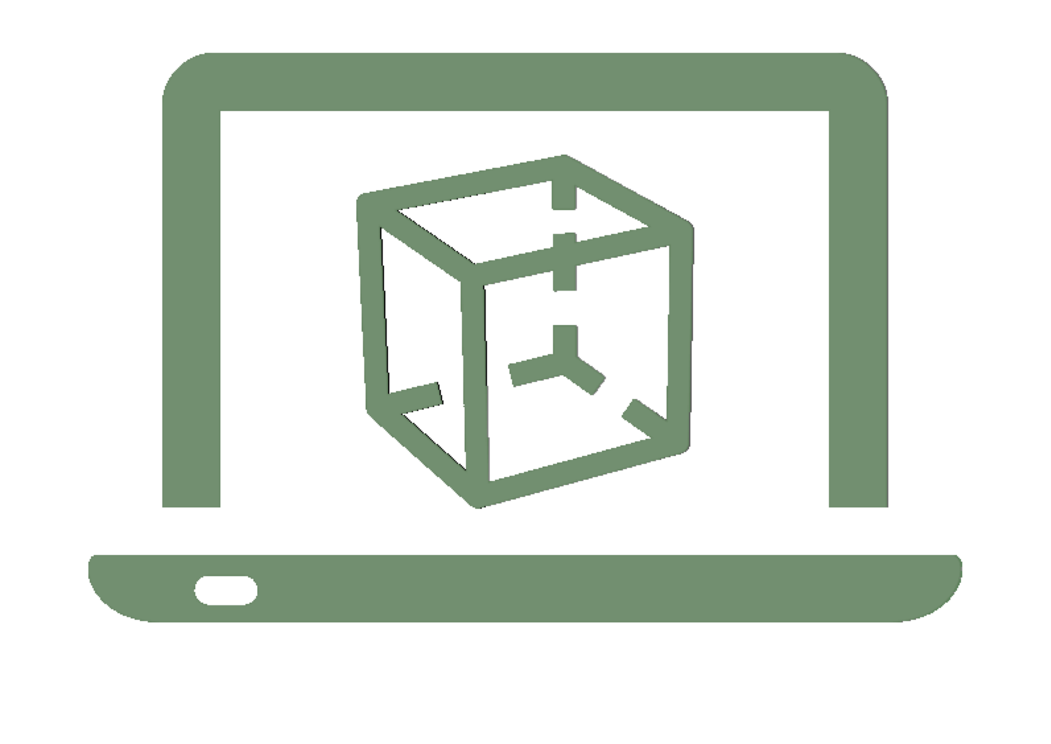


Online Flash Channel Modeling and Its Applications

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

Carnegie Mellon University, Seagate Technology

Flash as a Communication Channel

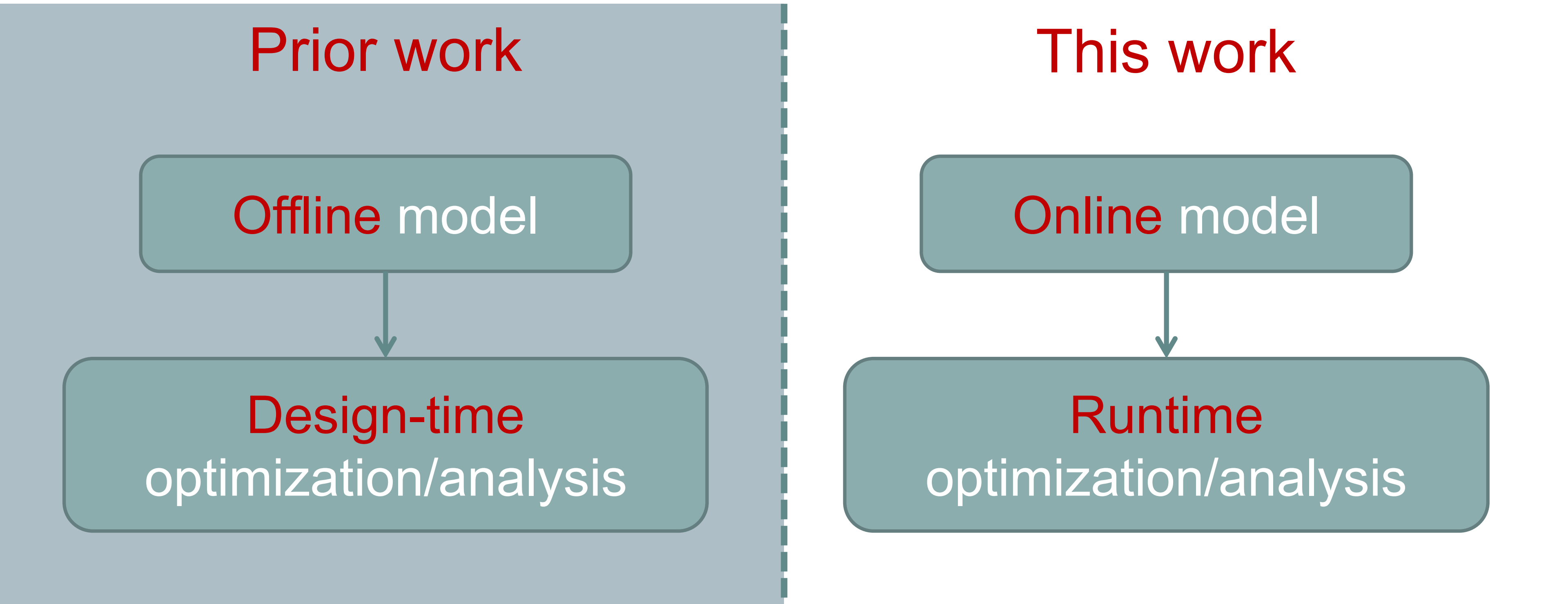
Motivation: Understanding flash channel can help minimize errors through the channel, or tolerate more errors efficiently

Prior Works on Distribution Models

- Design time analysis
  - Offline threshold voltage shift analysis [Cai+ DATE ’13]
  - Offline RBER analysis [Parnell+ GLOBECOM ’14]
- Design time optimization
  - Read reference voltage optimization [Papandreou+ GLSVLSI ’14]
  - ECC soft information optimization [Dong+ TCS ’13]
- Can’t be run online – none of these are both accurate *and* easy-to-compute

Why Online Modeling?

- Flash controllers becoming more powerful
- Can use idle cycles for background optimization
- Can adapt to real-world variation



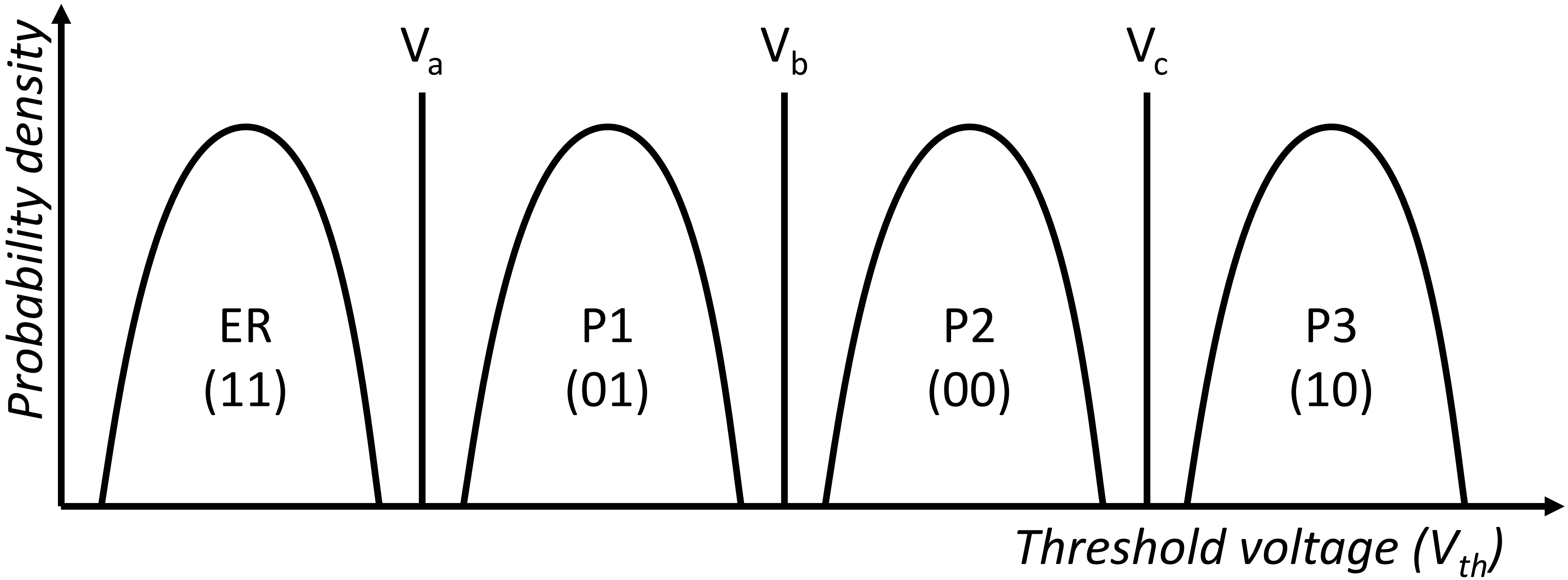
Goal

- Create online flash channel model
  - Helps with understanding flash channel
  - Enables runtime optimizations
  - Must be *accurate* and *easy to compute*
- Develop model-driven applications
  - Work to reduce or tolerate flash errors

Outline

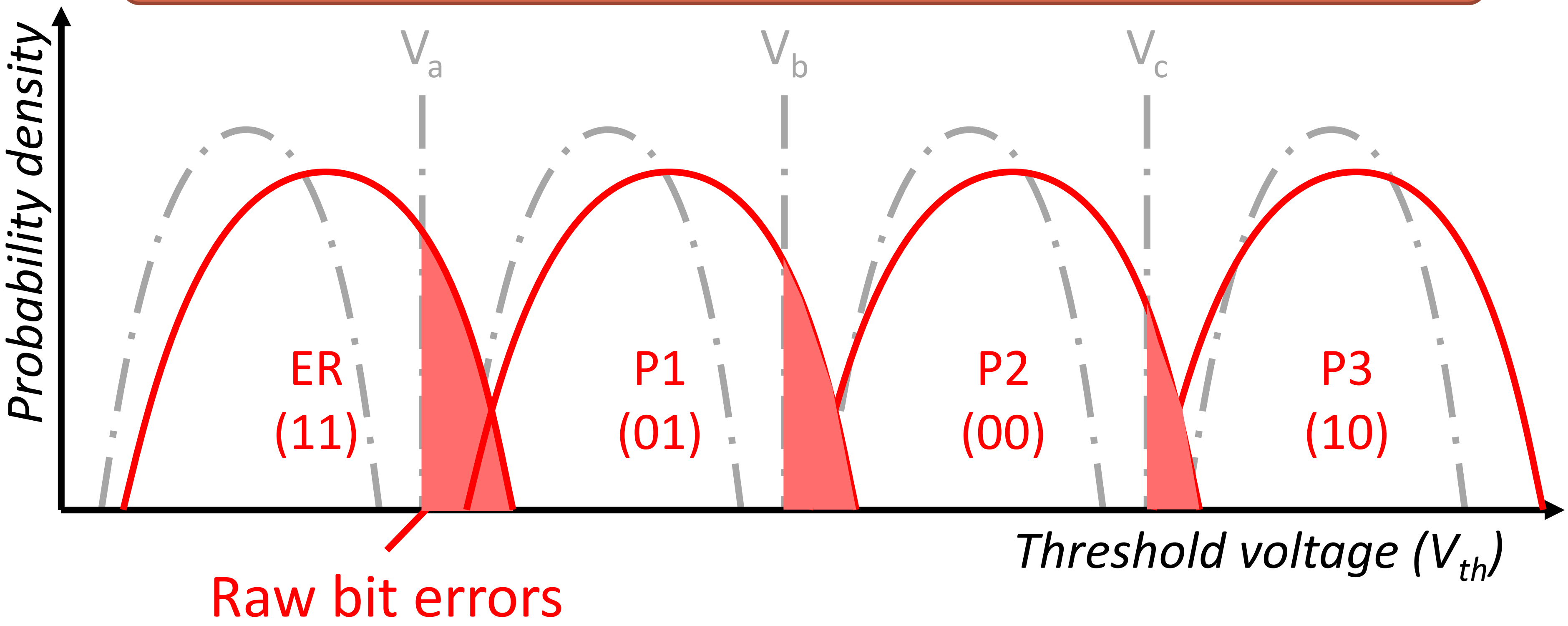
- What do we model?
  - Program variation noise
  - Program/erase cycling noise
- How do we model it?
  - Static flash channel model → program variation
  - Dynamic flash channel model → P/E cycling noise
- Applications of Online Flash Channel Model

Program Variation Noise



Program/Erase Cycling Noise

Distribution shifts increase raw bit errors



Flash Memory SUMMIT

Outline

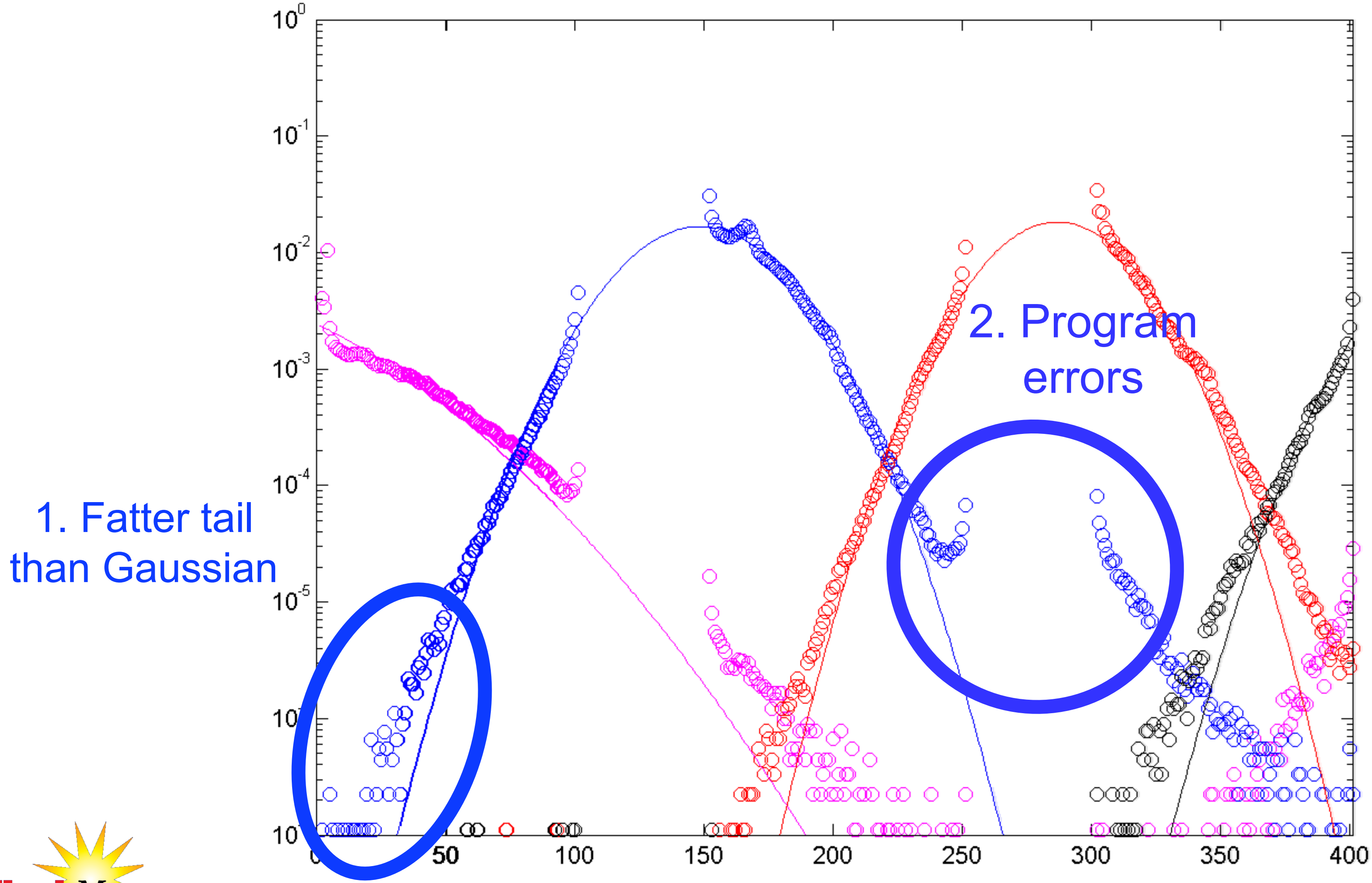
- What do we model?
  - Program variation noise
  - Program/erase cycling noise
- How do we model it?
  - Static flash channel model → program variation
  - Dynamic flash channel model → P/E cycling noise
- Applications of Online Flash Channel Model

Static Flash Channel Model

- Program variation noise
- Threshold voltage distribution @ N P/E cycles

- Program variation noise should be normally distributed → Why don't we use a Gaussian model?

Gaussian Model Isn't Accurate Enough



Student's t-Distribution

- Real distribution has larger tail than Gaussian
- Student's t has degree of freedom: v
  - v→∞: t-distribution → Gaussian
  - v→1: largest tail

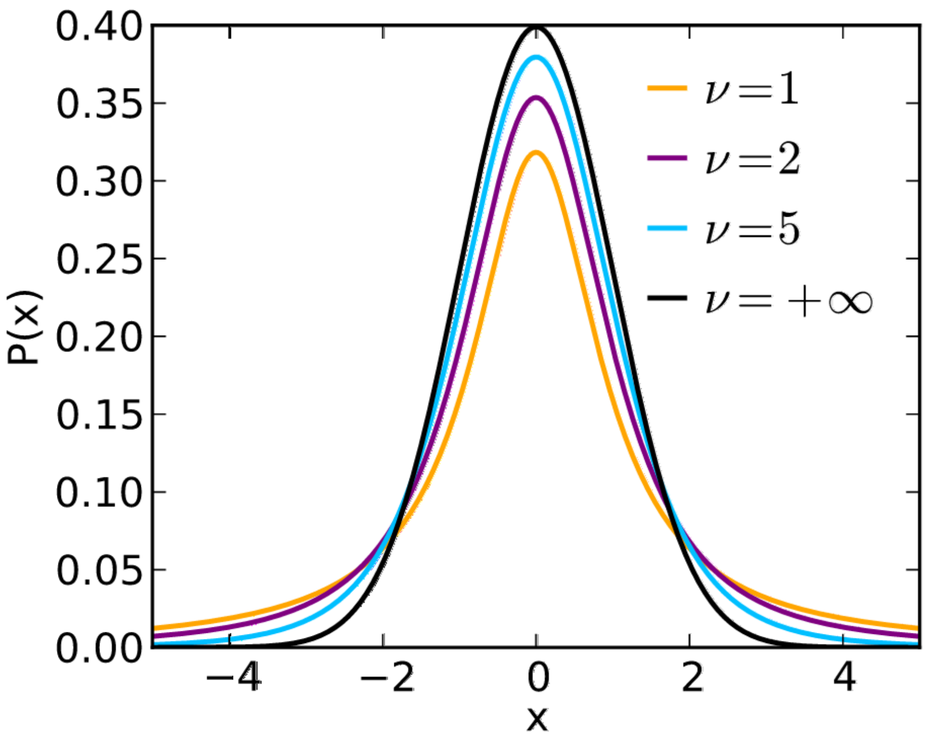
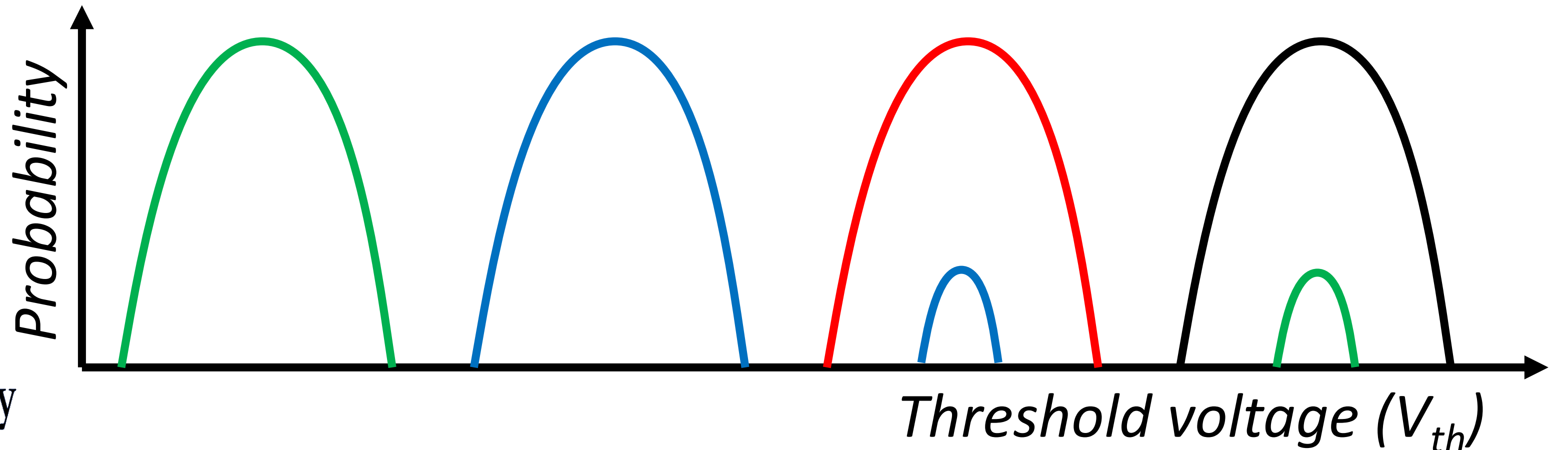


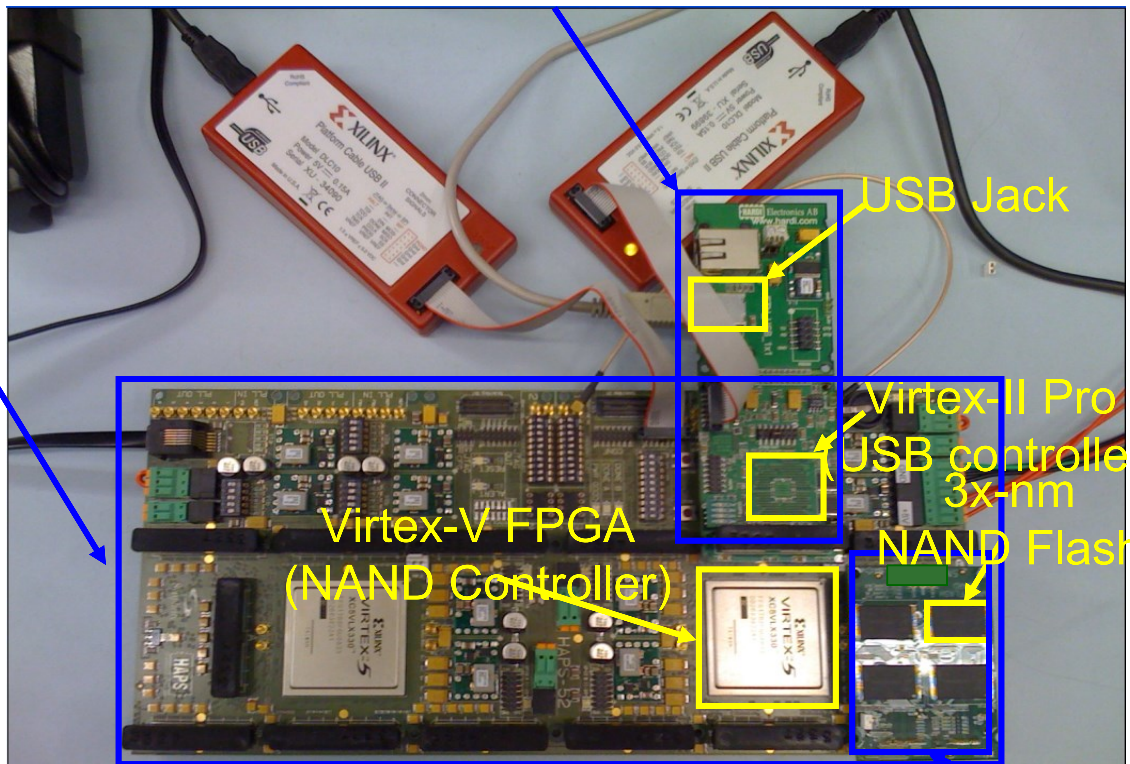
Image source: https://en.wikipedia.org/wiki/Student%27s_t-distribution

Modifications to Student's t-Distribution

- Generalize distribution
  - Allows for shifting and scaling
  - x → Z = $\frac{x-\mu}{\sigma}$
- Support asymmetric tail sizes: v → α(right), β(left)
- Superposition of two distributions
  - Cause: Two-step programming errors



Characterization Methodology

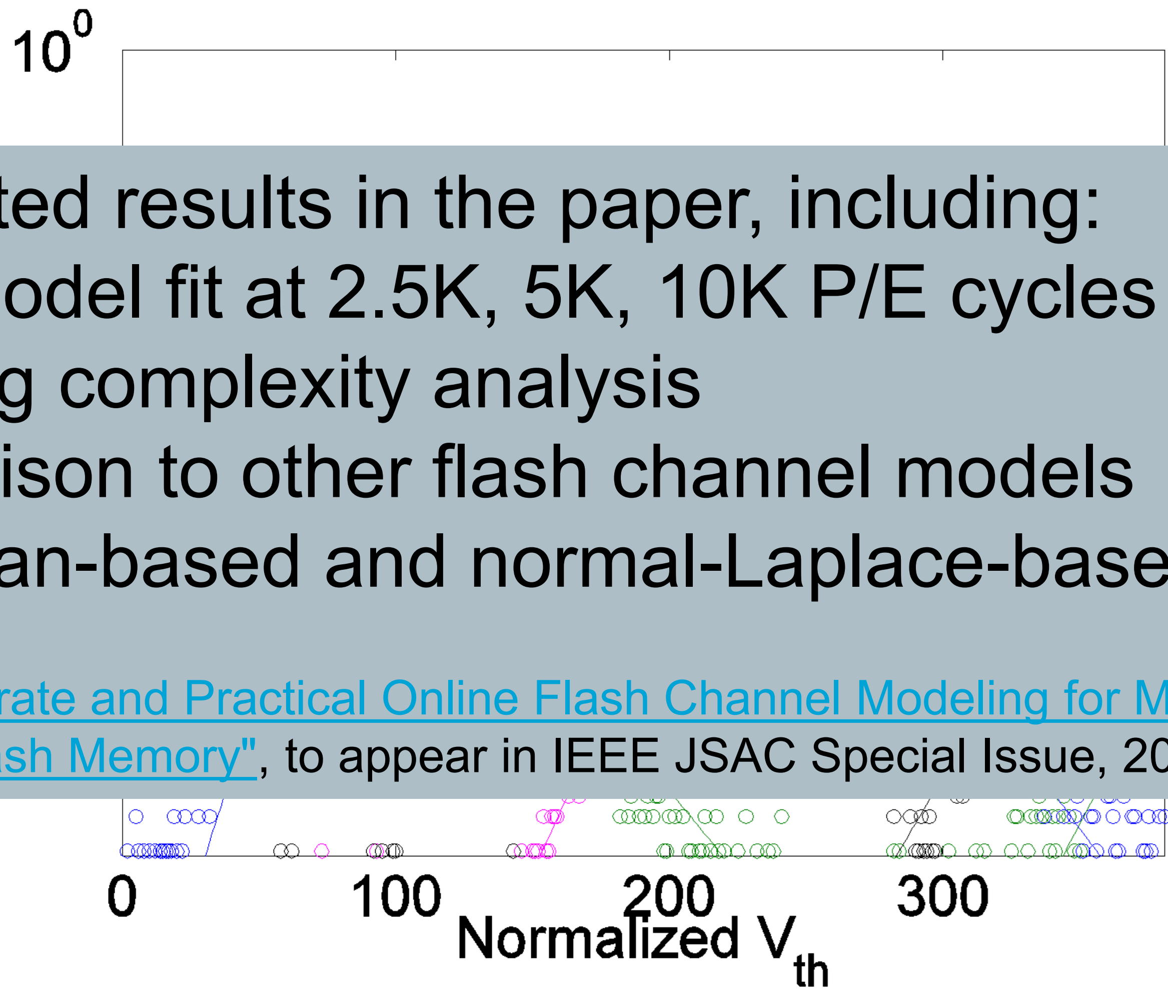


[Cai+, FCCM 2011, DATE 2012, ICCD 2012, DATE 2013, ITJ 2013, ICCD 2013, SIGMETRICS 2014, DSN 2015, HPCA 2015]

Cai et al., FPGA-based Solid-State Drive prototyping platform, FCCM 2011.

Static Modeling Results

- Our model (curve) vs. characterized (circle) @ 20K P/E cycle



More related results in the paper, including:

- Static model fit at 2.5K, 5K, 10K P/E cycles
- Modeling complexity analysis
- Comparison to other flash channel models (Gaussian-based and normal-Laplace-based)

"Enabling Accurate and Practical Online Flash Channel Modeling for Modern MLC NAND Flash Memory", to appear in IEEE JSAC Special Issue, 2016

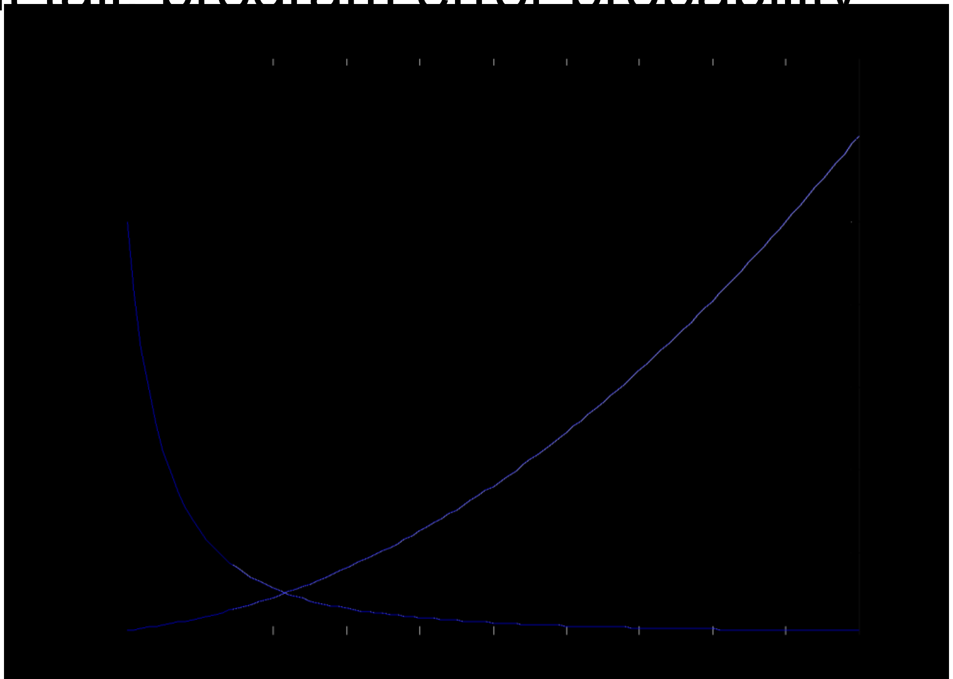
Outline

- What do we model?
  - Program variation noise
  - Program/erase cycling noise
- How do we model it?
  - Static flash channel model → program variation
  - Dynamic flash channel model → P/E cycling noise
- Applications of Online Flash Channel Model

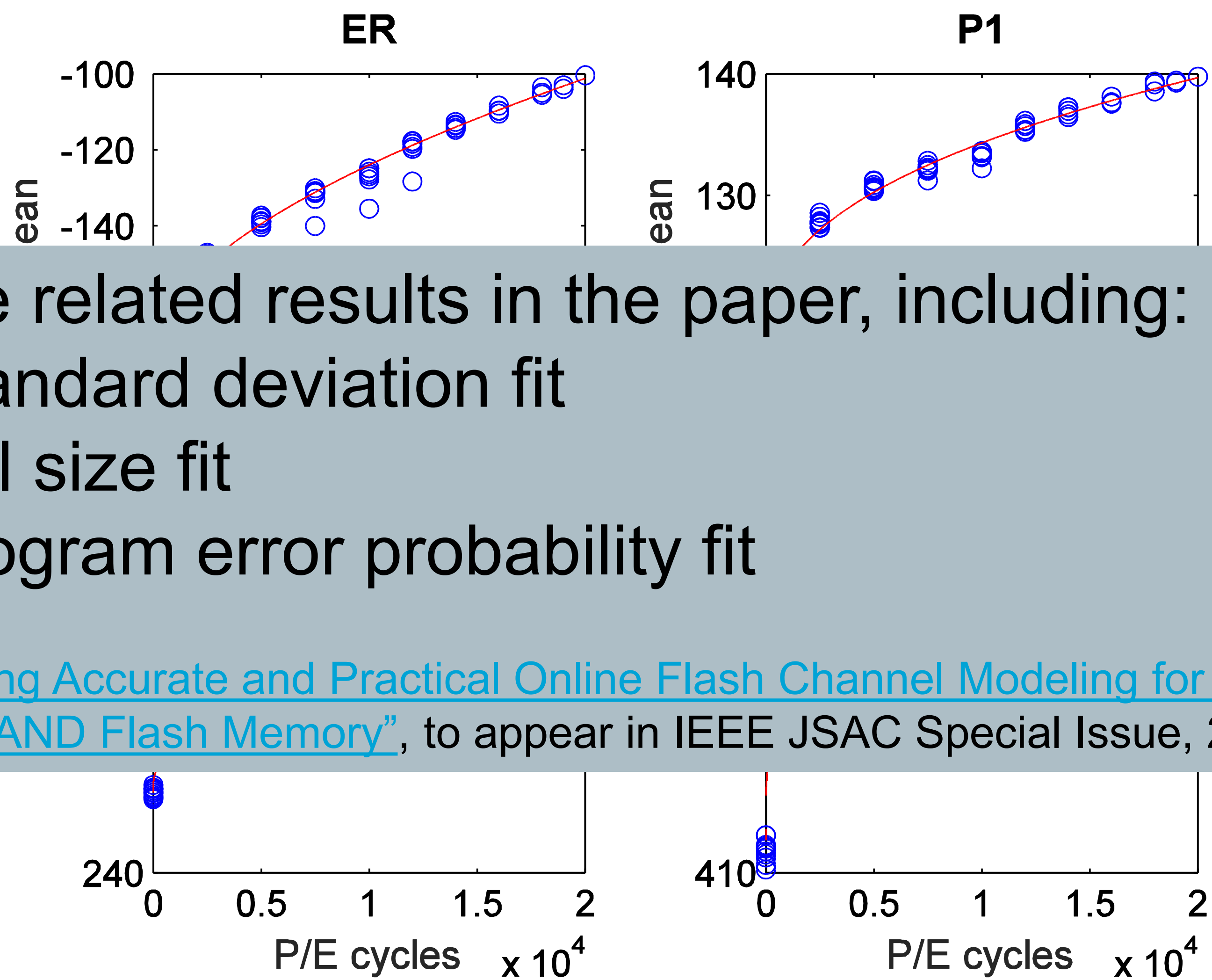
Dynamic Flash Channel Model

- P/E cycling noise
- Threshold voltage distribution shift

- *Dynamic model* modifies static model's parameters: mean, variance, left/right tail, program error probability
- Power-law model
  - $Y = a * x^b + c$



Flash Channel Model Results (Dynamic)



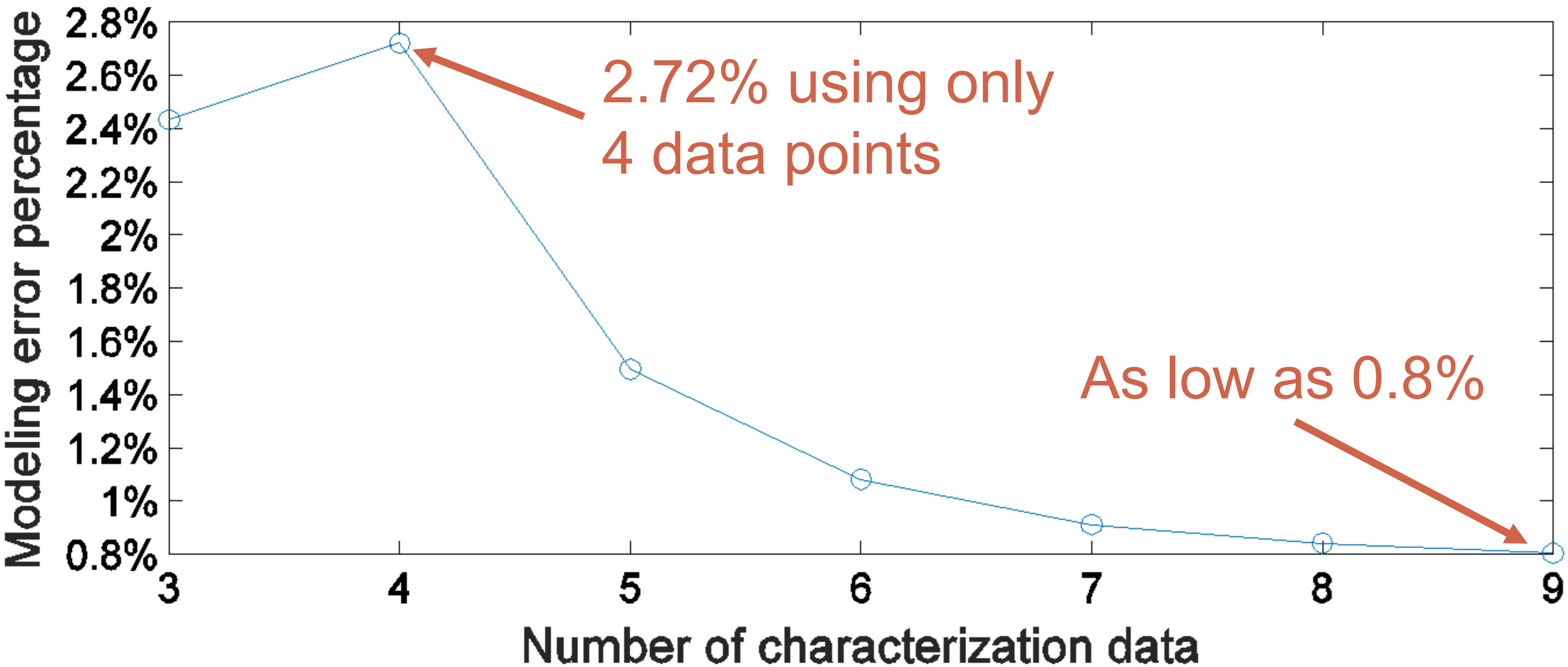
More related results in the paper, including:

- Standard deviation fit
- Tail size fit
- Program error probability fit

"Enabling Accurate and Practical Online Flash Channel Modeling for Modern MLC NAND Flash Memory", to appear in IEEE JSAC Special Issue, 2016

Flash Channel Model Results (Dynamic)

- Using N prior characterizations to predict flash channel @ 20K P/E cycle



Outline

- What do we model?
  - Program variation noise
  - Program/erase cycling noise
- How do we model it?
  - Student's t-based model → program variation
  - Power law-based model → P/E cycling noise
- Applications of Online Flash Channel Model
- Results

Optimal Read Reference Voltage Prediction

- Improves flash lifetime
  - 48.9% longer flash lifetime
- Minimizes number of read-retries
- Faster soft ECC decoding

Expected Lifetime Estimation

- Safely go beyond manufacturer-specified lifetime
  - 69.9% higher flash lifetime usage

Other Applications of Our Model

- Raw Bit Error Rate Estimation
  - Predict ECC margin, apply variable ECC strength
- Soft Information Estimation for LDPC Codes
  - Improves coding efficiency

Outline

- What do we model?
  - Program variation noise
  - Program/erase cycling noise
- How do we model it?
  - Student's t-based model → program variation
  - Power law-based model → P/E cycling noise
- Applications of Online Flash Channel Model

Conclusion

- Goal: Develop an online flash channel model, and utilize this model to improve flash reliability
- Static flash channel model
  - 0.68% modeling error
  - Amortized read latency overhead <50 ns
- Dynamic flash channel model
  - 2.72% modeling error
  - Using only 4 data points (even lower overhead)
- Example applications of online model
  - 48.9% longer flash lifetime, or 69.9% higher flash usage
  - Hopefully inspires other reliability/performance improving techniques to use our online model

Online Flash Channel Modeling

Yixin Luo
yixinluo@cs.cmu.edu
http://www.cs.cmu.edu/~yixinluo/

This presentation is based on a paper to appear in IEEE JSAC Special Issue, 2016:
"Enabling Accurate and Practical Online Flash Channel Modeling for Modern MLC NAND Flash Memory",
Yixin Luo, Saugata Ghose, Yu Cai, Erich F. Haratsch, Onur Mutlu

Our Other FMS 2016 Talks

- “Software-Transparent Crash Consistency for Persistent Memory”
  - Onur Mutlu (ETH Zurich & CMU) August 8 @ 11:40am
  - PreConference Seminar C: Persistent Memory
- "A Large-Scale Study of Flash Memory Errors in the Field”
  - Onur Mutlu (ETH Zurich & CMU) August 10 @ 3:50pm
  - Study of flash-based SSD errors in Facebook data centers over the course of 4 years
  - First large-scale field study of flash memory reliability
  - Forum F-22: SSD Testing (Testing Track)
- "WARM: Improving NAND Flash Memory Lifetime with Write-hotness Aware Retention Management”
  - Saugata Ghose (CMU Researcher) August 10 @ 5:45pm
  - Forum C-22: SSD Concepts (SSDs Track)