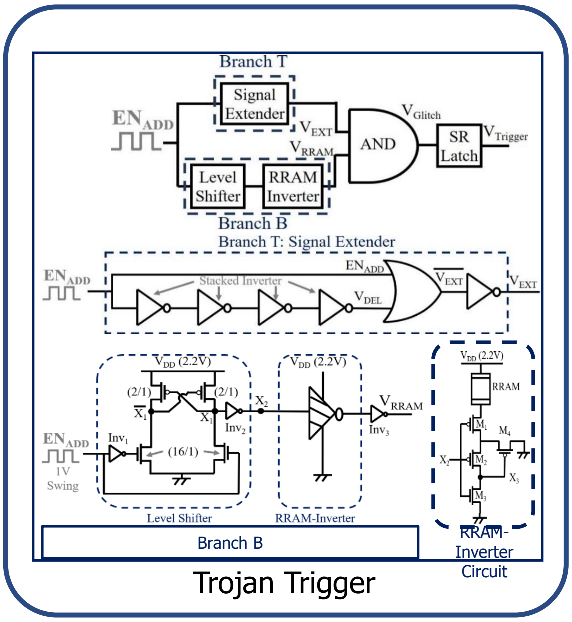
Trojan Attack Model

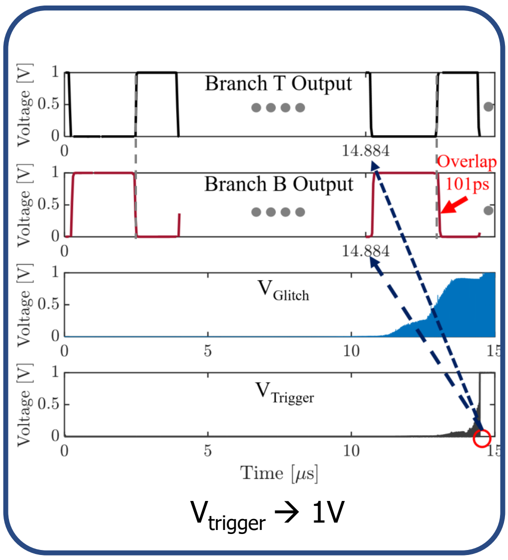
- Adversary adds Trojans (design or fabrication)
- Adversary deploy a program after chip is in the market
  - Program writes one particular L1 cache address N_{tr} times with specific data pattern → Trojan triggers
- Once triggered
  - Fault injected to victim's write/read
  - Victim's data is copied to adversary's address space



NVM Trojan Trigger

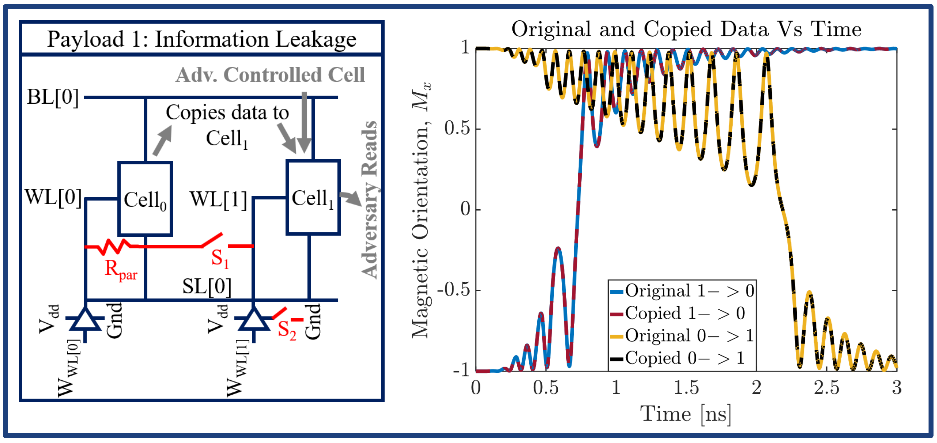


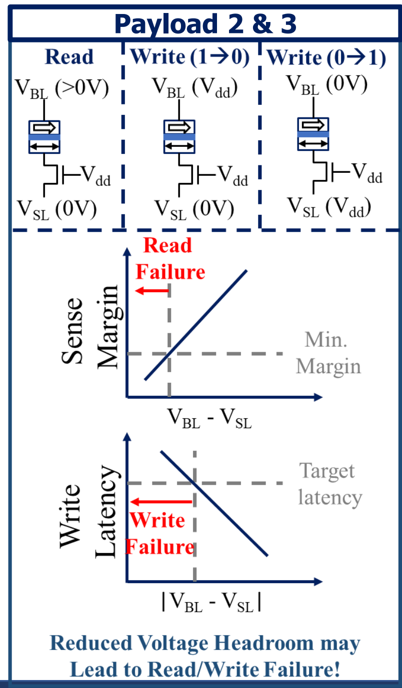




*Published in HOST 2019

NVM Trojan Payload





- Payload 1: Information leakage (copy data from one cell to another)
- Payload 2: Write failure (by injecting supply noise)
- Payload 3: Read failure (by injecting supply noise)

*Published in DATE 2019

Conclusions

- Hardware supply chain presents several new attack surfaces
- Conventional CMOS technologies offer limited randomness, variations and noise sources
- Emerging NVMs possess novel ingredients suitable for security
- We reviewed multiple techniques and their security applications
- We also covered security challenges

Open Research Problems

- Various new flavors of devices
  - SOT-MRAM
  - PMA-MTJ
  - Skyrmionic memory
- Application areas of hardware primitives
  - Data non-repudiation
  - System security issues e.g., buffer overflow
  - Machine learning

Thank you!

- Acknowledgements
  - Collaborators: Anupam Chattopadhyay, Shivam Bhasin, Jongsun Park, Rashmi Jha

NSF SRC Semiconductor Research Corporation DARPA DATA SCIENCE PSU intel