Comparison of ECC Performance on Multilevel Flash Memories

Paul H. Siegel
with Brian Butler, Scott Kayser, Eitan Yaakobi and Xiaojie Zhang

Center for Magnetic Recording Research
University of California, San Diego

Asilomar Conference on Signals, Systems, and Computers

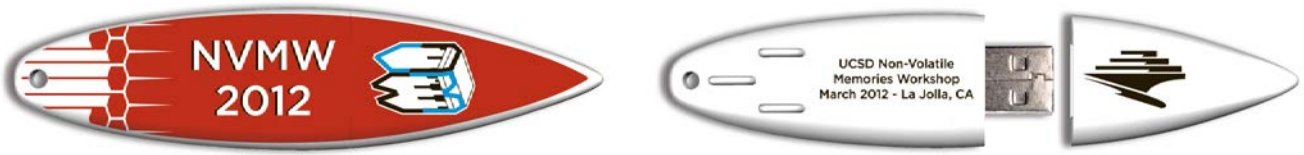
November 6, 2012

Outline

- Flash Memory Structure
- Error Characterization
- Error Correction Code (ECC) Comparison
  - Bit-level (binary) codes
  - Cell-level (multilevel) codes
- Observations and Future Directions

Flash Memory Basics

- A flash memory contains arrays (“blocks”) of floating-gate transistors (“cells”).
- A cell can support q voltage levels , e.g. q = 2, 4, 8.
- Increasing the voltage level of a cell is easy to do.
- To decrease a cell level, we must first erase its entire block, then re-program all cells to reflect the intended values.
- Such a block erase is costly in time, power, and cell wear.



NVMW 2013

4th Non-Volatile Memories Workshop

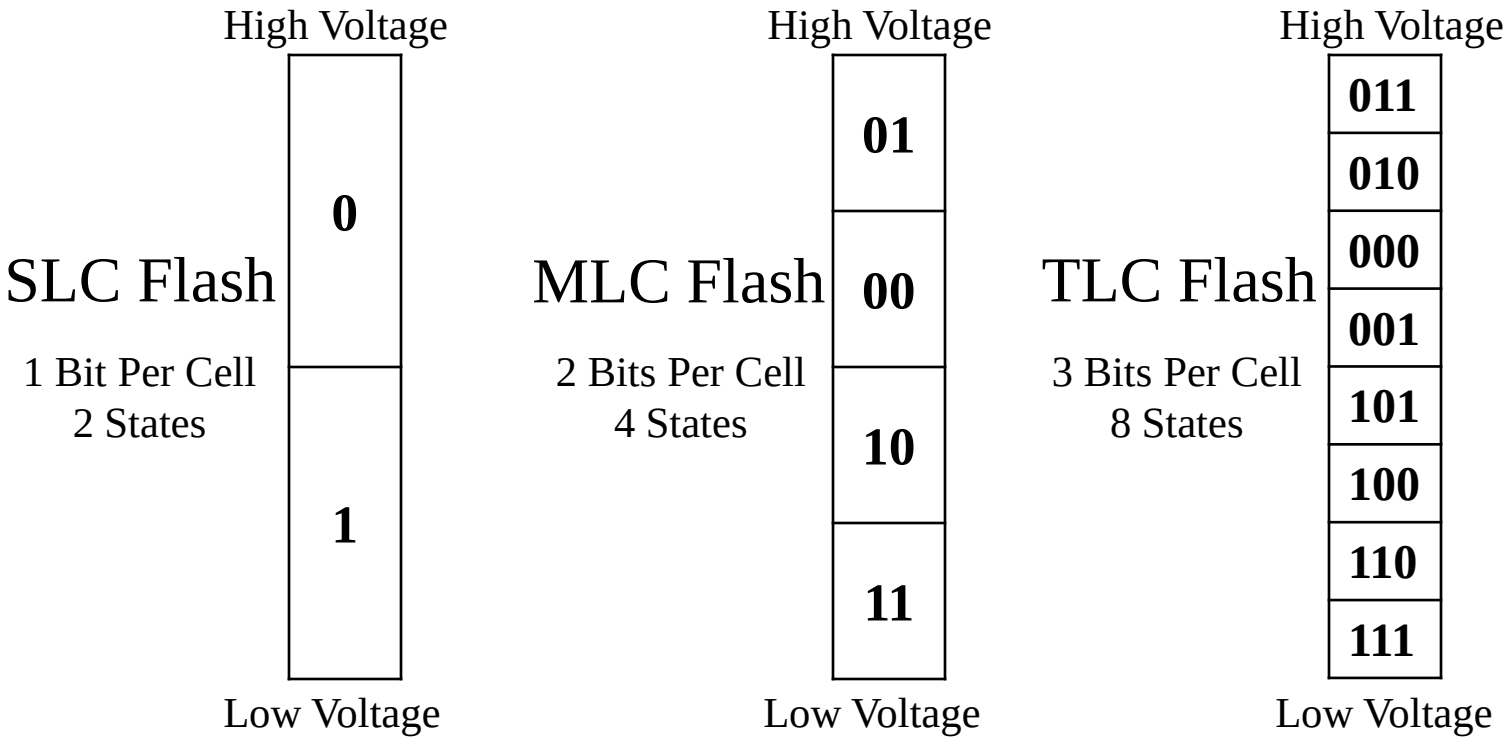
March 3-5, 2013
UC San Diego
La Jolla, California USA
http://nvmw.ucsd.edu

UCSD

NVSL
Non-volatile Systems Laboratory

CMRR
Center for Magnetic Recording Research

SLC, MLC, and TLC Flash



Flash Memory Structure - SLC

- A group of cells (2KB) constitute a page.
- A group of pages constitute a block.

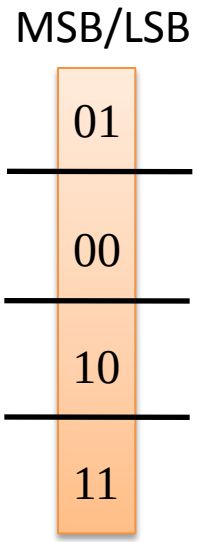
Typical SLC Block Layout

page 0	page 1
page 2	page 3
page 4	page 5
. . .	. . .
page 62	page 63

Flash Memory Structure - MLC

- The two bits in a cell belong to **different** pages: the **MSB page** and the **LSB page**

Row index	**MSB** of **first** 2^{14} cells	**LSB** of **first** 2^{14} cells	**MSB** of **last** 2^{14} cells	**LSB** of **last** 2^{14} cells
0	page 0	page 4	page 1	page 5
1	page 2	page 8	page 3	page 9
2	page 6	page 12	page 7	page 13
3	page 10	page 16	page 11	page 17
⋮	⋮	⋮	⋮	⋮
30	page 118	page 124	page 119	page 125
31	page 122	page 126	page 123	page 127



Flash Memory Structure - TLC

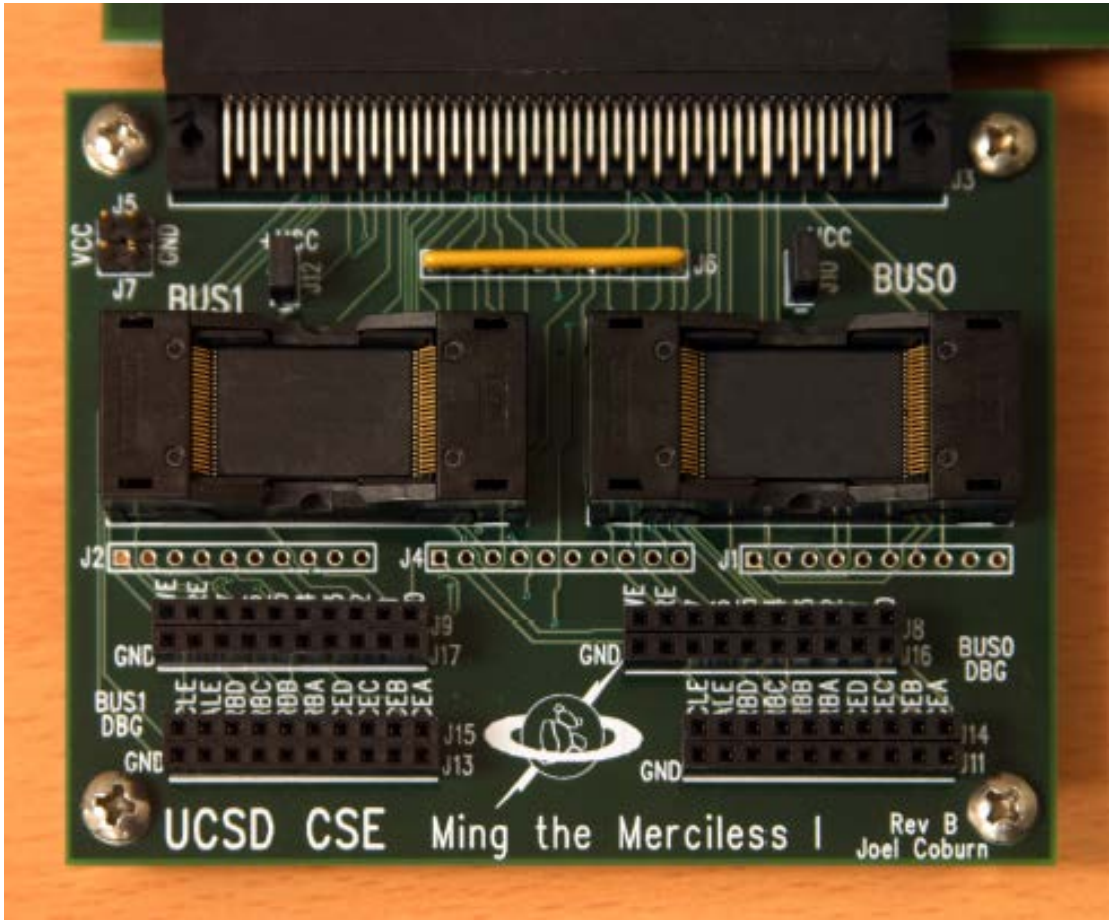
Row index	**MSB** of **first** 2^{16} cells	**CSB** of **first** 2^{16} cells	**LSB** of **first** 2^{16} cells	**MSB** of **last** 2^{16} cells	**CSB** of **last** 2^{16} cells	**LSB** of **last** 2^{16} cells
0	page 0			page 1		
1	page 2	page 6	page 12	page 3	page 7	page 13
2	page 4	page 10	page 18	page 5	page 11	page 19
3	page 8	page 16	page 24	page 9	page 17	page 25
4	page 14	page 22	page 30	page 15	page 23	page 31
⋮	⋮		⋮	⋮		⋮
62	page 362	page 370	page 378	page 363	page 371	page 379
63	page 368	page 376		page 369	page 377	
64	page 374	page 382		page 375	page 383	
65	page 380			page 381		

Error Characterization

- We tested several blocks on MLC and TLC chips.
- For each block the following steps were repeated:
  - The block is **erased.**
  - Pseudo-random data are **programmed** to the block.
  - The data are **read** and **errors** are identified.

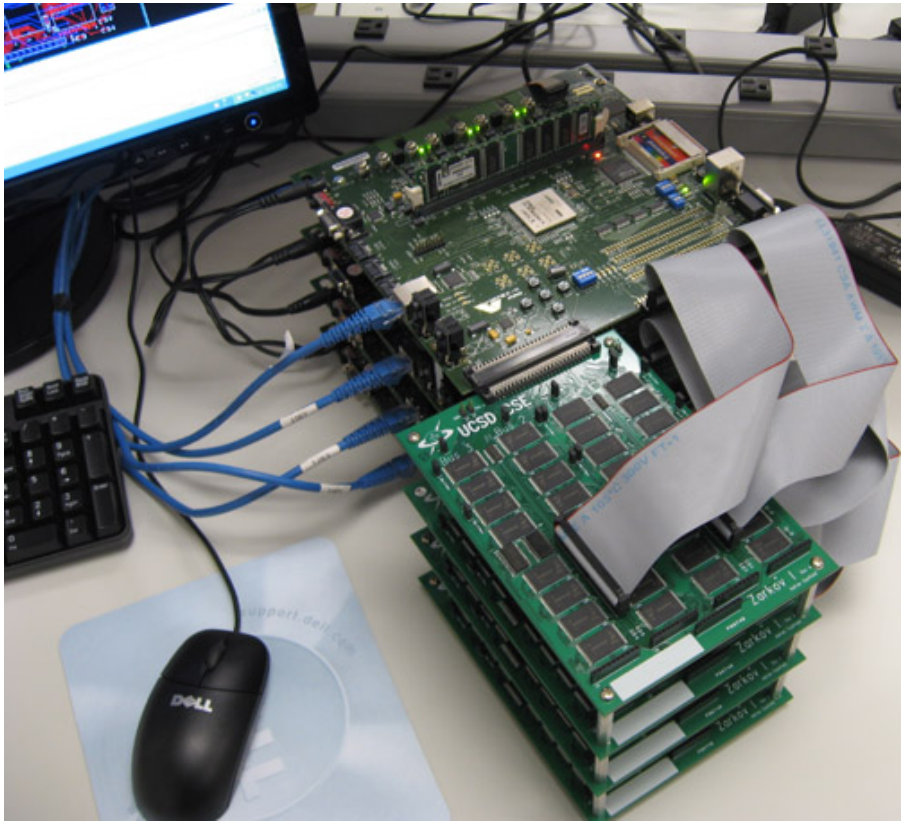
- **Disclaimers**:
  - We measured many more P/E cycles than the manufacturer's guaranteed lifetime of the device.
  - The experiments were done in laboratory conditions and related factors such as temperature change, intervals between erasures, or multiple readings before erasures were not considered.

Test Board (Ming I)



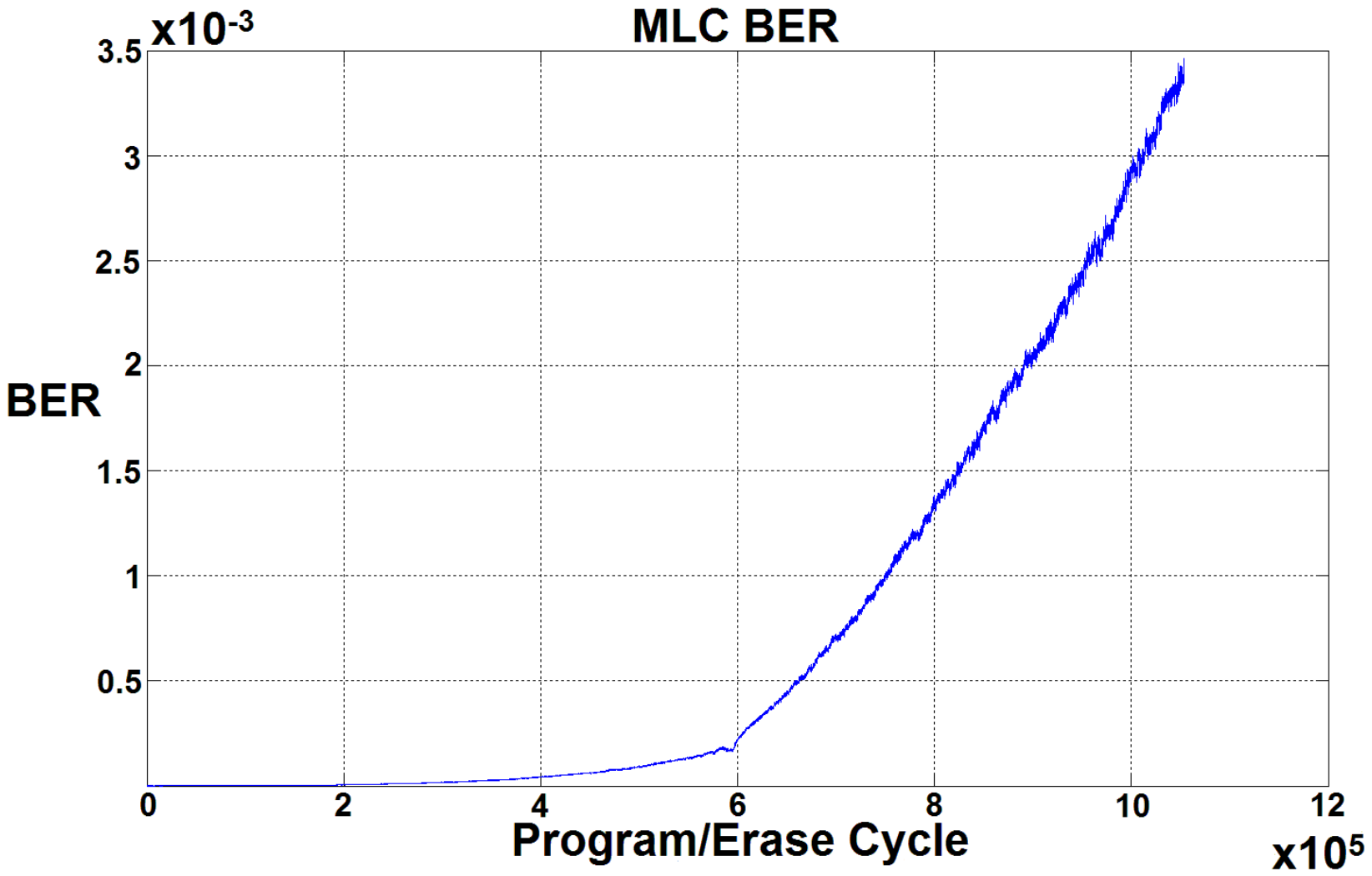
Courtesy Non-volatile Systems Laboratory, UCSD

Device Control and Data Collection

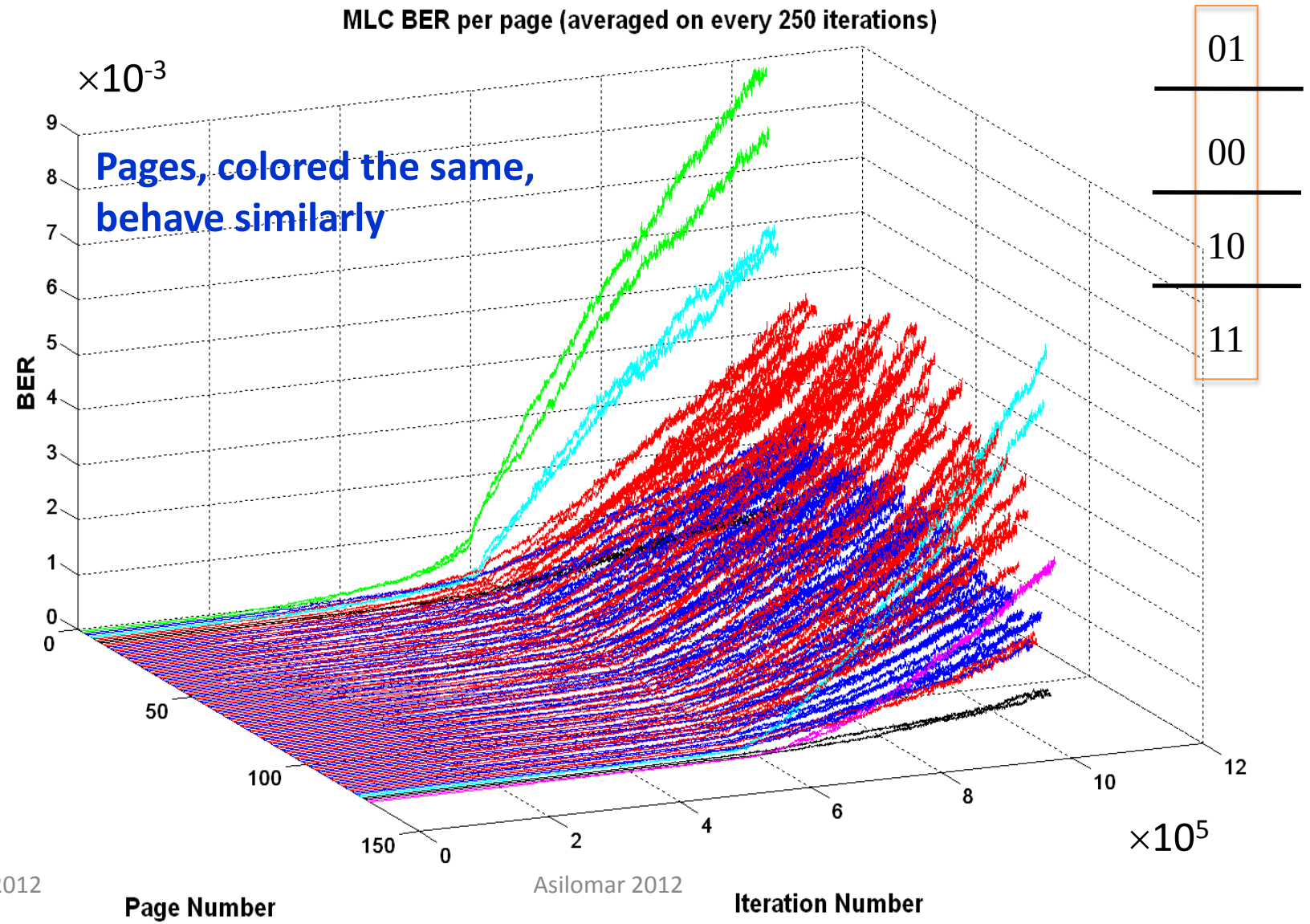


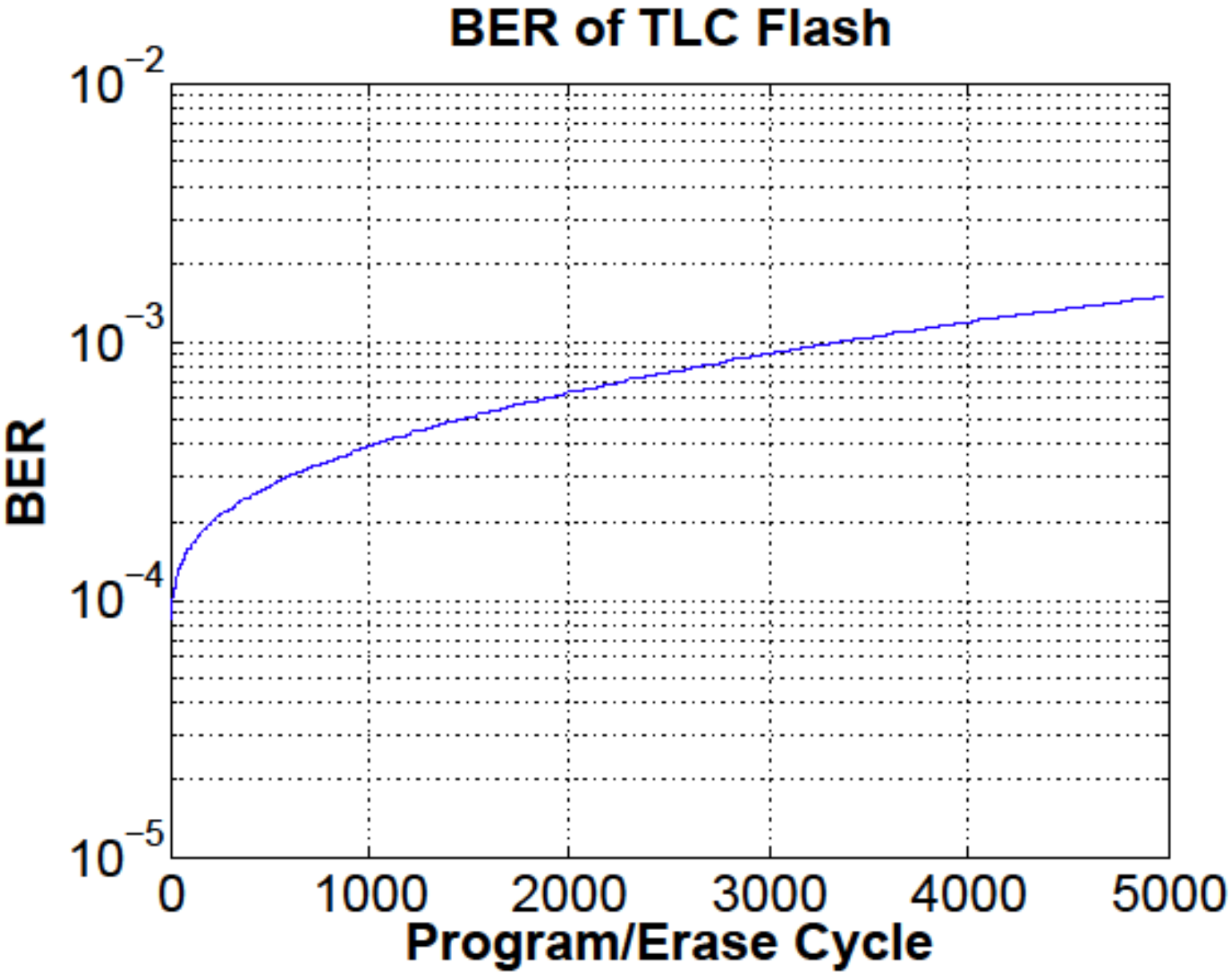
Courtesy Non-volatile Systems Laboratory, UCSD

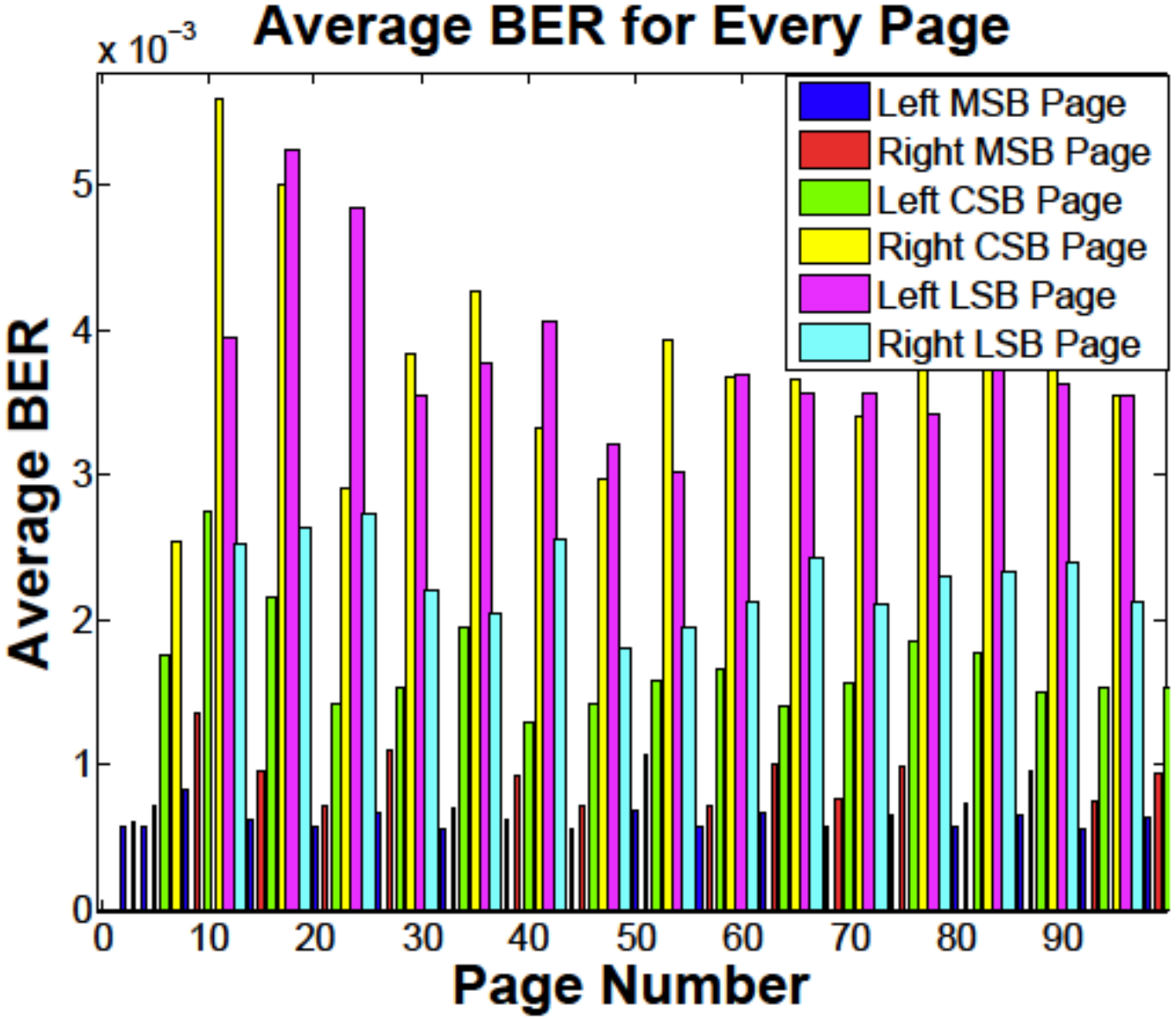
NVSL
Non-volatile Systems Laboratory


MLC BER
x10-3
3.5
3
2.5
2
1.5
1
0.5
BER
0
2
4
6
8
10
12
Program/Erase Cycle
x10^5

BER per page - MLC




BER of TLC Flash
BER
10^{-2}
10^{-3}
10^{-4}
10^{-5}
0
1000
2000
3000
4000
5000
Program/Erase Cycle



M/C/L

011
010
000
001
101
100
110
111

ECC Comparison for TLC flash

- BCH Codes
- LDPC Codes
  - Gallager codes (3,k)-regular, R=0.8, 0.9, 0.925, length 2^{16}
  - AR4JA protograph-based codes, R=0.8, lengths 1280, 5120, 20480
  - MacKay codes variable-regular degree (3 or 4) ; no 4-cycles, R=0.82, 0.87, 0.93; lengths 4095, 16383, 32000.
  - IEEE 802.3an* (10Gb/s Ethernet), R ≈0.84, length 2048.
- Cell-based algebraic codes

* Djurdjevic et al., *IEEE Commun. Letters*, July 2003.

ECC Comparison for TLC flash

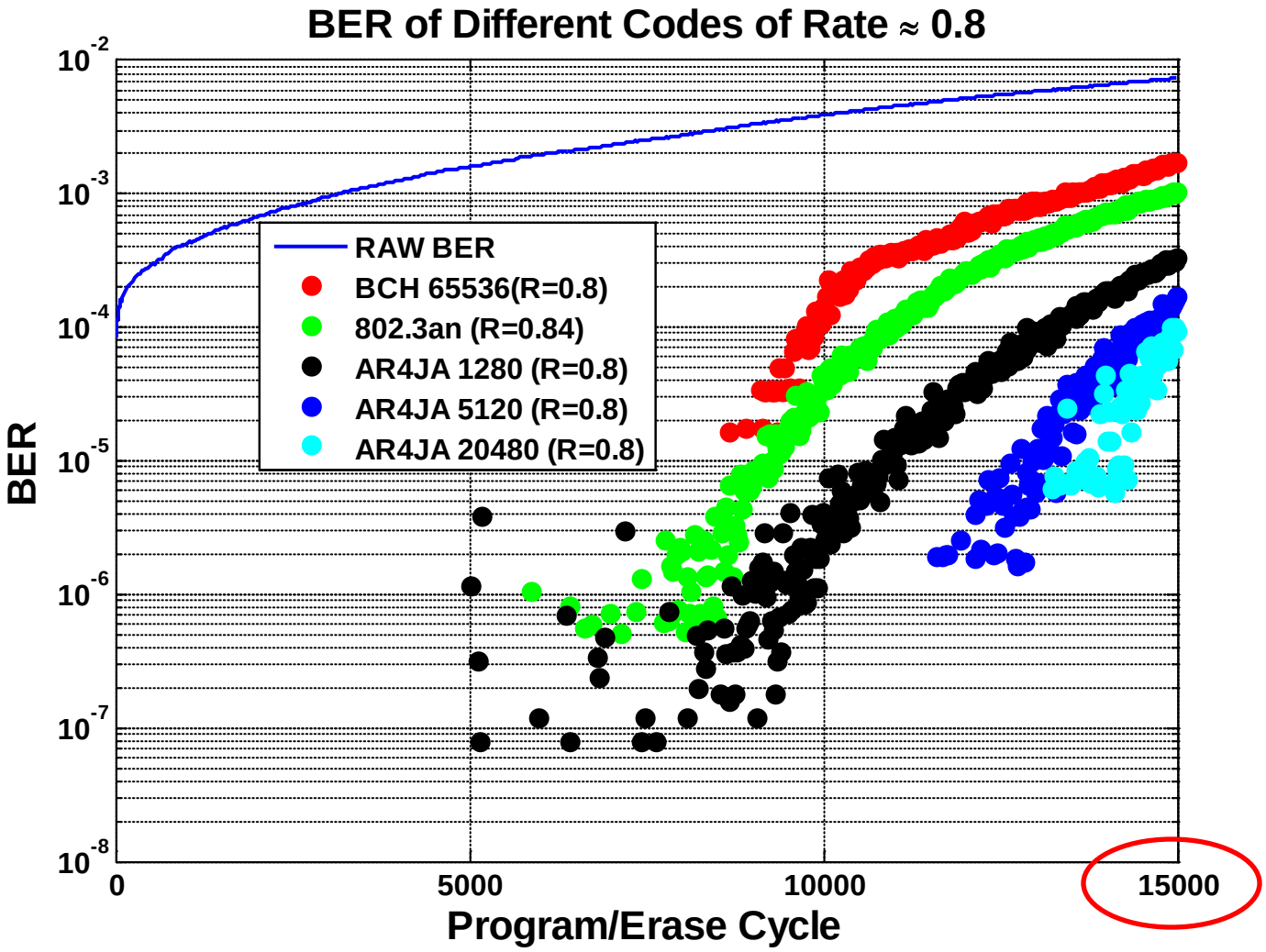
- BER computed for the first 100 iterations, then every 25^{th} iteration from then on.
- Data averaged over 6 TLC blocks.
- BCH decoder: corrects error patterns with up to t errors; detects and leaves unchanged more than t errors.
- LDPC decoders: assume binary symmetric channel model BSC(p), with empirical error probability p.

LDPC Decoders

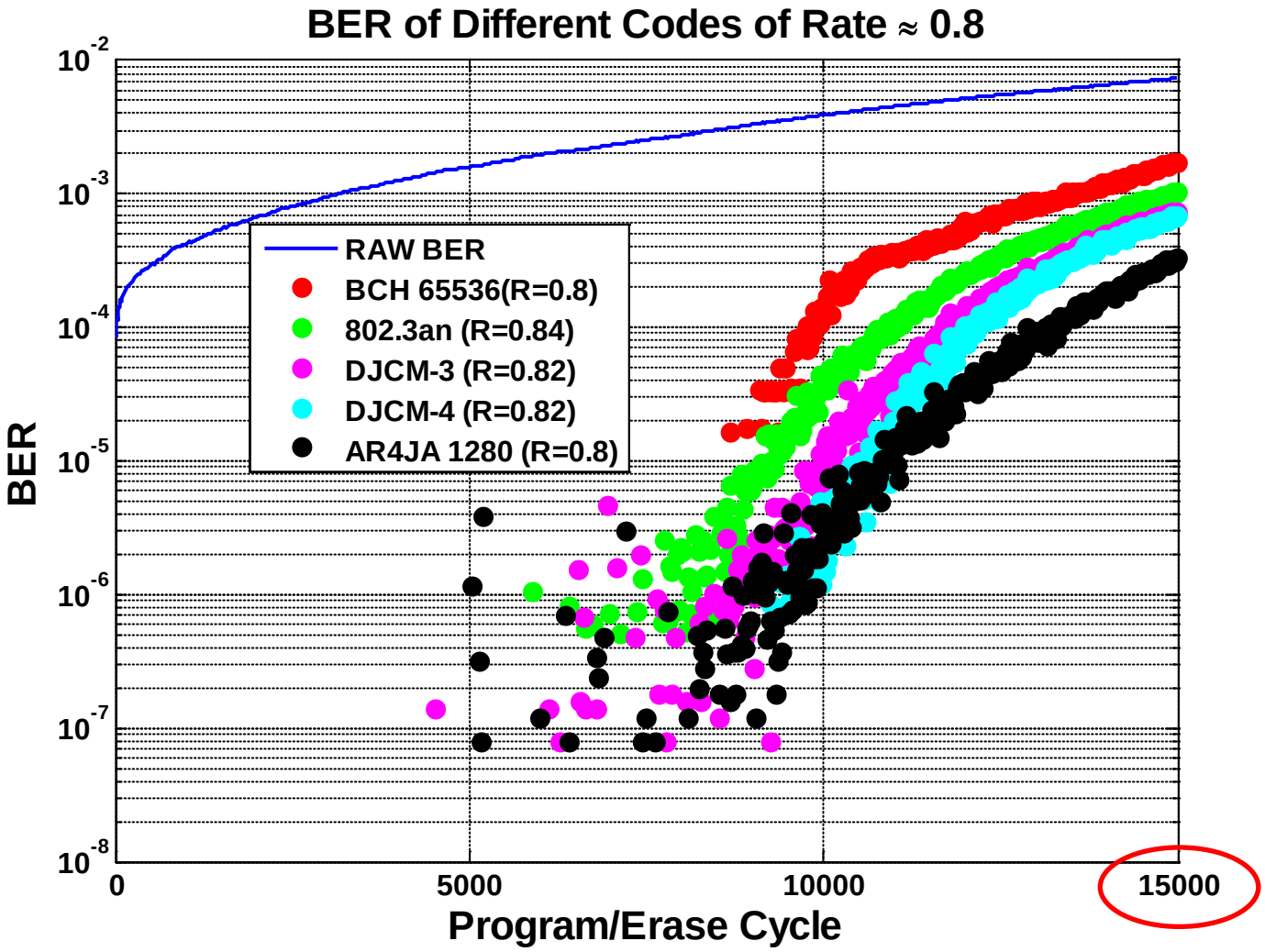
- Sum-product algorithm (SPA)
  - Floating-point, max iterations 200
  - (5+1)-bit quasi-uniform quantization
- Min-sum algorithm (MSA)
  - No LLR limits, max iterations 200
- Linear programming (LP) decoding
  - Alternating Direction Method of Multipliers (ADMM)* with new fast “projection step”

* Barman, et al., *Proc. 46th Allerton Conference*, Sept. 2011.

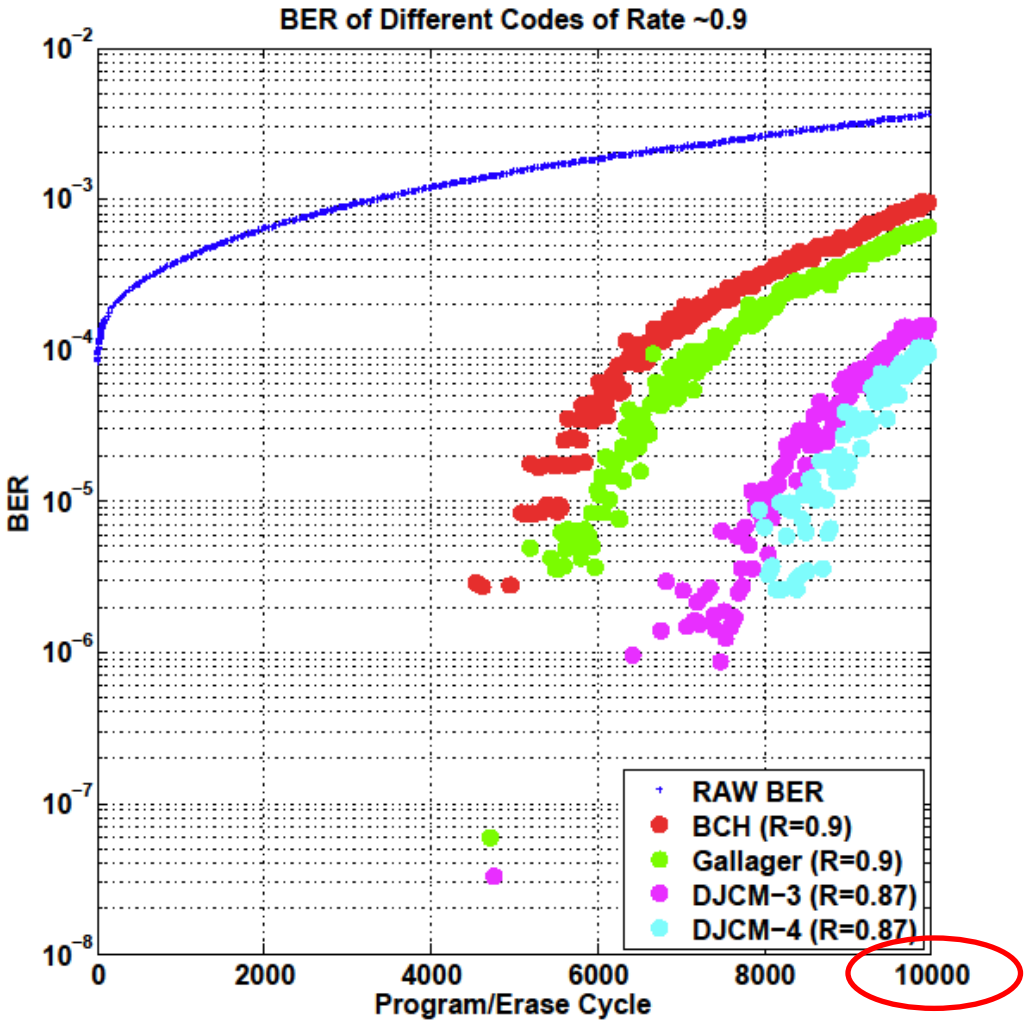
R≈0.82, LDPC with SPA Decoding



R≈0.8, LDPC with SPA Decoding

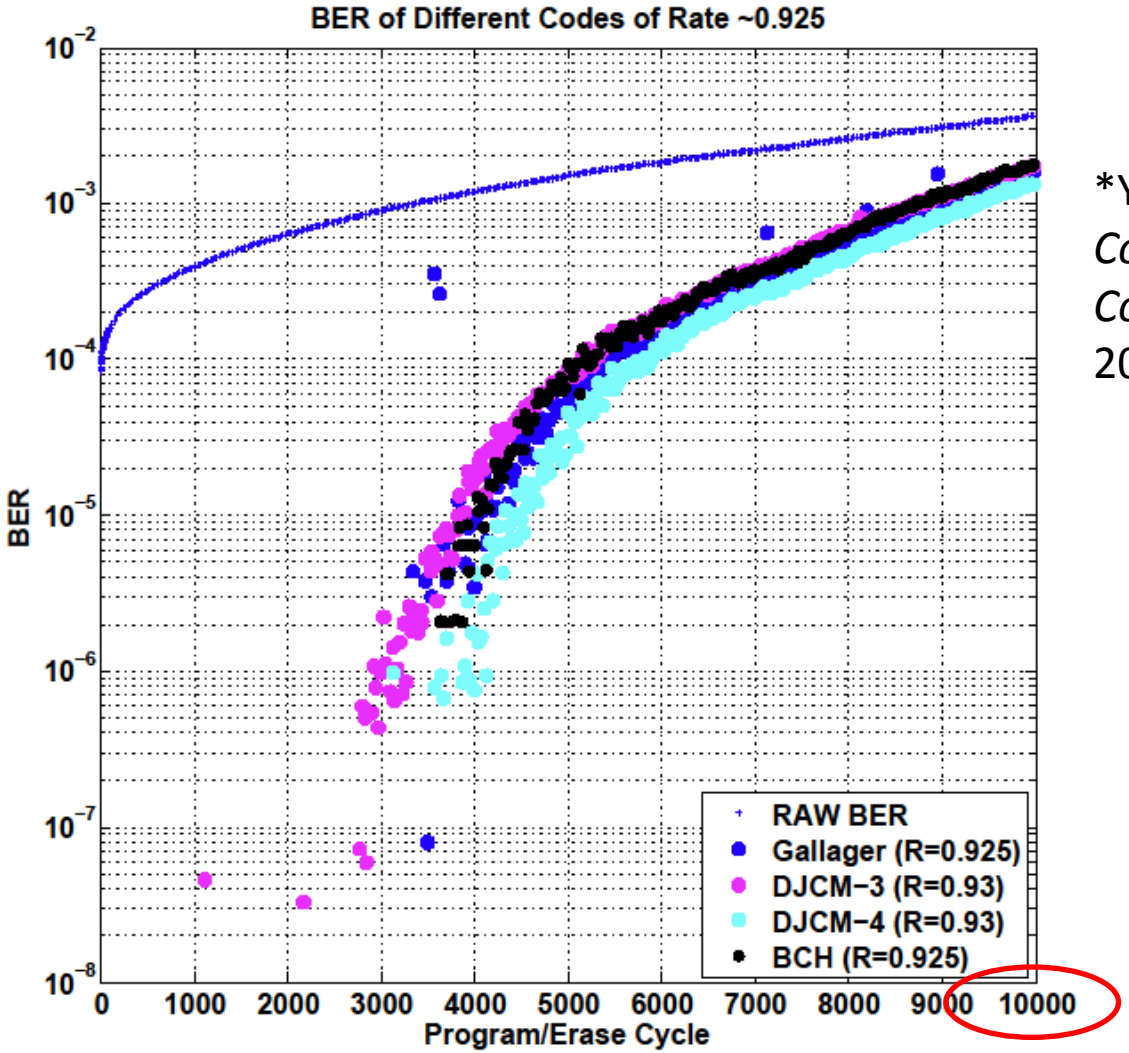


R≈0.9, LDPC with SPA Decoding*



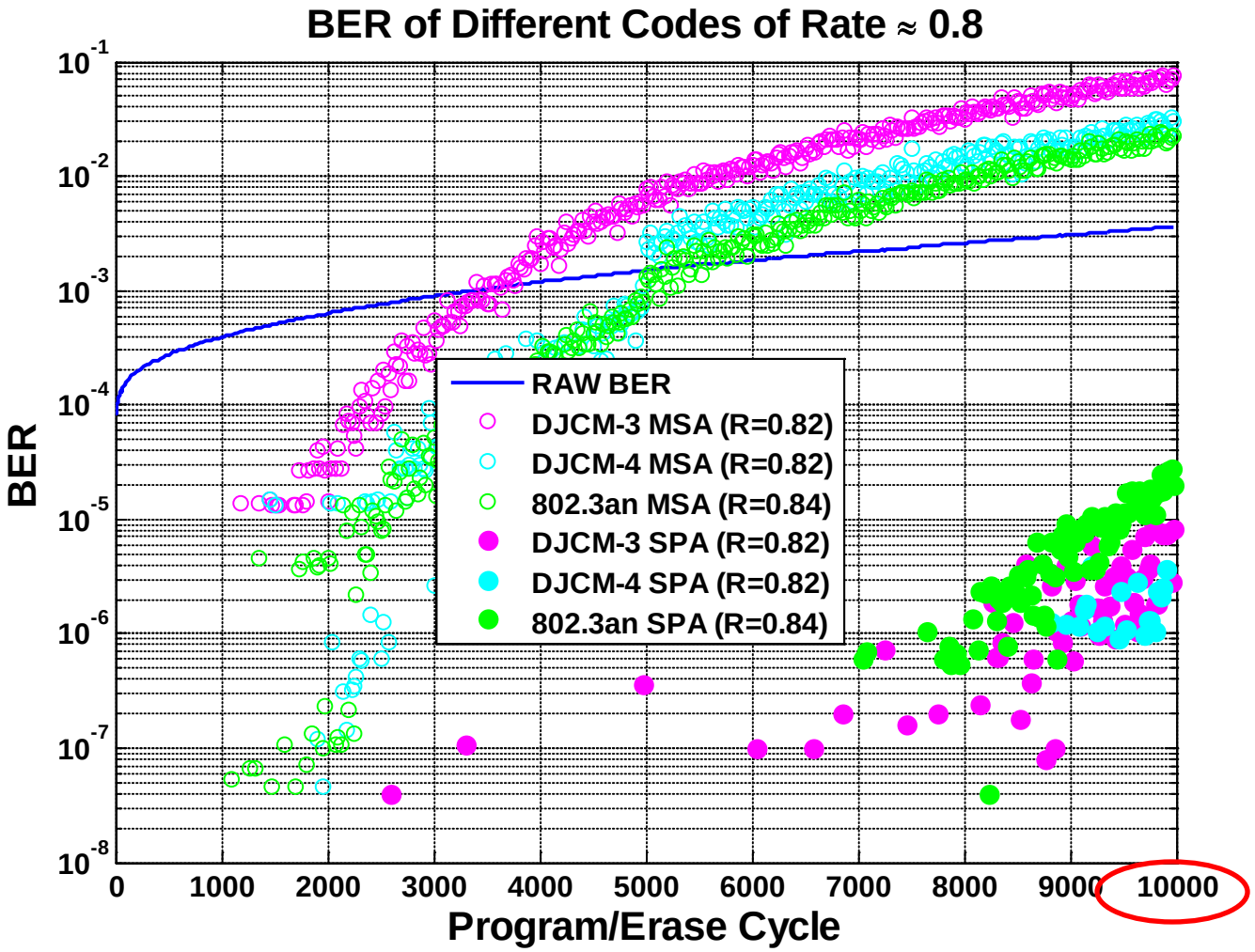
*Yaakobi, et al., *Proc. Int. Conf. on Comp., Network. Commun. (ICNC),* Jan.-Feb. 2012.

R≈0.925, LDPC with SPA Decoding*



*Yaakobi, et al., *Proc. Int. Conf. on Comp., Network. Commun. (ICNC),* Jan.-Feb. 2012.

R≈0.8, MSA vs. SPA Decoding



Linear Programming Decoding

- Linear Programming (LP) decoding of LDPC codes was introduced by Feldman in 2003.
- LP decoding with Alternating Direction Method of Multipliers (ADMM) recently proposed to speed up LP decoding (Barman, et al., 2011)
- We show results using a more efficient scheme based upon Adaptive LP decoding (ALP) with fast Cut-Search Algorithm (CSA) to further speed up the key “projection step” in LP-ADMM.

(q+1)-bit Quasi-uniform Quantization*

* Zhang, et al., *Proc. IEEE SPCOM,* July 2012

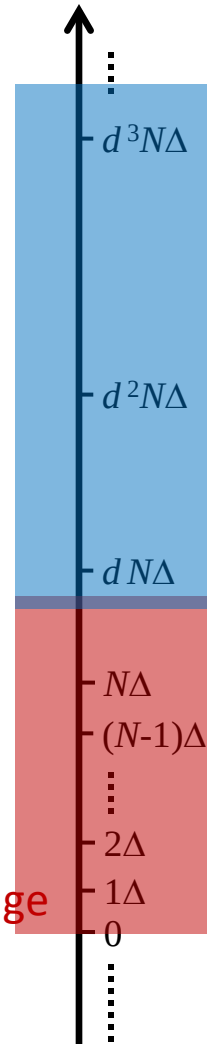
Non-uniform quantization with exponential step size: $(d^r - d^{r-1})N\Delta$

$$Q(L) = \begin{cases} (1, r), & \text{if } d^r N\Delta \le L < d^{r+1} N\Delta \\ (1, -r), & \text{if } -d^{r+1} N\Delta < L \le -d^r N\Delta \\ (1, N+1), & \text{if } L \ge d^{N+1} N\Delta \\ (1, -N-1), & \text{if } L \le -d^{N+1} N\Delta \\ (0, l), & \text{if } l\Delta - \frac{\Delta}{2} < L \le l\Delta + \frac{\Delta}{2} \\ (0, N), & \text{if } N\Delta - \frac{\Delta}{2} < L < dN\Delta \\ (0, -N), & \text{if } -dN\Delta < L \le -N\Delta + \frac{\Delta}{2} \end{cases}$$

q-bit uniform quantization with step size: Δ

exponential range

uniform range

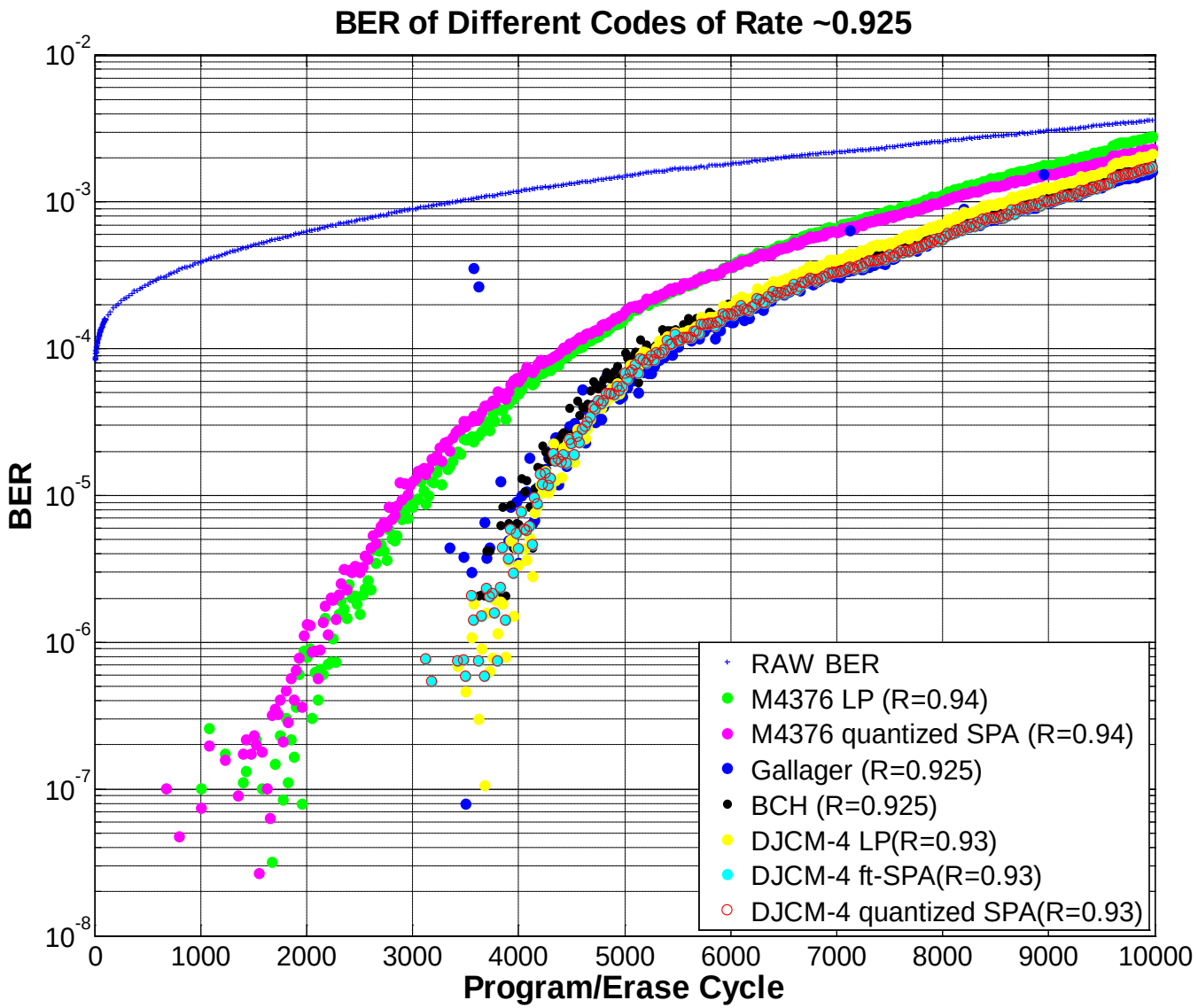


where $N = 2^{q-1} - 1$, $-N + 1 \le l \le N - 1$, $1 \le r \le N$ and d is a quantization parameter within the range $(1, d_v - 1]$.

Legend Notation

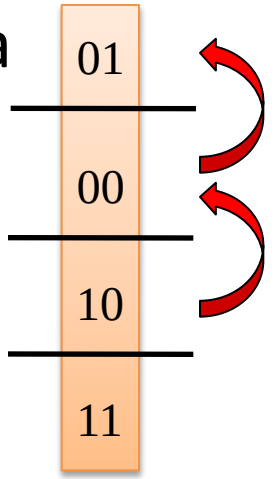
- M4376: MacKay code, length 4376 and rate 0.9356
- DJCM-4: MacKay code, length 3200 and rate 0.93
- LP: ADMM-based LP decoder, max iterations 200
- ft-SPA: floating-point SPA
- Quantized SPA: (5+1)-bit quasi-uniform quantized SPA

R≈0.925, LP vs. SPA Decoding on TLC



Cell-based ECC

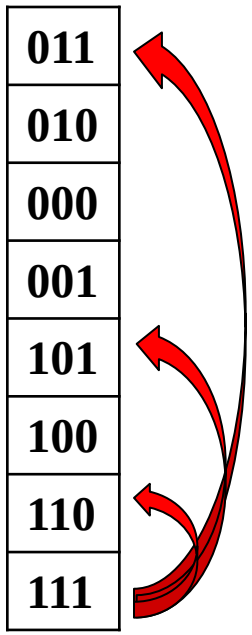
- Experiments have shown that certain specific cell-error types are dominant in MLC and TLC flash memories.
- The dominant cell errors in MLC involved a change in cell voltage by only one level: 10 to 00 or 00 to 01.
- An algebraic code that targets such errors by sharing redundancy between MSB and LSB pages showed improved BER vs. P/E.*



* Yaakobi, et al., *Proc. Globecom 2010*, Dec. 2010.

ECC Scheme for TLC Flash

- If a TLC cell is in error, then with high probability only one of the three bits in the cell is in error.
- The probability of a bit being in error does not depend on the target cell level.
- Algebraic coding schemes that target such errors offer potential BER improvements.[1,2]

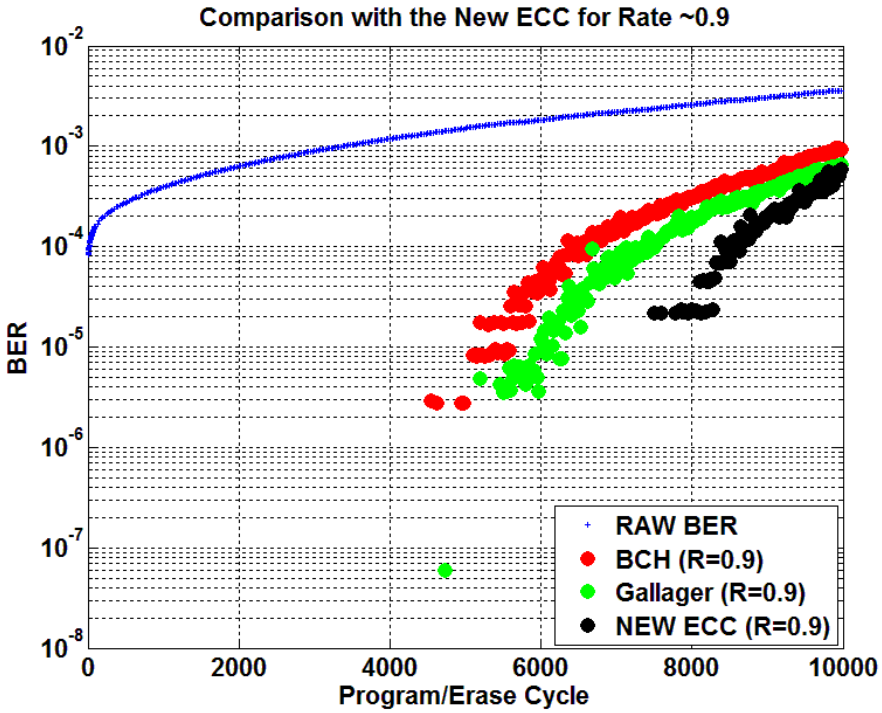


[1] Yaakobi, et al., *Proc. Int. Conf. on Comp., Network. Commun. (ICNC),* Jan.-Feb. 2012.

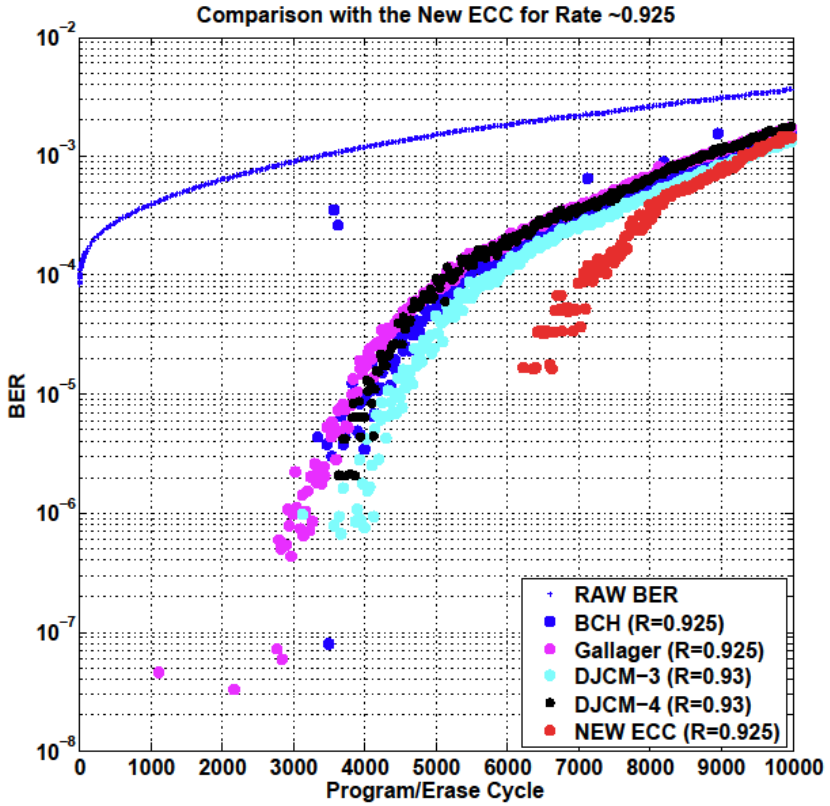
[2] Gabrys, et al., *Proc. IEEE Int. Symp. Inf. Theory*, July 2012.

BER for Cell-based Code for TLC Flash*

R≈0.9



R≈0.925

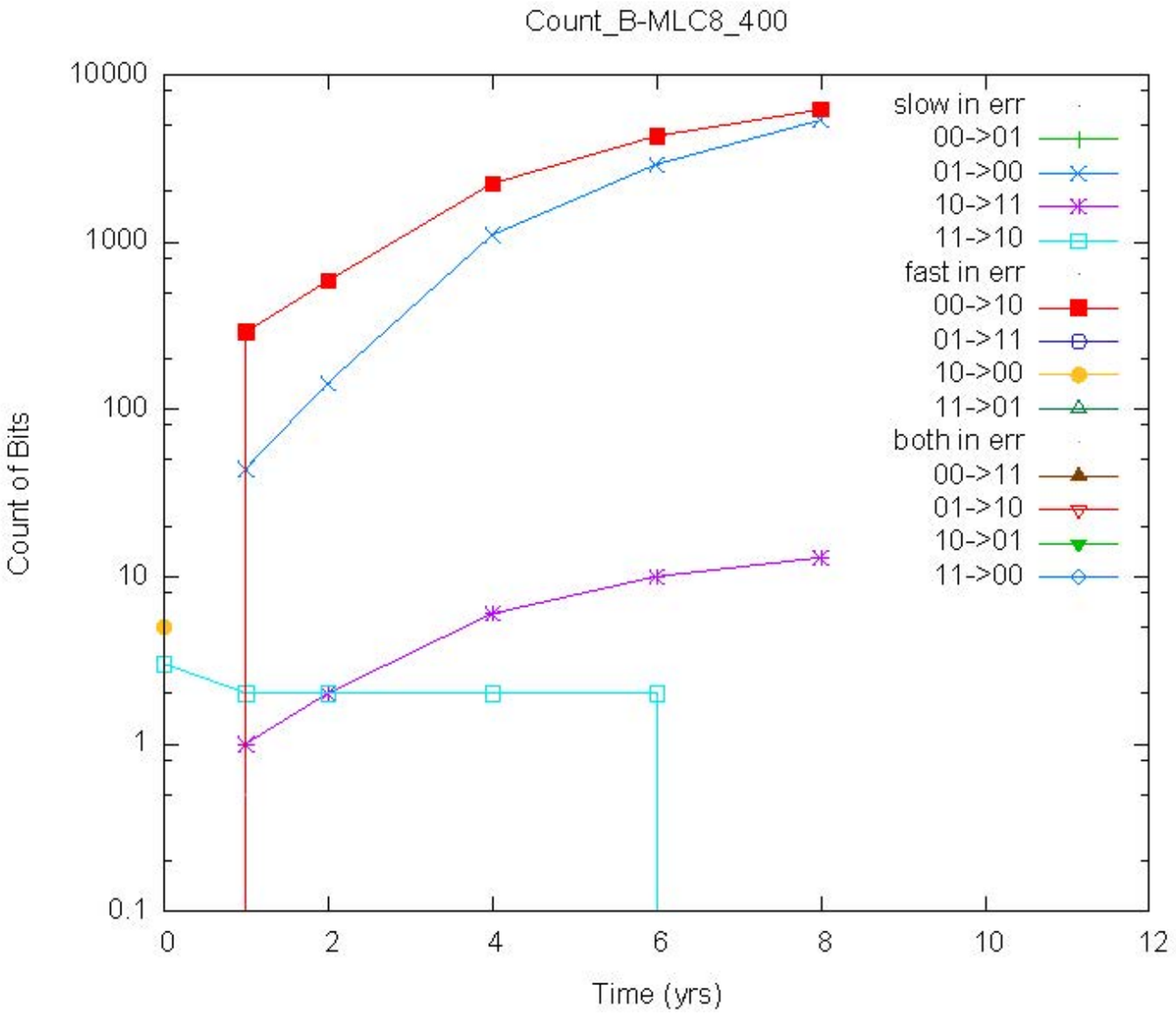


*Yaakobi, et al., *Proc. Int. Conf. on Comp., Network. Commun. (ICNC),* Jan.-Feb. 2012.

General Observations

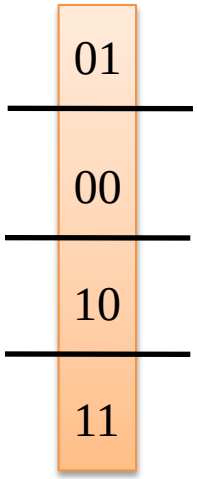
- Best LDPC performance surpasses BCH at all code rates R≈ 0.8, 0.9, 0.925.
- R≈0.8 LDPC codes at 15k cycles has BER comparable to R≈0.9 LDPC codes at 10k cycles.
- MSA was inferior to SPA decoding at R≈0.8.
- LP-ADMM was comparable to SPA decoding at R≈0.925, with slightly steeper slope.
- (5+1)-bit quasi-uniform quantized SPA (not optimized) matches floating-point SPA.
- Algebraic codes that target dominant cell-error types can offer improved performance.

MLC Data Retention



Cycle chip to 400% of lifetime.

Bake at 125°C for 9hrs20mins per year of aging.



Future Directions

- Further characterization of dominant cell-error types in multilevel flash memory is needed.
- ECC, in combination with constrained codes, will need to address effects of inter-cell and inter-page interference, as well as effects of device aging, power cutoff, and other factors.
- Soft-decision decoding should offer further challenges and potential performance gains.

Acknowledgments

Thank you

and please come to

NVMW 2013