2017 VLSI TECHNOLOGY SYMPOSIUM KYOTO

Embedded Memory Design

Memories differentiate Microcontoller Solutions

Thomas Jew

NXP

2017 VLSI TECHNOLOGY SYMPOSIUM KYOTO

Outline

- What are Embedded Memories
- How are Embedded Memories Used
- Why are Embedded Memories Important
- Examples of Differentiating Embedded Memory Design Considerations
- Emerging Memories for Embedded Microcontrollers
- Summary

2017 VLSI TECHNOLOGY SYMPOSIUM KYOTO

Outline

- **What are Embedded Memories**
- How are Embedded Memories Used
- Why are Embedded Memories Important
- Differentiating Embedded Memory Design Considerations
- Emerging Memories for Embedded Microcontrollers
- Summary

Embedded Memories

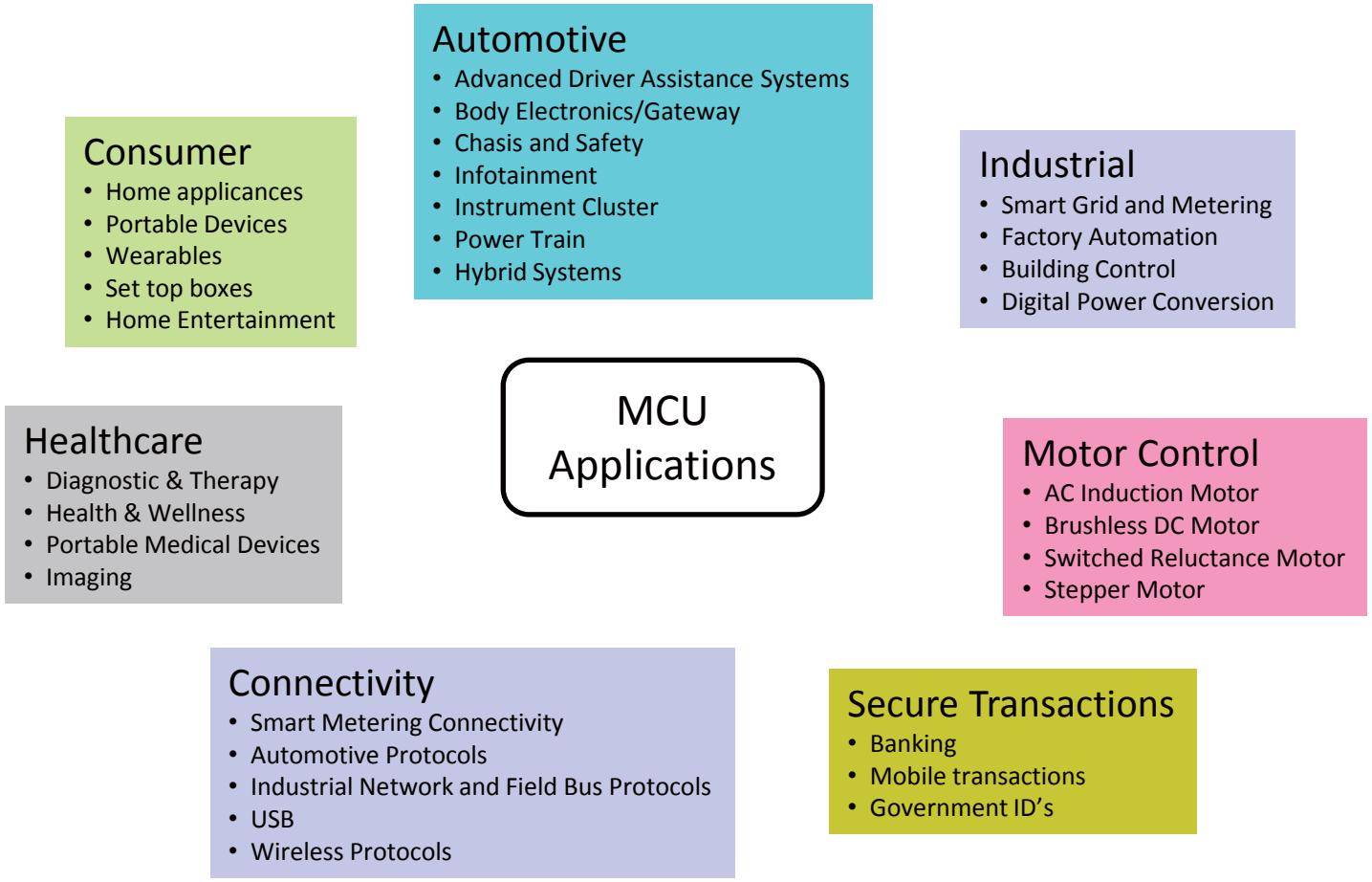
- Embedded memories are memories that are integrated into microcontroller devices
  - SRAM
  - ROM
  - eFuse
  - Flash

- Embedded Memories are found in a variety of microcontroller devices across a broad range of technology nodes

2017 VLSI TECHNOLOGY SYMPOSIUM KYOTO

Microcontroller Applications

Many microcontrollers (MCU's) available for a wide range of applications



Outline

- What are Embedded Memories
- **How are Embedded Memories Used**
- Why are Embedded Memories Important
- Differentiating Embedded Memory Design Considerations
- Emerging Memories for Embedded Microcontrollers
- Summary

Embedded Memory Usage

Nonvolatile memories are the “brains” of the MCU

- For devices with no embedded flash
  - eFuses are used to store part configuration and ID information
  - ROM may be used to store boot and application code
- For Flash-based devices
  - Flash contains the learning (instructions and data) that defines the MCU’s function when it is active and retains that learning when powered off
  - Boot and application code, diagnostics, data tables, application parameters, logs, passcodes, ID’s, etc. are stored in Flash
  - Offers the ability to make running application updates during production and field updates

Embedded Memory Usage

SRAM is the high speed “working” memory of the MCU

- Working data set of the MCU application is stored in SRAM
  - Scratch pad for computations and maintaining copies of the application state during context changes
  - Queues for incoming and outgoing message data
  - Buffer for data sampling and measurements
- SRAM is used to bridge high speed CPU Cores and slower flash memories to enable more processing capability
  - Instruction and Data Caches are used to hold commonly used code and data on demand for CPU’s to consume
  - Look up tables are stored in SRAM local to the CPU (Tightly Coupled RAM) for faster access or in general system RAM for shared system level access
  - Virtual data structures are also maintained in SRAM

2017 VLSI TECHNOLOGY SYMPOSIUM KYOTO

Outline

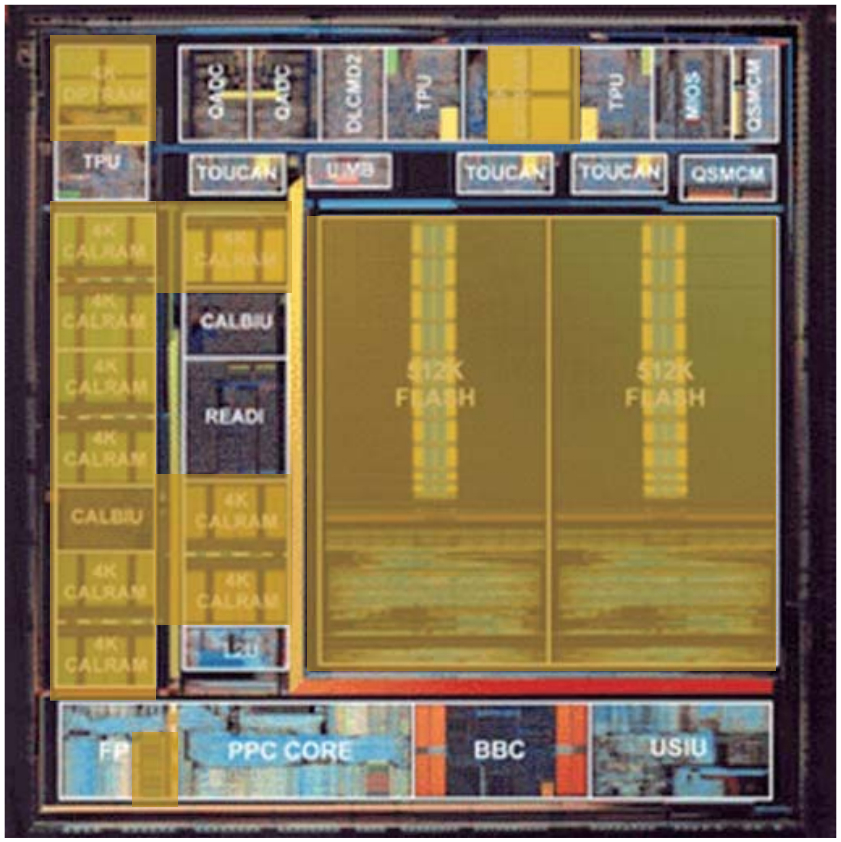
- What are Embedded Memories
- How are Embedded Memories Used
- **Why are Embedded Memories Important**
- Differentiating Embedded Memory Design Considerations
- Emerging Memories for Embedded Microcontrollers
- Summary

Memories Enable the Application

- Amount of embedded memories limit the capacity of what device can do
  - Code storage footprint defines the scope of the microcontroller device
  - General system RAM determines the complexity of the working stack
  - Need sufficient SRAM for temporary storage and nonvolatile memory for persistent memory storage

- Memory characteristics can also limit device application
  - Power consumption
  - Read access time / data rate
  - Write/Erase times and endurance of flash
  - Wake up times
  - Robustness of memories to application conditions

Embedded Memories Are Expensive



- Memories are a predominant part of the cost of microcontroller devices
  - Embedded memories occupy a significant percentage of device silicon area
  - Nonvolatile memories come with additional processing costs
  - Memories test cost on a high performance logic tester can be expensive
  - Use of novel memory technologies can come with IP/Technology royalty costs

2017 VLSI TECHNOLOGY SYMPOSIUM KYOTO

Outline

- What are Embedded Memories
- How are Embedded Memories Used
- Why are Embedded Memories Important
- **Differentiating Embedded Memory Design Considerations**
- Emerging Memories for Embedded Microcontrollers
- Summary

Embedded Memory Cases

- Internet of Things
  - Focus on “end node” application requirements
  - Power / Performance
- Automotive
  - Extreme operating conditions
  - Zero Defects
  - Design for Test
  - Functional Safety
- Security
  - System isolation is natural for embedded flash solution

IOT Design Considerations

IOT Microcontroller Device Application Characteristics

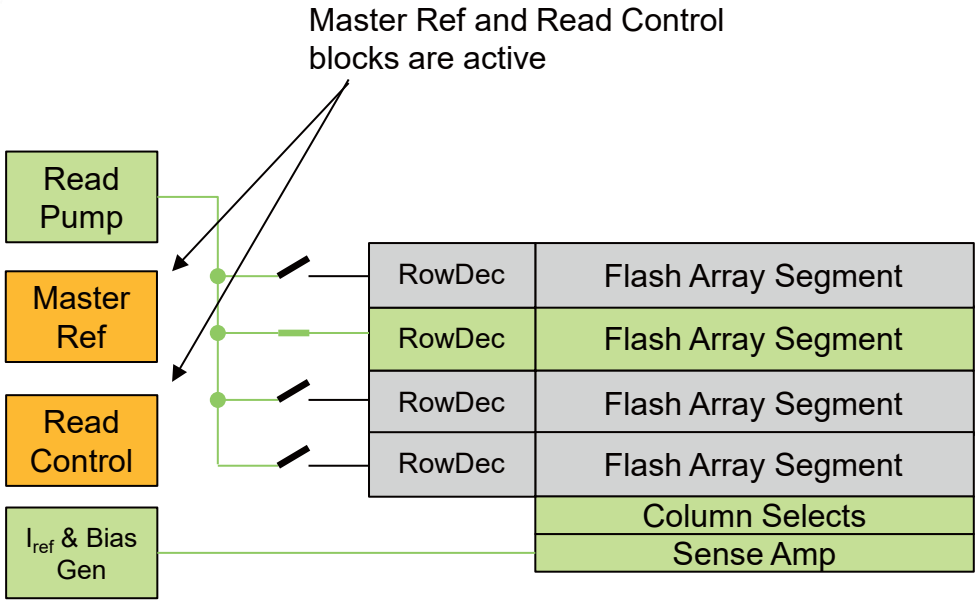
- Spend majority of product lifetime in deep power down mode
- Periodically wake up and check to see if something needs to be done with quick on/off and low performance
- Performs primary application task with high priority and performance with premeditated wake up for fast completion and go back to sleep

Task	Memory	Performance	Power	Wake-up / Power Down
Waiting	SRAM	None	Deep power down / State Retention	Fast power down
	Flash	None	Deep power down	Fast Power Down
Checking	SRAM	Low	Very low power	Fast wake up
	Flash	Low	Very low power	Fast wake up
Active Tasks	SRAM	High Performance Run	Run power	Moderate wake up
	Flash	Fast reads with P/E of data	Run power	Moderate wake up

Flash Low Power Modes

- Deep Power Down
  - Power down all flash circuits
- Very Low Power Mode
  - Maintain essential circuits on for reading
  - Keep array nodes in discharged state between reads
  - Power up only array segments for reads when requested

Very Low Power Mode for Flash



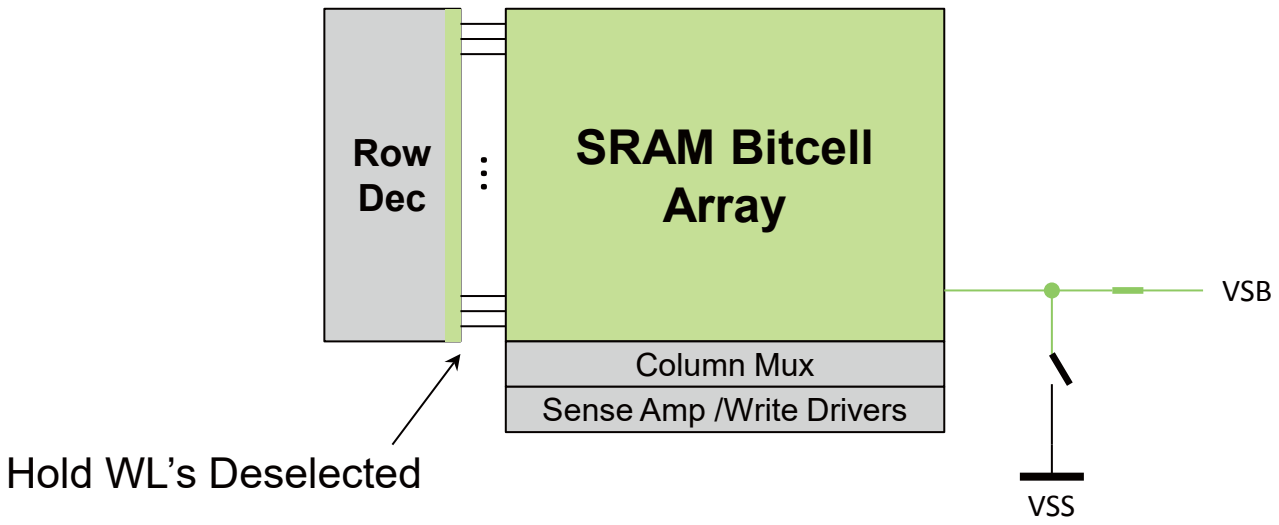
Very Low Power Read Operation

1. Switch on read pump, Iref/Bias Gen
2. Connect addressed flash array segment RowDec block to Read Pump
3. Energize Column Selects and Sense Amp's
4. Wait for Read Pump and Iref/Bias Gen outputs to stabilize
5. Execute read
6. Shut down read pump, Iref/Bias Gen, and associated rea
7. Remove power from column selects and sense amp's

- Minimal circuits in flash are powered for "checking" tasks while idle
- Only required circuits are energized to read flash
- By minimizing circuits to be energized for flash reads, wake up time can be very fast

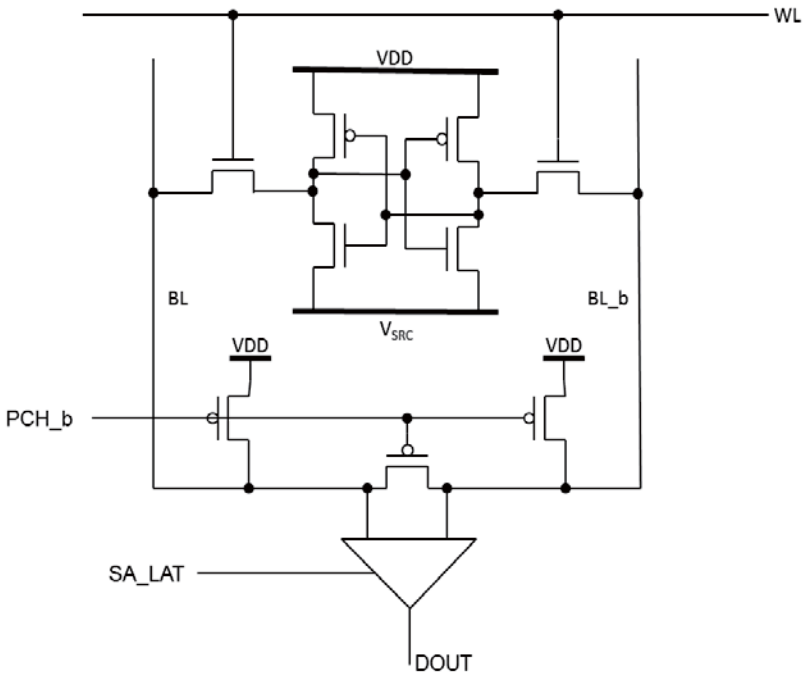
- All circuits in flash are continuously powered for high speed run modes

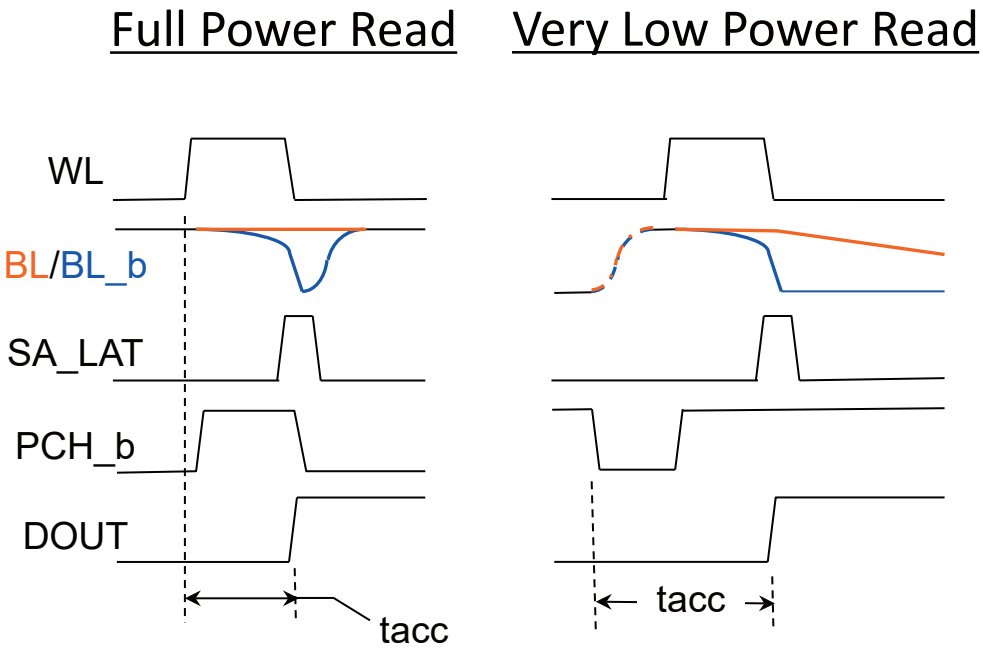
State Retention for SRAM



- By holding all WL's deselected, SRAM array content can be protected so that SRAM periphery can be powered down
- By controlling the SRAM array source voltage in state retention mode, power can be further reduced while retaining SRAM memory state

Very Low Power Mode for SRAM





- Automatic precharge of BL/BL_b can be disabled in low power read modes, trading off leakage power vs. read access time
- Depending on read performance requirements in low power modes, additional power mitigations like managing array source (VSRC) biasing during reads can further reduce power

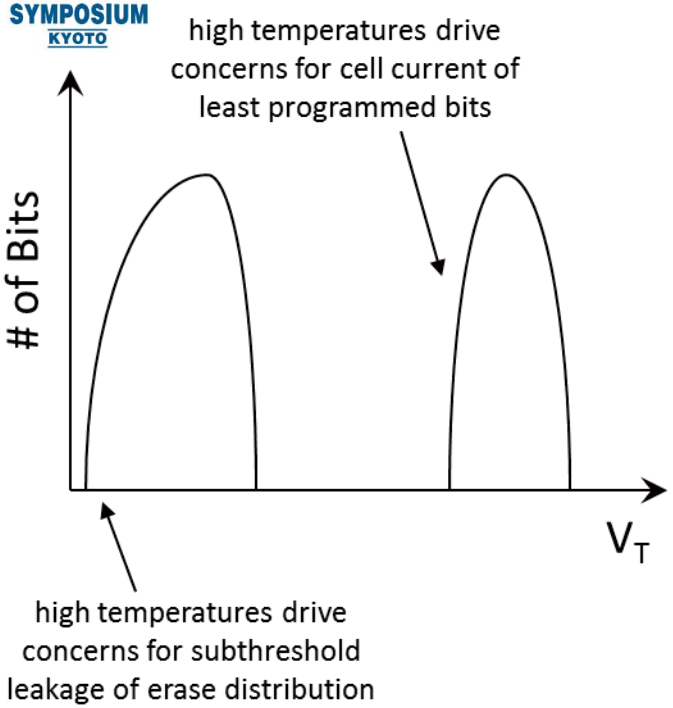
Automotive Requirements

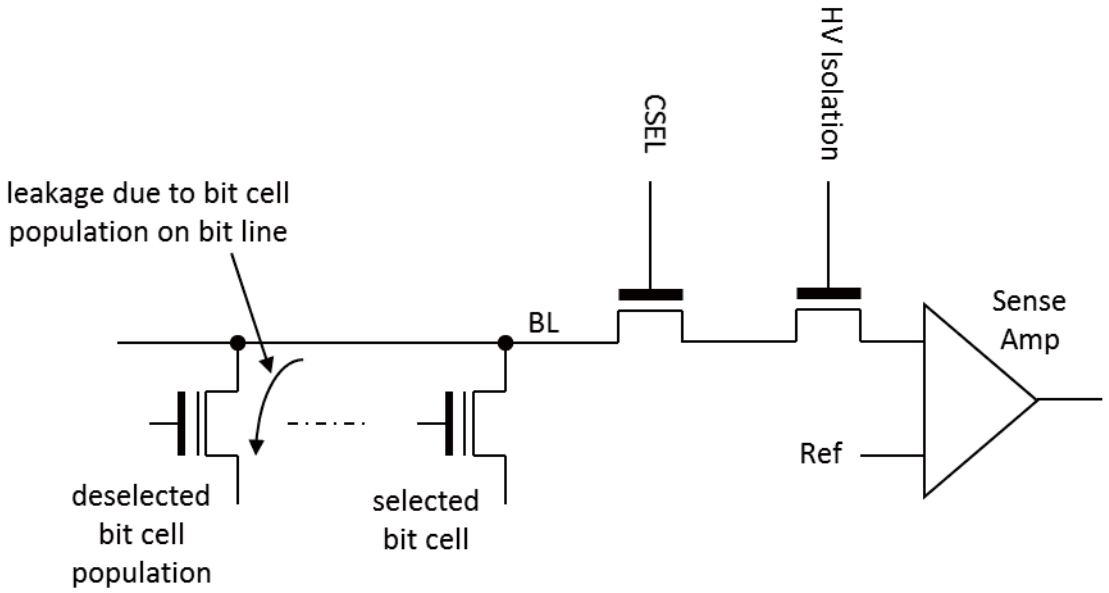
- Extreme operating temperature
  - AEC-Q100 specification calls for different operating temperature ranges

Temperature Grade	$T_{ambient}$ (Min)	$T_{ambient}$ (Max)
Grade 0	-40C	150C
Grade 1	-40C	125C
Grade 2	-40C	105C
Grade 3	-40C	85C
Grade 4	0C	70C

  - Self-heating can mean $T_{J,Max}$ = 175C or more for Grade 0 product circuit operation

Increased Temperature BL Leakage



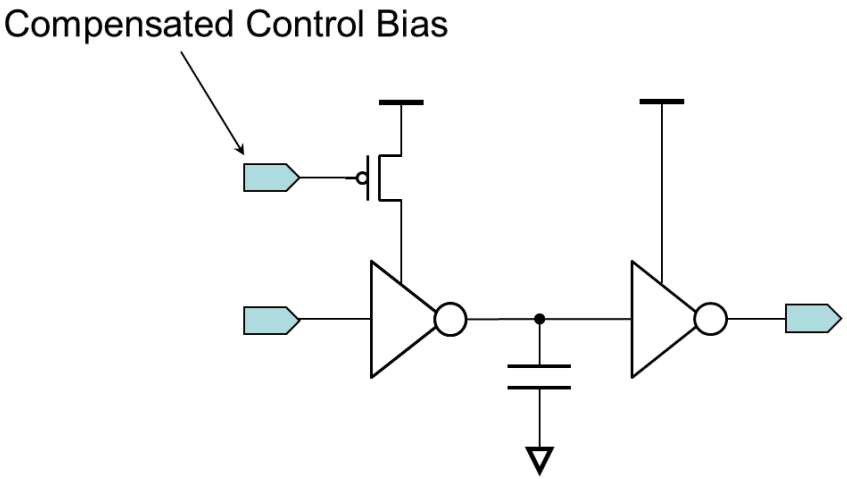


- Bit line leakage from unselected bit cells increases with higher temperatures for both read and program operation in Flash
  - Reduce number of bits on a bit line
  - Use negative deselect biases for unselected rows
  - Apply bit line leakage compensation techniques, such as current cancellation
  - Push Vt distributions higher and/or compact erase Vt distribution more

Hostile SoC Environments

Critical self-timed circuits have increased variation due to wider operating temperature range

- standard “delay element” timing generators cause marginalities at extreme temperature/voltage operating conditions
- temperature compensated RC based timing circuits
- bandgap controlled timing circuits



Zero Defects

- Automotive Zero Defect Requirements mean no device operation failures in the field over product lifetime!

- Margin design for aging effects
  - Transistor HCI, NBTI, PBTI effects on g_m and V_T

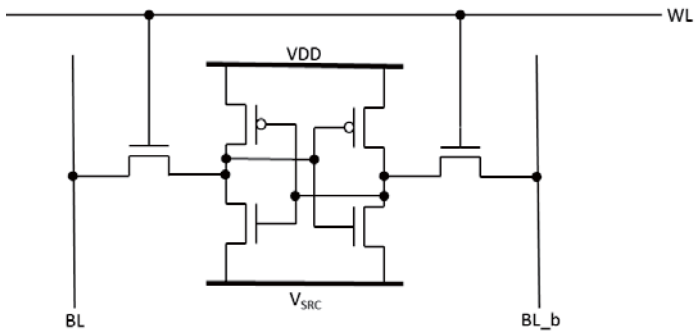
- Minimize TDDB on transistor oxides as well as Interconnect dielectrics
  - Minimize overdrive conditions
  - Apply rules for voltage dependent physical design rules, especially for overdrive condition
    - Metal-to-Metal spacing, Active-to-Active spacing, Active-to-Gate spacing, minimum L, Gate-to-Gate spacing, guardring / ties, etc.

Zero Defects

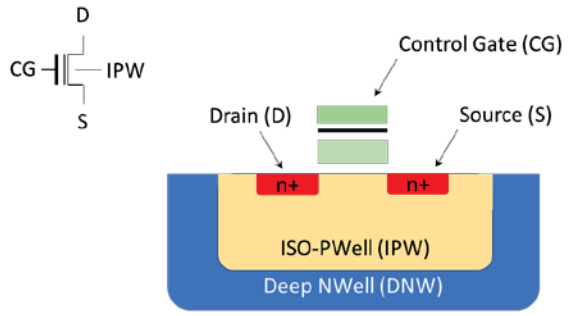
- Account for physical layout effects for all critical circuits
- Observe design for manufacturability guidelines
  - Don’t use minimum devices in sensitive circuits
  - Redundant vias and contacts
  - Redundancy repair
- Eliminate variation!

Design for Test

- Implement test modes to stress memory devices to eliminate latent defects



- Drive voltage on BL / BL_b and WL to stress pass devices
- Modulate bitcell VDD and control BL or BL_b to stress pull up and pull down devices in the bitcell



- Drive high voltage across CG and IPW to stress poly dielectric stack
- Drive high voltage on D to stress D-to-Gate dielectrics and D junctions
- Drive high voltage on S to stress S-to-Gate dielectrics and S junctions

- Include measurement modes to measure effects of stresses

Zero Defects

Apply ECC!

Automotive Functional Safety

- ISO 26262 Standard
  - International Standard for Functional Safety applicable to safety-related automotive systems
  - Provides a risk-based method for safety system development
  - Uses an FMEDA (Failure Modes, Effects and Diagnostic Analysis) tool
    - FMEDA is similar to an FMEA but adds concepts of failure rate and detection of failures

- Functional Safety is a layer of design measures implemented in a system to avoid unreasonable risks due to hazards caused by malfunctioning behavior of an electronic system

Auto Safety Integrity Level
A - Lowest ASIL Level
B
C
D – Highest ASIL Level

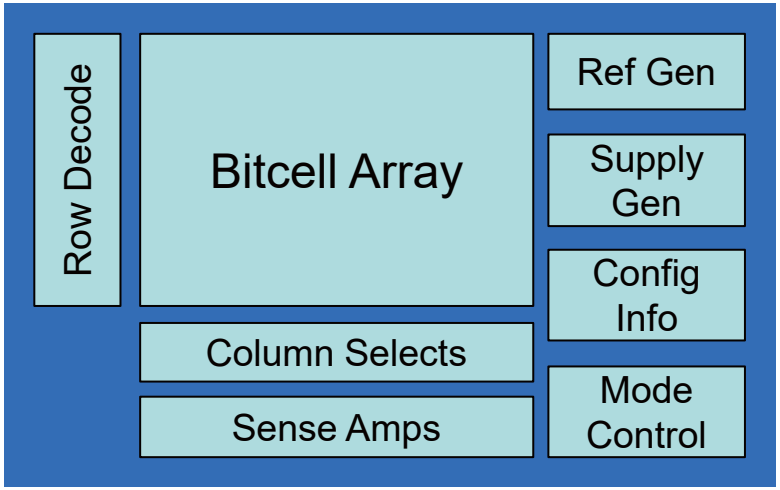
Fault Types of Concern

- Need to comprehend safety concept and whether memory function is safety relevant or not
- If memory is safety relevant, assess memory subsystem for SPF, LF, and CCF using FMEDA process
- Implement design changes and apply safety countermeasures to eliminate failure

Failures	Behavior	Handling
Single Point Failure (SPF)	Immediate potential to cause a hazard	Quick detection
Latent Failure (LF)	Can become a hazard, if activated	Periodic detection
Common Cause Failure (CCF)	Causes multiple circuits to cause a hazard	Quick detection

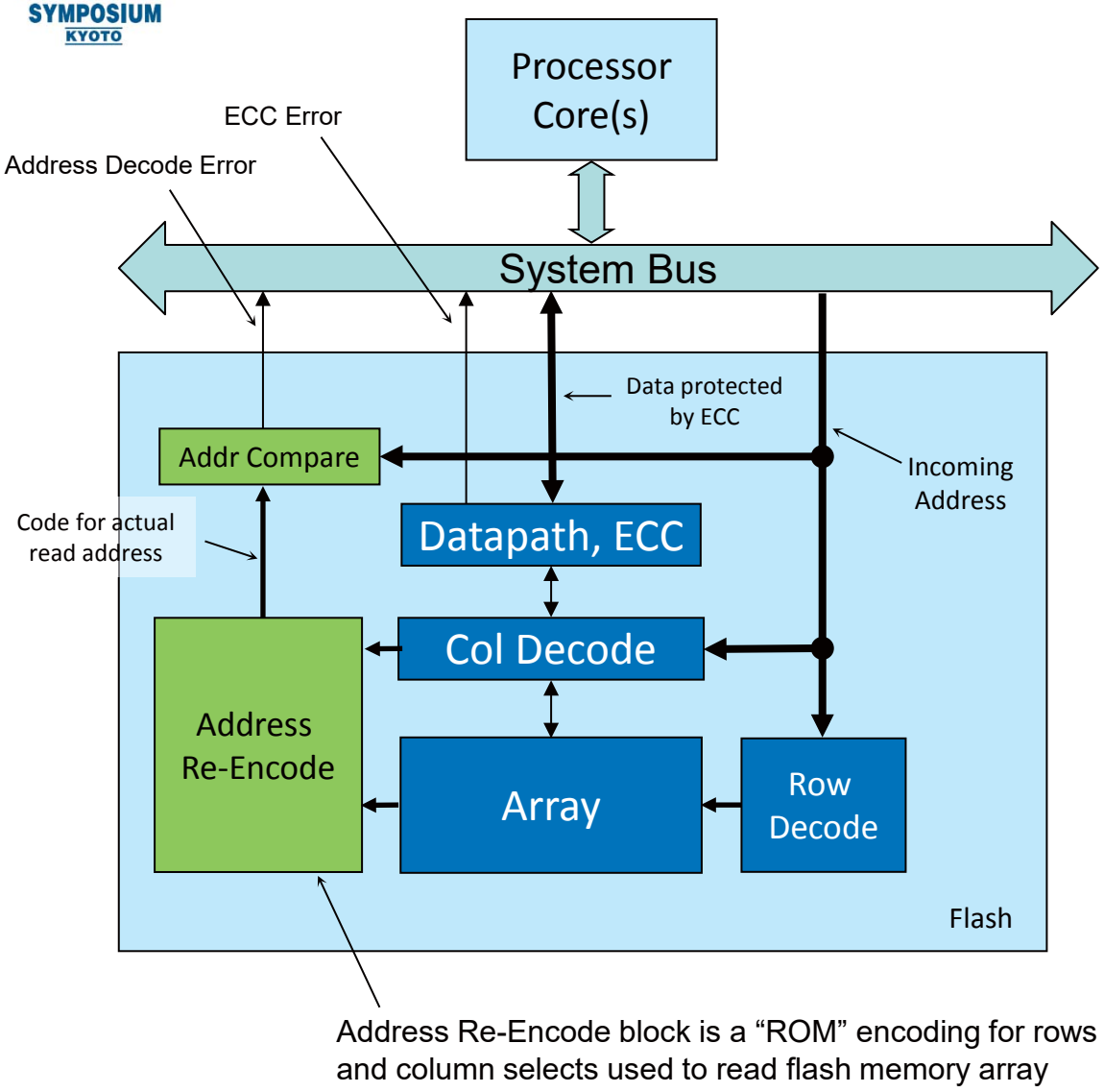
Failure Mode Analysis

- Apply SPF, LF, and CCF models for analysis of memory failures



Failure	Potential Design Mitigations
Decode failure	Decode Checker, Periodic Signature Diagnostic
Bit failure	Parity, ECC, MISR, Memory BIST, Margin Checks
Sense amp failure	Parity, ECC, MISR, Memory BIST
Config bit failure	Voting Flip-Flops
Supply failure	Supply detector
Reference failure	Reference monitor
Mode Control failure	Voting control logic, LBIST

Safety Measure Examples



Hardware Safety Measures

- ECC protects Flash Content
- Address Re-encode values of the selected row & column are compared against requested address to detect addressing faults

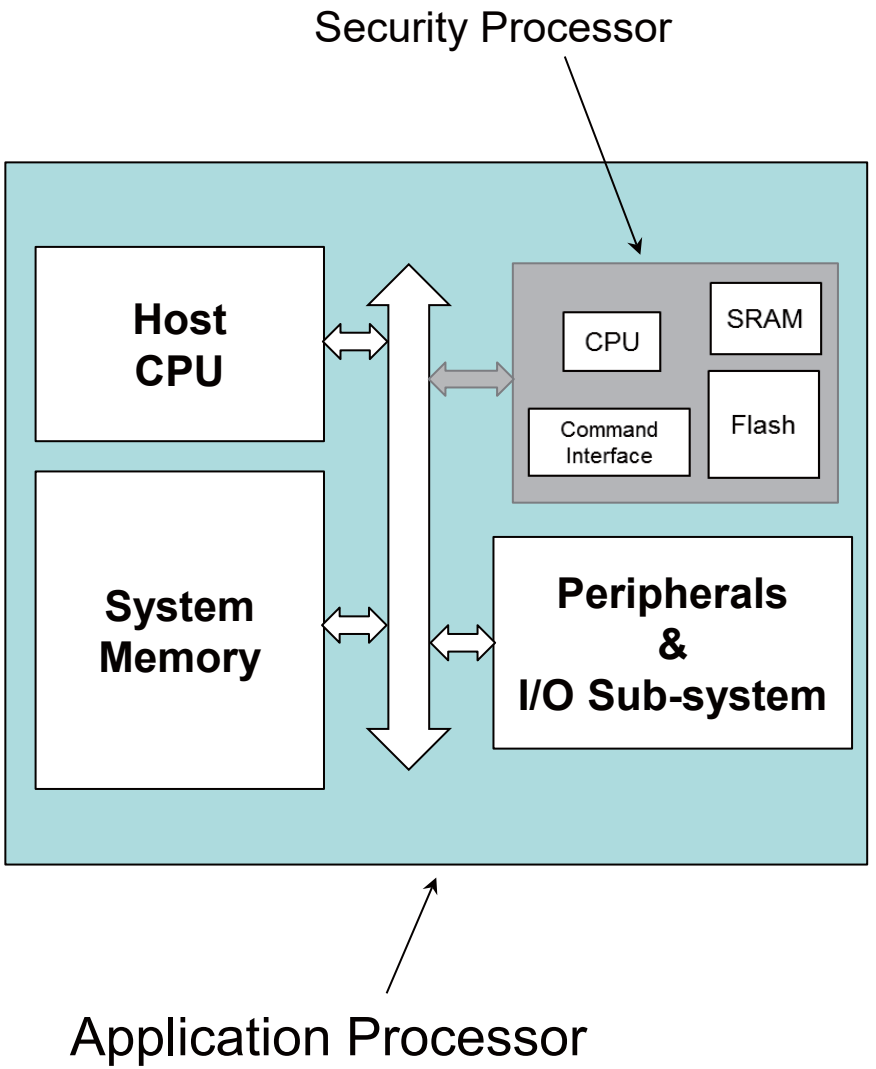
Flash reads are protected during code execution against SPF's associated to read failures

Similar schemes can be applied to SRAM reads

Secure Products

Embedded flash is key to enabling secure element devices

- Embedded flash allows for a self contained, autonomous security processing system to monitor application
- In system program/erase allows for dynamic decisions and event logging
- Charge storage memories are not optically readable and retain state through power cycling



2017 VLSI TECHNOLOGY SYMPOSIUM KYOTO

Secure Products (cont.)

Security Concerns	Memory Design Requirements	Potential Solutions
Integrity of Memory Content	SRAM and Flash content is robust	• Cell stability and operating margin to protected operating conditions • ECC protection of data content
Observability of memory operation	Manage detectability of memory contents and content changes during operation	• Mask power signatures for read and write operation • Fast atomic flash program/erase operation • Mask passive probing through shielding of critical signals within memory
Ensure proper security code execution	Detect erroneous code fetches from flash	• Monitor critical biases and control signals required for proper read operations • Monitor address decoding faults • Protect read data with ECC
Protect memory implementation details	Firewall memory design process	• Implement secure environment for memory design activities

Outline

- What are Embedded Memories
- How are Embedded Memories Used
- Why are Embedded Memories Important
- Differentiating Embedded Memory Design Considerations
- **Emerging Memories for Embedded Microcontrollers**
- Summary

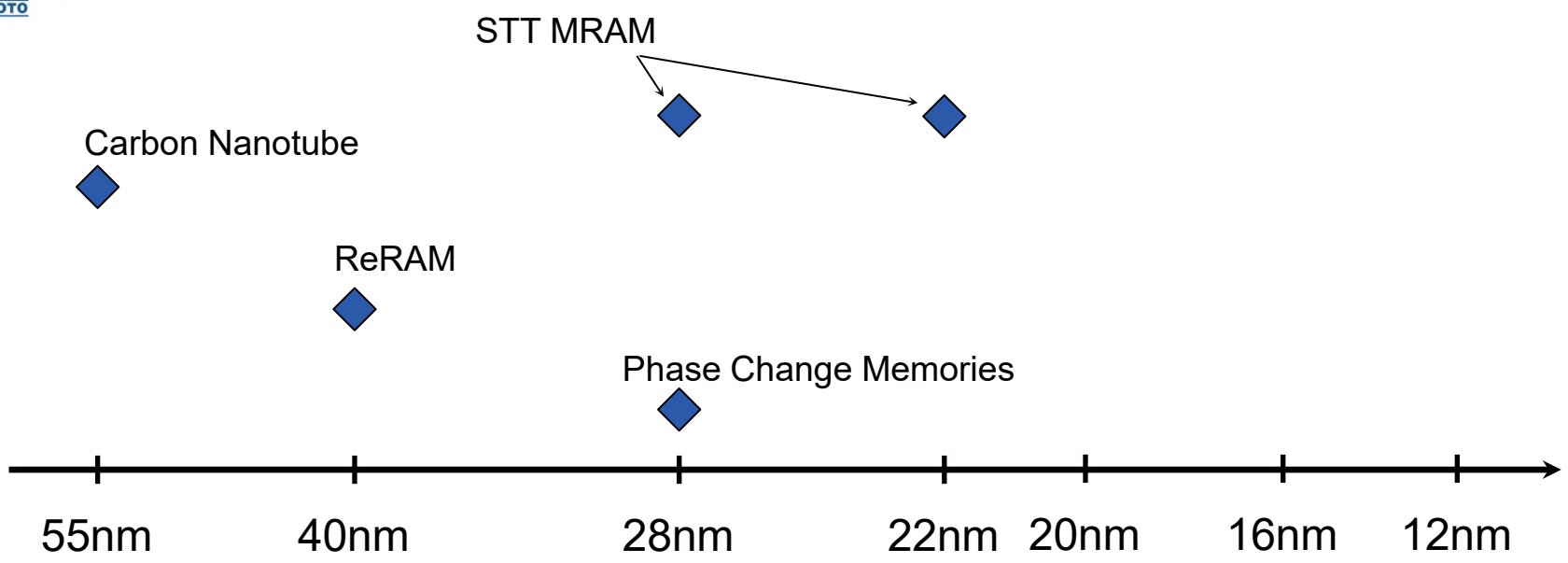
2017 VLSI TECHNOLOGY SYMPOSIUM KYOTO

Embedded Emerging Memories

- Current microcontroller embedded memory offerings are still based on traditional 6T SRAM and charge storage Flash solutions

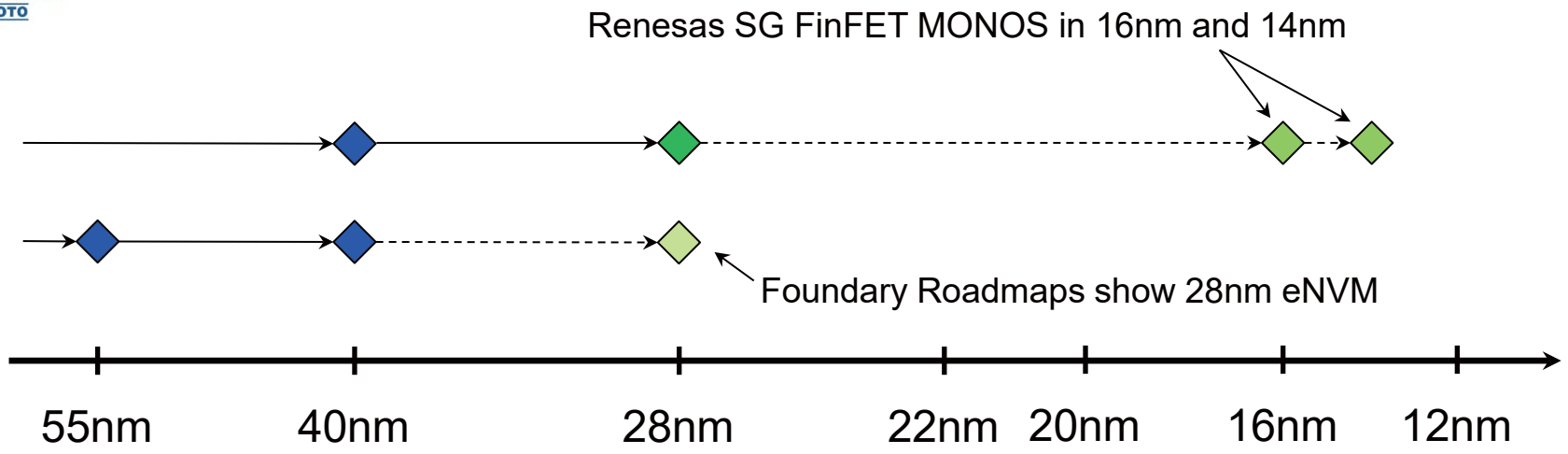
- Industry is investing heavily into new memory solutions based on assertions that:
  - Embedding conventional charge storage flash technology are becoming prohibitive in high performance CMOS logic processes, like high-K metal gate last and Fin-FET processes
  - Conventional 6T SRAM scaling is getting more difficult, especially as density requirements are growning

Embedded Emerging Memories



- Many announcements in 2016 for emerging memories for embedded MCU applications
  - Primary focus is non-volatility to replace flash
  - Secondary focus is to provide high density memories with limited non-volatility to address area & power challenges with 6T SRAM

Challenges for Emerging Memories



- Conventional embedded SRAM and Flash memory technologies are pushing ahead
- Emerging Memories will have to compete with established solutions and cost structures

2017 VLSI TECHNOLOGY SYMPOSIUM KYOTO

Challenges for Emerging Memories

- STT MRAM appears to be getting a lot of industry attention
  - Many consider it to be the leading emerging memory for embedded non-volatile storage
- MRAM's challenge is that it is neither Flash nor SRAM
- Design solutions to address MRAM behaviors present challenges
  - Write power is still high compared to SRAM for persistent memory solutions
  - MTJ read window is very small – driving high area overhead requirements for sensing
  - 2-bit correcting ECC is needed to address probabilistic write failures associated to high endurance RAM functions AND read marginalities
  - Data retention during solder reflow may be a challenge

2017 VLSI TECHNOLOGY SYMPOSIUM KYOTO

Outline

- What are Embedded Memories
- How are Embedded Memories Used
- Why are Embedded Memories Important
- Differentiating Embedded Memory Design Considerations
- Emerging Memories for Embedded Microcontrollers
- **Summary**

Summary

- Know your application targets and market requirements
  - Understand product use cases and mission profiles
- Design considerations of technology behaviors are becoming more essential for successful products
- Circuit elements and devices used in other IP blocks on the Microcontroller may be fair game to use in Memory design
- Pay attention to the details

2017 VLSI TECHNOLOGY SYMPOSIUM KYOTO

Thank You for Your Kind Attention

2017 VLSI TECHNOLOGY SYMPOSIUM KYOTO

References

[1] T. Jew, “Embedded Microcontroller Memories - Application Memory Usage”, Memory Workshop (IMW), 2015 IEEE International

[2] K. Baker, “Embedded Nonvolatile Memories: A Key Enabler for Distributed Intelligence”, Memory Workshop (IMW), 2012 4th IEEE International

[3] Wanli Chang, “Memory-Aware Embedded Control Systems Design”, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (Volume: 36, Issue: 4, April 2017)

[4] James Pallister, “Optimizing the flash-RAM energy trade-off in deeply embedded systems”, 2015 IEEE/ACM International Symposium on Code Generation and Optimization (CGO), Feb. 2015

[5] Yoshisato Yokoyama, “40nm Ultra-low leakage SRAM at 170 deg.C operation for embedded flash MCU”, Quality Electronic Design (ISQED), 2014 15th International Symposium on Quality Electronic Design, March 2014

[6] Vita Pi-Ho Hu, et. al., “Analysis of GeOI FinFET 6T SRAM Cells With Variation-Tolerant WLUD Read-Assist and TVC Write-Assist”, IEEE TRANSACTIONS ON ELECTRON DEVICES, VOL. 62, NO. 6, JUNE 2015

[7] Hulfang Qin, ”SRAM leakage suppression by minimizing standby supply voltage”, Quality Electronic Design, 2004. Proceedings. 5th International Symposium on Signals, Circuits and Systems. Proceedings, SCS 2003. (Cat. No.03EX720)

[8] Yasuhiko Taito, “A 28 nm Embedded Split-Gate MONOS (SG-MONOS) Flash Macro for Automotive Achieving 6.4 GB/s Read Throughput by 200 MHz No-Wait Read Operation and 2.0 MB/s Write Throughput at Tj of 170° C”, IEEE Journal of Solid-State Circuits (Volume: 51, Issue: 1, Jan. 2016)

[9] S. Tsuda, “First demonstration of FinFET split-gate MONOS for high-speed and highly-reliable embedded flash in 16/14nm-node and beyond”, Electron Devices Meeting (IEDM), 2016 IEEE International

[10] Elena I. Vatajelu, “Adaptive Source Bias for Improved Resistive-Open Defect Coverage during SRAM Testing”, 2013 22nd Asian Test Symposium

2017 VLSI TECHNOLOGY SYMPOSIUM KYOTO

References

[11] S. Hamdioui, R. Wadsworth, J. Delos Reyes, A. J. van de Goor, "Importance of dynamic faults for new SRAM technologies, " IEEE European Test Workshop, 2003, pp. 29-34

[12] Md Imranur Rahman, “Performance evaluation and read stability enhancement of SRAM bit-cell in 16nm CMOS”, 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)

[13] M.J.M. Pelgrom, “Matching properties of MOS transistors”, IEEE Journal of Solid-State Circuits (Volume: 24, Issue: 5, Oct 1989)

[14] M. Miyamura, “SRAM critical yield evaluation based on comprehensive physical / statistical modeling, considering anomalous non-Gaussian intrinsic transistor fluctuations“, VLSI Technology, 2007 IEEE Symposium on VLSI Technology

[15] T. Andre, “ST-MRAM fundamentals, challenges, and applications”, Custom Integrated Circuits Conference (CICC), 2013 IEEE

[16] J. Slaughter, “Technology for reliable spin-torque MRAM products”, Electron Devices Meeting (IEDM), 2016 IEEE International

[17] J.K. Jerry Lee, “Protecting against emerging vmin failures in advanced technology nodes”, Test Conference (ITC), 2014 IEEE International

[18] Sheyang Ning, “Investigation and Improvement of Verify-Program in Carbon Nanotube-Based Nonvolatile Memory”, IEEE Transactions on Electron Devices (Volume: 62, Issue: 9, Sept. 2015)

[19] Glen Rosendale, “Storage element scaling impact on CNT memory retention and ON/OFF window”, Memory Workshop (IMW), 2014 IEEE 6th International

[20] Yuri Tkachev, “A new method for analysis of cycling-induced degradation components in split-gate flash memory cells”, Integrated Reliability Workshop Final Report (IRW), 2013 IEEE International