CMOS-Compatible Logic Embedded High-K Charge-Trap Multi-Time-Programmable Memory

Janakiraman Viraraghavan, Derek Leu, **Balaji Jayaraman**, Alberto Cestero, Ming Yin, John Golz, Rajesh R. Tummuru, Ramesh Raghavan, Dan Moy, Thejas Kempanna, Faraz Khan, Toshiaki Kirihata, Subramanian Iyer

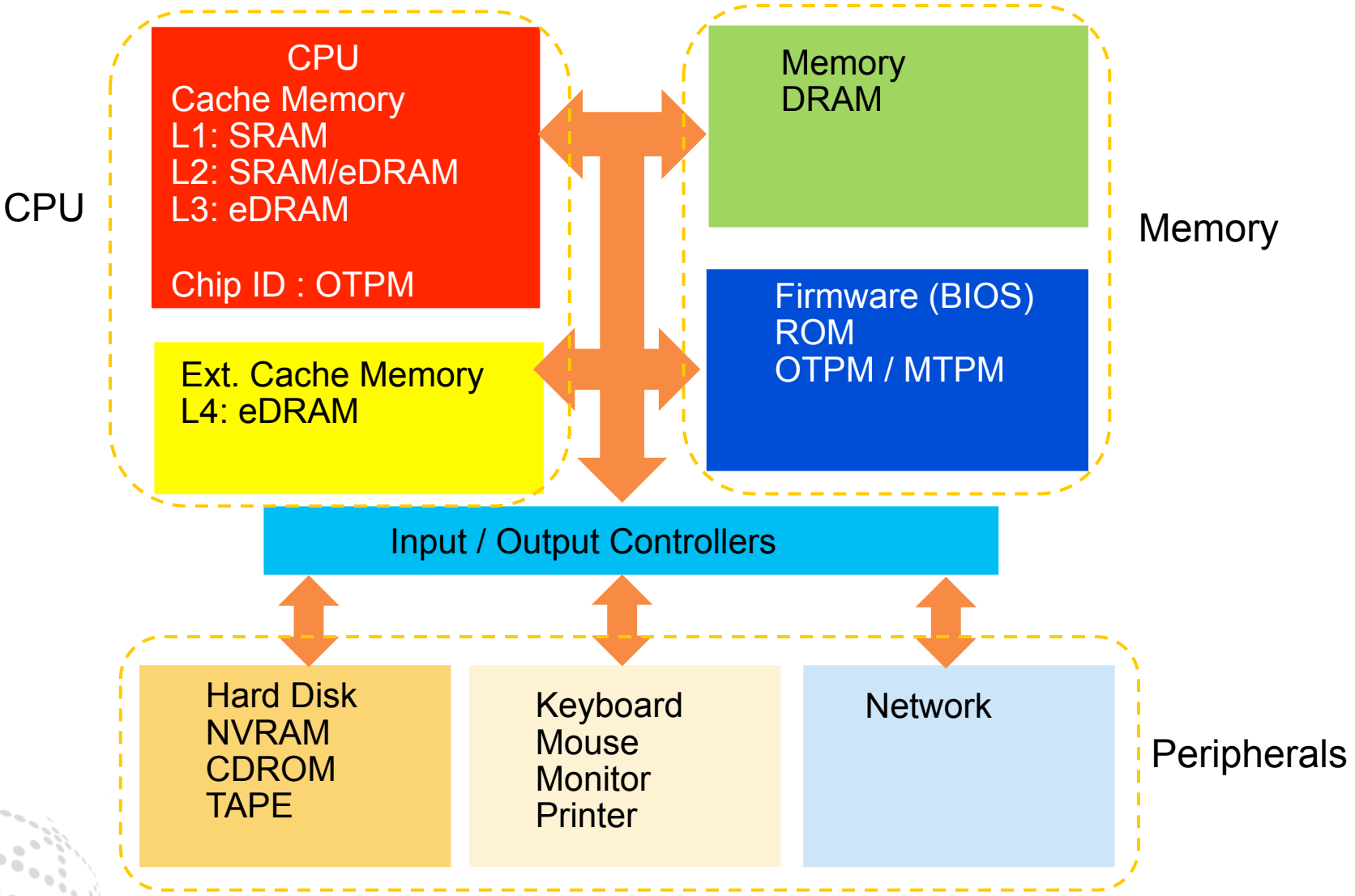
UCLA CHIPS
CENTER FOR HETEROGENEOUS INTEGRATION AND PERFORMANCE SCALING

GLOBALFOUNDRIES®

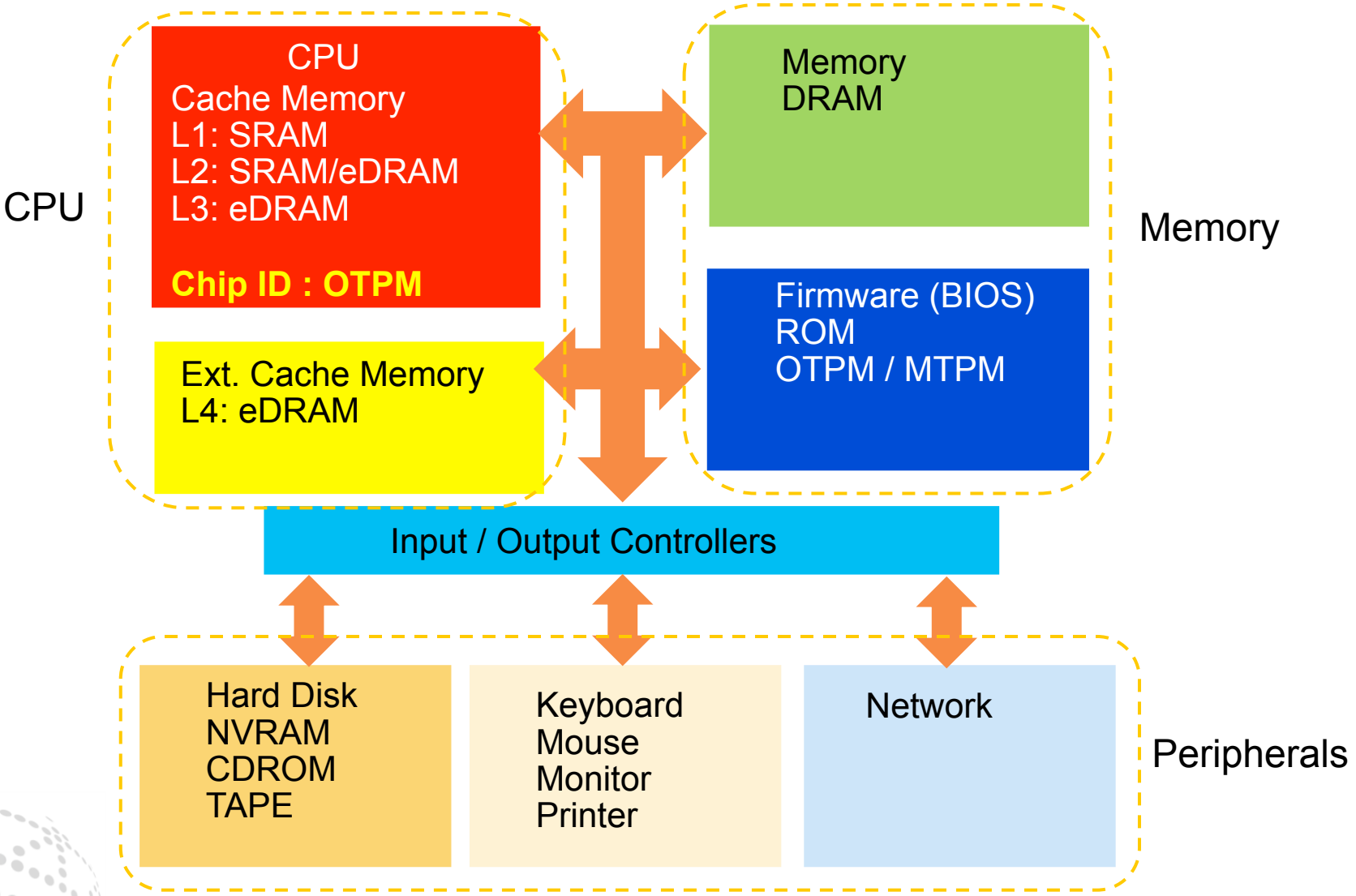
Outline

- Introduction
- Embedded Nonvolatile Memory Applications
- Charge Trap Transistor (CTT) Technology
- Macro Overview
- Hardware Results
  - SOI
  - 14nm Bulk FinFET
- Future scope
- Summary

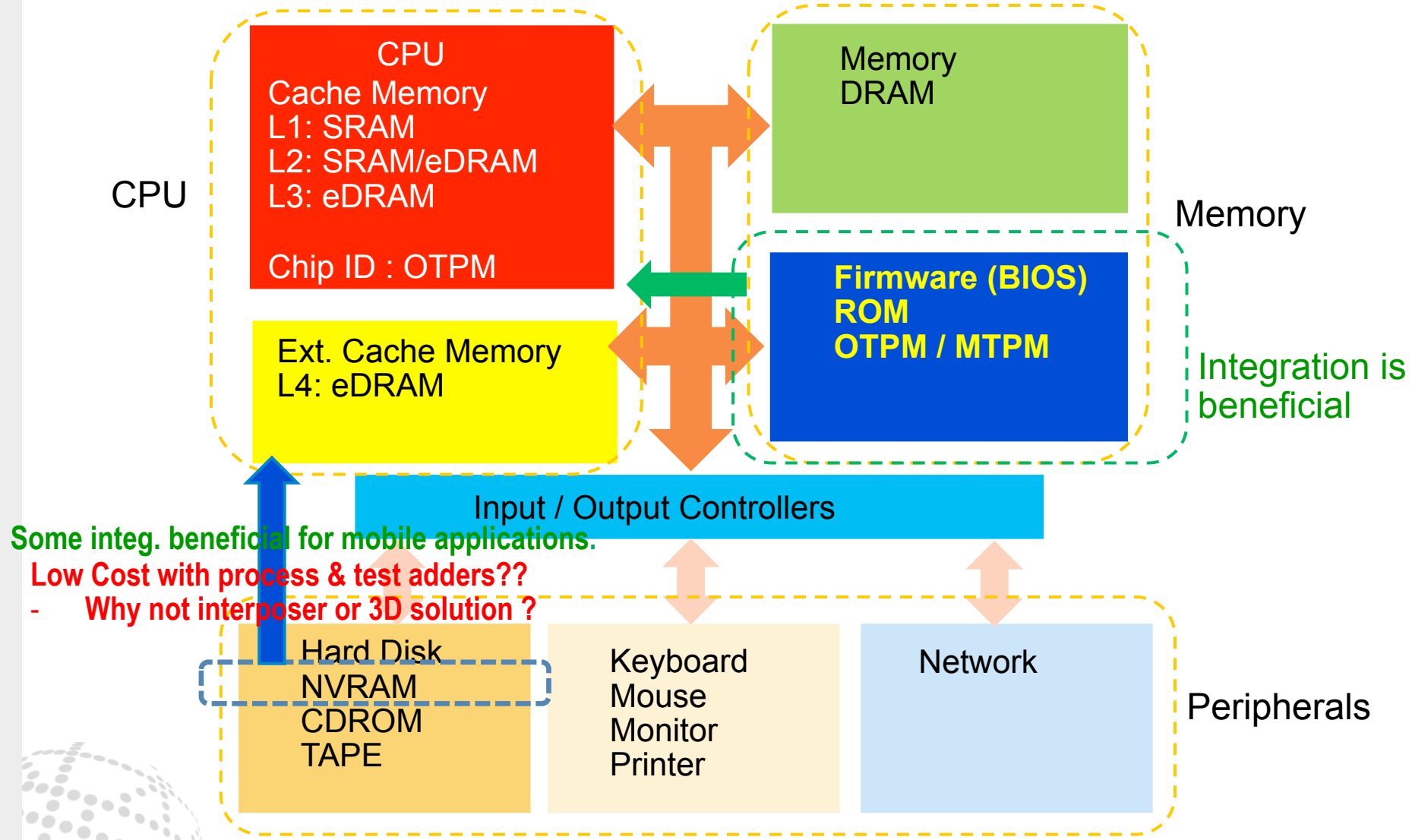
Introduction – Computer Architecture



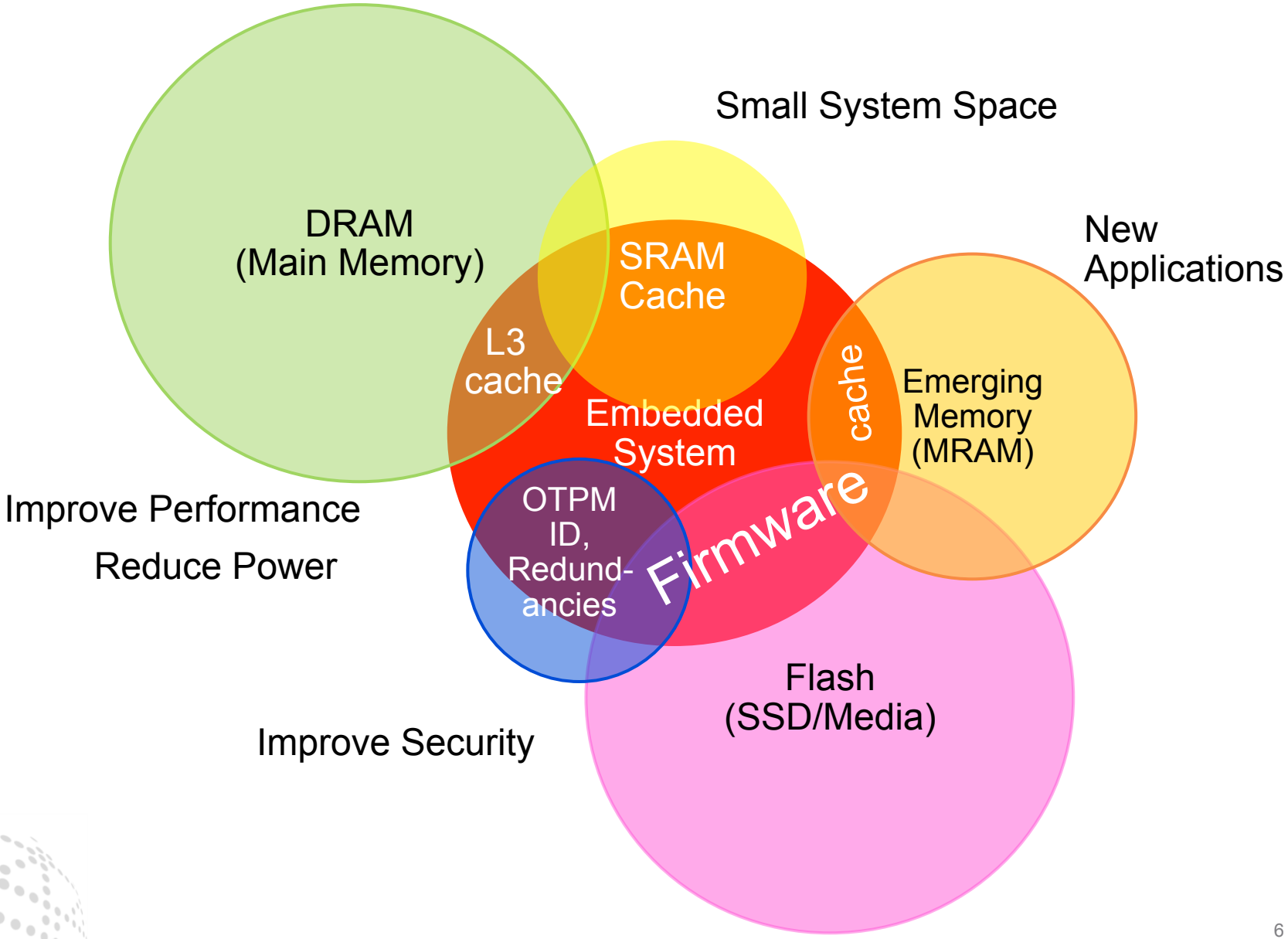
Introduction – Computer Architecture



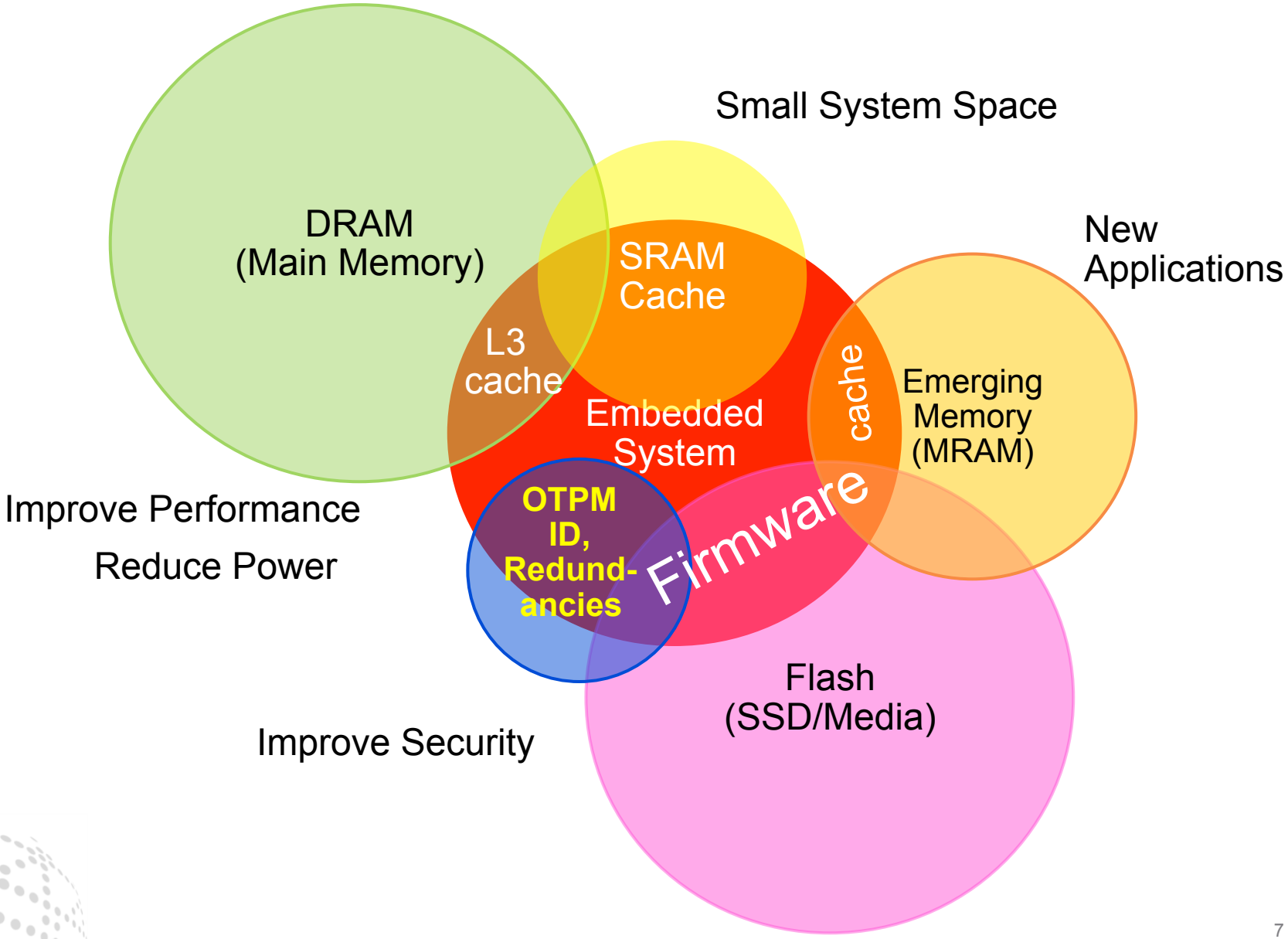
Introduction – Computer Architecture



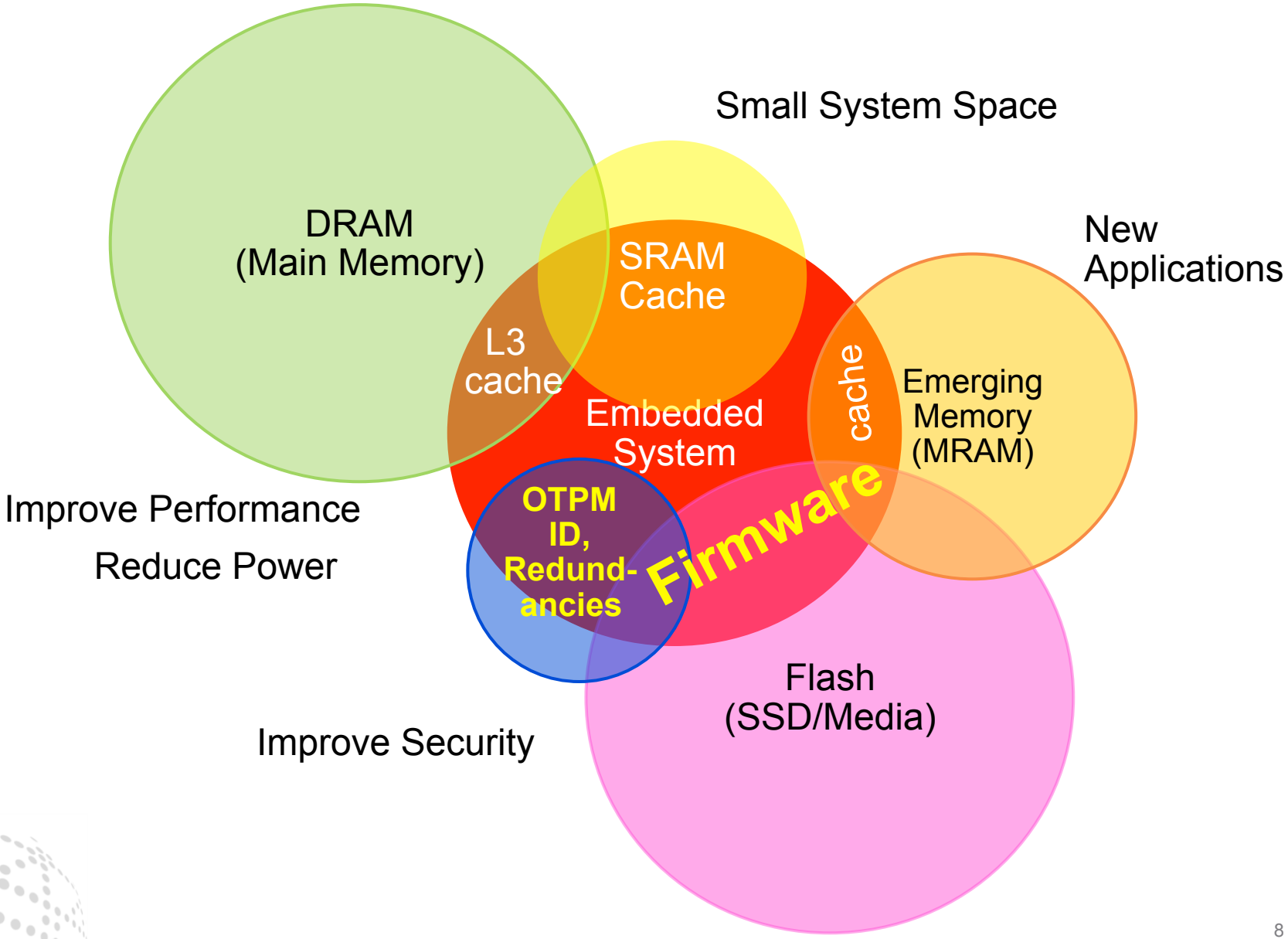
Embedded Memories



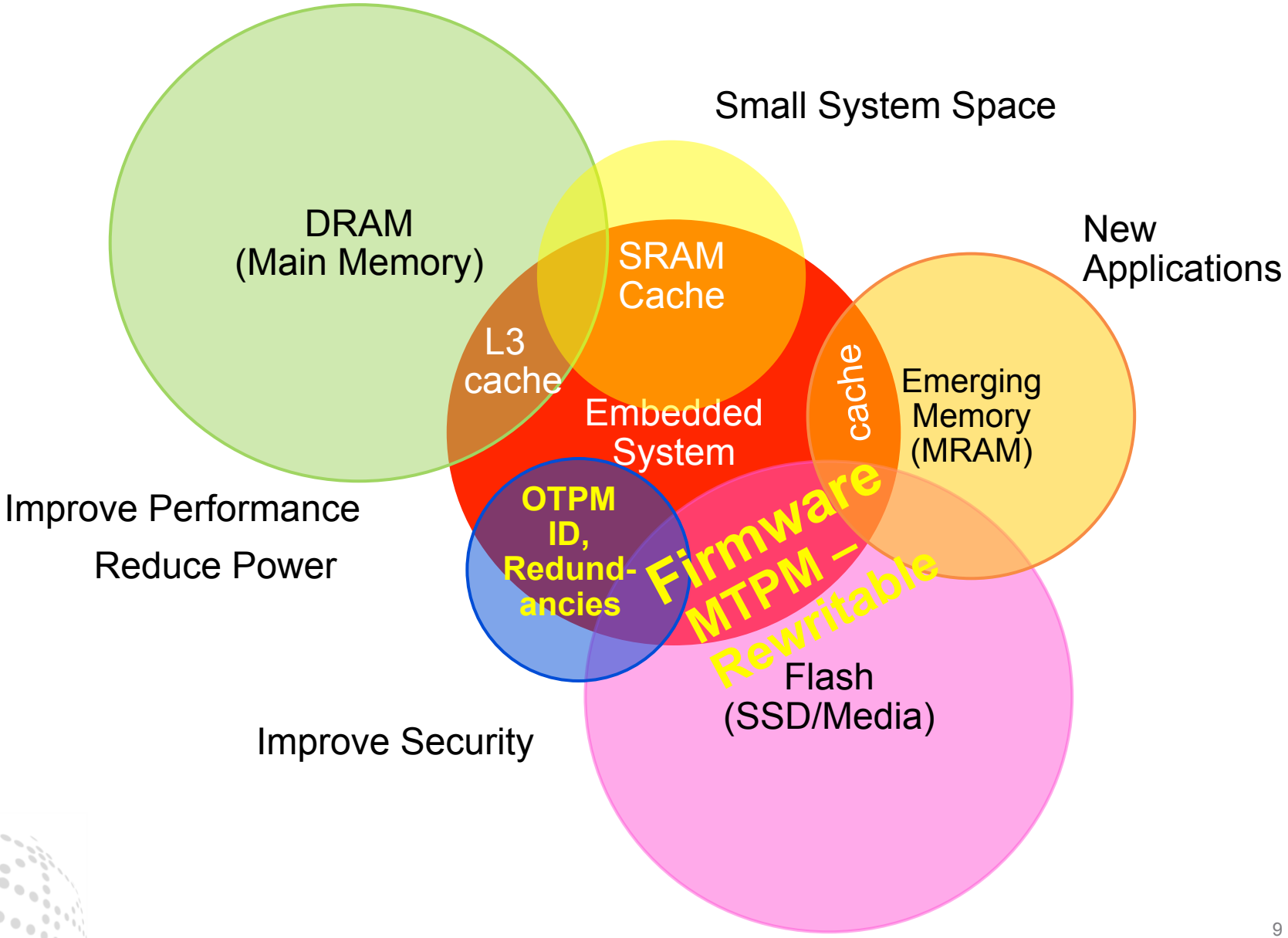
Embedded Memories



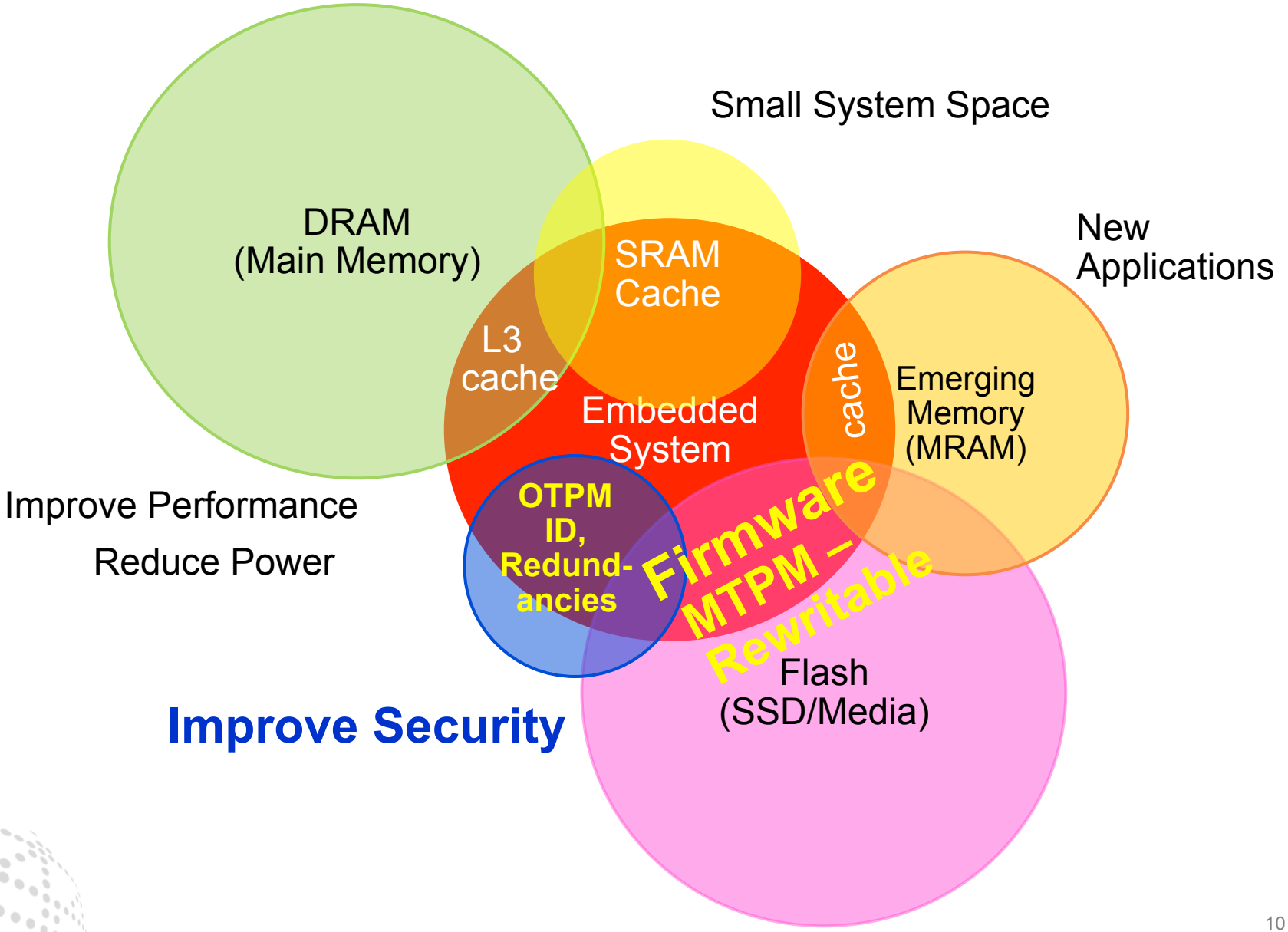
Embedded Memories



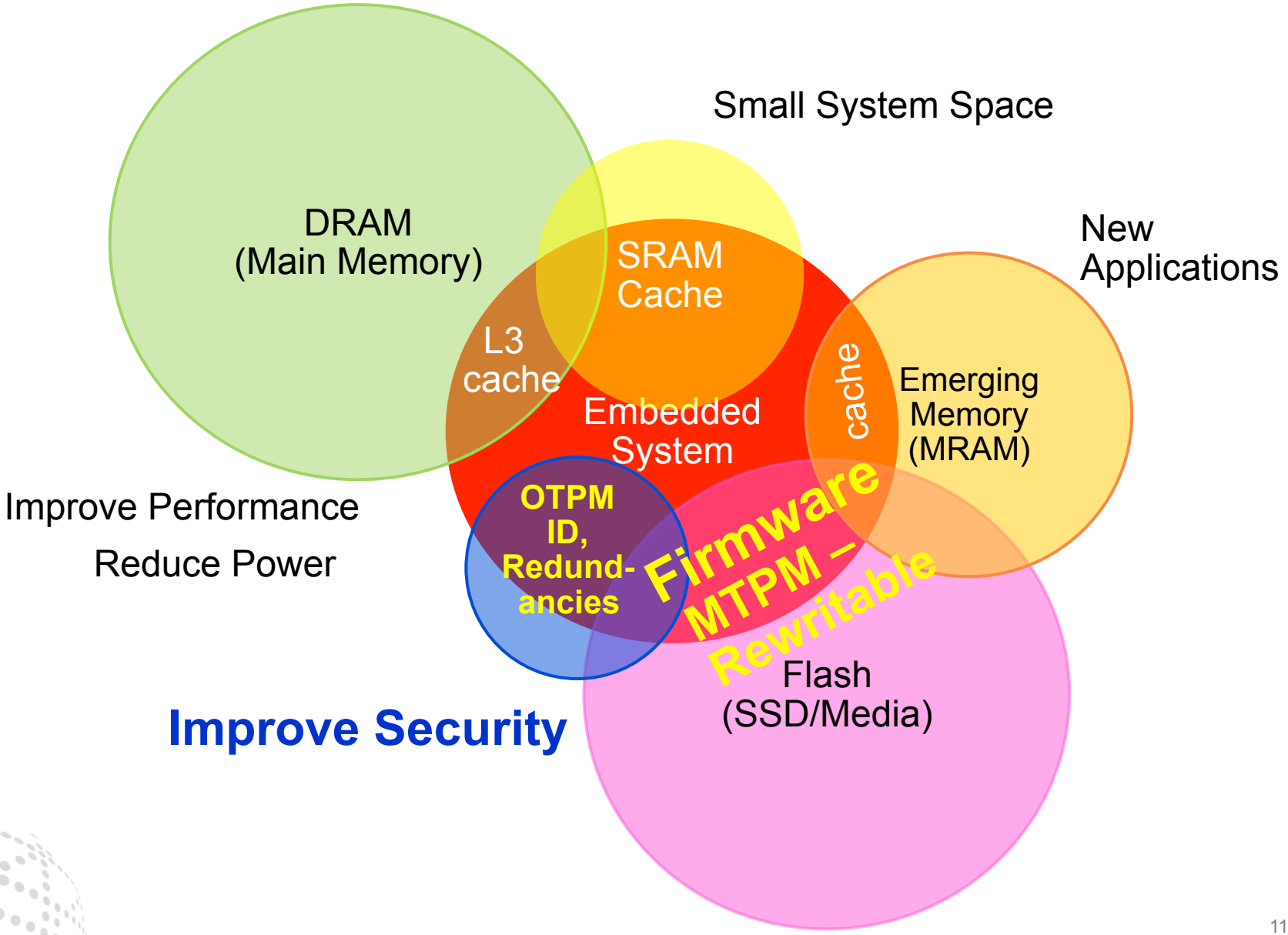
Embedded Memories



Embedded Memories



Embedded Memories



Outline

- Introduction
- Embedded Nonvolatile Memory Applications
- Charge Trap Transistor (CTT) Technology
- Macro Overview
- Hardware Results
  - SOI
  - 14nm Bulk FinFET
- Future scope
- Summary

Embedded Nonvolatile Memories

Applications

Type	Tech-nique	Density	Re-writability	ID	Security	2D Red.	3D Red.	Firmware	Data	Cache
OTPM	eFUSE	Low	One Time	●	▲	●	✖	✖	✖	✖
Dense OTPM	Anti-fuse	Medium	Emulation	●	●	●	●	●	✖	✖
MTPM	Charge Trap	Medium	Medium	●	●	●	●	●	✖	✖
Flash	FG	High	High	✖	●	▲	●	●	●	✖
MRAM	Magnetic	High	High	✖	●	▲	●	●	●	●

- Integrating Firmware on chip is beneficial
  - Cost reduction → If done with zero-mask adder to logic process.
  - Enhanced security.

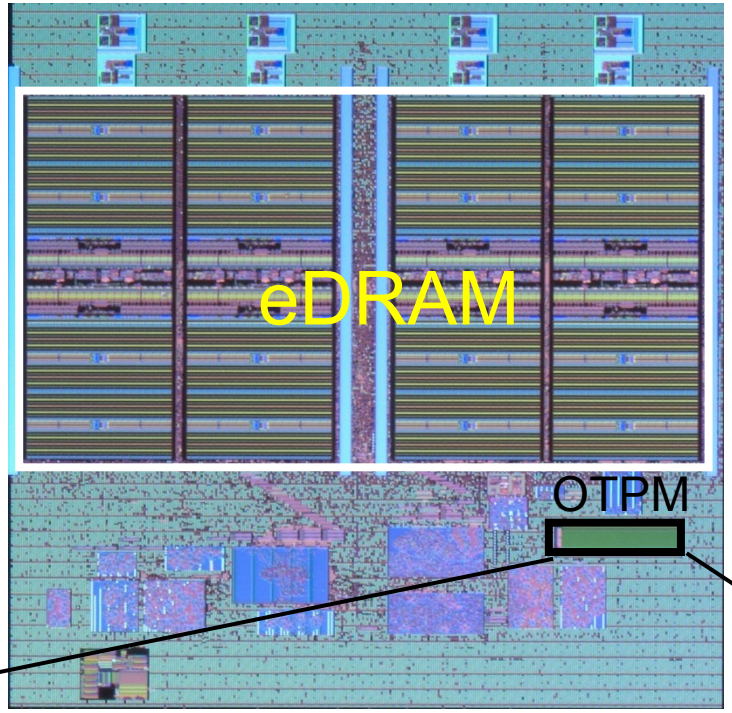
● Ideally suited

● Fairly suited

▲ Somewhat suited

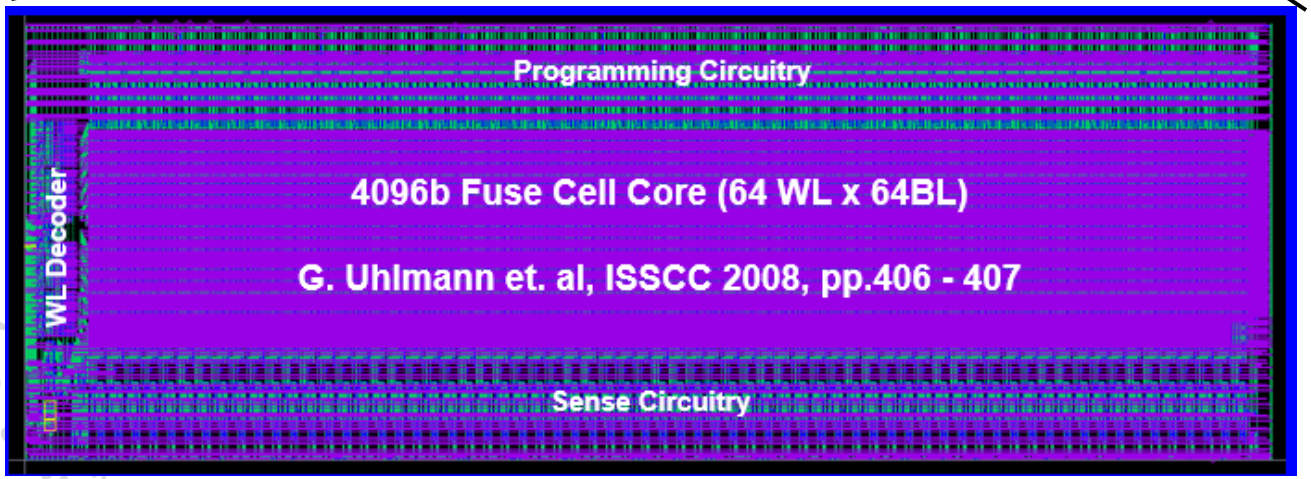
✖ Not suited

eNVM For Redundancy

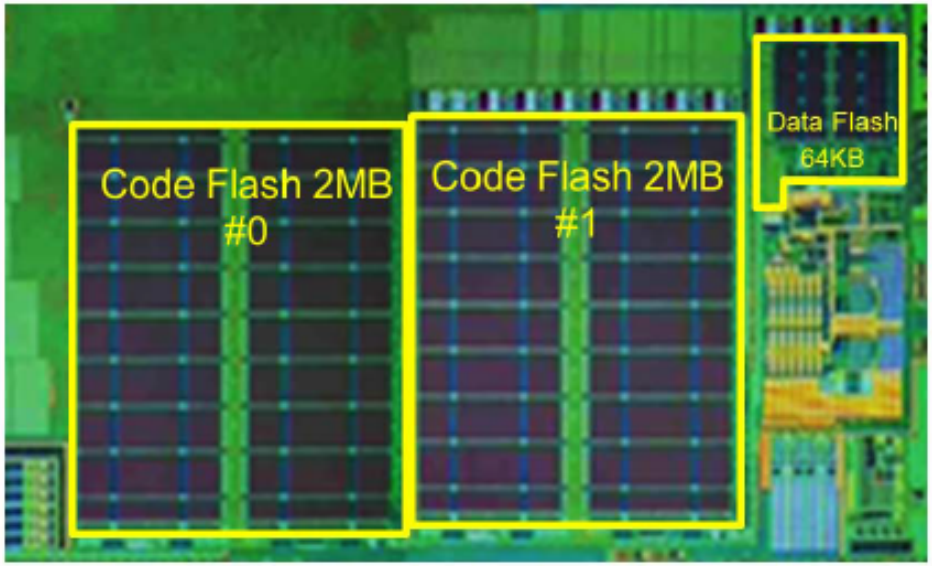


3D memory requires significantly more OTPMs

P. Klim et. al, VLSI 2008



eNVM for Automotive application

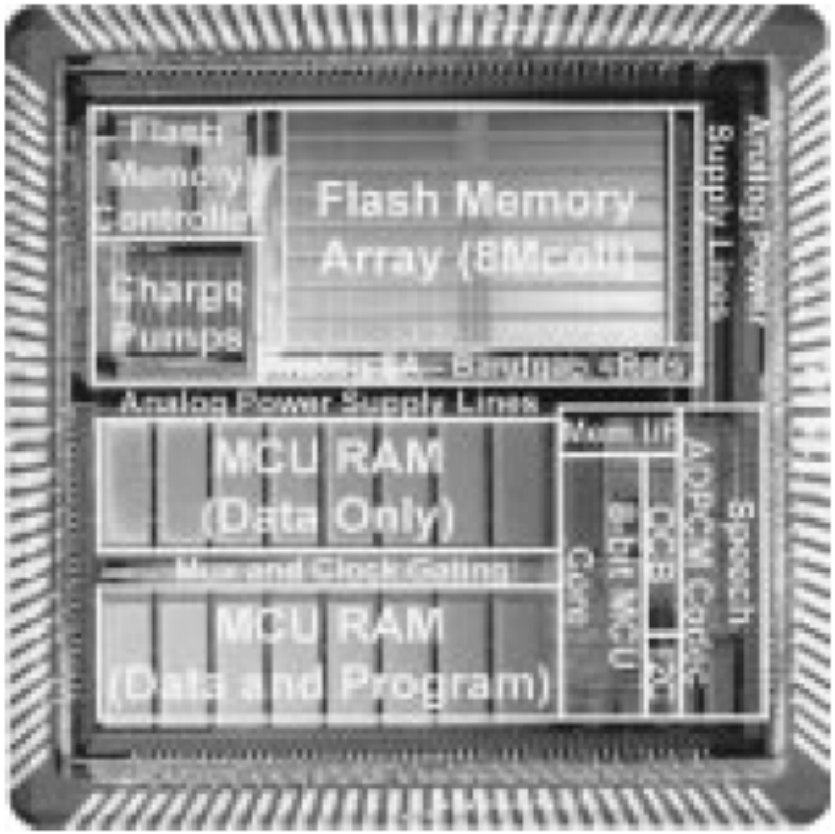


	Code Flash	Data Flash
Technology	28nm SG-MONOS	
Memory Cell Size	0.053μm²	
Memory Capacity	2MB X 2	64KB
Power Supply	Core(VDD) 1.1V±0.1V, I/O(VCC) 2.7-5.5V	
Operating Temp.	-40~170°C (Tj)	
Read I/O number	(128bit + 10bit) X 2	32bit + 7bit
Random Read Freq.	200MHz	10MHz
Program Speed	2.0MB/s	150us/4B
Erase Speed	0.91MB/s	1.5ms/64B
P/E endurance	10k cycles	>1M cycles
Maximum Capacity	32MB	512KB

- 4MB code flash + 64KB data flash integrated on-chip.
- Targeted for automotive MCUs for engine control and driver assistance applications.

Y. Taito et al., ISSCC 2015

eNVM for Media application

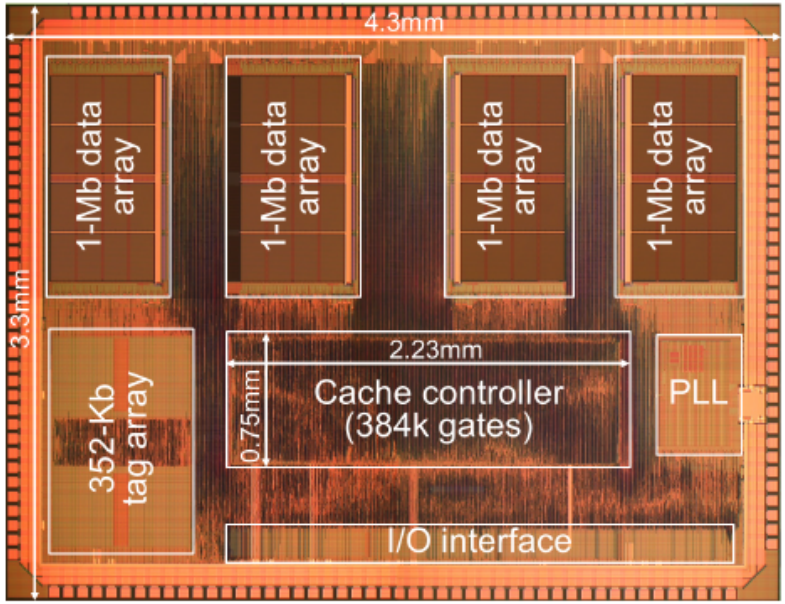


Embedded Flash memory in 0.5um CMOS for voice-storage application:

- 32Mb, 4-level embedded flash to store 64 minutes of voice.
- On-chip Memory BIST is included.

M. Borgatti et al., IEEE JSSC 2001

eNVM for other applications

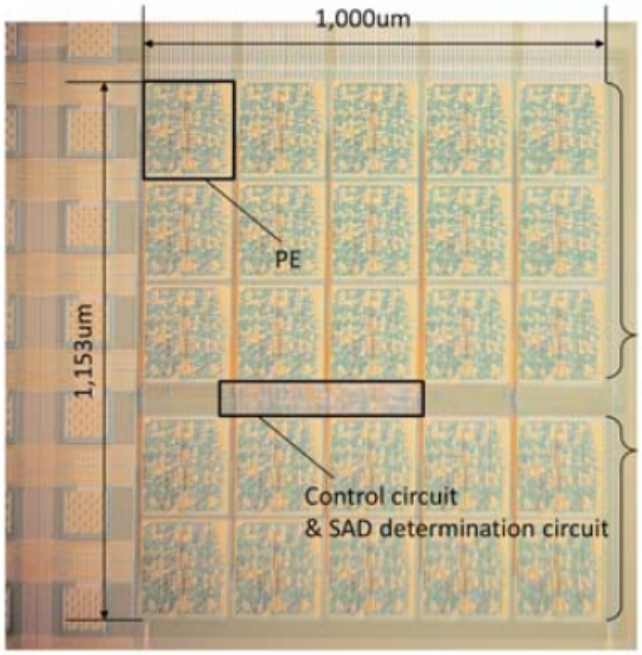


4Mb cache
65-nm CMOS
352-Kb SRAM tag
4×1-Mb MRAM data
4-way set associative
256-bit line width
4k sets
8-entry input queue
8-entry output queue
8-entry miss buffer
4-entry fill buffer
4-entry write-back buffer

STT-MRAM based cache memory in 65nm node,
H. Noguchi et al., ISSCC 2016

Memory cell : 2T 2MTJ

Nonvolatile logic-in-memory array processor using MTJ/MOS in 90nm node,
M. Natsui et al., ISSCC 2013



Process	90nm MTJ/MOS
Area (Core)	1.153 mm²
Num. of MOS Trs.	474,019
Num. of MTJs	13,400
Supply voltage	1.0 V

.

Outline

- Introduction
- Embedded Nonvolatile Memory Applications
- Charge Trap Transistor (CTT) Technology
- Macro Overview
- Hardware Results
  - SOI
  - 14nm Bulk FinFET
- Future scope
- Summary

This talk focuses on a CTT based eNVM targeted for :

- Secure OTPM ID
- 2D and 3D Redundancies
- Firmware applications

Comparison of Embedded NVRAM Solutions

Floating Gate
- Dual poly [1]

Process Overhead
High Voltage

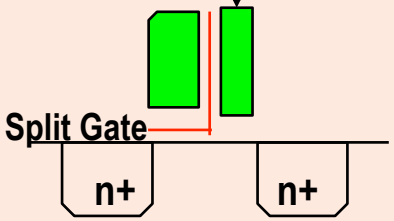
BEOL NVM
- ReRAM [3]
- STT MRAM[4]

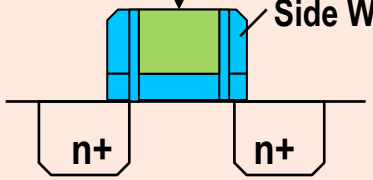
Process Overhead

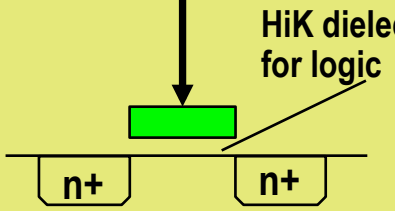
Charge Trap NVM

OTPM
- eFuse [5]
- Anti Fuse [6]

One Time Programmable







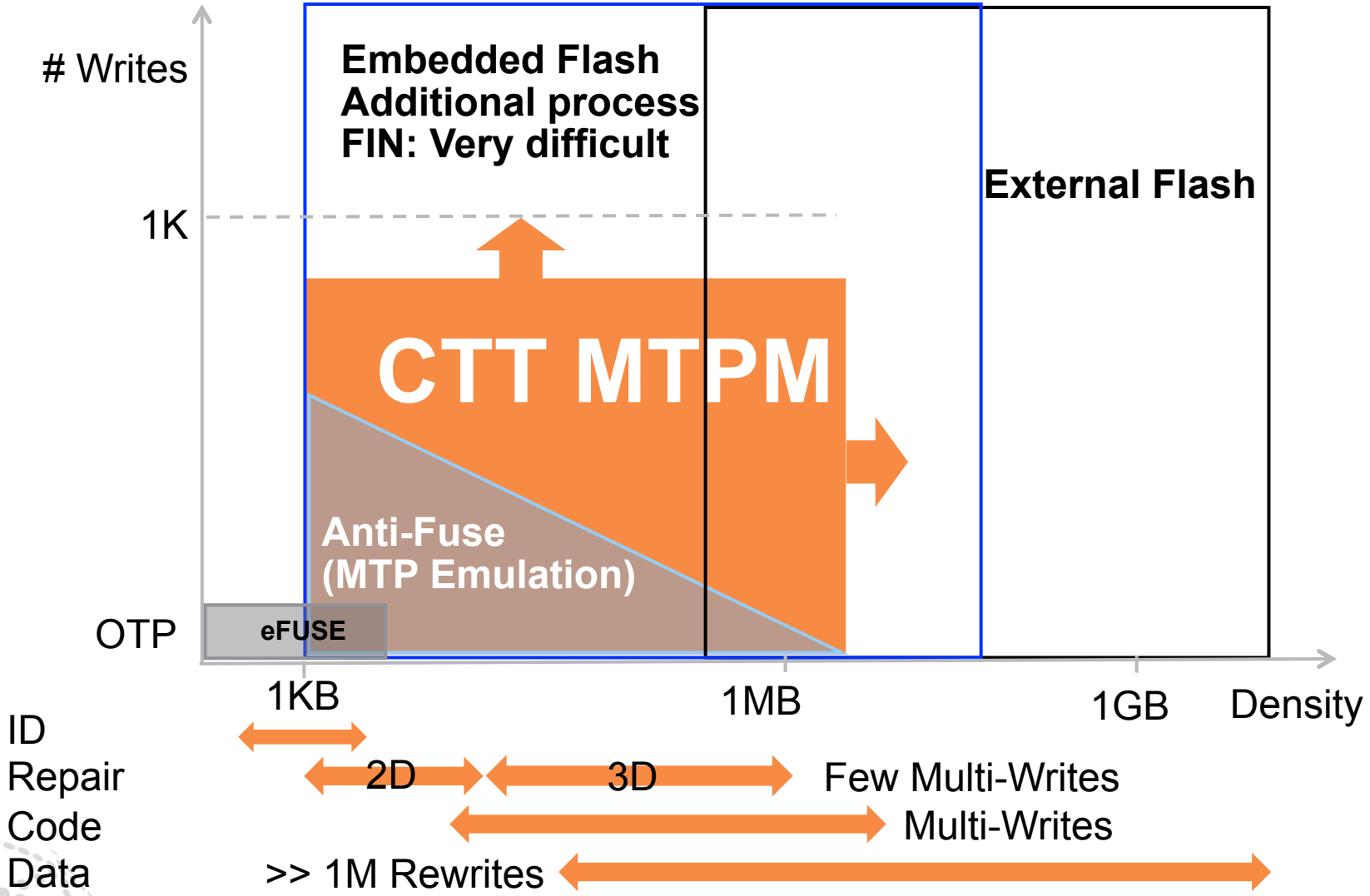
Split Gate SGMONOS [2]
- CHE in Si_3N_4 in Split Gate
- Voltage ~ 7-10V

OTP from NSCore [9]
- CHE in Si_3N_4 Side wall spacer
- Voltage ~ 5-7V

CTT [This Work]
- CHE in HiK HfO_2
- Voltage ~ 1.5-2V

CTT – Dense, No mask adder, Bulk/ SOI FIN scalable, logic voltage compatible MTPM

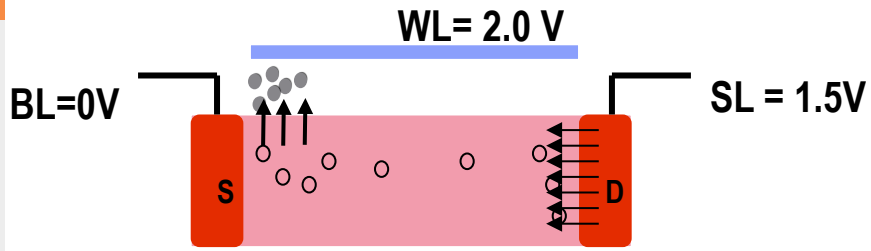
CTT MTPM Applications



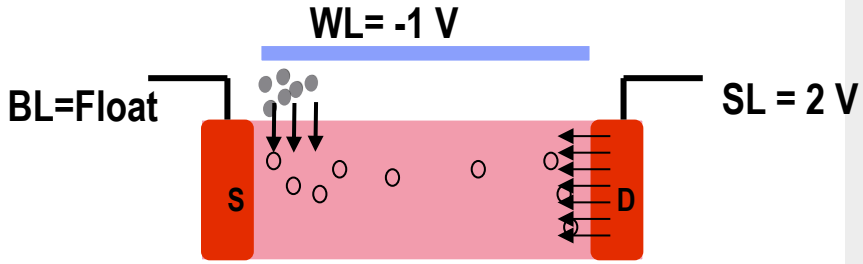
Outline

- Introduction
- Embedded Nonvolatile Memory Applications
- Charge Trap Transistor (CTT) Technology
- Macro Overview
- Hardware Results
  - SOI
  - 14nm Bulk FinFET
- Future scope
- Summary

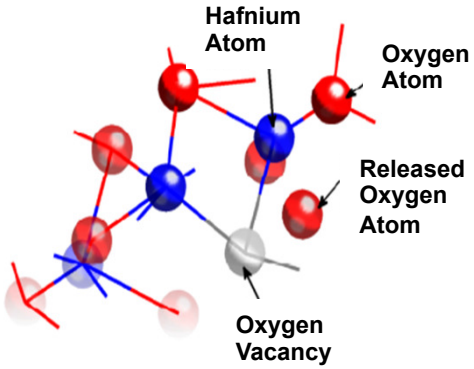
Charge Trap Transistor (CTT) Technology

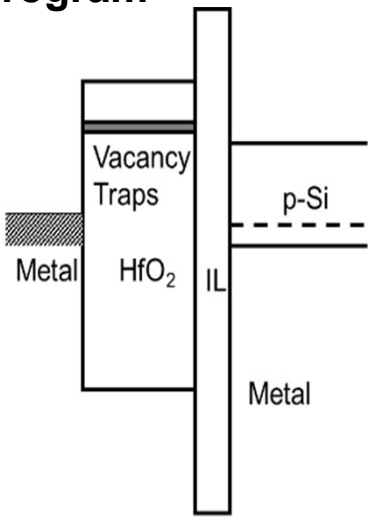


HiK Logic NMOS - Program



HiK Logic NMOS Transistor - Erase





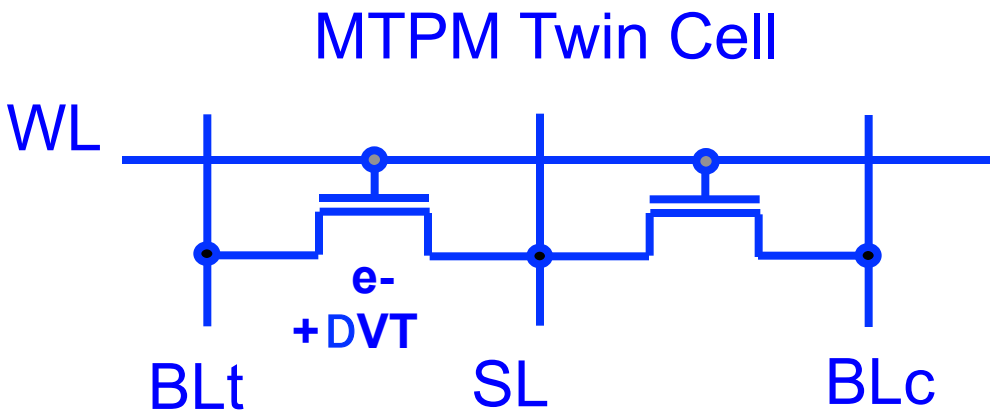
	WL	BL	SL
Write	~2V (VPP)	0V	~1.5V (VSL1)
Read	1V (VDD)	Floating	1V (VDD)
Erase	~-1V (VWL)	Floating	2V (VSL2)
Standby	0V	Floating	0V

Charge trapping mechanism:
Intrinsic Oxygen vacancies in HfO_2 [7]

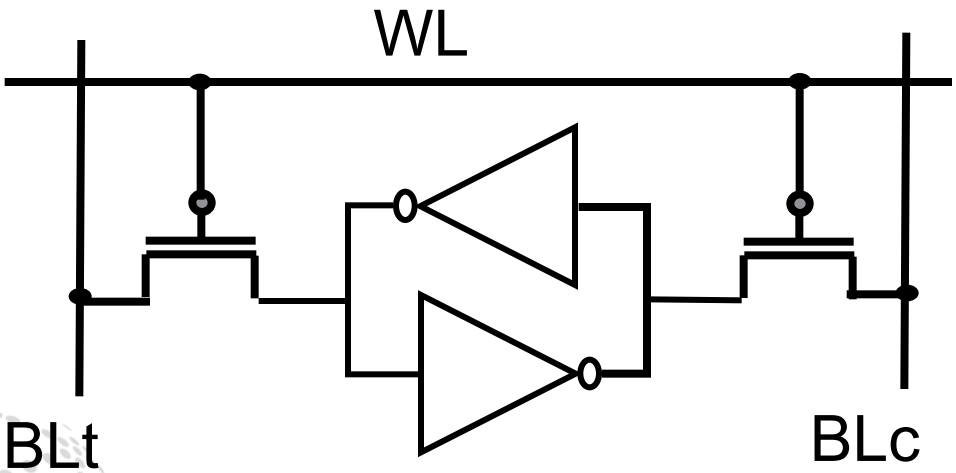
Intrinsic Oxygen vacancies in HfO_2 HiK dielectrics

Charge Trapping done at logic process compatible voltages

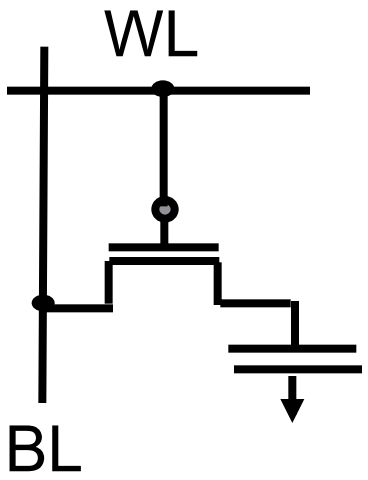
OTPM / MTPM Twin Cell



0.108mm² in 32nm node



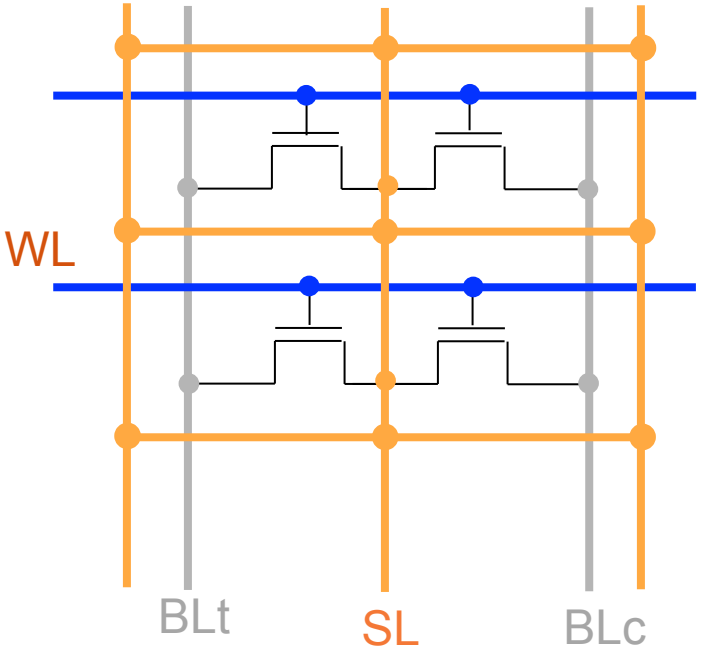
6T SRAM 0.169mm² in 32nm node



1T1C DRAM 0.039mm² in 32nm node

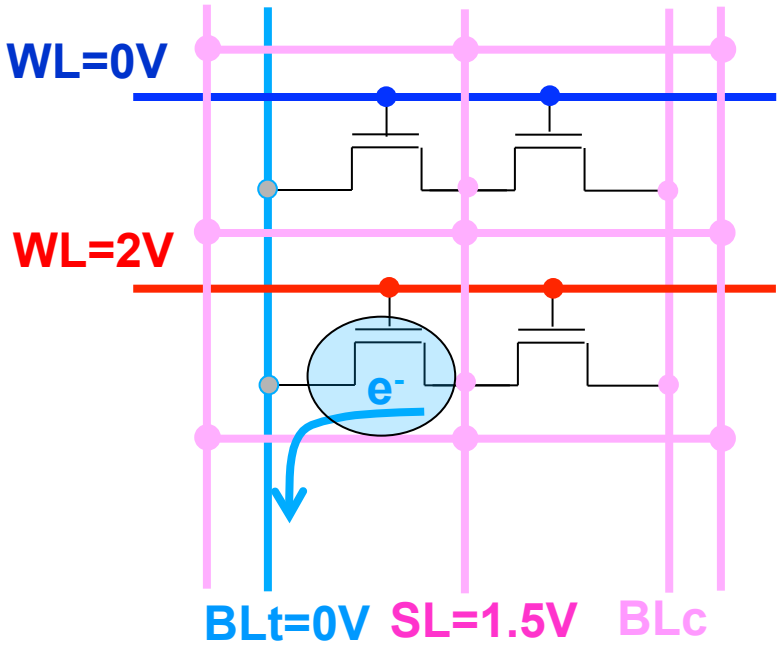
1/4 Standby – Initial State Before Programming

Mode	WL	BL	SL
Standby	0V	float	0V
Program	2V	0V	~1.5V
Read	1V	Signal	1V
Erase	-1V	float	~2V



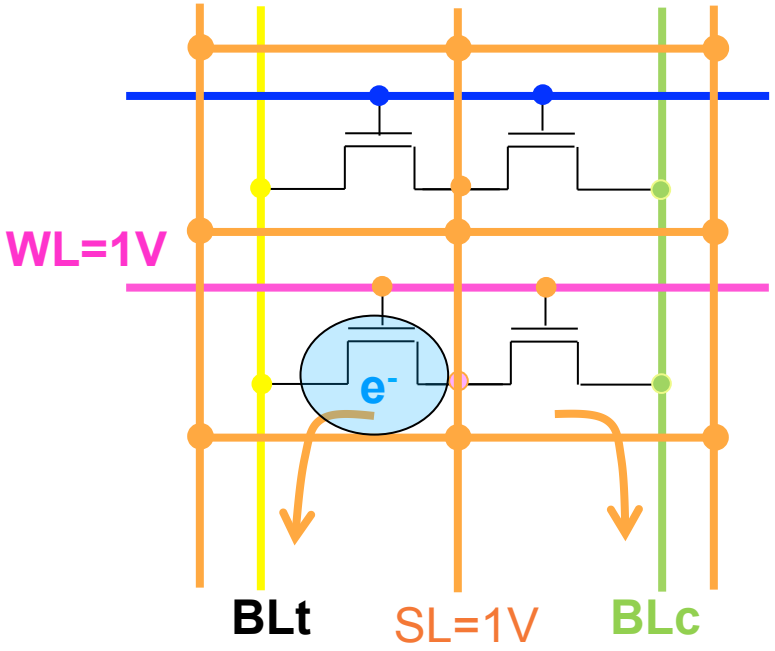
2/4 Program

Mode	WL	BL	SL
Standby	0V	float	0V
Program	2V	0V	~1.5V
Read	1V	Signal	1V
Erase	-1V	float	~2V



3/4 Read

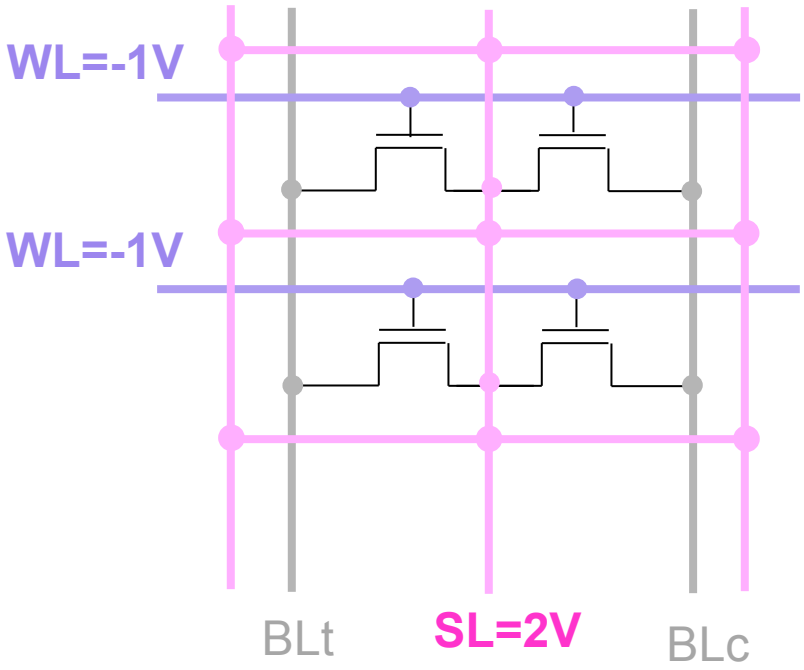
Mode	WL	BL	SL
Standby	0V	float	0V
Program	2V	0V	~1.5V
Read	1V	Signal	1V
Erase	-1V	float	~2V



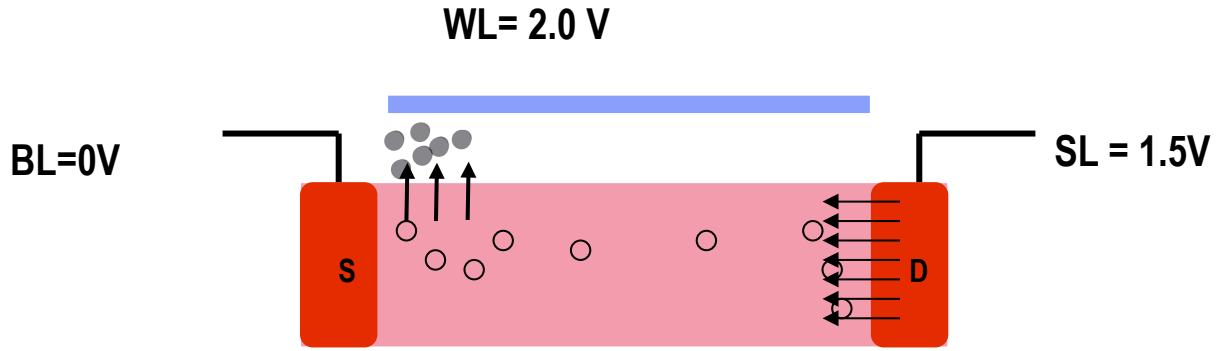
DSA = $V_{BLt} - V_{BLc}$ OR $I_{BLt} - I_{BLc}$

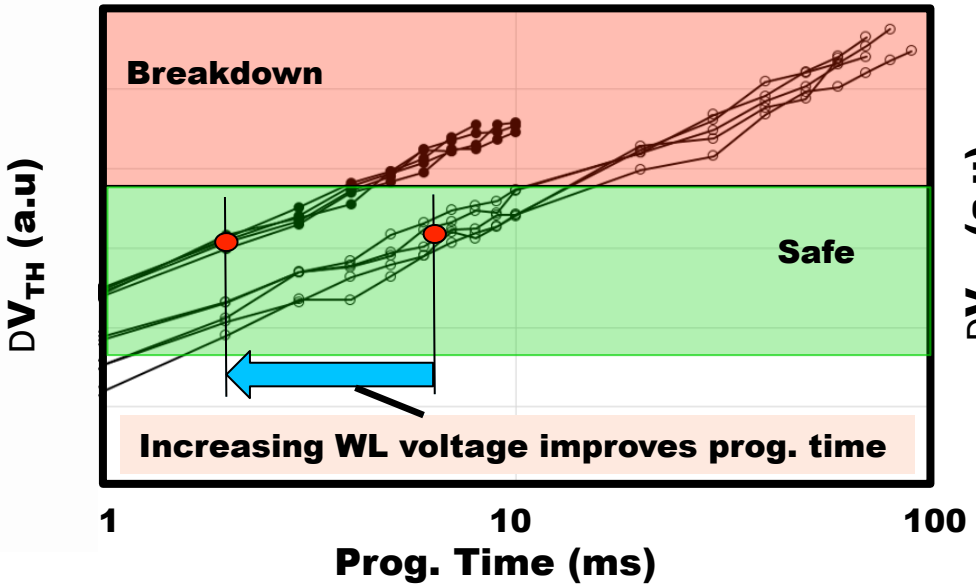
4/4 Erase

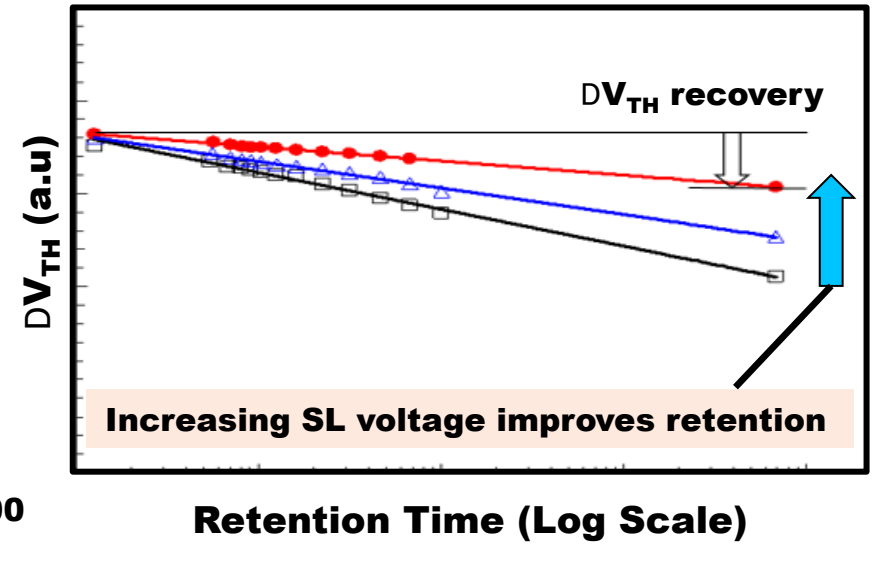
Mode	WL	BL	SL
Stand-by	0V	float	0V
Program	2V	0V	~1.5V
Read	1V	Signal	1V
Erase	-1V	float	~2V



CTT : Operating zones

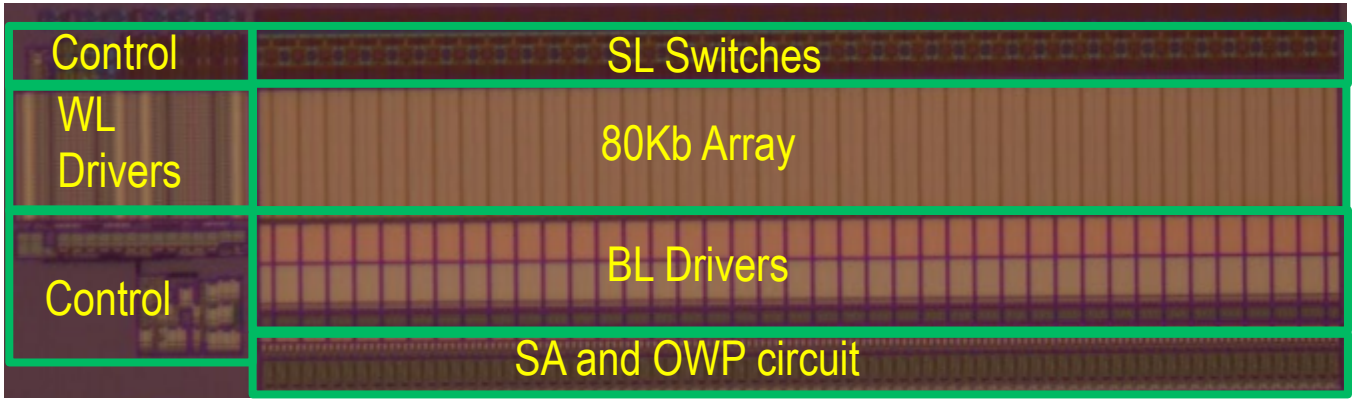






Circuit Assist Techniques needed to operate in the "Safe" Zone

HiK Charge Trap Transistor Enables Process Free Multi-time Programmability



	eFUSE	Anti Fuse	MTPM
Mb/(mm^2)	<0.1	~1	~1
Rewritable	No	No	Yes

1st Write

2nd Write

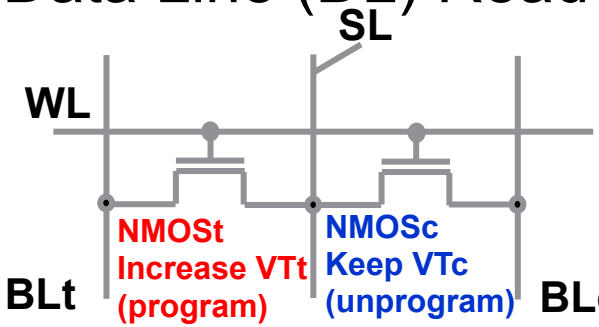
Bitmap Image

Outline

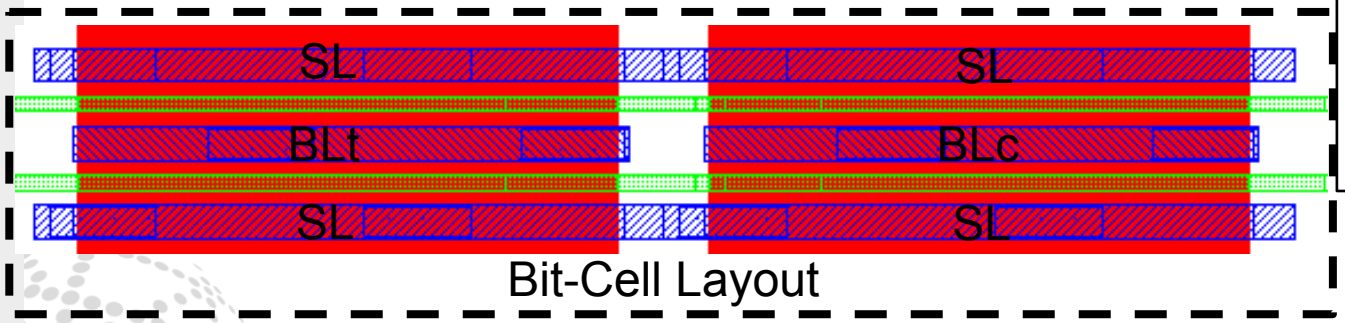
- Introduction
- Embedded Nonvolatile Memory Applications
- Charge Trap Transistor (CTT) Technology
- Macro Overview
- Hardware Results
  - SOI
  - 14nm Bulk FinFET
- Future scope
- Summary

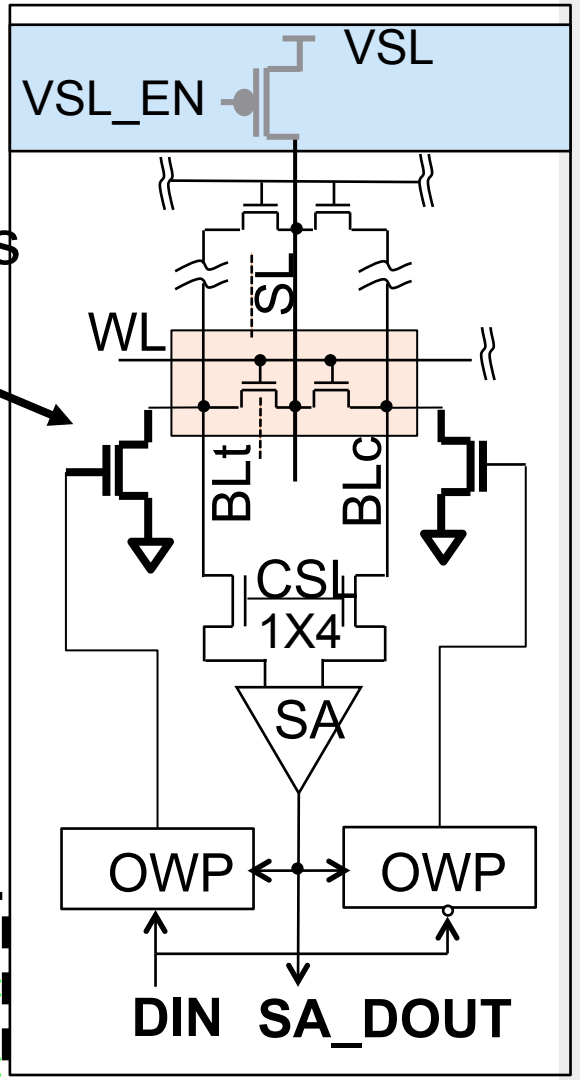
80Kb Twin Cell Macro Architecture

80 Data Line (DL) Read Out(80X4 = 320 Bit Lines)



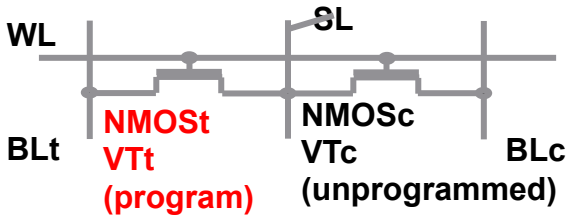
Detect DIFVT
DIFVT = VTt -VTc





Protection devices not shown

Overwrite Protection



DIFVT = VTt - Vtc

Before Programming

After Programming

Overwrite Protection

DIFVT Distribution (Normalized)

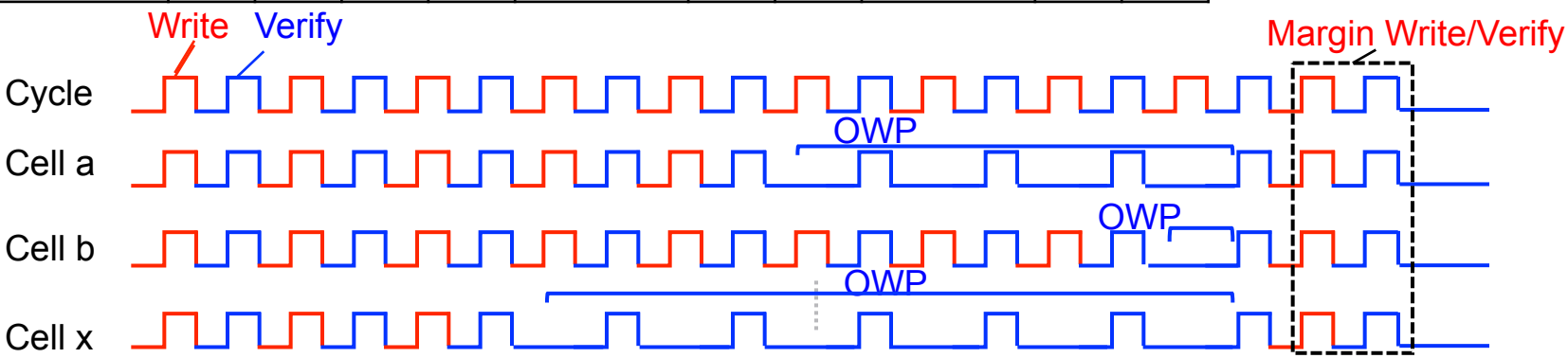
Overwrite:
Too much VT Shift
(TDDB Risk)

Insufficient write
Too small VT Shift
(Retention Risk)

Circuit Assist – Over Write Protection

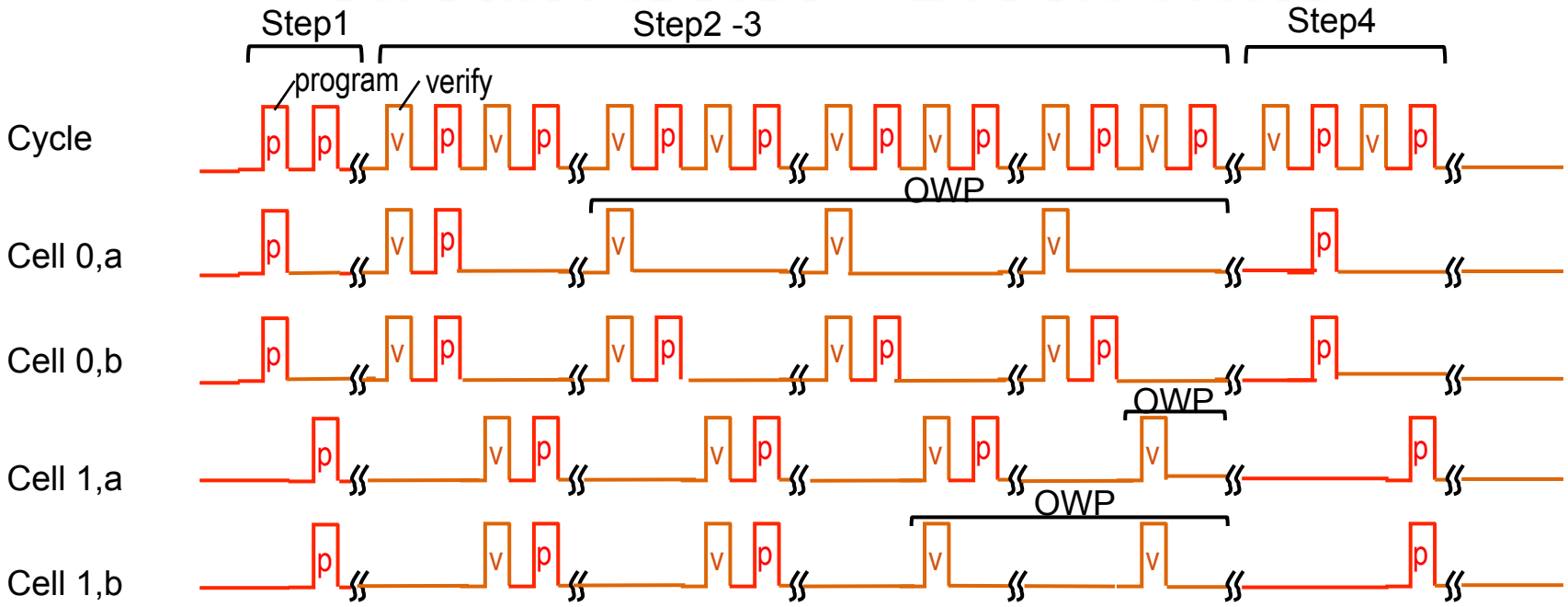
			Program Cycle 1		Read Cycle 1	Program Cycle 2		Read Cycle 2	Margin Write	
Cell a DIN=1					DOUT=1 Matches DIN	MASK		DOUT= 1 Match		
Cell b DIN = 1					DOUT=0 Mismatch with DIN			DOUT=1 Match		
	T	C	VTt	VTc		VTt	VTc		VTt	VTc

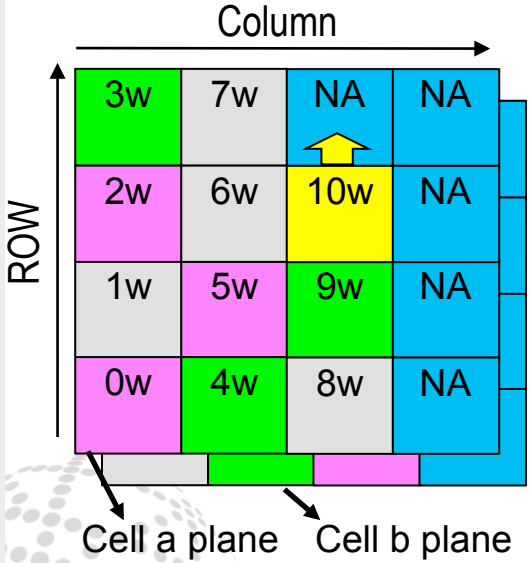
Comparable V_{TH} difference = (VTt – Vtc)



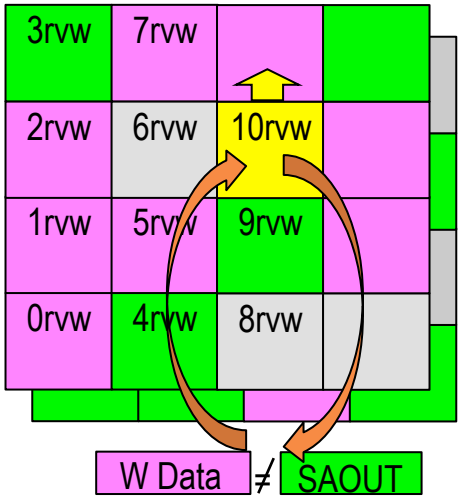
Write in multi-steps with OWP

Circuit Assist – Block Write

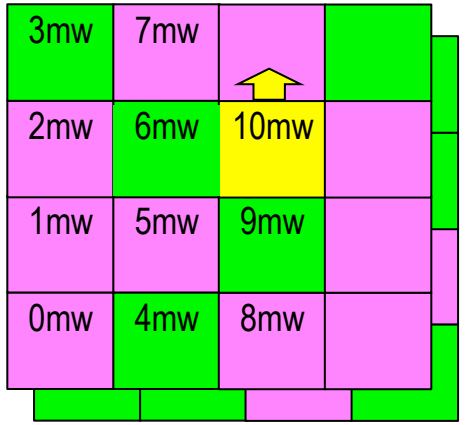




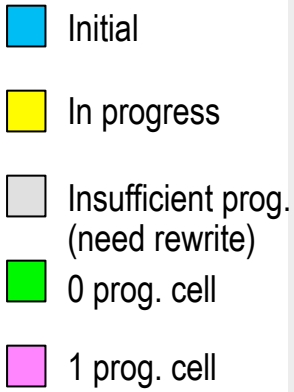
Step 1: Initial write (w)



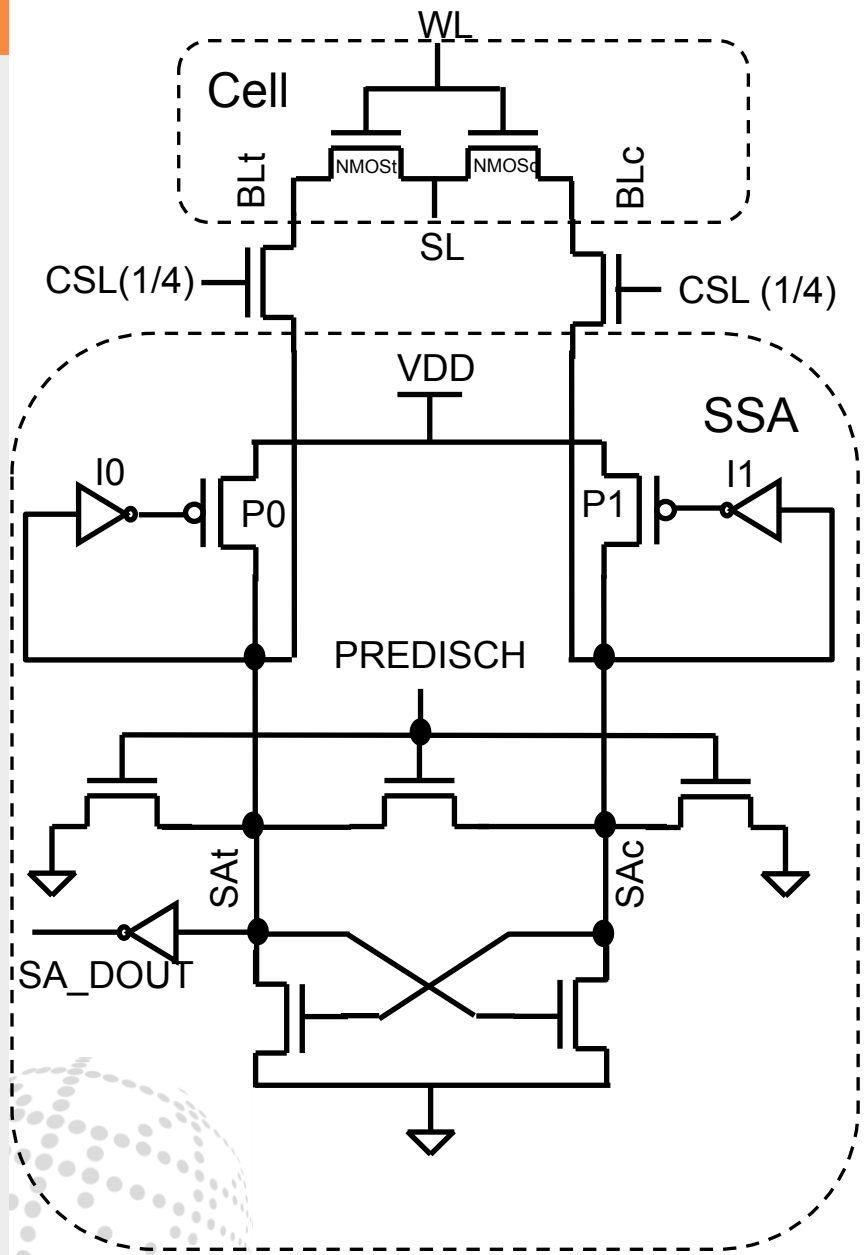
Step 2&3: read+verify+write (rvw)
(rewrite if not verified correctly)



Step 4: margin write (mw)

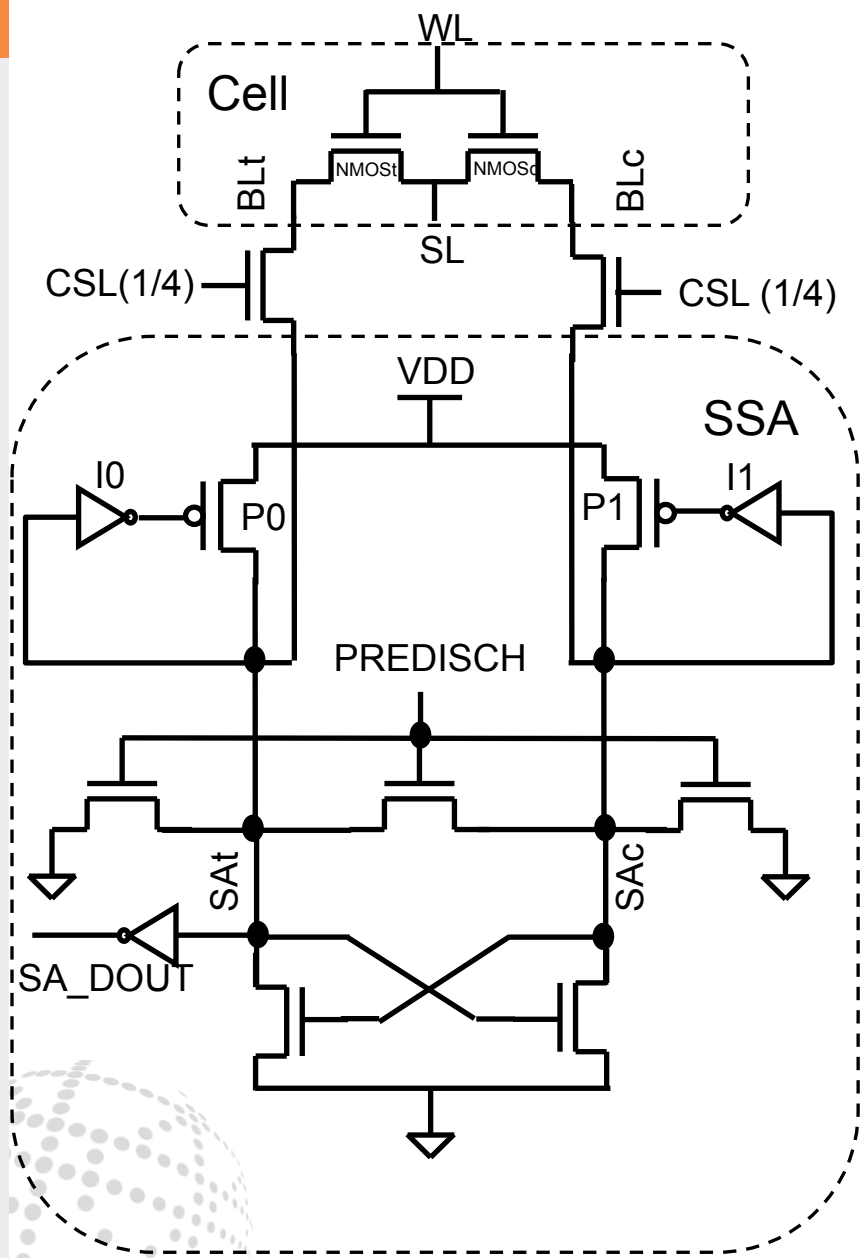


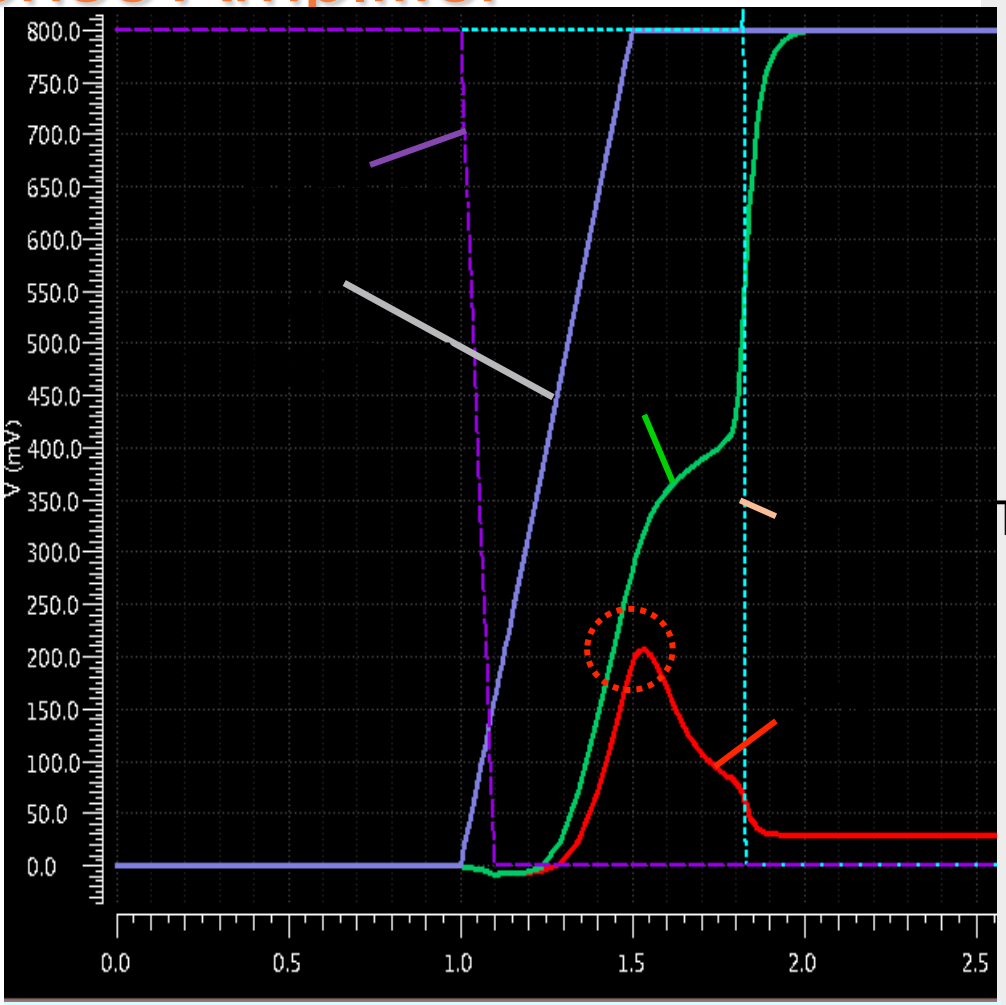
Circuit Assist – Slew Sense Amplifier



- Turn on column switch
  - Connect cell to SA
- Ramp the WL slowly to VDD
  - Rise time ~500ps
- Differential charging of SAt and SAc
- Self timed Sense Amp
- Capacitance mismatch
  - Fixable using reduced slew

Circuit Assist – Slew Sense Amplifier

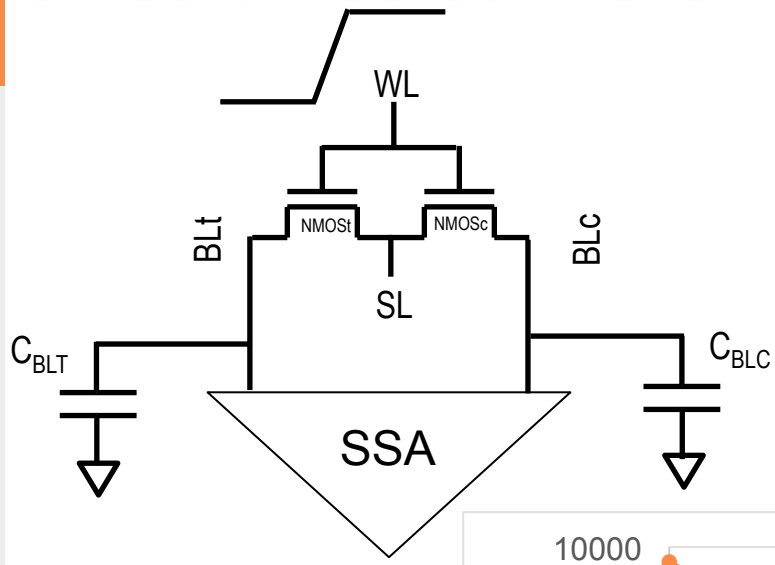


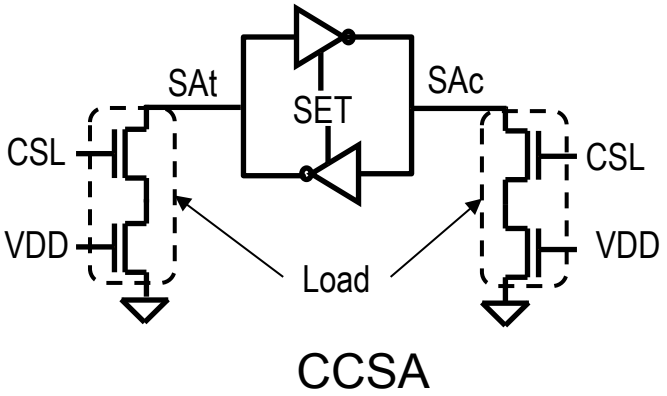


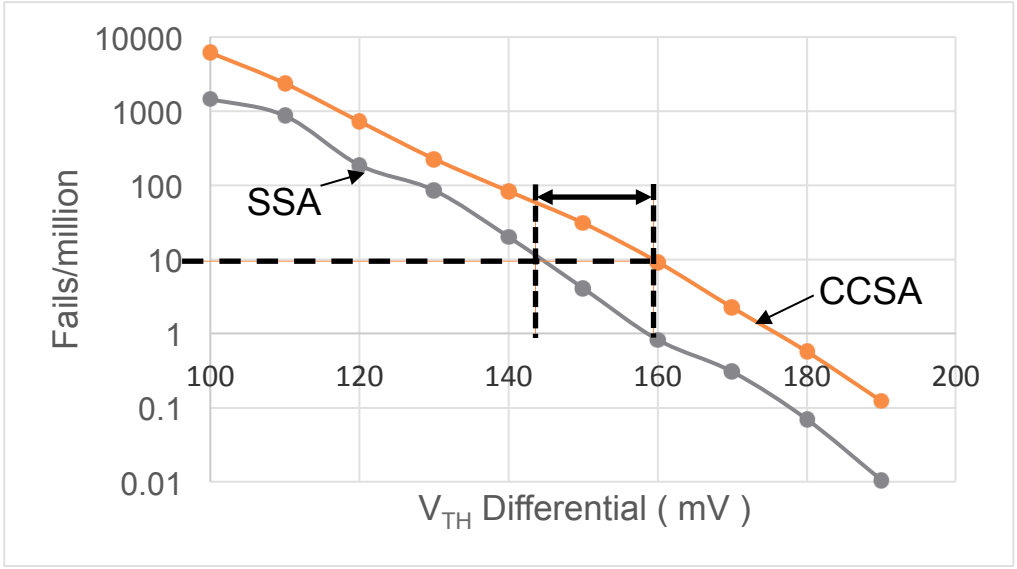
NMOSc is still OFF

NMOSt turns ON in Saturation

Circuit Assist – Slew Sense Amplifier





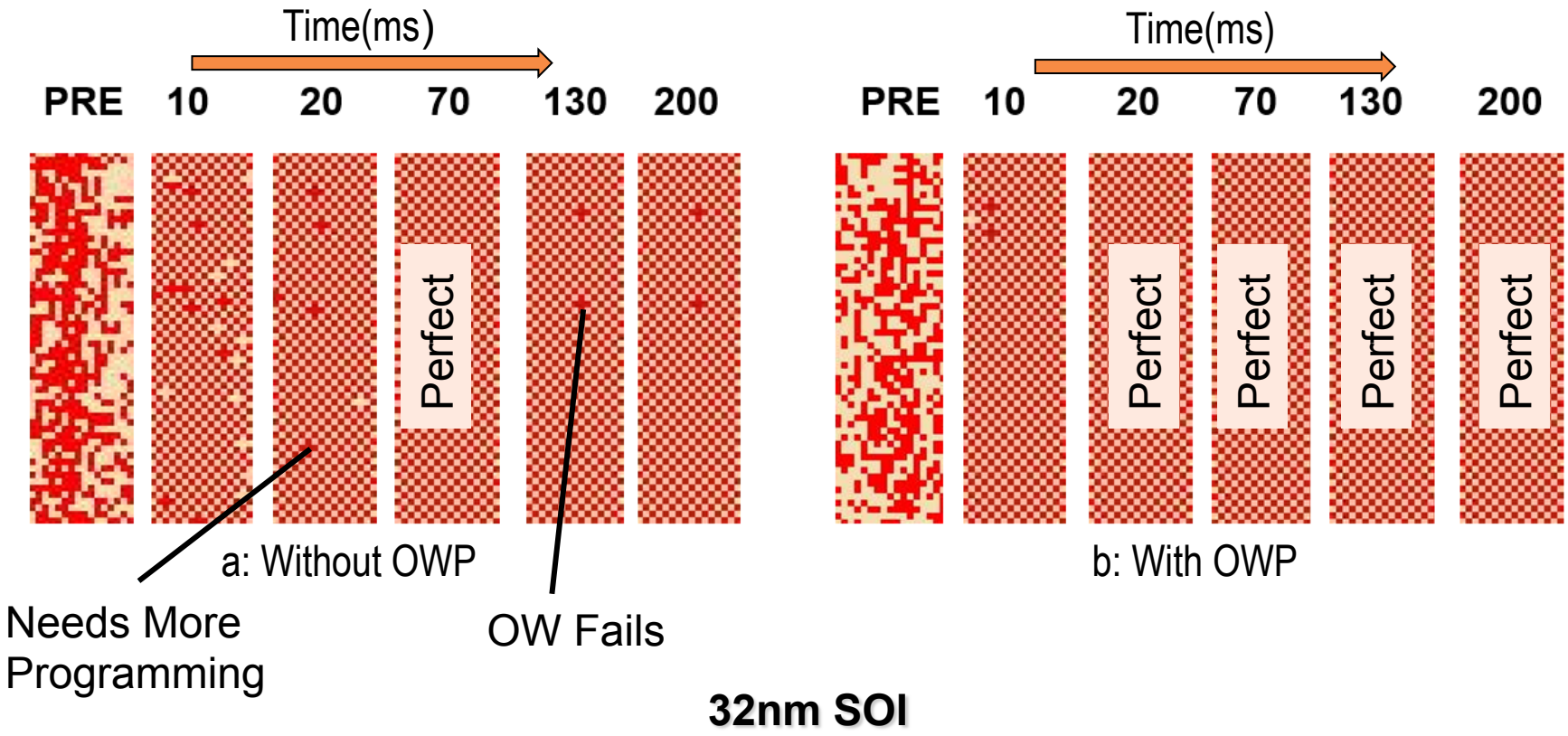


SSA can sense 10% less VT shift compared to CCSA

Outline

- Introduction
- Embedded Nonvolatile Memory Applications
- Charge Trap Transistor (CTT) Technology
- Macro Overview
- Hardware Results
  - SOI
  - 14nm Bulk FinFET
- Future scope
- Summary

SOI Hardware Results - OWP

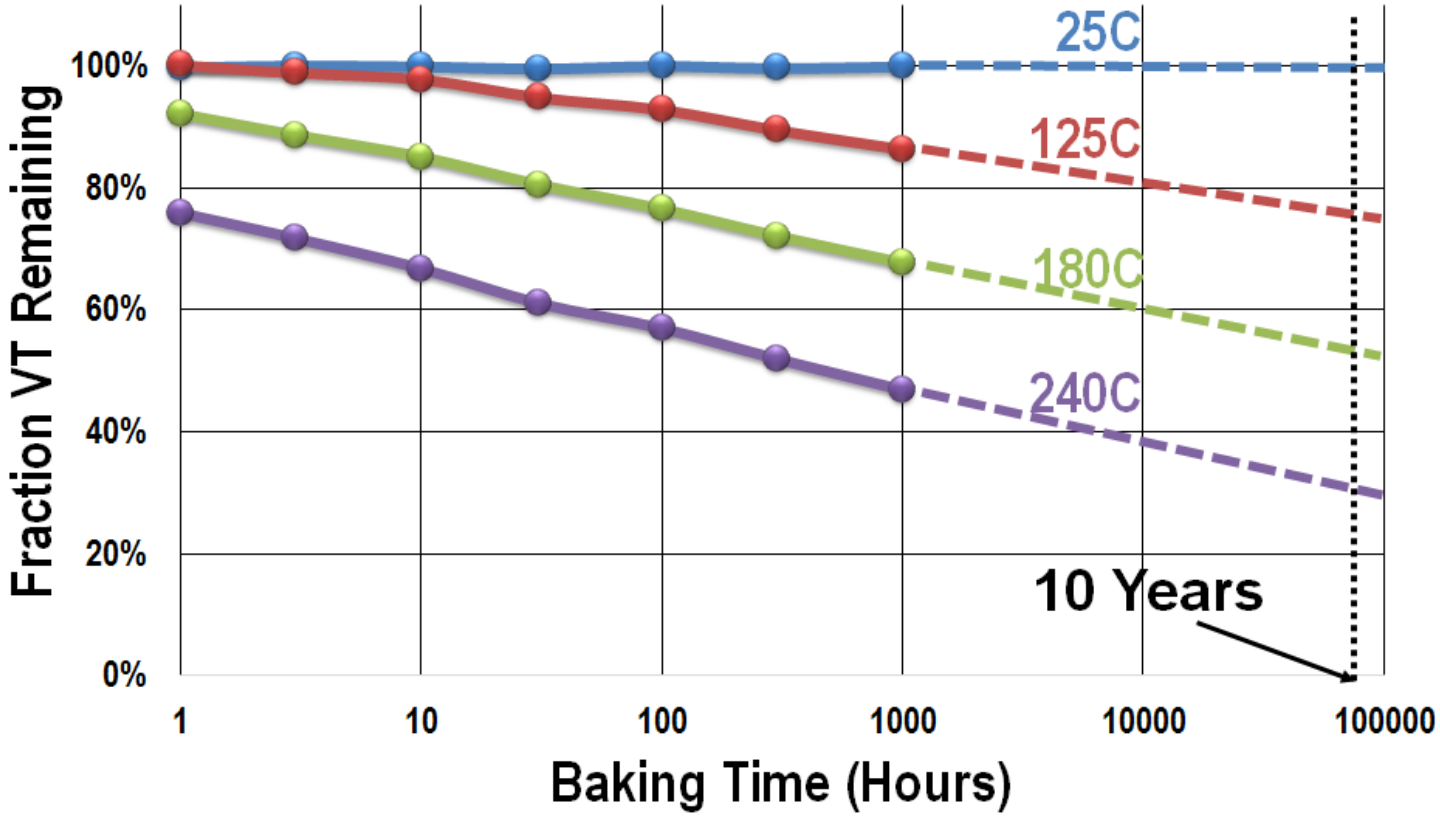


32nm SOI

Over-Write Protection ensures operation in the "safe" zone

J. Viraraghavan et al., IEEE VLSI Symp. 2016

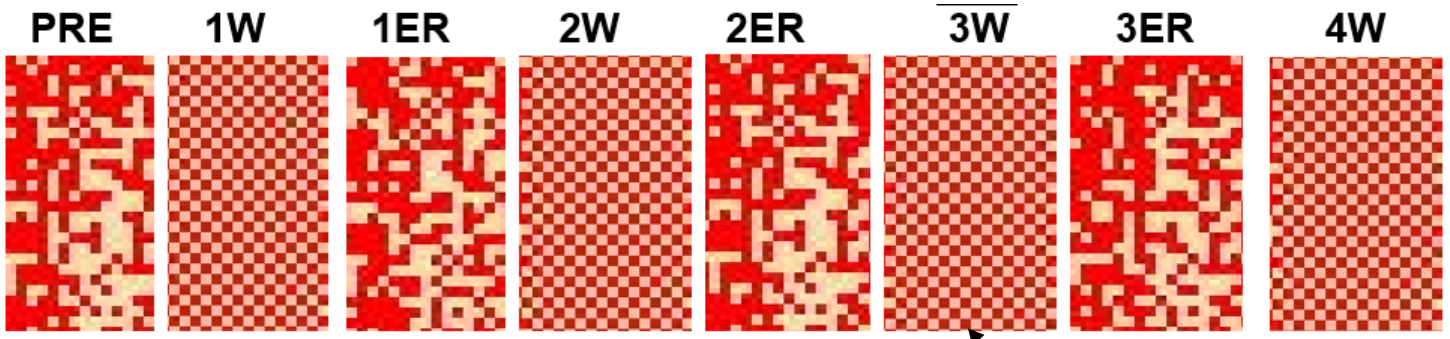
SOI Hardware Results – HTS Stress



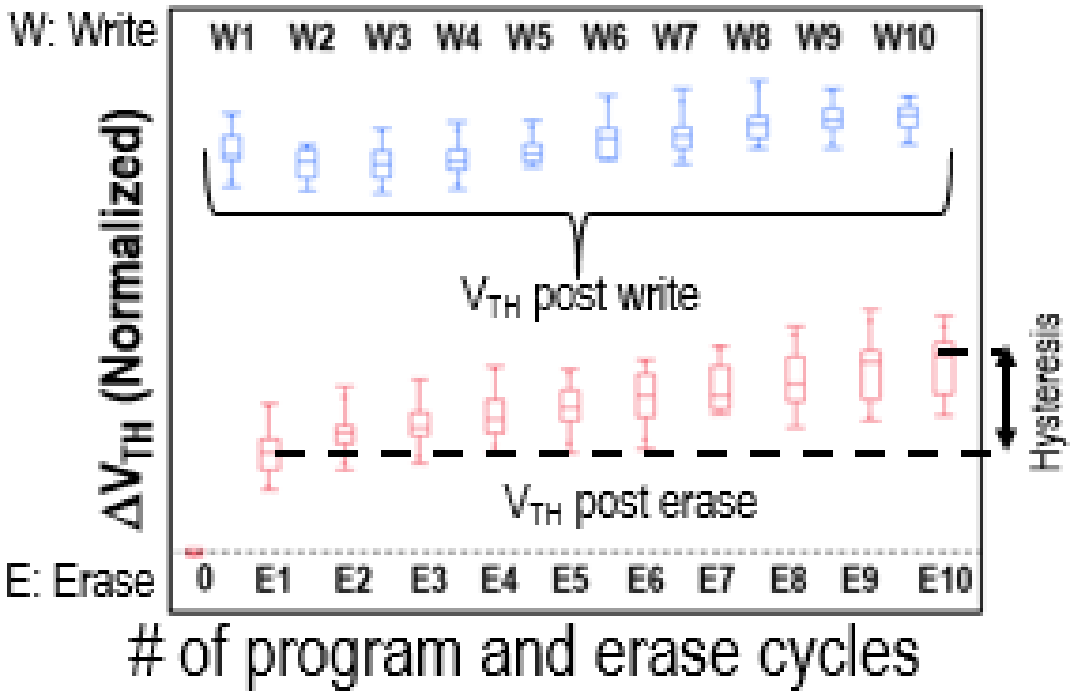
32nm SOI

Bake Tests Indicate 30% V_{TH} Degradation in 10 years @ 125^{O}C

SOI Hardware Results – Multi-Time Programming



a: Multiple Write (4X) with OWP



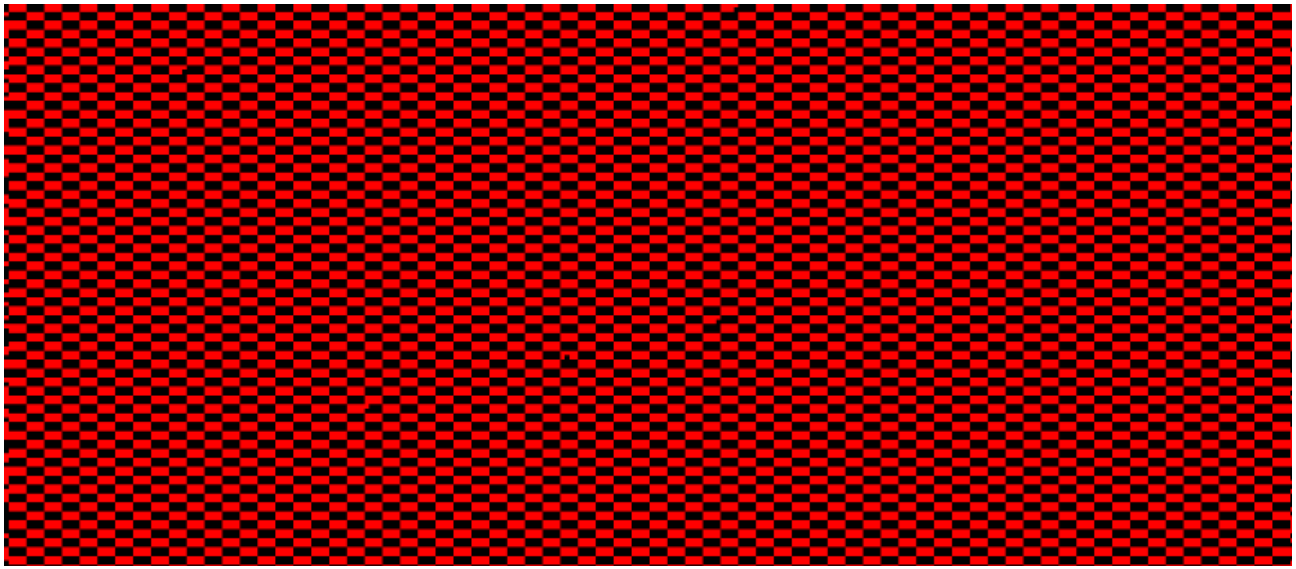
J. Viraraghavan et al., IEEE VLSI Symp. 2016

Outline

- Introduction
- Embedded Nonvolatile Memory Applications
- Charge Trap Transistor (CTT) Technology
- Macro Overview
- Hardware Results
  - SOI
  - 14nm bulk FinFET
- Future scope
- Summary

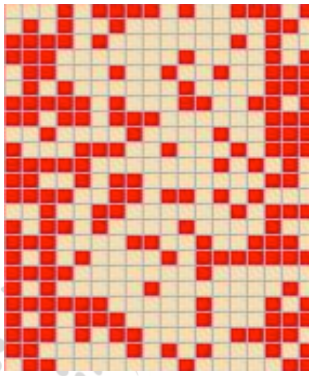
14nm bulk FinFET Hardware Results

40Kb Checker Board Programming

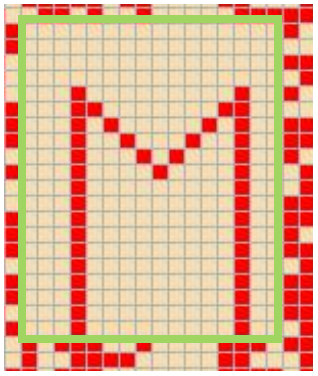


256b Multi Programming

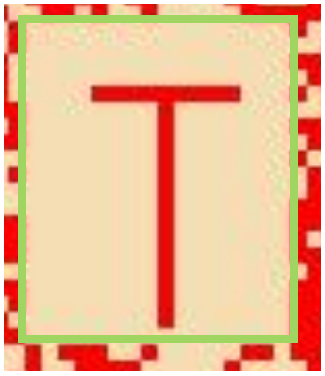
Pre



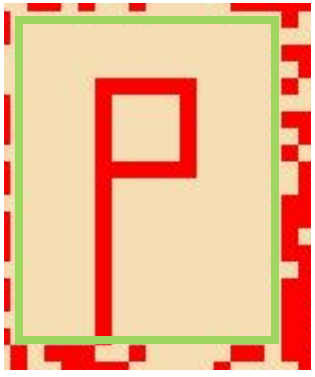
1st Write



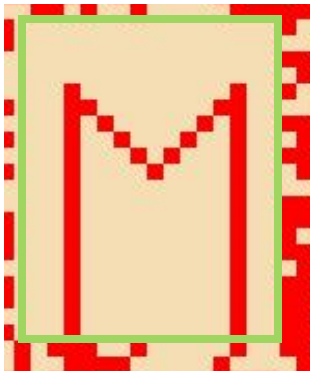
2nd Write



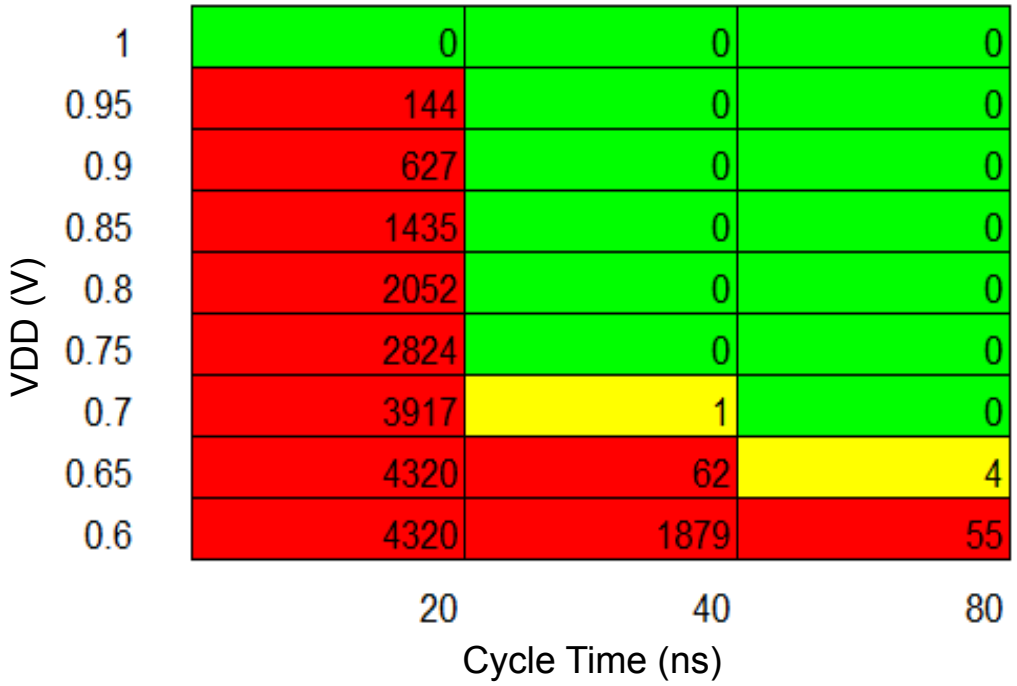
3rd Write



4th Write



14nm Hardware Results - Schmoo and retention



Prototype is functional using VDD = 0.7V

Projected Charge loss for 10 year product life: <35%

Outline

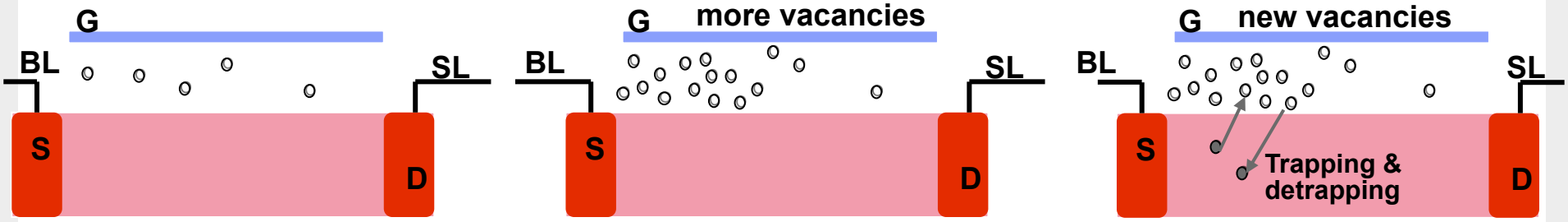
- Introduction
- Embedded Nonvolatile Memory Applications
- Charge Trap Transistor (CTT) Technology
- Macro Overview
- Hardware Results
  - SOI
  - 14nm bulk FinFET
- Future scope
- Summary

Future : Endurance Improvement Concept

Initially few vacancies

Higher gate voltage & long time to create more vacancies

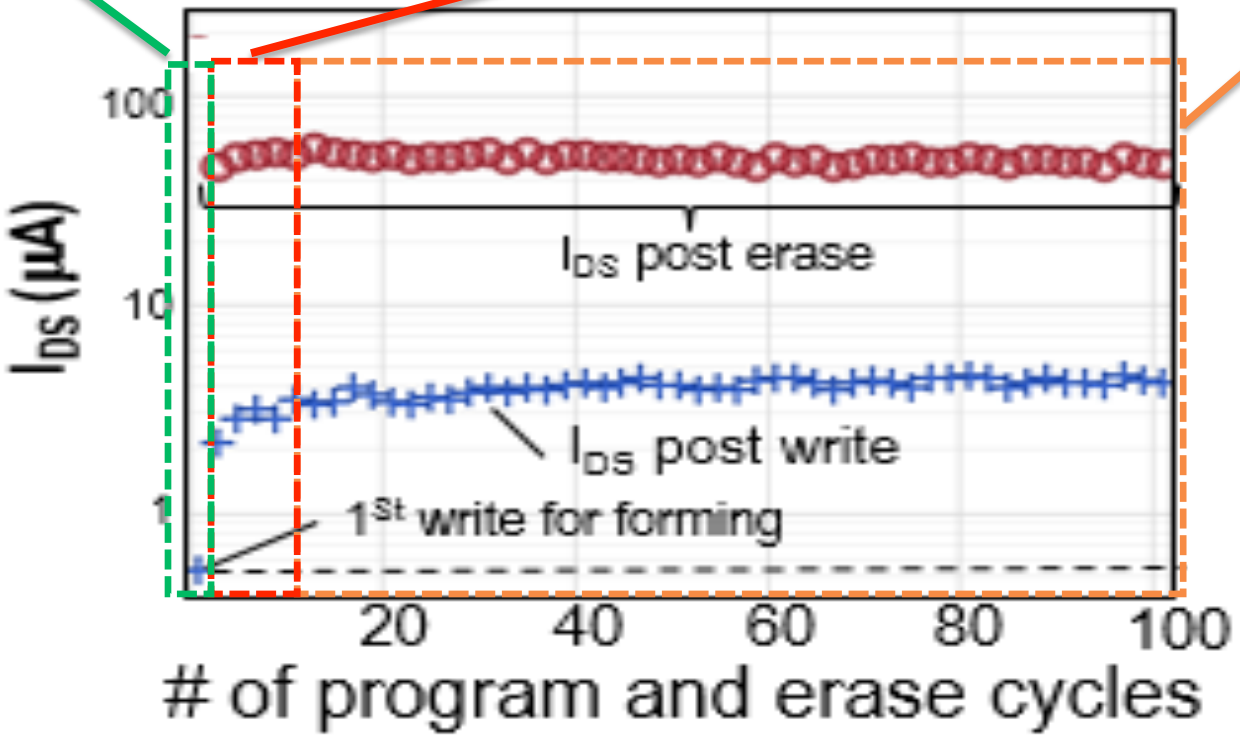
Low gate voltage & short time to avoid new vacancies



Process Phase

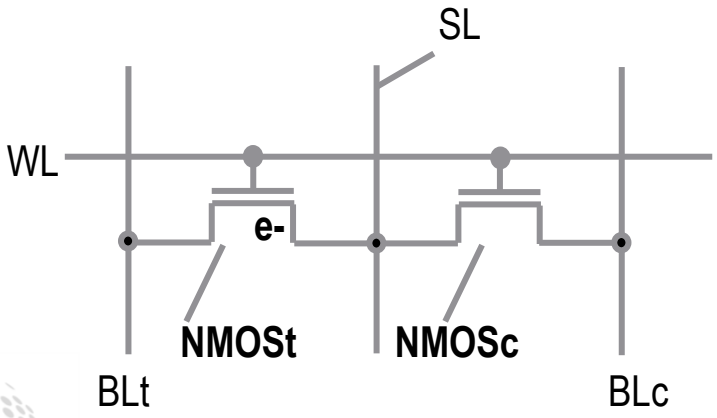
Manufacturing Test Phase

User Phase

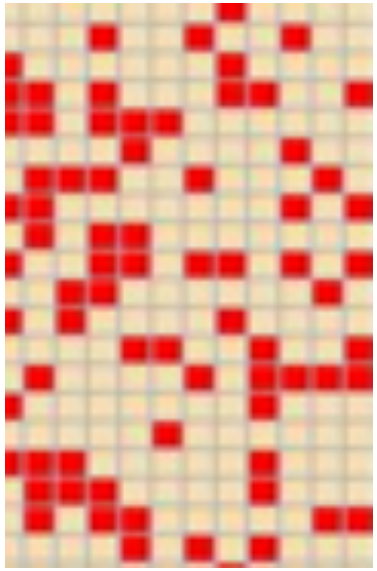


Future : Hardware-Based Security

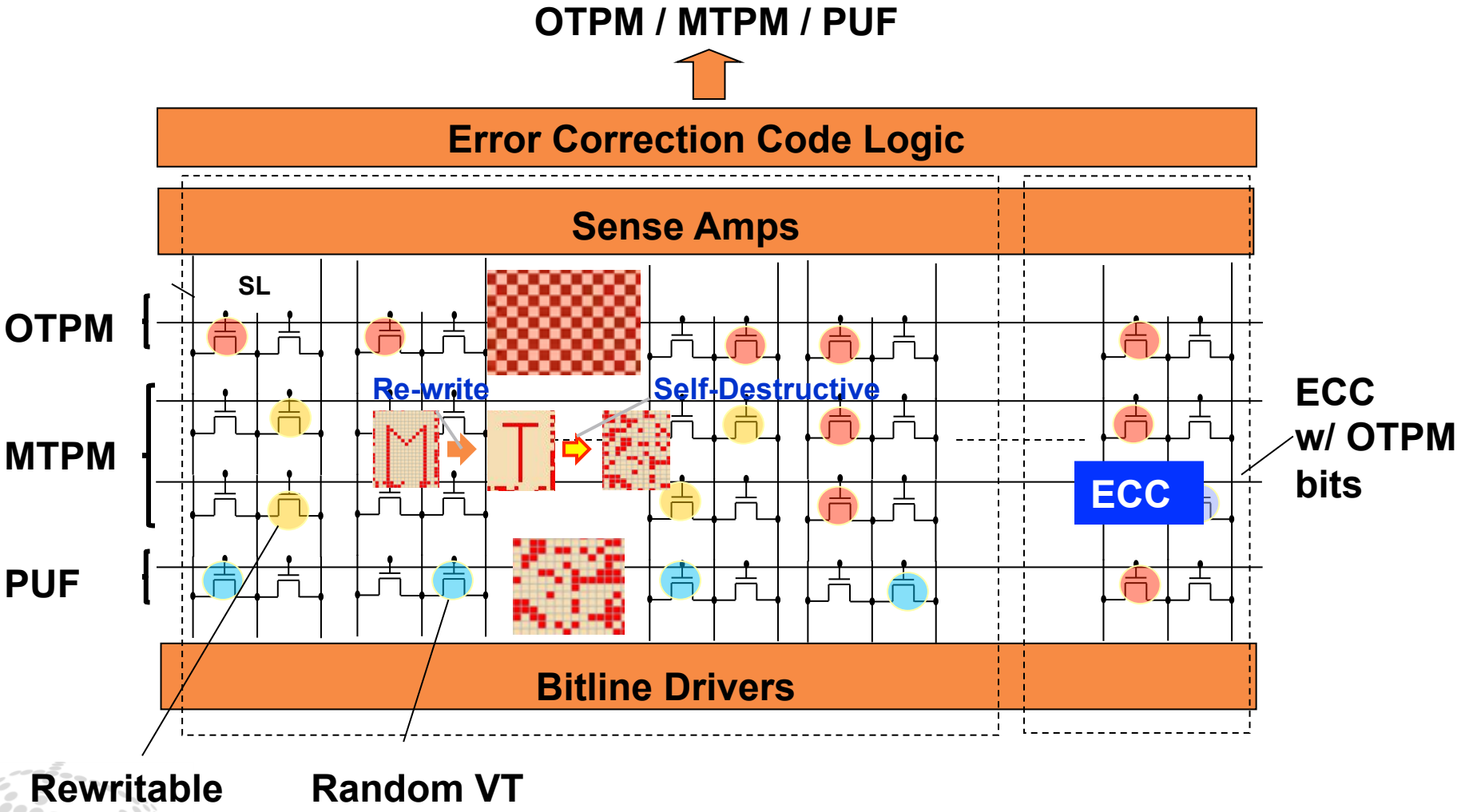
- Programmed bits physically invisible
- Enable Physically Unclonable Fuse (PUF) for a secure product source solution.
- Create a unified memory array, using these three modes of operation
  - OTPM Mode: Used for stable non-volatile bits
  - MTPM Mode: Self Destructive Re-writable Memory non-volatile bits
  - Intrinsic ID Mode: Physically Unclonable Fuse (PUF).



Twi Cell Memory Cell



Future : Unified memory using CTT MTPM



Outline

- Introduction
- Embedded Nonvolatile Memory Applications
- Charge Trap Transistor (CTT) Technology
- Macro Overview
- Hardware Results
  - SOI
  - 14nm bulk FinFET
- Future scope
- Summary

Summary

- MTPM demonstrated in both SOI and Bulk FIN
- Scalable standard Hi-K logic process
  - Process as it is (ex. No mask adder)
  - Operated in logic compatible voltages
  - ~30X more dense than eFUSE
- Future : MTPM fundamental concepts
  - Endurance improvement, using forming approach
  - Unified memory, using OTPM, self-destructive MTPM, PUF.

Technology	32nm SOI	22nm SOI	14nm FIN Bulk
Cell	0.109μm^2 with 1.4nm Gox NMOS	0.144μm^2 with 1.2nm Gox NMOS	0.1411μm^2 with FIN NMOS
Macro Density	80Kb	64Kb	40Kb
Density/mm^2	~2Mb/mm^2	~2.5Mb/mm^2	~1.3*Mb/mm^2
Activation Energy	~1.4eV	-	~1.6eV

References

[1] H. Kojima et. al., IEDM, 2007, pp. 677–680.
[2] Y. Taito et. al., ISSCC, pp. 132-133, Feb. 2015.
[3] M. Jefremow et. al., ISSCC, pp. 216-217, Feb. 2013.
[4] M. Ueki et. al, VLSI Tech., pp. 108-109, 2015.
[5] G. Uhlmann et. al., ISSCC, pp. 406-407, Feb. 2008.
[6] Zicheng Liu et. al., ISNE pp. 1-3, 2015
[7] C. Kothandaraman et. al., IRPS, 2015 pp. MY2.1-MY2.4
[8] F. Khan et. al., IEEE EDL 37 pp. 88-91, Jan. 2016.
[9] Kenji Noda, Using Hot Carrier Injection for Embedded Non-volatile Memory NSCore, Inc. White Paper http://www.nscore.com/images/WhitePaper_081002.pdf

Acknowledgements

Darren Anand
John Fifield
Sami Rosenblatt
Xiang Chen
Krishnan Rengarajan
Giuseppe Larosa
Norman Robson
Jim Pape
Daniel Berger
Dan Moy
Robert Katz
Yoann Mamy Randriamihaja
Zakariae Chbili
Andreas Kerber
Raman Kodhandaraman

This work at UCLA is partially supported by the Defense Advanced Research ProjectsAgency (DARPA). The views, opinions, and/or findings contained in this article are those of the author(s) and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government

Thank You