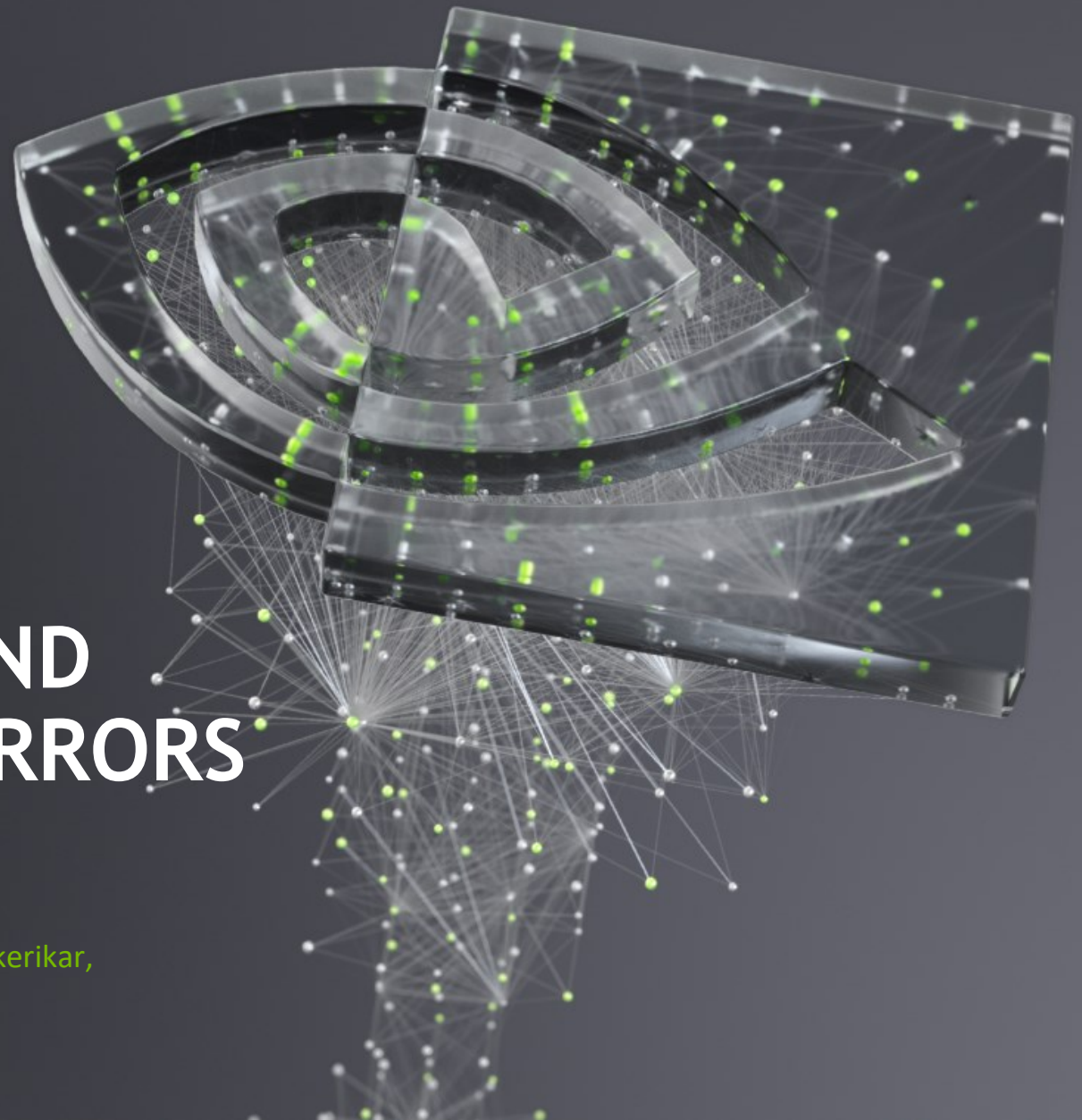


NVIDIA

CHARACTERIZING AND MITIGATING SOFT ERRORS IN HBM2 DRAM

Michael B. Sullivan, Nirmal Saxena,
Mike O'Connor, Donghyuk Lee, Paul Racunas, Saurabh Hukerikar,
Timothy Tsai, Siva Kumar Sastry Hari, Stephen W. Keckler

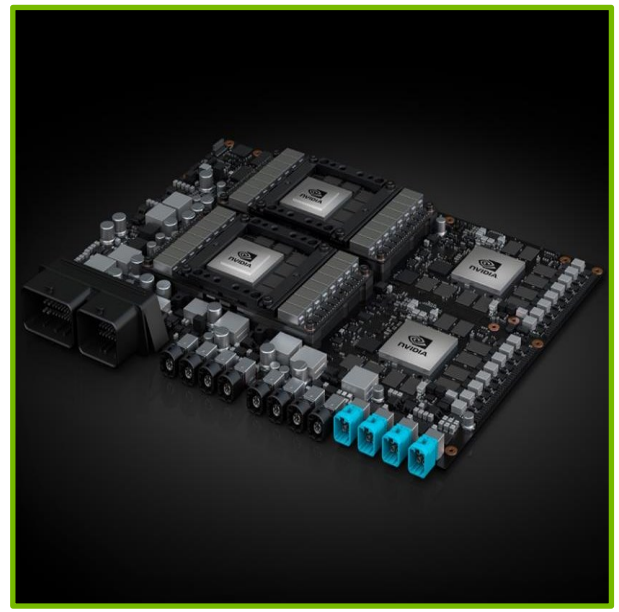
GPUs for High-Reliability Applications

Many Areas Demand High-Reliability Operation

HPC

Datacenter

Autonomous Vehicles



Unprotected DRAM is generally the most susceptible component to errors.

Error-correcting codes (ECC) detect/correct common error patterns.

Soft Error Concerns

Avoid Crashes (DUEs), Silent Data Corruption

Terrestrial radiation can lead to soft errors, where induced charge leads to data corruption.

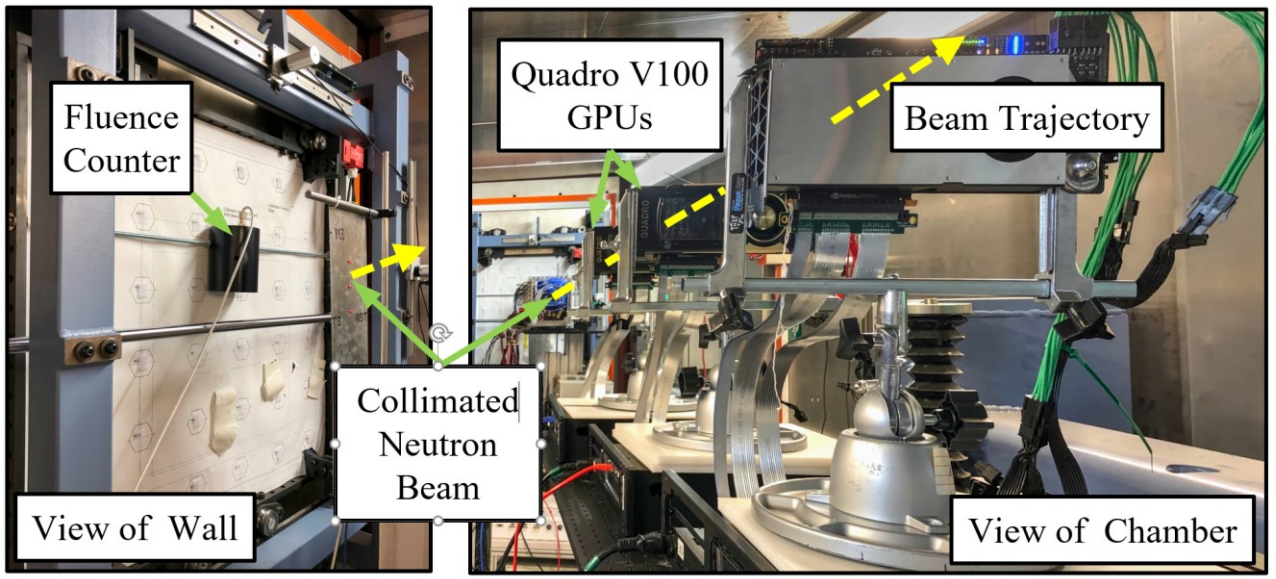
A soft error can result in:

1. A corrected error is caught and corrected by the ECC.
2. An error-related crash from a detected-yet-uncorrected (DUE) ECC result.
3. Silent data corruption (SDC). Either an error remains undetected, or ECC mis-corrects the wrong data.

To get the best protection, ECC should be tailored to the most prevalent error patterns!

Beam Testing for Characterization

High-Energy Neutron Beam Testing with a Specialized Microbenchmark



ChipIR Neutron Beam Experimental Setup

High-energy neutron beam.

Neutron energy spectrum tuned to that of terrestrial radiation.

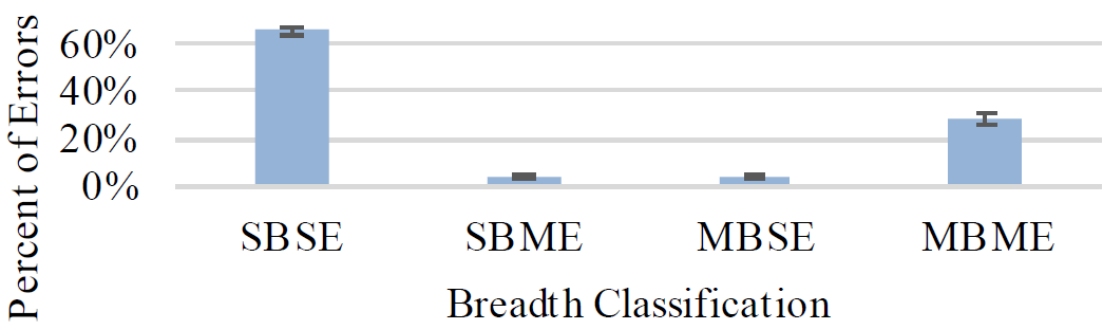
GPUs are placed in the beam.

We run specialized microbenchmarks on the exposed GPUs.

Acceleration factor: 252,000,000×.

Error Patterns: Breadth and Severity

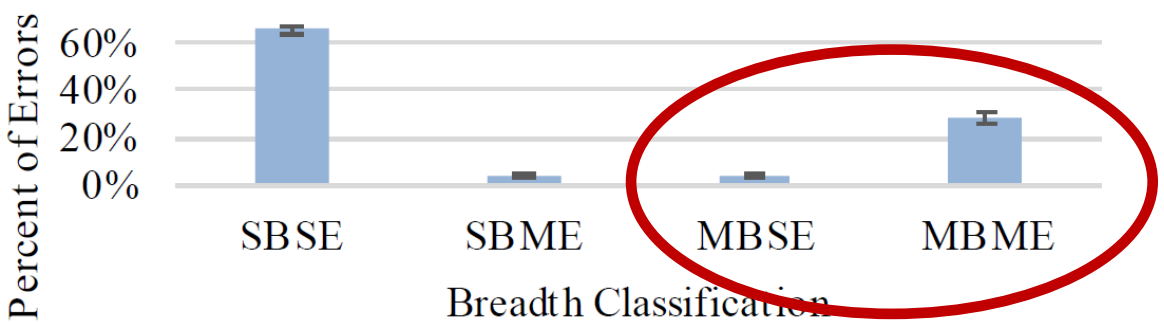
~26% of Errors Multi-Bit!



1. SBSE: Single-Bit, Single-Entry
2. SBME: Single-Bit, Multi-Entry
3. MBSE: Multi-Bit, Single-Entry
4. MBME: Multi-Bit, Multi-Entry

Error Patterns: Breadth and Severity

~26% of Errors Multi-Bit!

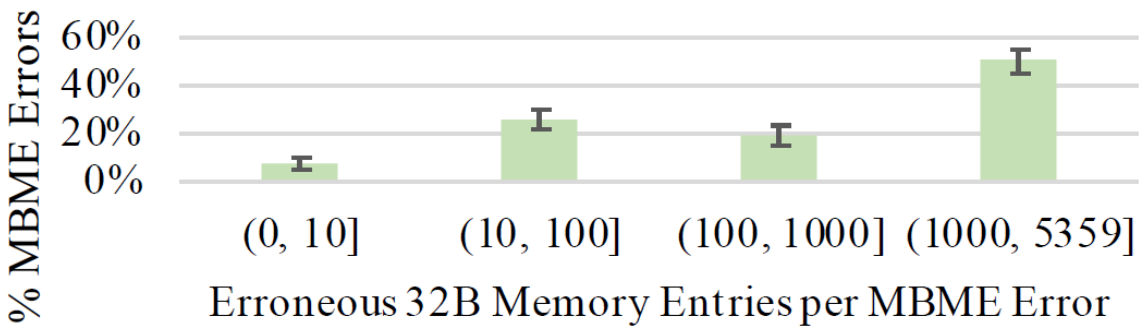


If not corrected → Crash!
If not detected → SDC!

1. SBSE: Single-Bit, Single-Entry
2. SBME: Single-Bit, Multi-Entry
3. MBSE: Multi-Bit, Single-Entry
4. MBME: Multi-Bit, Multi-Entry

Error Patterns: Breadth of Multi-Entry Errors

Some Single Events Corrupt Thousands of Entries!



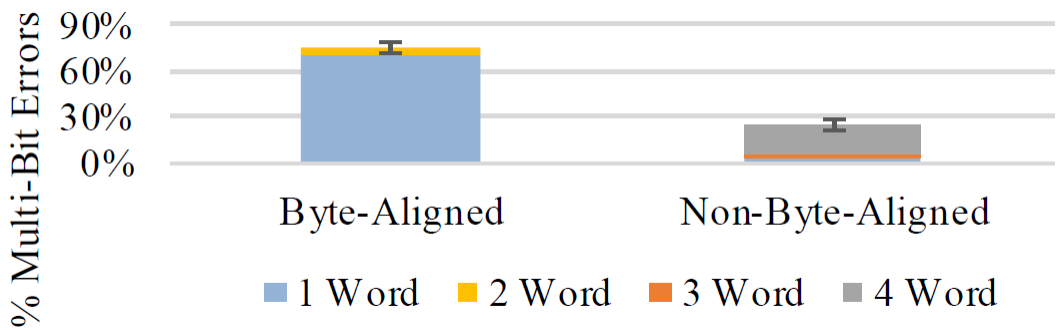
The breadth (number of erroneous entries) per MBME event.

MBME errors with thousands of erroneous entries common!

Demonstrates again the potential multi-bit error SDC risk.

Error Patterns: Byte Alignment

Most Multi-Bit Errors Affect a Single Byte



~74.6% of multi-bit errors are confined to a logically-contiguous byte.

Byte errors generally affect one byte per entry.

Non-byte-aligned multi-bit errors can affect the entire 32B memory entry.

Observed Soft Error Patterns

Multi-Bit Errors are Common! Especially Byte and Whole-Entry Errors

Tab. 1: Soft Error Pattern Probabilities.

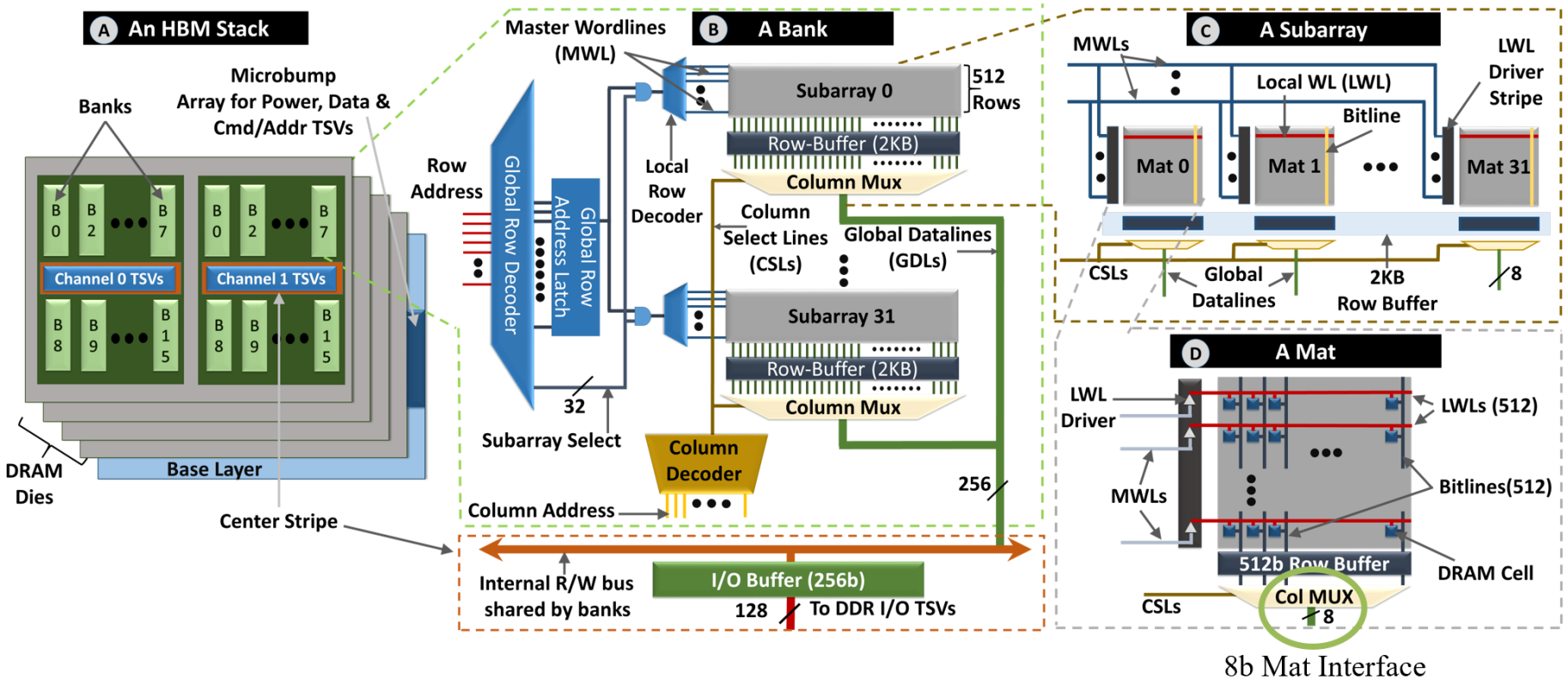
Severity	Bits	Probability
1 Bit	1	73.98%
1 Pin	2–4	0.19%
1 Byte	2–8	22.56%
2 Bits	2	0.11%
3 Bits	3	0.03%
1 Beat	4–64	0.90%
1 Entry	4–256	2.23%

~26% of events affect multiple bits!

"Byte" and "Whole-Entry" errors are especially prevalent.

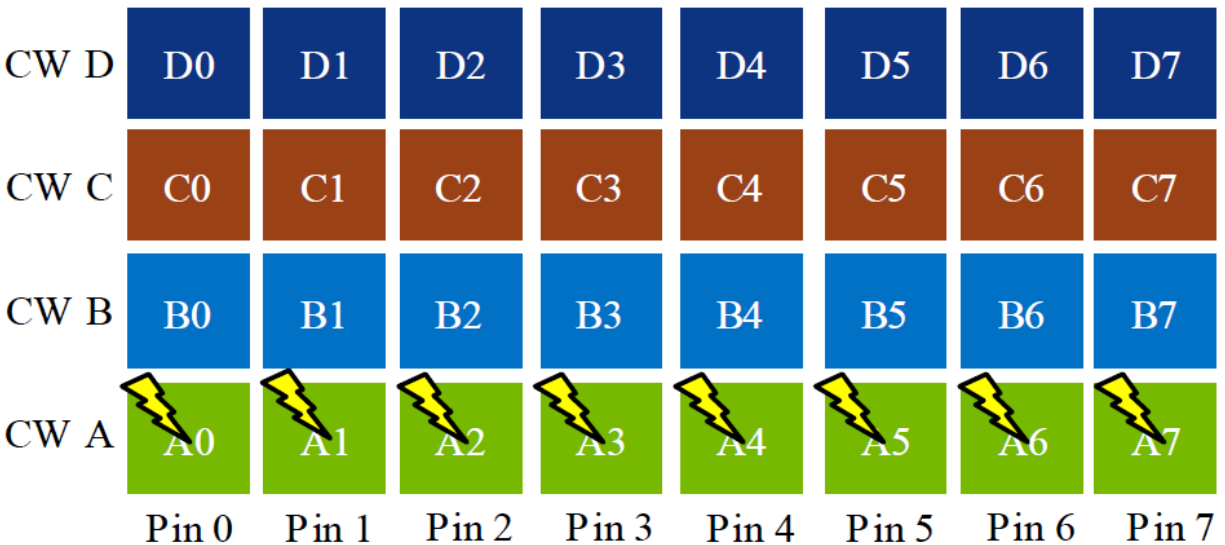
A Sketch of DRAM

Byte Errors are likely from Mat-Local Errors in DRAM



Baseline: SEC-DED ECC

4x Single-Bit Correcting, Double-Bit Detecting Codewords



(a) SEC-DED with a Byte Error

If any of the four codewords DUE: DUE for whole entry.

Else, if any of the four codewords SDC: SDC for whole entry.

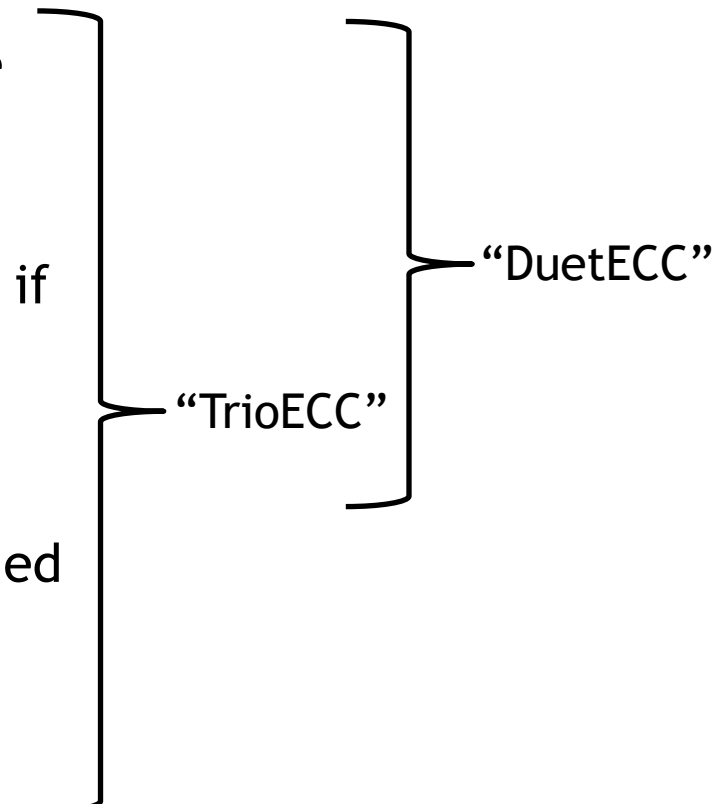
Issues:

1. Byte errors affect a single codeword.
2. Beat and whole-entry errors have high SDC risk.

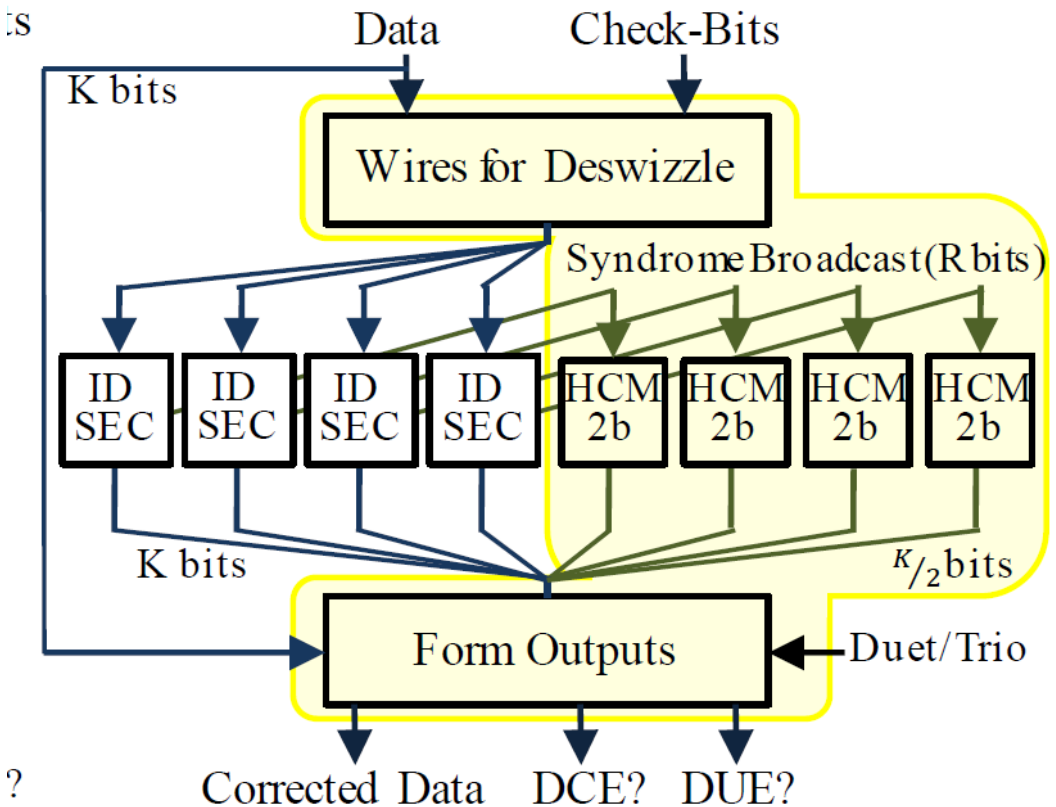
Optimizations

Three Complementary Optimizations

1. Interleaving: interleave the four codewords to spread a byte or beat error among all four! Gives four chances to report a DUE.
2. “Correction Sanity Check”: only allow correction to proceed if all corrected errors are in a likely pattern (same pin, same byte). Sacrifices a scant amount of opportunistic correction, but reduces the SDC risk by orders of magnitude.
3. 2b Symbol Correction: move to a code that can correct aligned 2b-adjacent errors. Sacrifices some error detection but provides byte correction when interleaved!



Hardware for SEC-DED, DuetECC, TrioECC



Reed-Solomon code with 8b symbols, roviding single-byte correct with 2B of edundancy.

ecoding Steps:

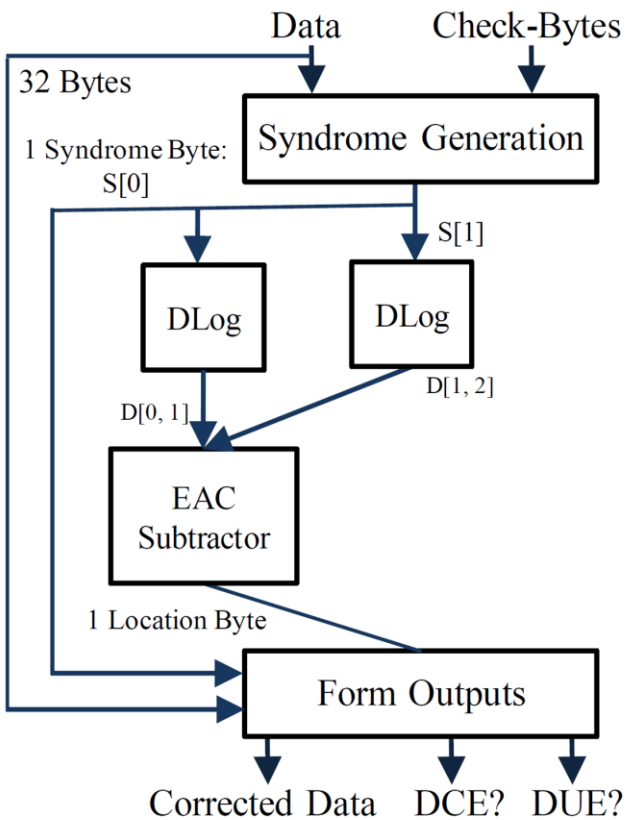
2. Two syndrome bytes are generated.
2. Error location is performed using the pair of syndromes.

   DLog = discrete logarithm (used for low-cost polynomial division)

   EAC = end-around-carry (mod 255) modular subtractor.

3. The error location and first syndrome byte are used to correct the byte error.

Alternative Baseline: Single-Byte Correction



Reed-Solomon code with 8b symbols, providing single-byte correct with 2B of redundancy.

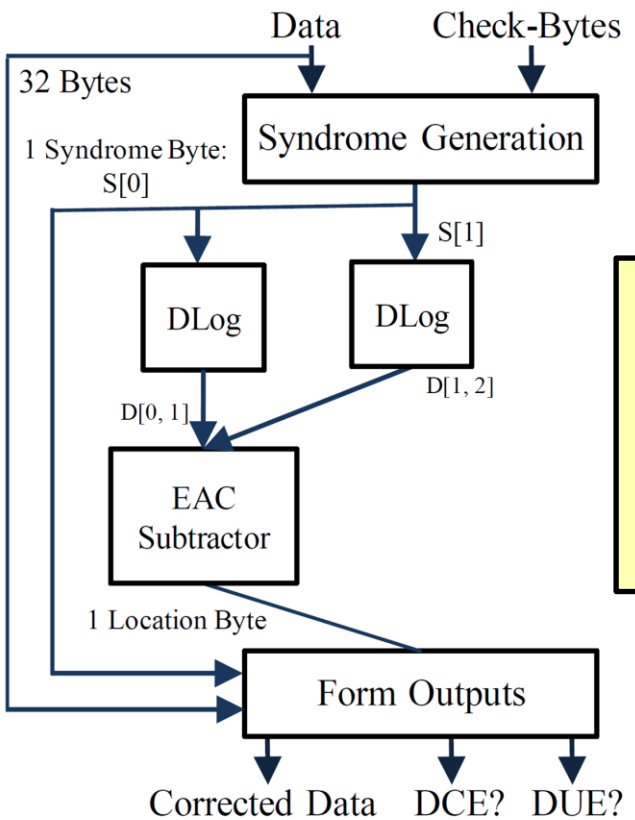
Decoding Steps:

1. Two syndrome bytes are generated.
2. Error location is performed using the pair of syndromes.

   DLog = discrete logarithm (used for low-cost polynomial division)

   EAC = end-around-carry (mod 255) modular subtractor.
3. The error location and first syndrome byte are used to correct the byte error.

SSC: Single-Symbol (Byte) Correction



Reed-Solomon code with 8b symbols, providing single-byte correct with 2B of redundancy.

Decoding Steps:

1. Two syndrome bytes are generated.

Note: also amenable to the interleaving and correction sanity check optimizations, using 2x SSC codewords.

But error coverage is not as good as an improved byte error correcting code (next…)

pair of

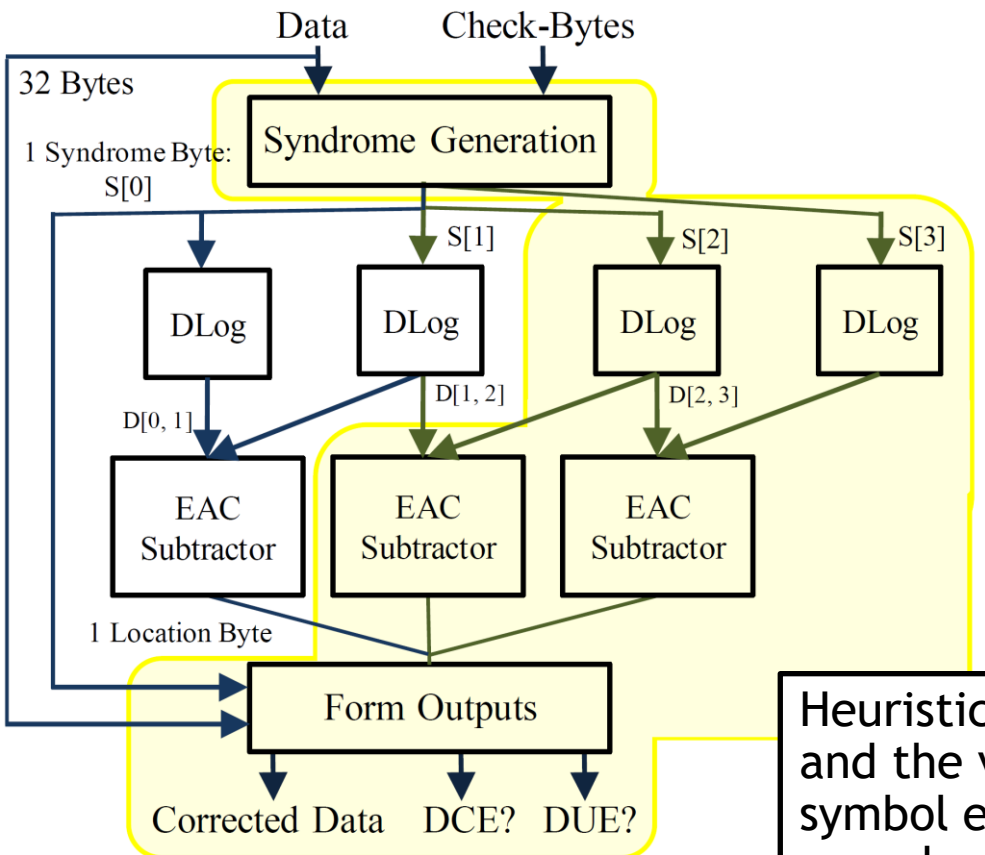
-cost polynomial

EAC = end-around-carry (mod 255) modular subtractor.

3. The error location and first syndrome byte are used to correct the byte error.

SSC-DSD+ vs the SSC Decoder

Modest New Hardware Additions…



New or modified structures highlighted in yellow.

Decoding Steps:

1. **Four** syndrome bytes are generated.
2. Error location is performed using **each contiguous pair** of syndromes.
3. If the calculated error locations disagree, a detectable-uncorrectable error is reported.

Heuristic detects all double-symbol and the vast majority of triple-symbol errors without solving the error locator polynomial, enabling an efficient, single-cycle decoder.

Per-Pattern ECC Behavior

C = Corrected, D = Detected, % = SDC Risk

	SEC-DED (NI:SEC-DED)	I:SEC-DED	DuetECC (I:SEC-DED+CSC)	NI:SEC-2bEC	I:SEC-2bEC	TrioECC (I:SEC-2bEC+CSC)	I:SSC	I:SSC+CSC	SSC-DSD+
# CWs	4	4	4	4	4	4	1	1	1
Symbol	1b	1b	1b	2b	2b	2b	8b	8b	8b
1 Bit	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
1 Pin	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	D
1 Byte	22.6721%	D	D	39.4062%	**C**	**C**	**C**	**C**	**C**
2 Bits	**D**	**D**	**D**	5.0813%	5.0813%	5.0813%	9.6545%	9.6545%	**D**
3 Bits	3.4080%	3.4080%	3.4080%	14.9347%	14.9347%	4.7010%	16.8407%	3.8781%	**D**
1 Beat	28.5201%	0.6615%	0.0013%	42.2054%	3.1670%	0.0089%	0.4898%	0.0543%	**0.0002%**
1 Entry	0.6640%	0.6603%	0.0013%	3.1646%	3.1643%	0.0085%	0.4898%	0.0543%	**0.0002%**

Notation:
NI: Non-Interleaved
I: Interleaved
+CSC: With Correction Sanity Check

e.g. DuetECC is I:SEC-DED+CSC.

Per-Pattern ECC Behavior: SEC-DED

C = Corrected, D = Detected, % = SDC Risk

	SEC-DED (NI:SEC-DED)
# CWs	4
Symbol	1b
1 Bit	**C**
1 Pin	**C**
1 Byte	22.6721%
2 Bits	**D**
3 Bits	3.4080%
1 Beat	28.5201%
1 Entry	0.6640%

Issues:

Multi-Bit errors that affect a single codeword.

- High byte error rate
- High beat error rate

Per-Pattern ECC Behavior: DuetECC

C = Corrected, D = Detected, % = SDC Risk

	SEC-DED (NI:SEC-DED)	I:SEC-DED	DuetECC (I:SEC-DED+CSC)
# CWs	4	4	4
Symbol	1b	1b	1b
1 Bit	**C**	**C**	**C**
1 Pin	**C**	**C**	**C**
1 Byte	22.6721%	D	D
2 Bits	**D**	**D**	**D**
3 Bits	3.4080%	3.4080%	3.4080%
1 Beat	28.5201%	0.6615%	0.0013%
1 Entry	0.6640%	0.6603%	0.0013%

Interleaving provides:

- Perfect Byte Detection
- Improved Beat Detection (~43x better than NI:SEC-DED)

+CSC provides:

- Improved Beat Detection (~500x better than I:SEC-DED)
- Improved Entry Detection (~500x better than I:SEC-DED)

Per-Pattern ECC Behavior: DuetECC

C = Corrected, D = Detected, % = SDC Risk

	SEC-DED (NI:SEC-DED)	I:SEC-DED	DuetECC (I:SEC-DED+CSC)	NI:SEC-2bEC	I:SEC-2bEC
# CWs	4	4	4	4	4
Symbol	1b	1b	1b	2b	2b
1 Bit	**C**	**C**	**C**	**C**	**C**
1 Pin	**C**	**C**	**C**	**C**	**C**
1 Byte	22.6721%	D	D	39.4062%	**C**
2 Bits	**D**	**D**	**D**	5.0813%	5.0813%
3 Bits	3.4080%	3.4080%	3.4080%	14.9347%	14.9347%
1 Beat	28.5201%	0.6615%	0.0013%	42.2054%	3.1670%
1 Entry	0.6640%	0.6603%	0.0013%	3.1646%	3.1643%

2b Aligned Symbol Error Correction:

Alone: not useful! Prohibitive SDC risk.

Interleaved: improved against byte, beat and entry errors. But still worse than NI:SEC-DED…

Per-Pattern ECC Behavior: DuetECC

C = Corrected, D = Detected, % = SDC Risk

	SEC-DED (NI:SEC-DED)	I:SEC-DED	DuetECC (I:SEC-DED+CSC)	NI:SEC-2bEC	I:SEC-2bEC	TrioECC (I:SEC-2bEC+CSC)
# CWs	4	4	4	4	4	4
Symbol	1b	1b	1b	2b	2b	2b
1 Bit	**C**	**C**	**C**	**C**	**C**	**C**
1 Pin	**C**	**C**	**C**	**C**	**C**	**C**
1 Byte	22.6721%	D	D	39.4062%	**C**	**C**
2 Bits	**D**	**D**	**D**	5.0813%	5.0813%	5.0813%
3 Bits	3.4080%	3.4080%	3.4080%	14.9347%	14.9347%	4.7010%
1 Beat	28.5201%	0.6615%	0.0013%	42.2054%	3.1670%	0.0089%
1 Entry	0.6640%	0.6603%	0.0013%	3.1646%	3.1643%	0.0085%

TrioECC (I:SEC-2bEC+CSC) reduces the beat and entry SDC risk to useful levels.

Corrects byte errors, while maintaining a ~78x better beat and entry SDC risk than SEC-DED.

Per-Pattern ECC Behavior: DuetECC

C = Corrected, D = Detected, % = SDC Risk

Single-Symbol (Byte) Correction can correct byte errors, but with a prohibitive SDC risk against 2b/3b errors.

+CSC improves the resilience some, but not as drastic as with binary ECCs, due to reduced combinatorial space from 2 codewords and byte error location granularity.

Our improved SSC-DSD+ design provides the lowest overall SDC risk, while maintaining byte (but not pin) error correction.

	I:SSC	I:SSC+CSC	SSC-DSD+
# CWs	1	1	1
Symbol	8b	8b	8b
1 Bit	**C**	**C**	**C**
1 Pin	**C**	**C**	D
1 Byte	**C**	**C**	**C**
2 Bits	9.6545%	9.6545%	**D**
3 Bits	16.8407%	3.8781%	**D**
1 Beat	0.4898%	0.0543%	**0.0002%**
1 Entry	0.4898%	0.0543%	**0.0002%**

Per-Pattern ECC Behavior

C = Corrected, D = Detected, % = SDC Risk

	SEC-DED (NI:SEC-DED)	I:SEC-DED	DuetECC (I:SEC-DED+CSC)	NI:SEC-2bEC	I:SEC-2bEC	TrioECC (I:SEC-2bEC+CSC)	I:SSC	I:SSC+CSC	SSC-DSD+
# CWs	4	4	4	4	4	4	1	1	1
Symbol	1b	1b	1b	2b	2b	2b	8b	8b	8b
1 Bit	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**
1 Pin	**C**	**C**	**C**	**C**	**C**	**C**	**C**	**C**	D
1 Byte	22.6721%	D	D	39.4062%	**C**	**C**	**C**	**C**	**C**
2 Bits	**D**	**D**	**D**	5.0813%	5.0813%	5.0813%	9.6545%	9.6545%	**D**
3					9347%	4.7010%	16.8407%	3.8781%	**D**
1					1670%	0.0089%	0.4898%	0.0543%	**0.0002%**
1					1643%	0.0085%	0.4898%	0.0543%	**0.0002%**

DuetECC/TrioECC:

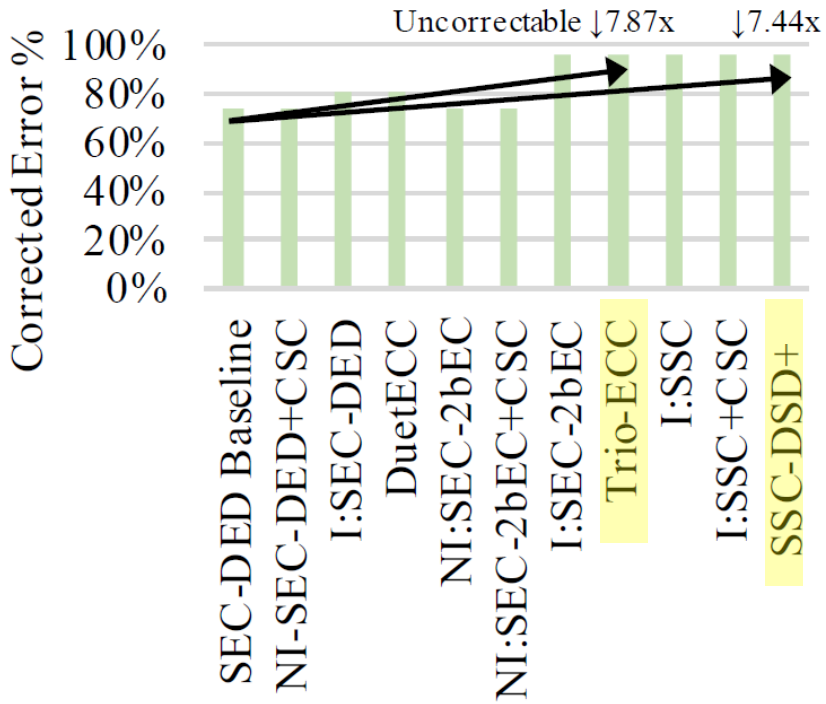
Detect (DuetECC) or correct (TrioECC) byte errors.
Highly efficient hardware design.

SSC-DSD+:

Correct byte errors, reduce SDC risk by >700,000x!
More costly hardware than DuetECC/TrioECC, but still single-cycle.

The Impact of ECC

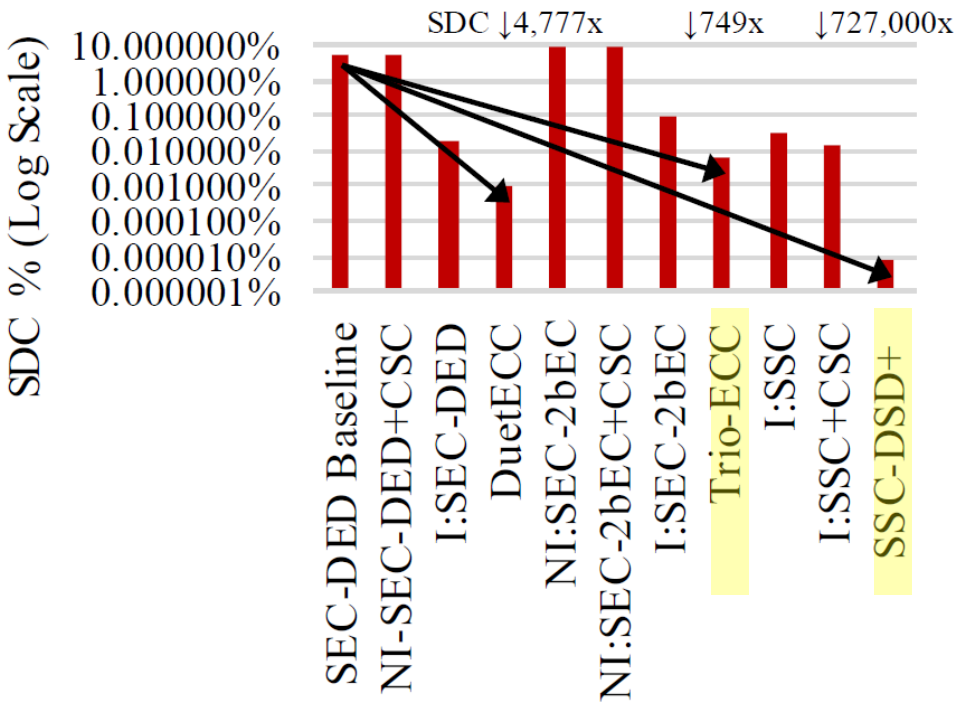
Tailored ECC Can Reduce the SDC Risk by Orders of Magnitude



Reduced Error-Related Crashes

The Impact of ECC

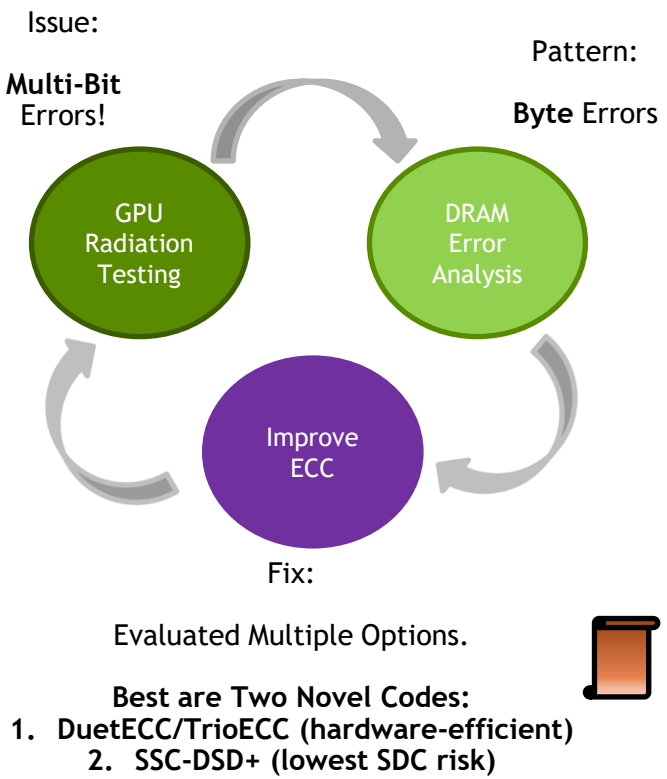
Tailored ECC Can Reduce the SDC Risk by Orders of Magnitude



Reduced SDC Risk

Conclusion

Characterizing and Mitigating Errors in HBM2 DRAM



Neutron beam testing results show byte and whole-entry errors prevalent.

ECC can be tailored to these patterns.

1. DuetECC: 3 orders-of-magnitude SDC risk reduction.
2. TrioECC: 2 orders-of-magnitude SDC risk reduction, 7.87x lower crash rate.
3. SSC-DSD+: 5 OOM SDC risk reduction, 7.44x lower crash rate.

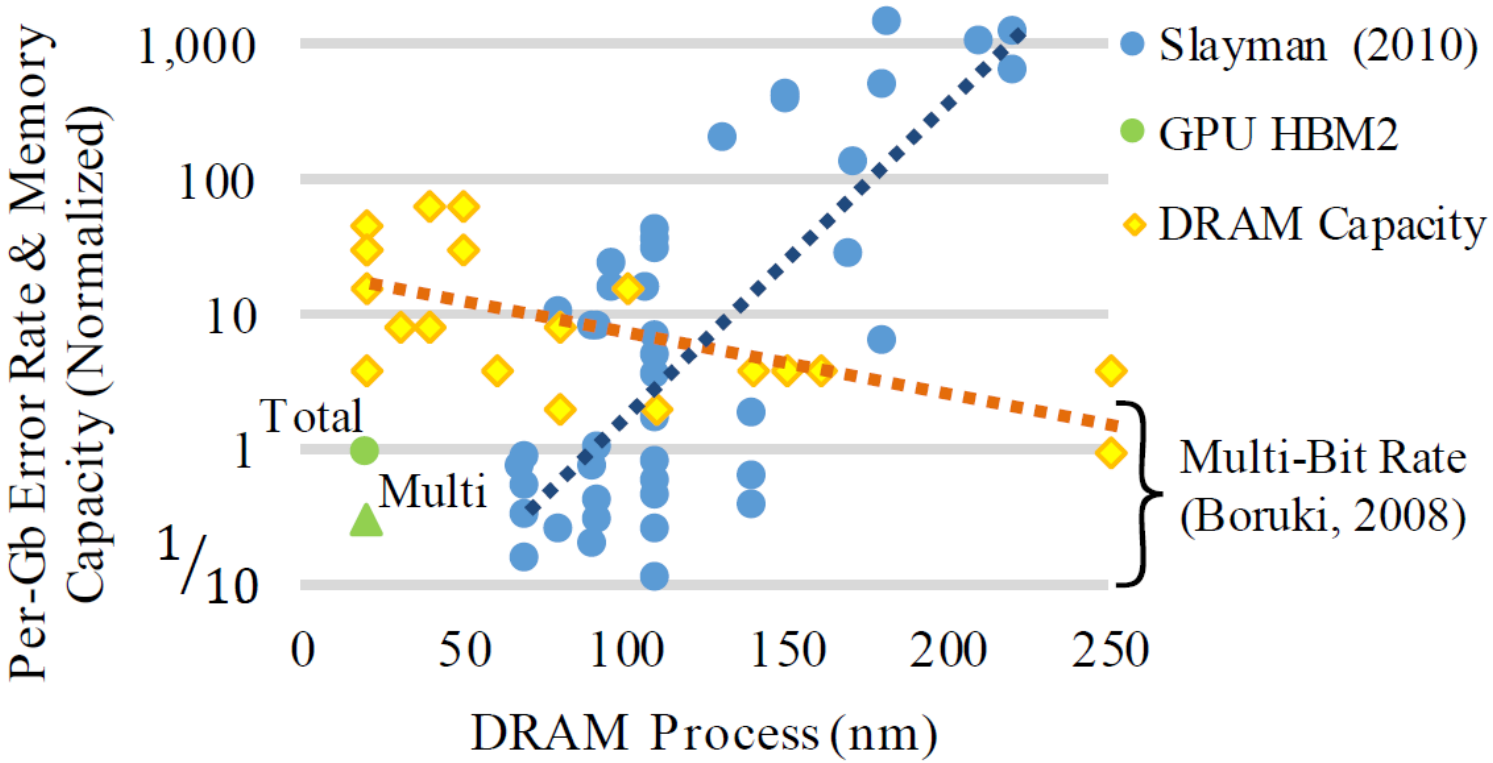
Demonstrates an error characterization and ECC design loop that can apply to any memory…

Thank you for listening!

Contact information: Michael Sullivan
misullivan@nvidia.com

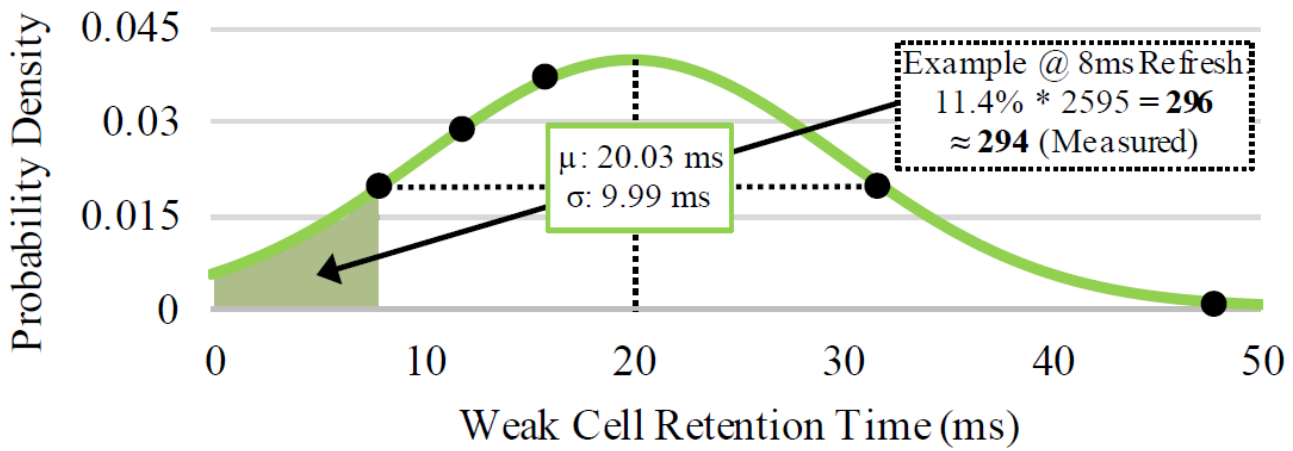
Historical Error Rates

Low Error Rate, but Relative Multi-Bit Error Rate Growing!



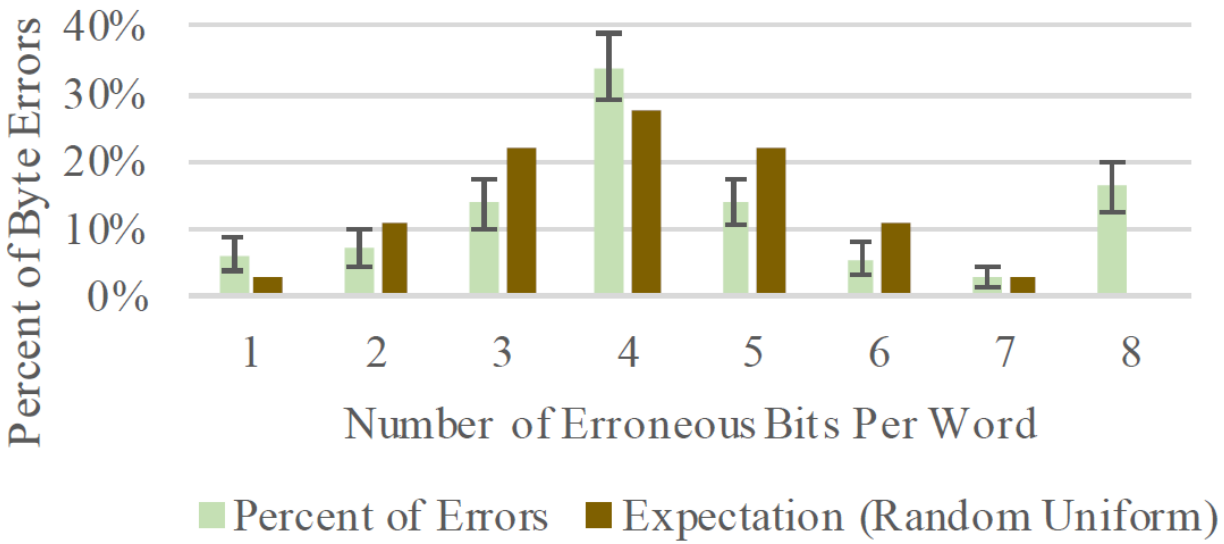
Intermittent Errors

Strong Evidence Due to Displacement Damage. Not a Field Effect!



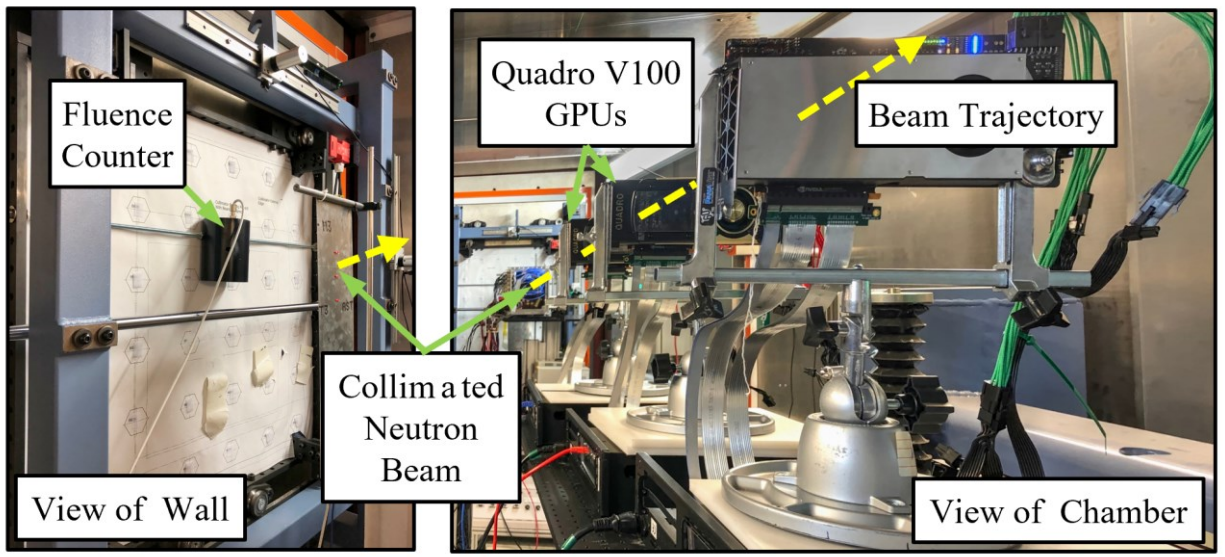
Error Patterns: Bit-Count per Multi-Bit Error

Random Uniform Model Fits. (And is Conservative.)

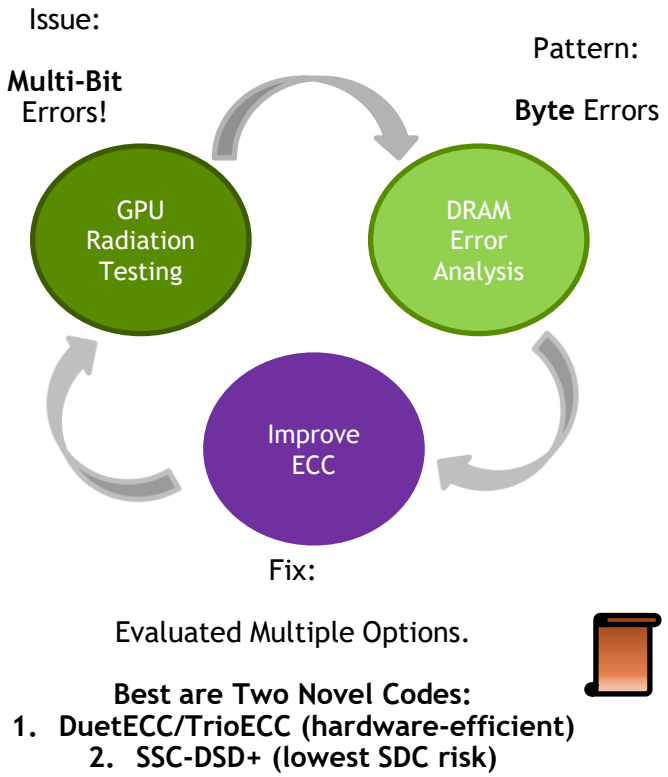


Beam Testing for Characterization

High-Energy Neutron Beam Testing with a Specialized Microbenchmark



ChipIR Neutron Beam Experimental Setup



Characterization/Design Loop

Beam Testing for Characterization

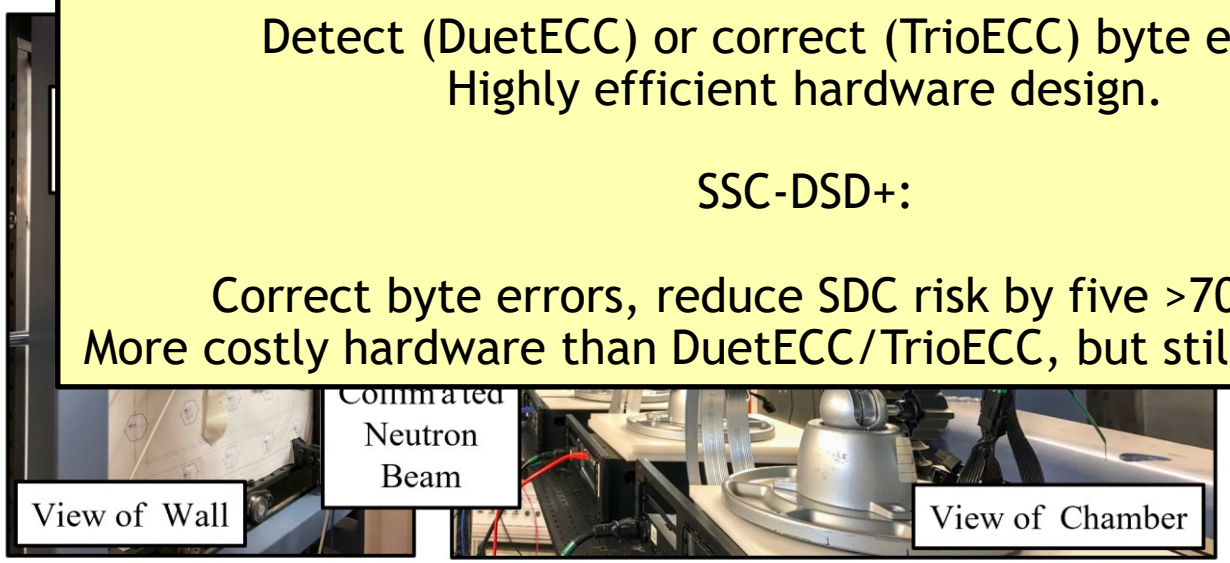
High-Energy Neutron Beam Testing with a Specialized Microbenchmark

DuetECC/TrioECC:

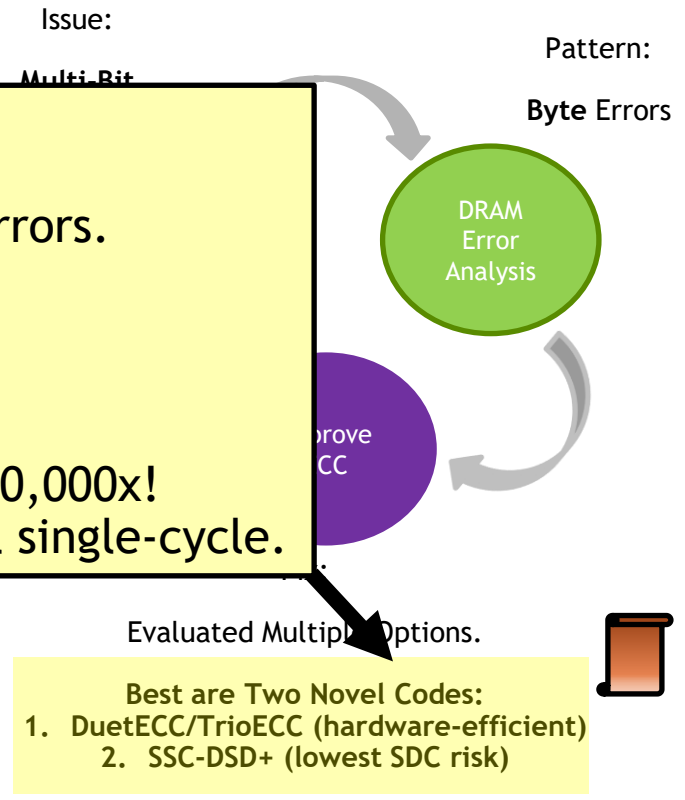
Detect (DuetECC) or correct (TrioECC) byte errors.
Highly efficient hardware design.

SSC-DSD+:

Correct byte errors, reduce SDC risk by five >700,000x!
More costly hardware than DuetECC/TrioECC, but still single-cycle.



ChipIR Neutron Beam Experimental Setup



Characterization/Design Loop