Flash Memory SUMMIT

Data Retention in MLC NAND Flash Memory: Characterization, Optimization, and Recovery

Yixin Luo
yixinluo@cmu.edu

(joint work with Yu Cai, Erich F. Haratsch, Ken Mai, Onur Mutlu)

Presented in the best paper session at HPCA 2015

SAFARI Carnegie Mellon SEAGATE DSSC

Flash Memory SUMMIT

Characterize retention loss in real NAND chip

Optimize read performance for old data

Recover old data after failure

TECHSPOT

TRENDING ▾ REVIEWS ▾ FEATURES ▾ DOWNLOADS ▾ PRODUCT FINDER ▾ FORUMS ▾ TEC

An unfortunate tale about Samsung's SSD 840

read performance degradation

An avalanche of reports emerged last September, when owners of the usually speedy Samsung SSD 840 and SSD 840 EVO detected the drives were no longer performing as they used to.

The issue has to do with older blocks of data: reading **old files** nsistently slower than norma **as slow as 30MB/s** whereas **newly-written files** ones used in benchmarks, perform as fast as new – aro **500 MB/s** for the well regarded SSD 840 EVO. The reason no one had noticed (we reviewed the drive back in September 2013) is that data has to be several weeks old to show the problem. Samsung promptly admitted the issue and proposed a fix.

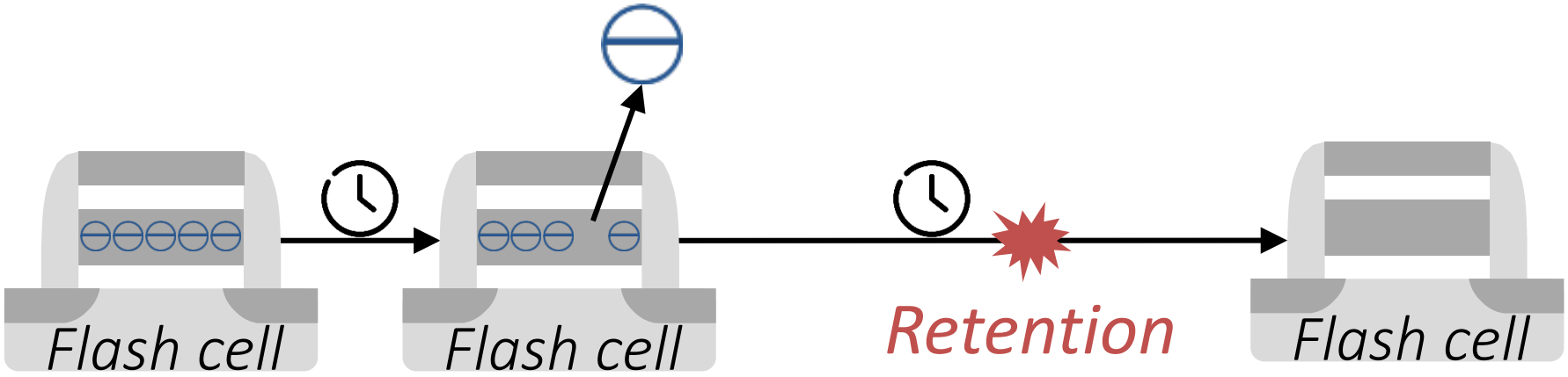
Reference: (May 5, 2015) Per Hansson, "*When SSD Performance Goes Awry*"
http://www.techspot.com/article/997-samsung-ssd-read-performance-degradation/

Why is old data slower?

Retention loss!

Retention loss

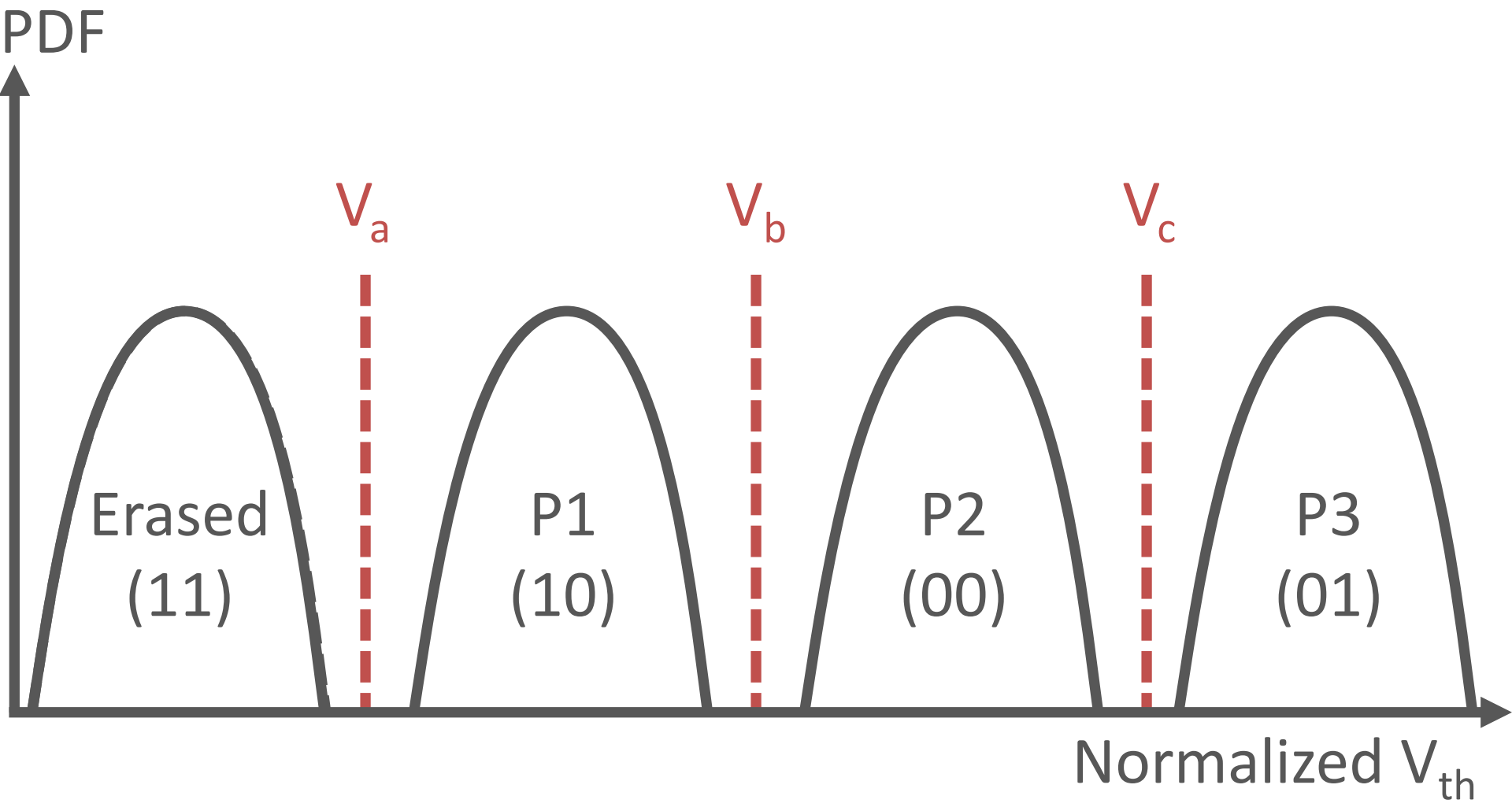
Charge leakage over time



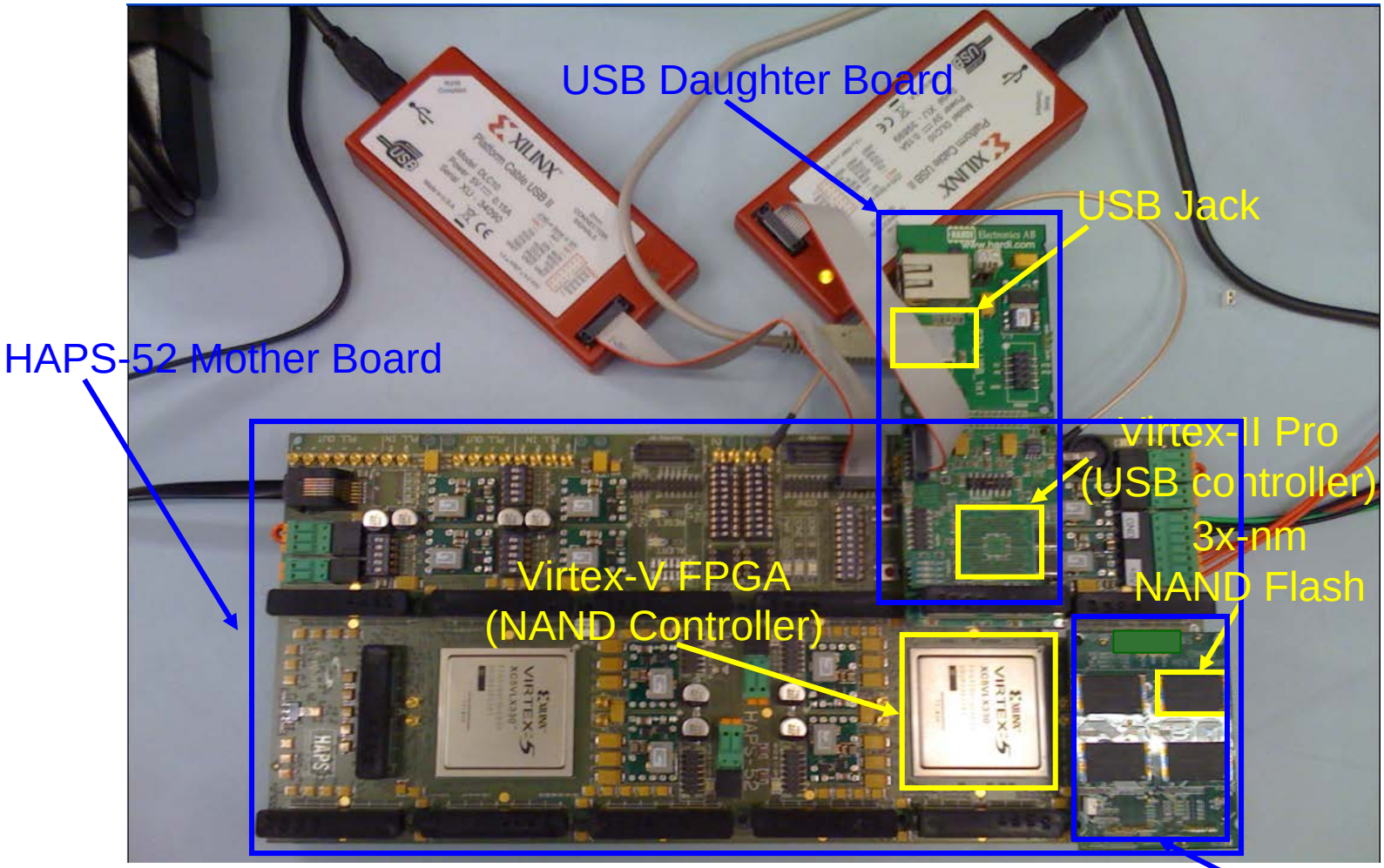
One dominant source of flash memory errors [DATE '12, ICCD '12]

Side effect: Longer read latency

Multi-Level Cell (MLC) threshold voltage distribution



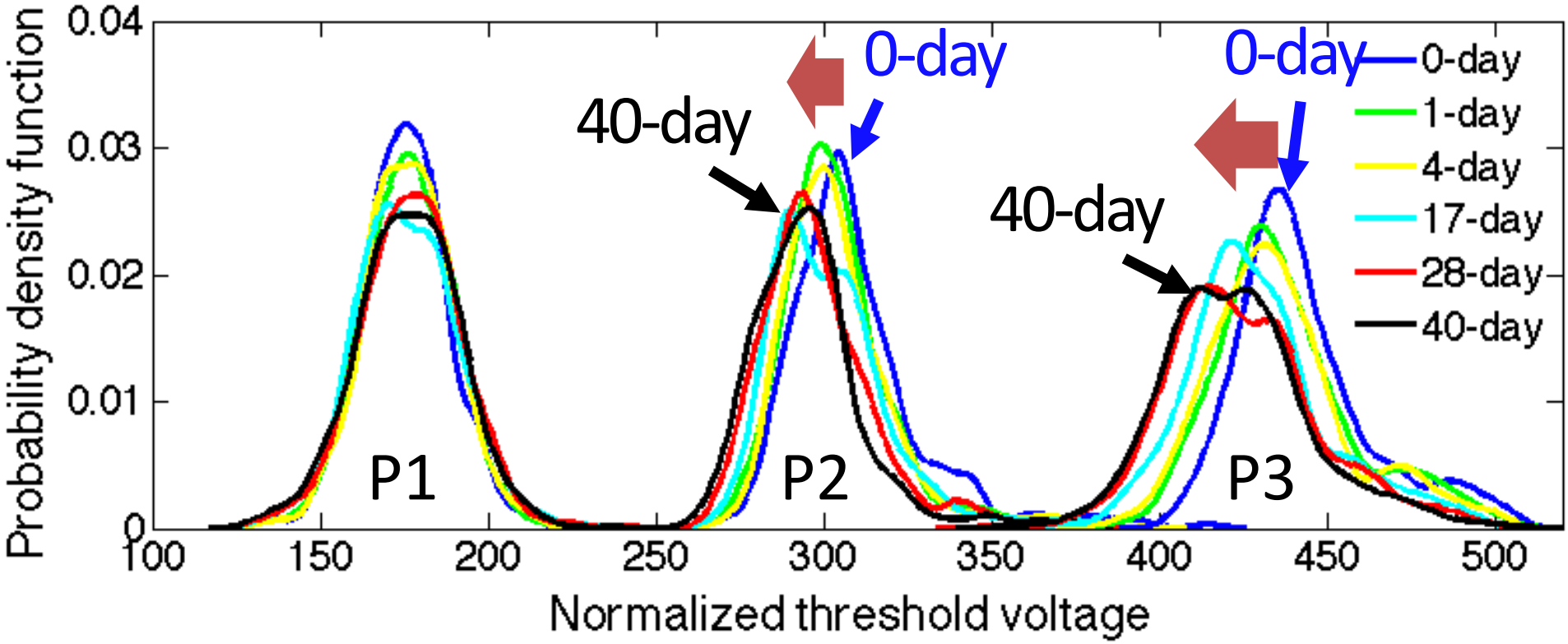
Experimental Testing Platform



[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.

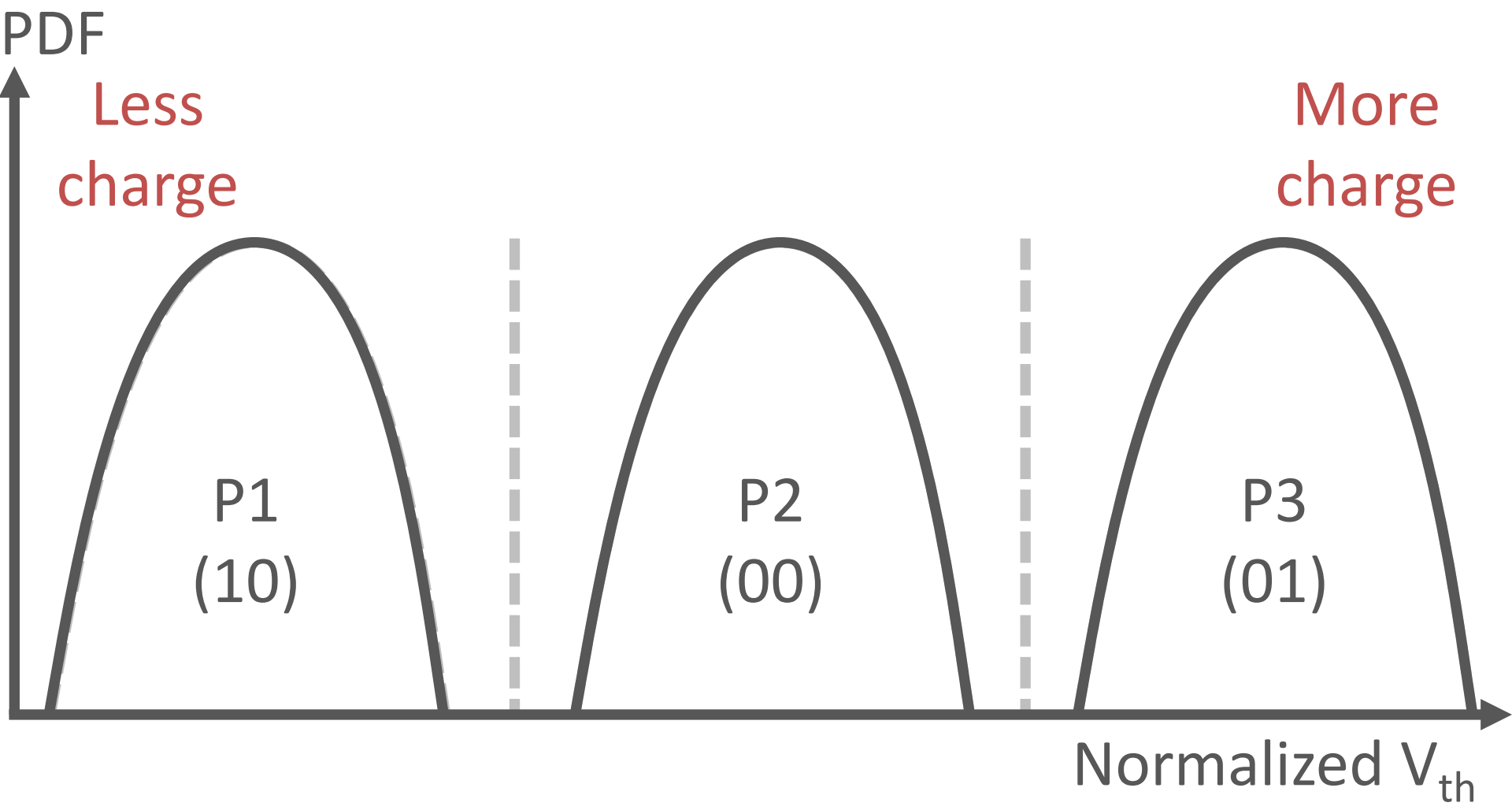
Characterized threshold voltage distribution



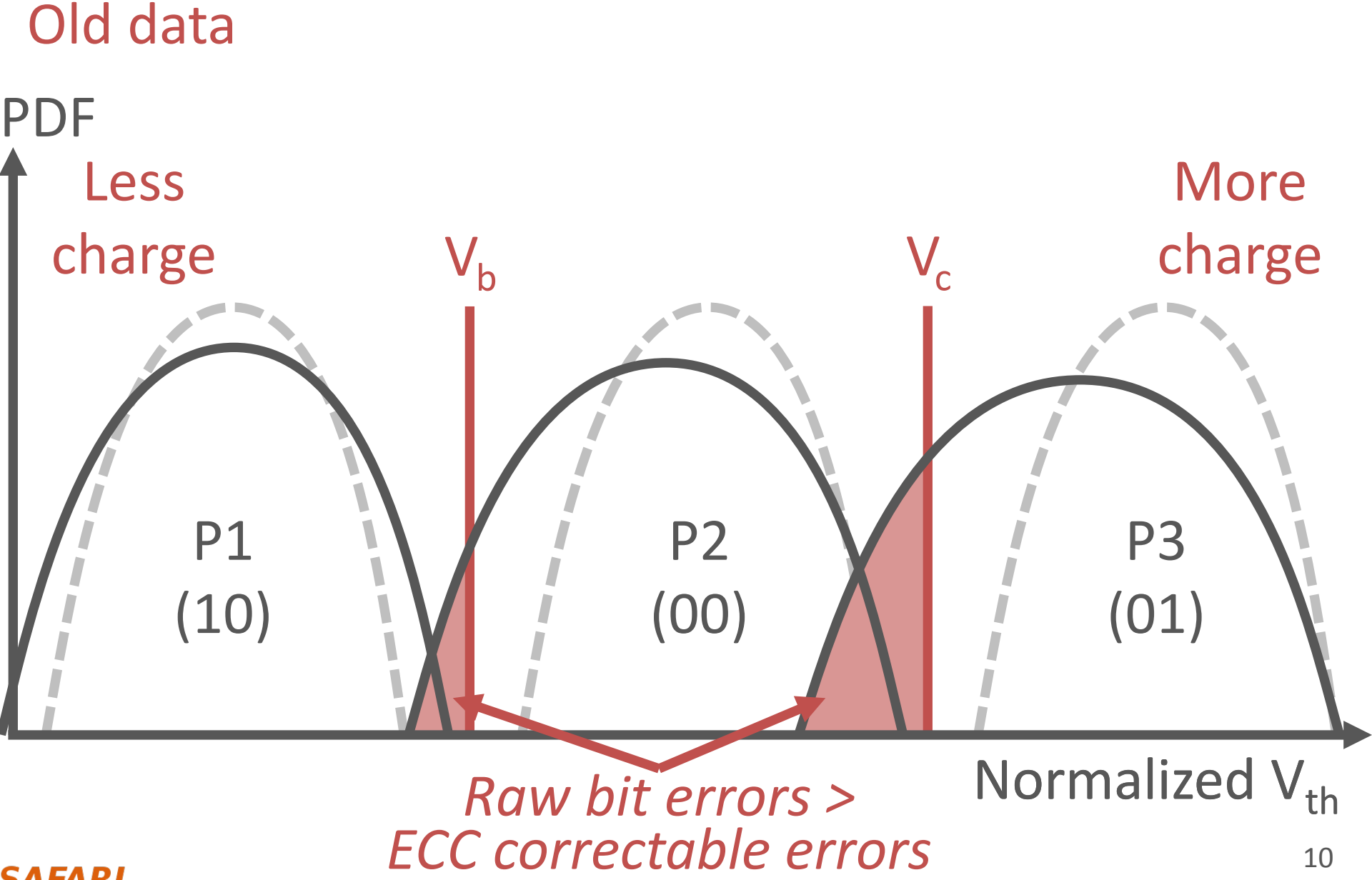
Finding: Cell's threshold voltage decreases over time

Threshold voltage reduces over time

Old data

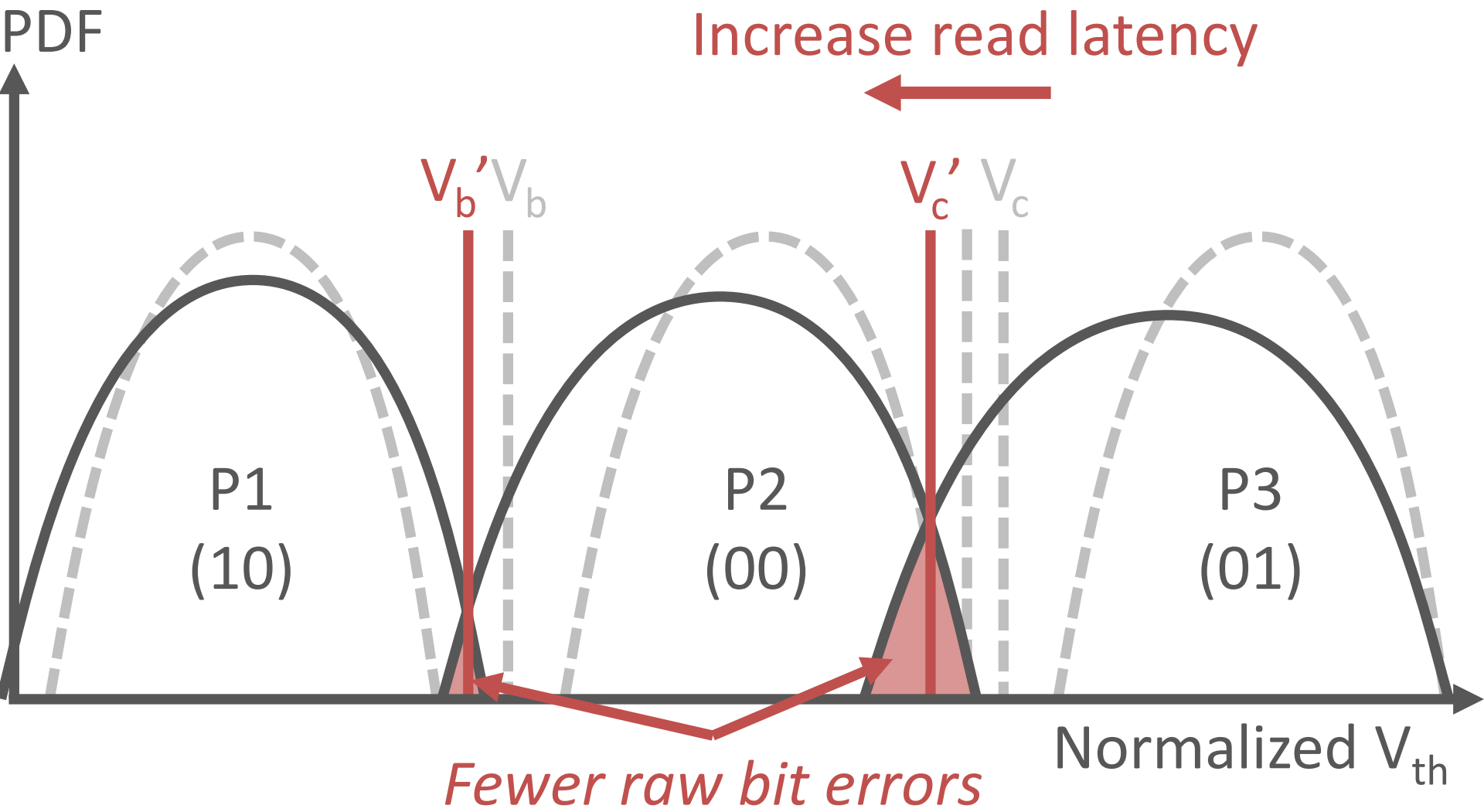


First read attempt fails



Read-retry

Old data



Why is old data slower?

Retention loss

→ *Leak charge over time*

→ *Generate retention errors*

→ *Require read-retry*

→ *Longer read latency*

Flash Memory SUMMIT

Characterize retention loss in real NAND chip

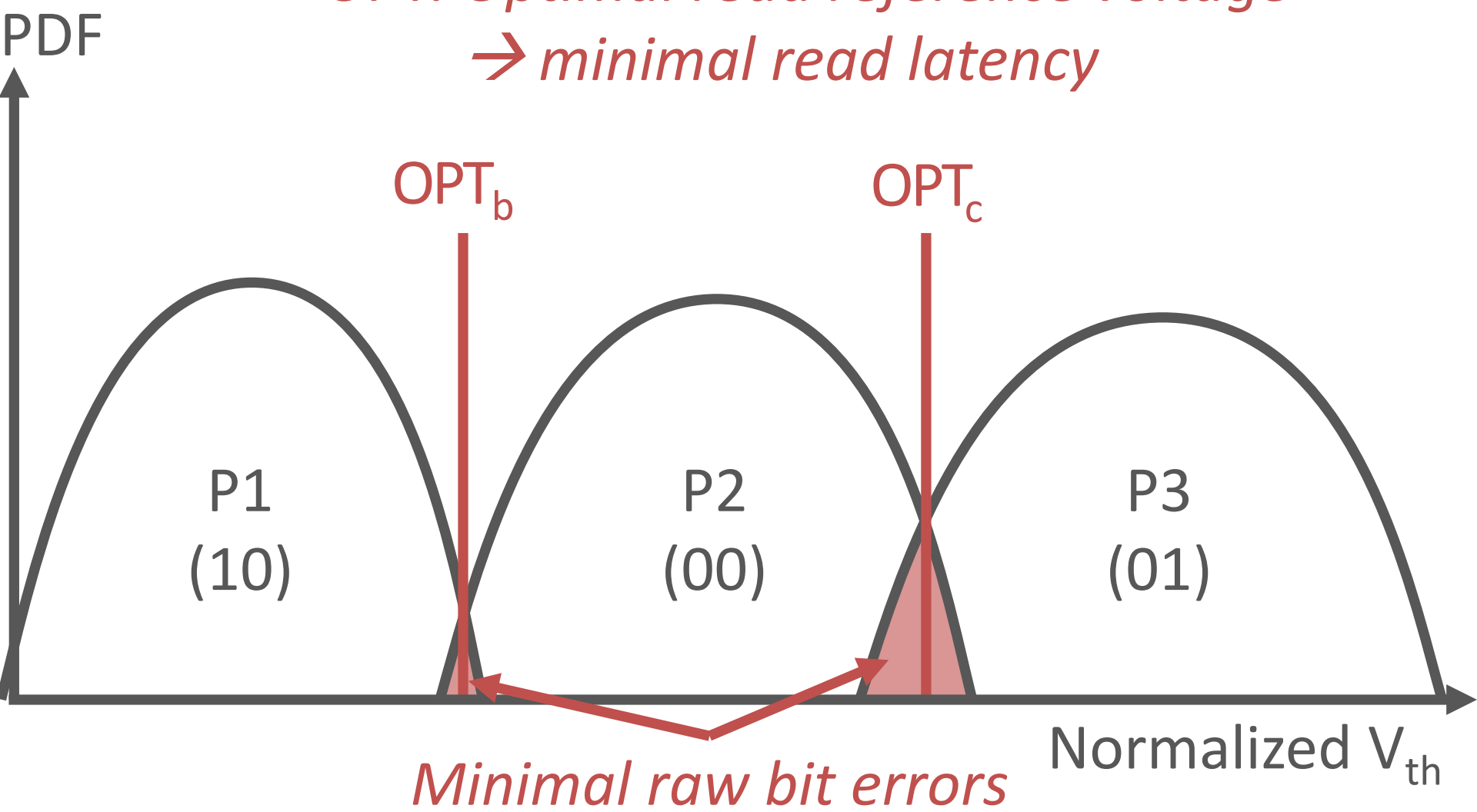
Optimize read performance for old data

Recover old data after failure

The ideal read voltage

Old data

OPT: Optimal read reference voltage
→ minimal read latency



In reality

- *OPT changes over time due to retention loss*

- *Luckily, OPT change is:*
  - Gradual
  - Uni-directional (decrease over time)

Retention Optimized Reading (ROR)

Components:

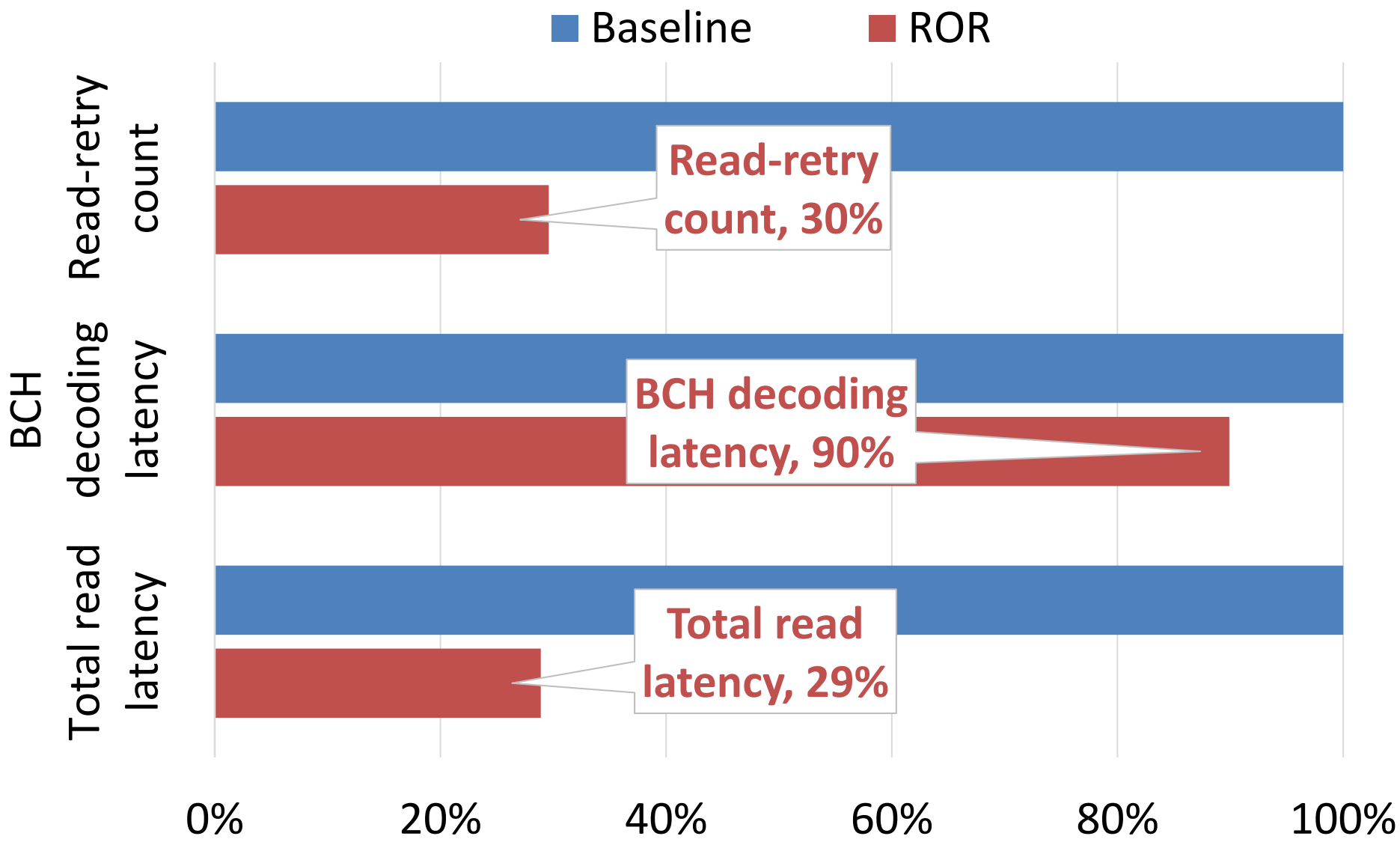
1. Online pre-optimization algorithm

- Learns and records OPT
- Performs in the background once every day

2. Simpler read-retry technique

- If recorded OPT is out-of-date, read-retry with *lower* voltage

ROR result


Baseline
ROR
Read-retry count
BCH decoding latency
Total read latency
Read-retry count, 30%
BCH decoding latency, 90%
Total read latency, 29%
0%
20%
40%
60%
80%
100%

Retention optimized reading

Retention loss → longer read latency

Optimal read reference voltage (OPT)

→ Shortest read latency

→ Decreases gradually over time (retention)

→ Learn OPT periodically

→ Minimize read-retry & RBER

→ Shorter read latency

Flash Memory SUMMIT

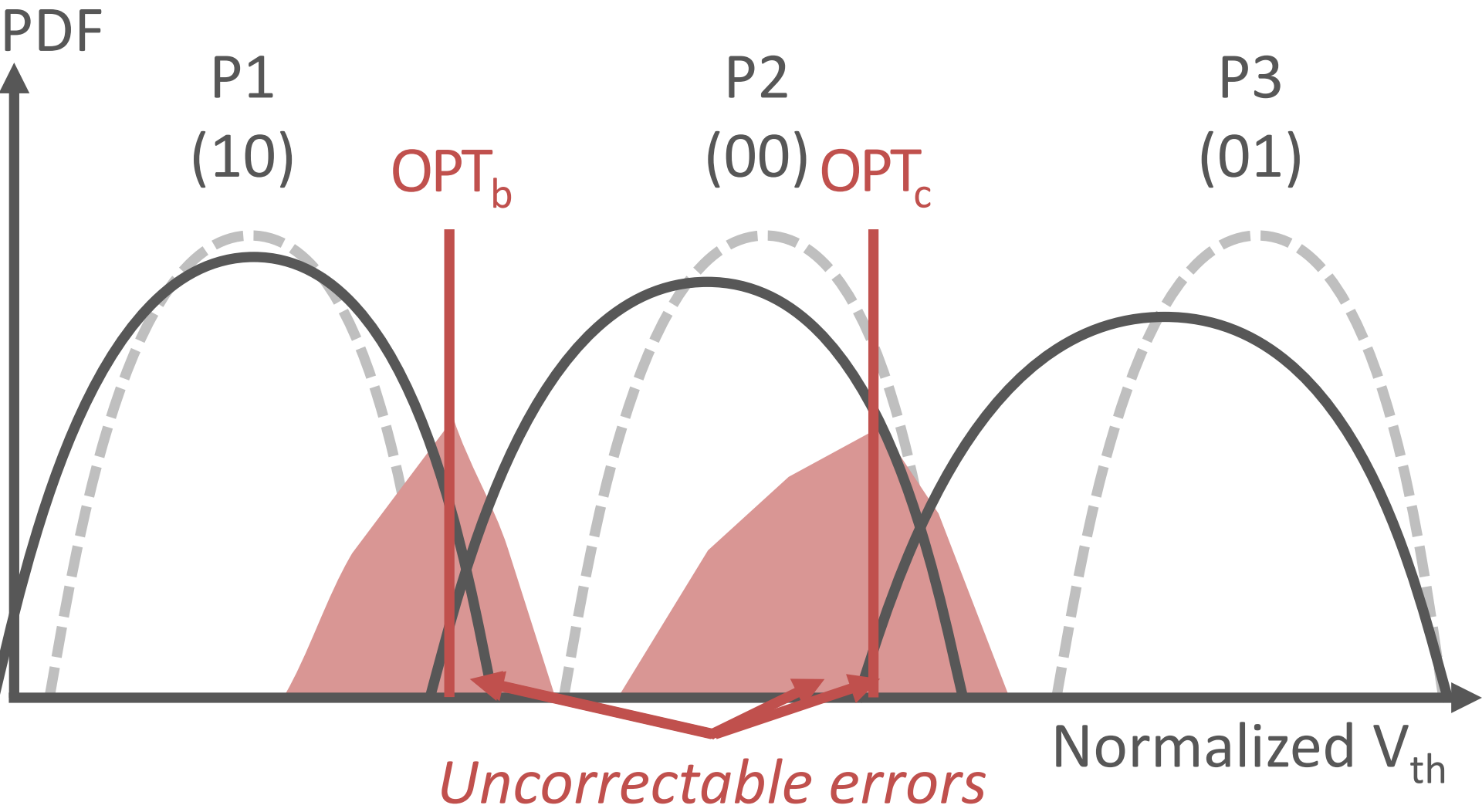
Characterize retention loss in real NAND chip

Optimize read performance for old data

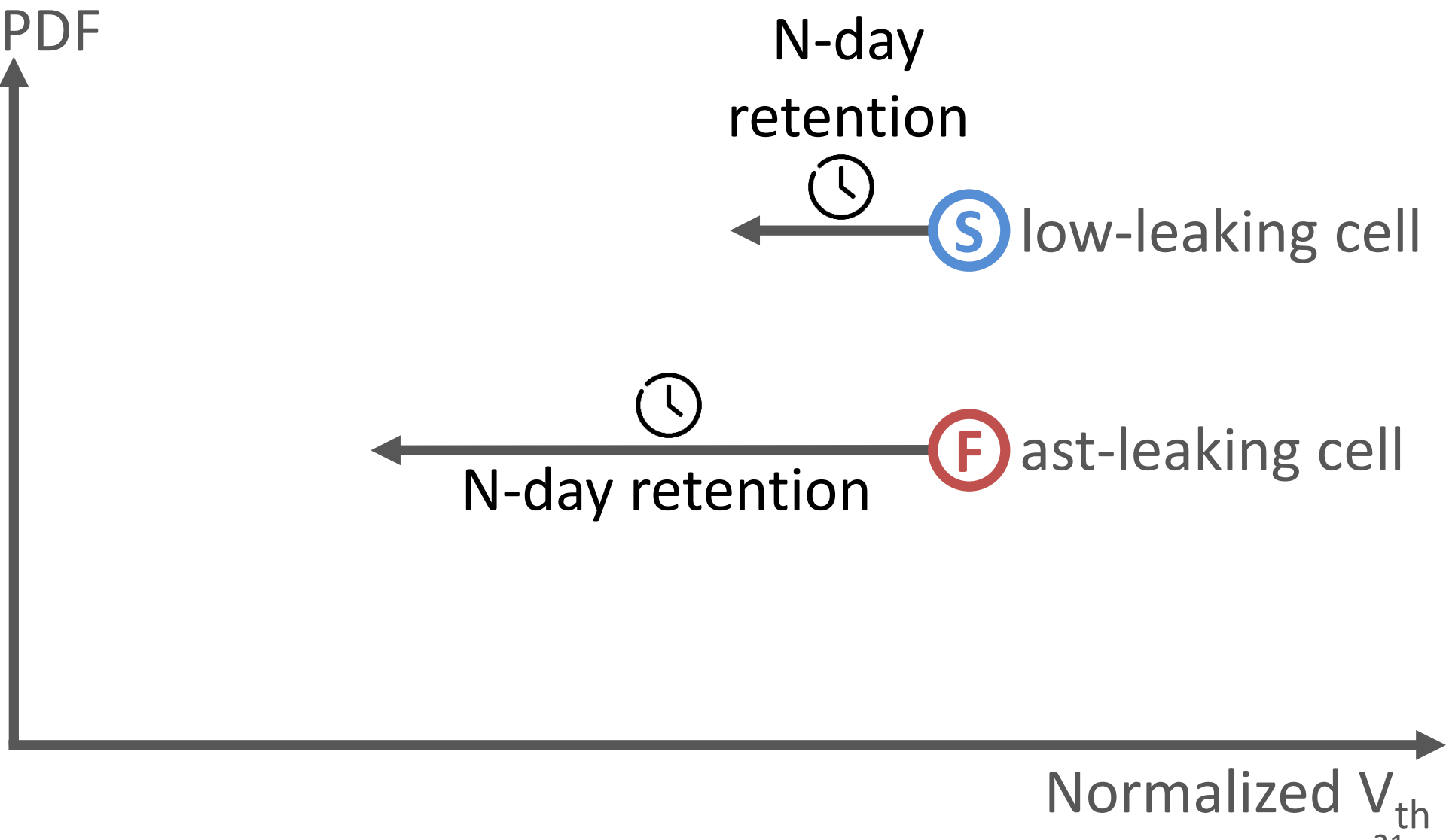
Recover old data after failure

Retention failure

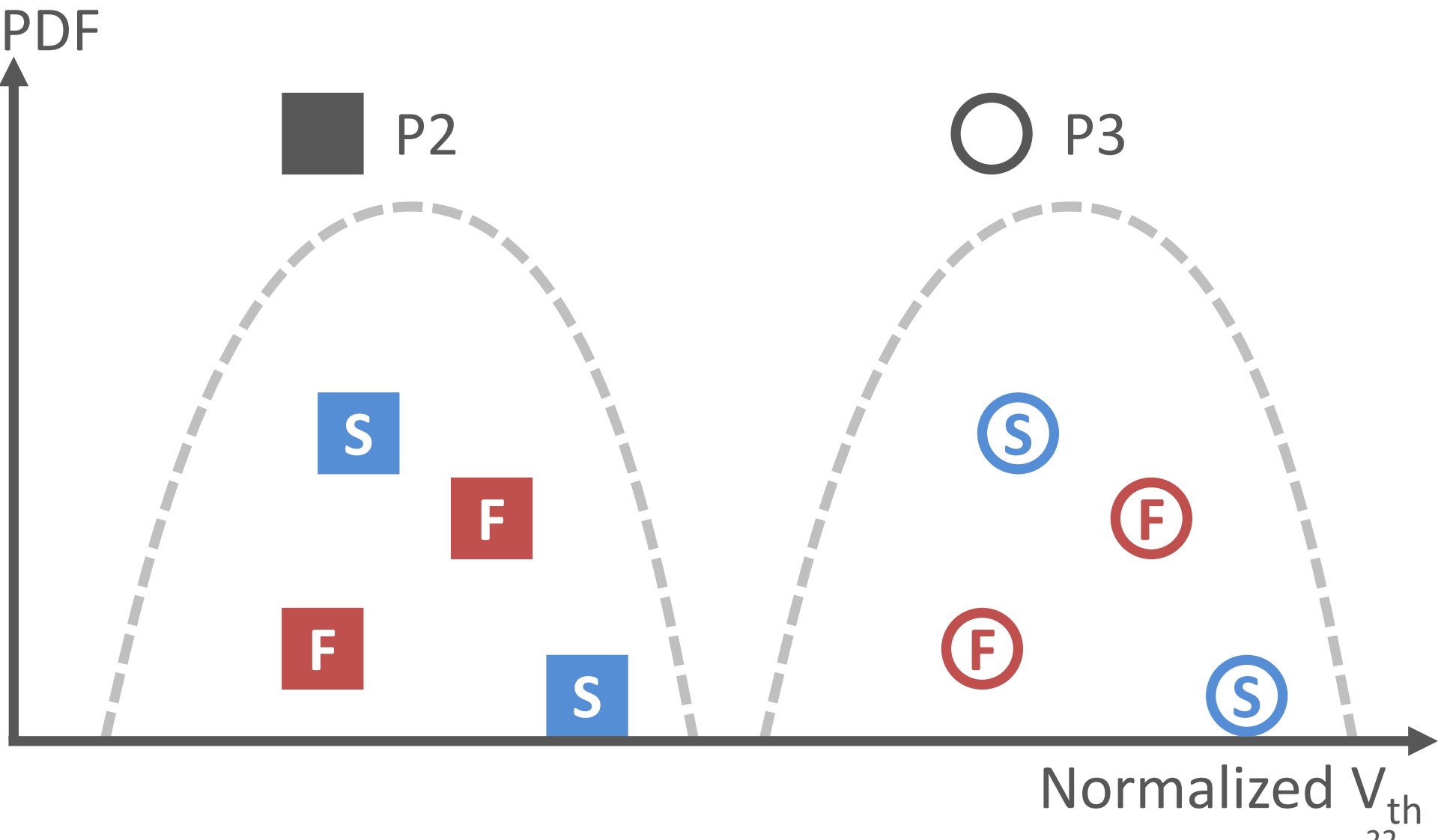
Very old data



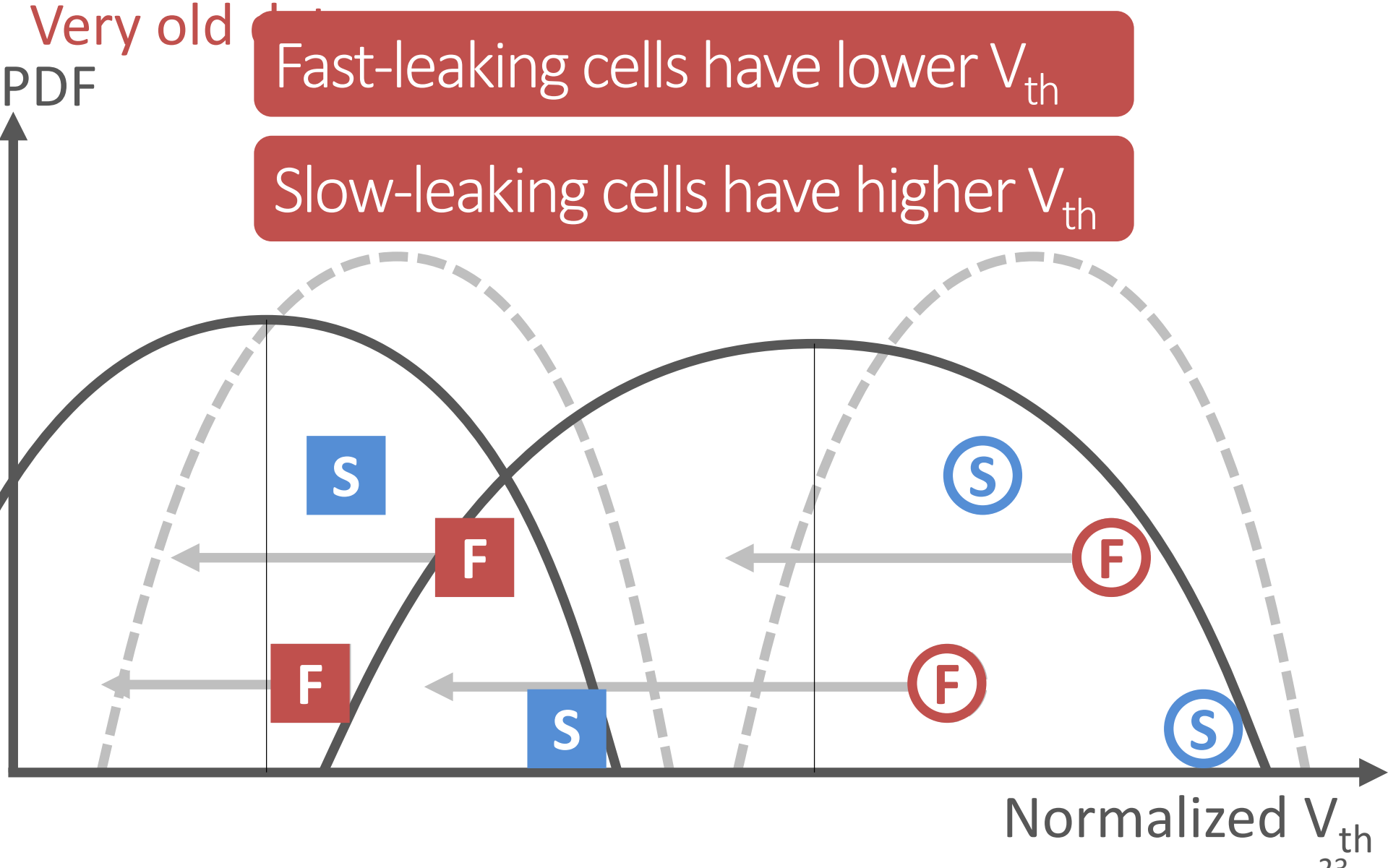
Leakage speed variation



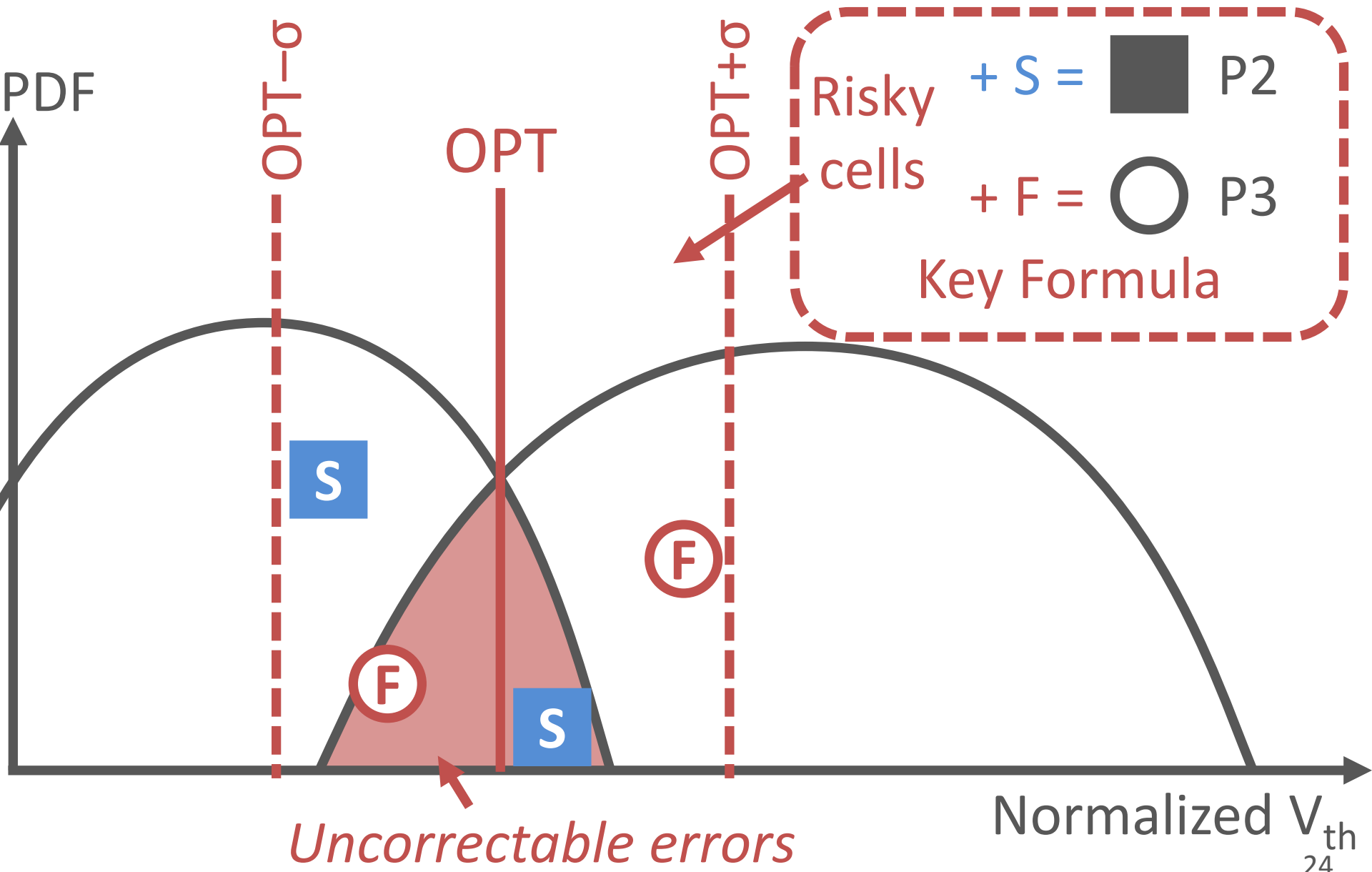
A simplified example



Reading very old data

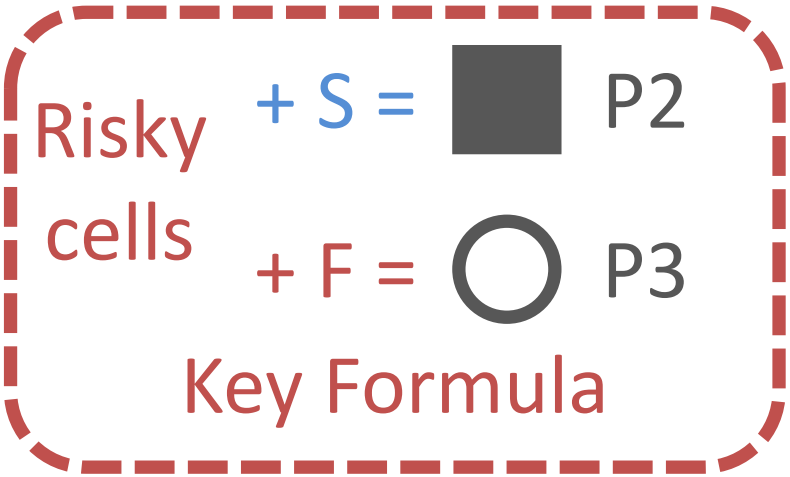


"Risky" cells


PDF
OPT−σ
OPT
OPT+σ
Risky cells
+ S = P2
+ F = P3
Key Formula
S
F
F
S
Uncorrectable errors
Normalized V_{th}

Retention Failure Recovery (RFR)

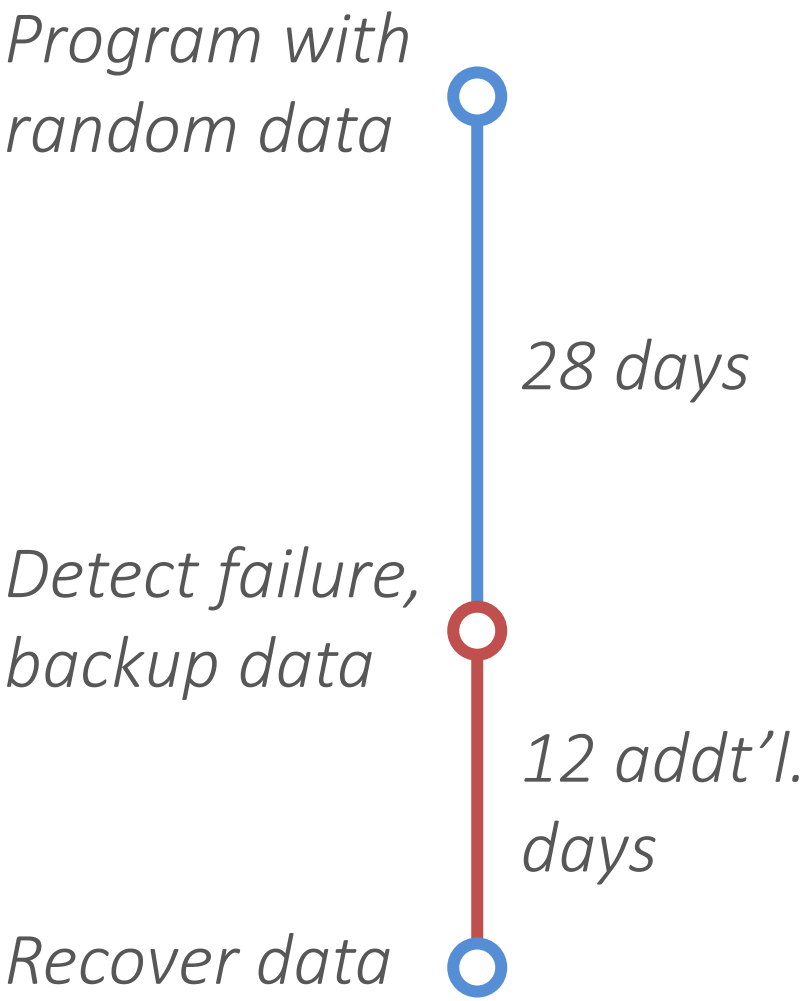
<u>Key idea:</u> Guess original state of the cell from its leakage speed property



Three steps

1. *Identify risky cells*
2. *Identify fast-/slow-leaking cells*
3. *Guess original states*

RFR Evaluation



- *Expect to eliminate 50% of raw bit errors*
- *ECC can correct remaining errors*

Characterize retention loss in real NAND chip

Optimize read performance for old data

Recover old data after failure

Conclusion

Retention loss → Longer read latency

Retention optimized reading (ROR)

→ *Learns OPT periodically*

→ *71% shorter read latency*

Retention failure recovery (RFR)

→ *Use leakage property to guess correct state*

→ *50% error reduction before ECC correction*

→ *Recover data after failure*

Our FMS Talks and Posters

- *Onur Mutlu, Error Analysis and Management for MLC NAND Flash Memory, FMS 2014.*
- *Onur Mutlu, Read Disturb Errors in MLC NAND Flash Memory, FMS 2015.*
- *Yixin Luo, Data Retention in MLC NAND Flash Memory, FMS 2015.*

- *FMS 2015 posters:*
  - WARM: Improving NAND Flash Memory Lifetime with Write-hotness Aware Retention Management
  - Read Disturb Errors in MLC NAND Flash Memory
  - Data Retention in MLC NAND Flash Memory

Our Flash Memory Works (I)

1. Retention noise study and management

1) Yu Cai, Gulay Yalcin, Onur Mutlu, Erich F. Haratsch, Adrian Cristal, Osman Unsal, and Ken Mai, Flash Correct-and-Refresh: Retention-Aware Error Management for Increased Flash Memory Lifetime, ICCD 2012.
2) Yu Cai, Yixin Luo, Erich F. Haratsch, Ken Mai, and Onur Mutlu, Data Retention in MLC NAND Flash Memory: Characterization, Optimization and Recovery, HPCA 2015.
3) Yixin Luo, Yu Cai, Saugata Ghose, Jongmoo Choi, and Onur Mutlu, WARM: Improving NAND Flash Memory Lifetime with Write-hotness Aware Retention Management, MSST 2015.

2. Flash-based SSD prototyping and testing platform

4) Yu Cai, Erich F. Haratsh, Mark McCartney, Ken Mai, FPGA-based solid-state drive prototyping platform, FCCM 2011.

Our Flash Memory Works (II)

3. Overall flash error analysis

5) Yu Cai, Erich F. Haratsch, Onur Mutlu, and Ken Mai, Error Patterns in MLC NAND Flash Memory: Measurement, Characterization, and Analysis, DATE 2012.

6) Yu Cai, Gulay Yalcin, Onur Mutlu, Erich F. Haratsch, Adrian Cristal, Osman Unsal, and Ken Mai, Error Analysis and Retention-Aware Error Management for NAND Flash Memory, ITJ 2013.

4. Program and erase noise study

7) Yu Cai, Erich F. Haratsch, Onur Mutlu, and Ken Mai, Threshold Voltage Distribution in MLC NAND Flash Memory: Characterization, Analysis and Modeling, DATE 2013.

Our Flash Memory Works (III)

5. Cell-to-cell interference characterization and tolerance

8) Yu Cai, Onur Mutlu, Erich F. Haratsch, and Ken Mai, Program Interference in MLC NAND Flash Memory: Characterization, Modeling, and Mitigation, ICCD 2013.

9) Yu Cai, Gulay Yalcin, Onur Mutlu, Erich F. Haratsch, Osman Unsal, Adrian Cristal, and Ken Mai, Neighbor-Cell Assisted Error Correction for MLC NAND Flash Memories, SIGMETRICS 2014.

6. Read disturb noise study

10) Yu Cai, Yixin Luo, Saugata Ghose, Erich F. Haratsch, Ken Mai, and Onur Mutlu, Read Disturb Errors in MLC NAND Flash Memory: Characterization and Mitigation, DSN 2015.

7. Flash errors in the field

11) Justin Meza, Qiang Wu, Sanjeev Kumar, and Onur Mutlu, A Large-Scale Study of Flash Memory Errors in the Field, SIGMETRICS 2015.

Referenced Papers and Talks

- *All are available at*
  http://users.ece.cmu.edu/~omutlu/projects.htm

Flash Memory SUMMIT

Thank you!

Feel free to email me with any questions & feedback

yixinluo@cmu.edu

http://www.cs.cmu.edu/~yixinluo

Flash Memory SUMMIT

Data Retention in MLC NAND Flash Memory: Characterization, Optimization, and Recovery

Yixin Luo
yixinluo@cmu.edu

(joint work with Yu Cai, Erich F. Haratsch, Ken Mai, Onur Mutlu)

SAFARI Carnegie Mellon SEAGATE DSSC

Flash Memory SUMMIT

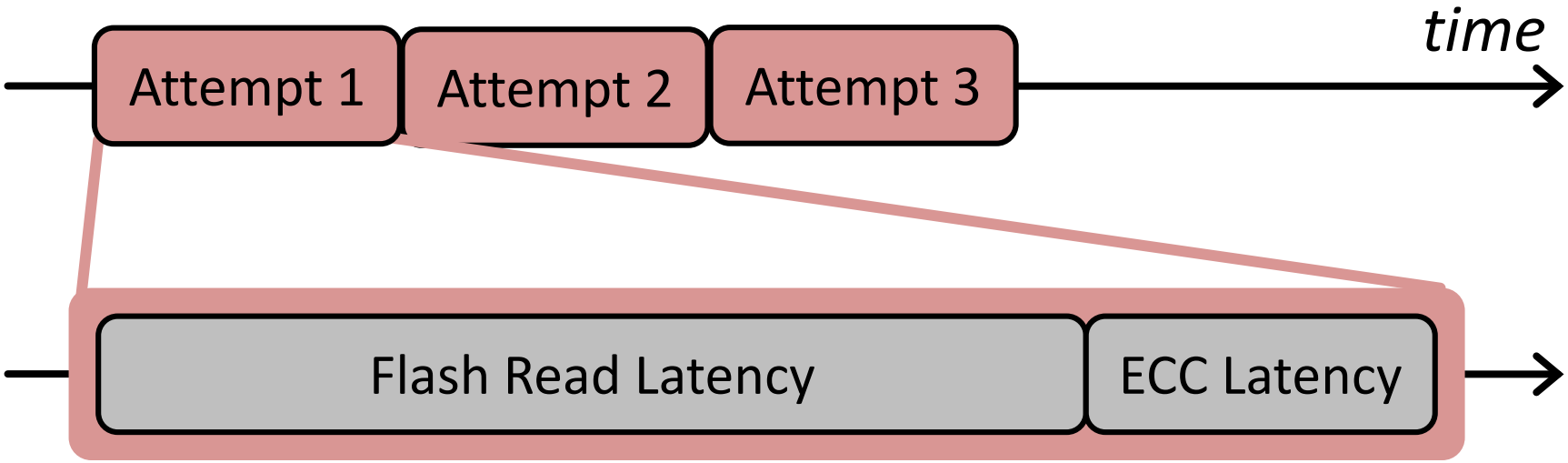
Backup Slides

ROR overheads

- *Power-on latency: 3, 15, and 23 seconds for flash memory with 1-day, 7-day, and 30-day equivalent retention age*
- *Per-day pre-optimization latency: 3 seconds*
- *Total storage overhead: 768 KB*

Read-Retry Latency Diagnosis

Read page A:



= Constant

∝ Raw bit error

ROR assumptions

- *We model a 512 GB flash-based SSD (composed of sixteen 256 Gbit flash memory chips) with an 8 KB page size, 256-page block size, and 100 µs read latency.*
- *We model a flash controller with an iterative BCH decoder that can correct 40 bit errors for every 1 KB of data [11] (i.e., it can tolerate an RBER of 10-3 during the flash lifetime).*

RFR Motivation

Data loss can happen in many ways

1. *High P/E cycle*
2. *High temperature → accelerates retention loss*
3. *High retention age (lost power for a long time)*

What if there are other errors?

Key: RFR does not have to correct all errors

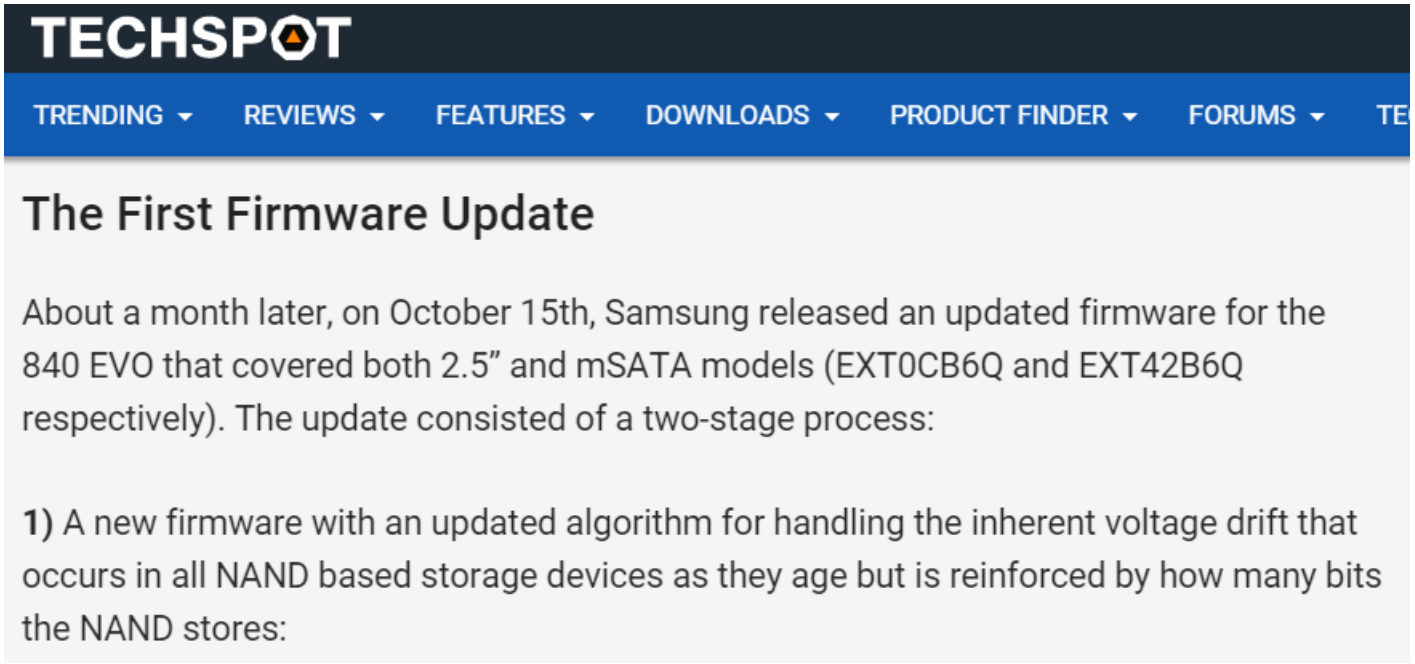
Example:

- *ECC can correct 40 errors in a page*
- *Corrupted page has 20 retention errors, 25 other errors (45 total errors)*
- *After RFR: 10 retention errors, 30 other errors (40 total errors → ECC correctable)*

Characterization methodology

- *FPGA-based flash memory testing platform*
- ***Real** 20- to 24-nm MLC NAND flash chips*
- *0- to 40-day worth of retention loss*
- *Room temperature (20^{o}C)*
- *0 to 50k P/E Cycles*

Firmware fix



TECHSPOT

TRENDING REVIEWS FEATURES DOWNLOADS PRODUCT FINDER FORUMS TE

The First Firmware Update

About a month later, on October 15th, Samsung released an updated firmware for the 840 EVO that covered both 2.5" and mSATA models (EXT0CB6Q and EXT42B6Q respectively). The update consisted of a two-stage process:

1) A new firmware with an updated algorithm for handling the inherent voltage drift that occurs in all NAND based storage devices as they age but is reinforced by how many bits the NAND stores:

Firmware fix

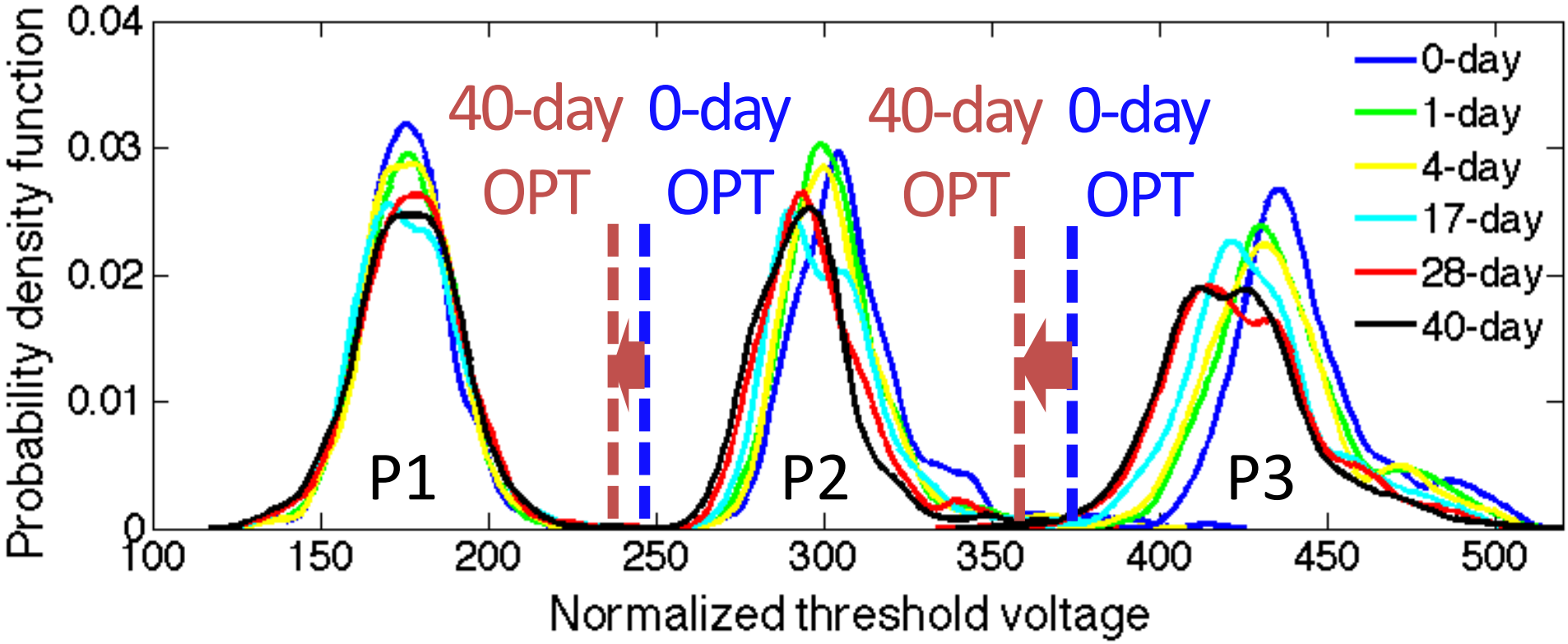
2) The second stage of Samsung's new firmware with the updated algorithm mandated that all data on the disk should be rewritten to restore performance on older data. Since it took around 8 weeks for the issue to become visible in the 840 EVO, this meant that we could not fully know if Samsung's firmware worked or not until some weeks later.

A Second Firmware Update: Reading Between The Lines

We couldn't know for sure if the firmware was a successful solution in the long term, and in fact the problem did come back. Samsung started to work on newer firmware (EXT0DB6Q), but this time with a different approach: instead of simply adjusting the algorithm for reading old data, the disk would also continuously rewrite old data in the background.

It's not an elegant fix, and it's also a fix that will degrade the lifetime of the NAND since the total numbers of writes it's meant to withstand is limited. But as we have witnessed in Tech Report's extensive durability test there is a ton of headroom in how NAND is rated, so in my opinion this is not a problem. Heck, the Samsung 840 even outlasted two MLC drives.

Optimal Read Reference Voltage (OPT)



Finding: OPT decreases over time

Retention Optimized Reading: Summary

Flash Read Techniques	Lifetime (P/E Cycle)	Performance (Read Latency)
Fixed V_{ref}	✗	✓
Sweeping V_{ref}	✓ 64% ↑	✗
ROR	✓ 64% ↑	✓ Nom. Life: 2.4% ↓ Ext. Life: 70.4% ↓

Observations

1. The optimal read reference voltage gradually decreases over time

Key idea: Record the old OPT as a prediction (V_{pred}) of the actual OPT

Benefit: Close to actual OPT → Fewer read retries

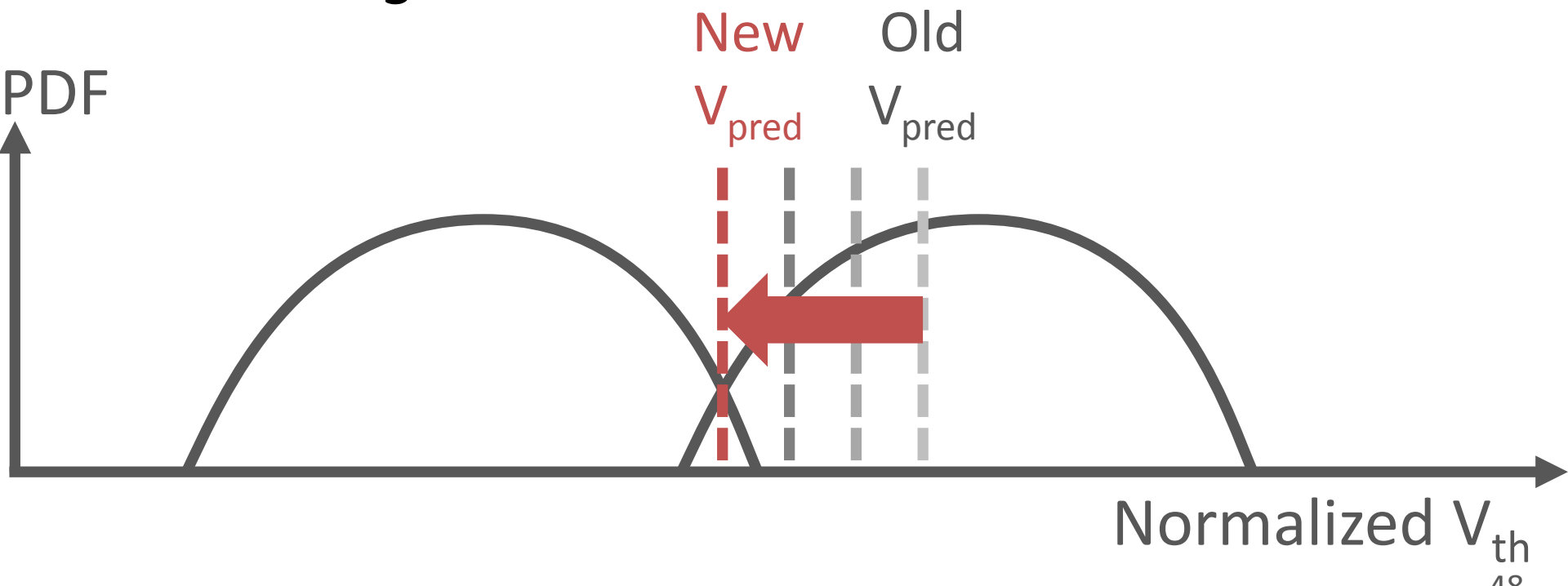
2. The amount of retention loss is similar across pages within a flash block

Key idea: Record only one V_{pred} for each block

Benefit: Small storage overhead (768KB out of 512GB)

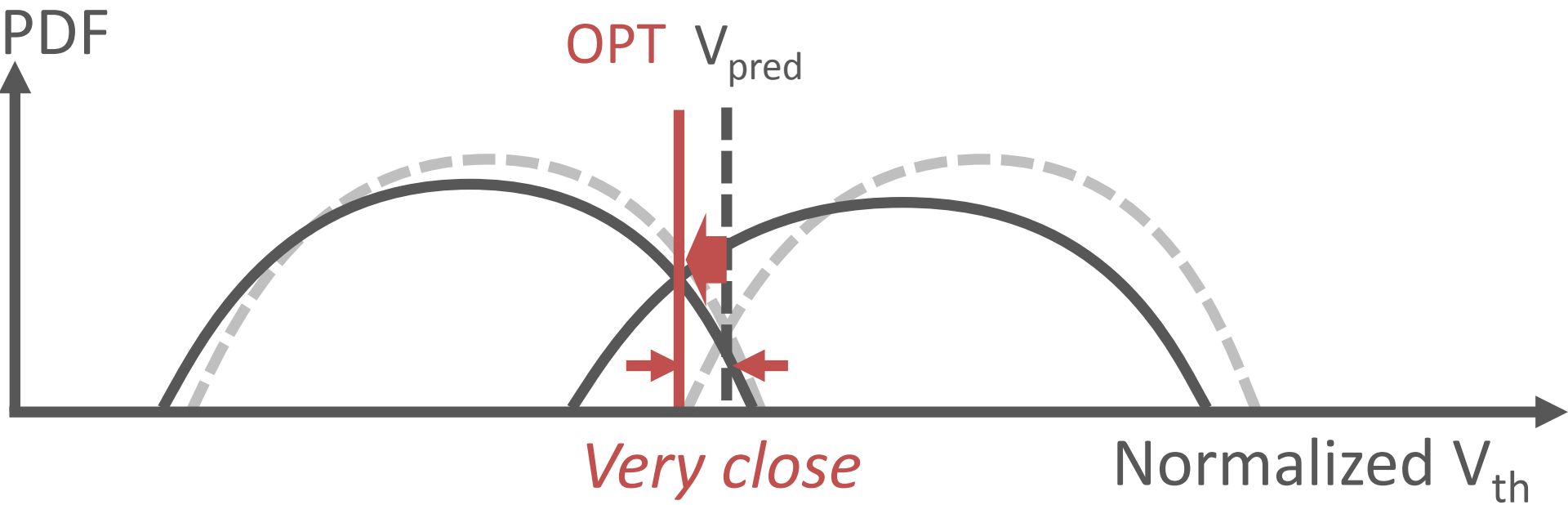
1. Online Pre-Optimization Algorithm

- *Triggered periodically (e.g., per day)*
- *Find and record an OPT as per-block V_{pred}*
- *Performed in background*
- *Small storage overhead*

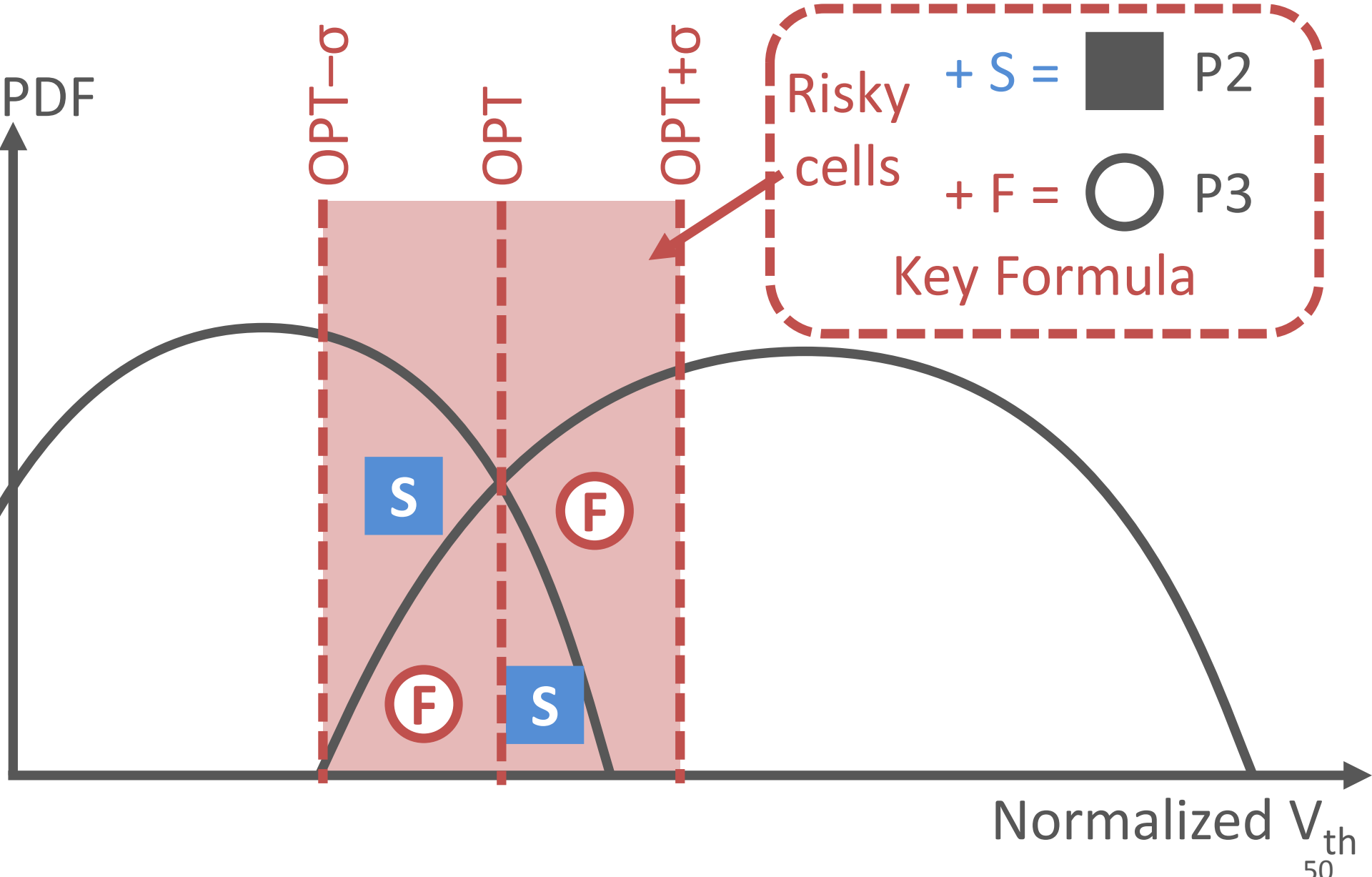


2. Improved Read-Retry Technique

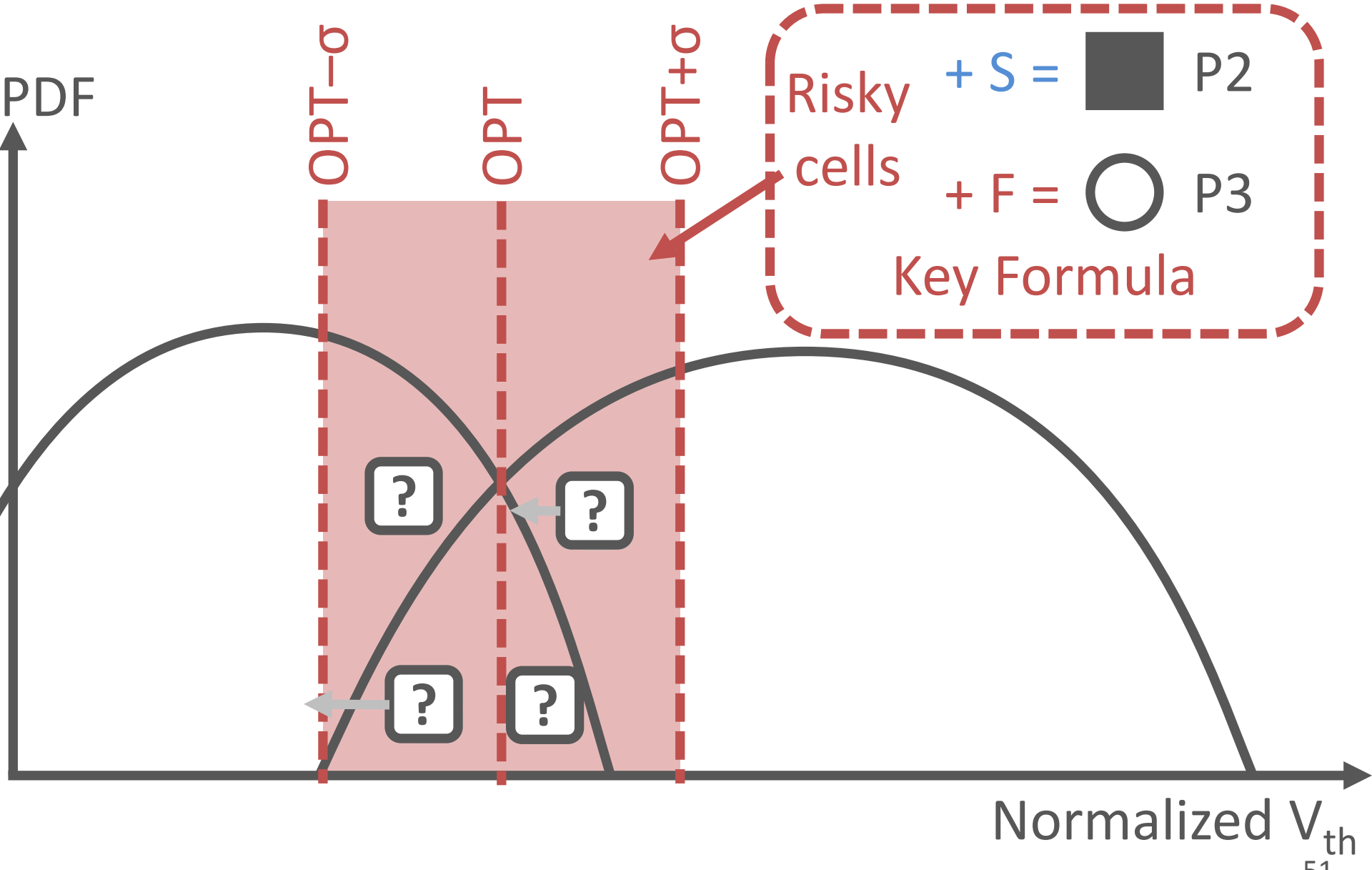
- *Performed as normal read*
- *V_{pred} already close to actual OPT*
- *Decrease V_{ref} if V_{pred} fails, and retry*



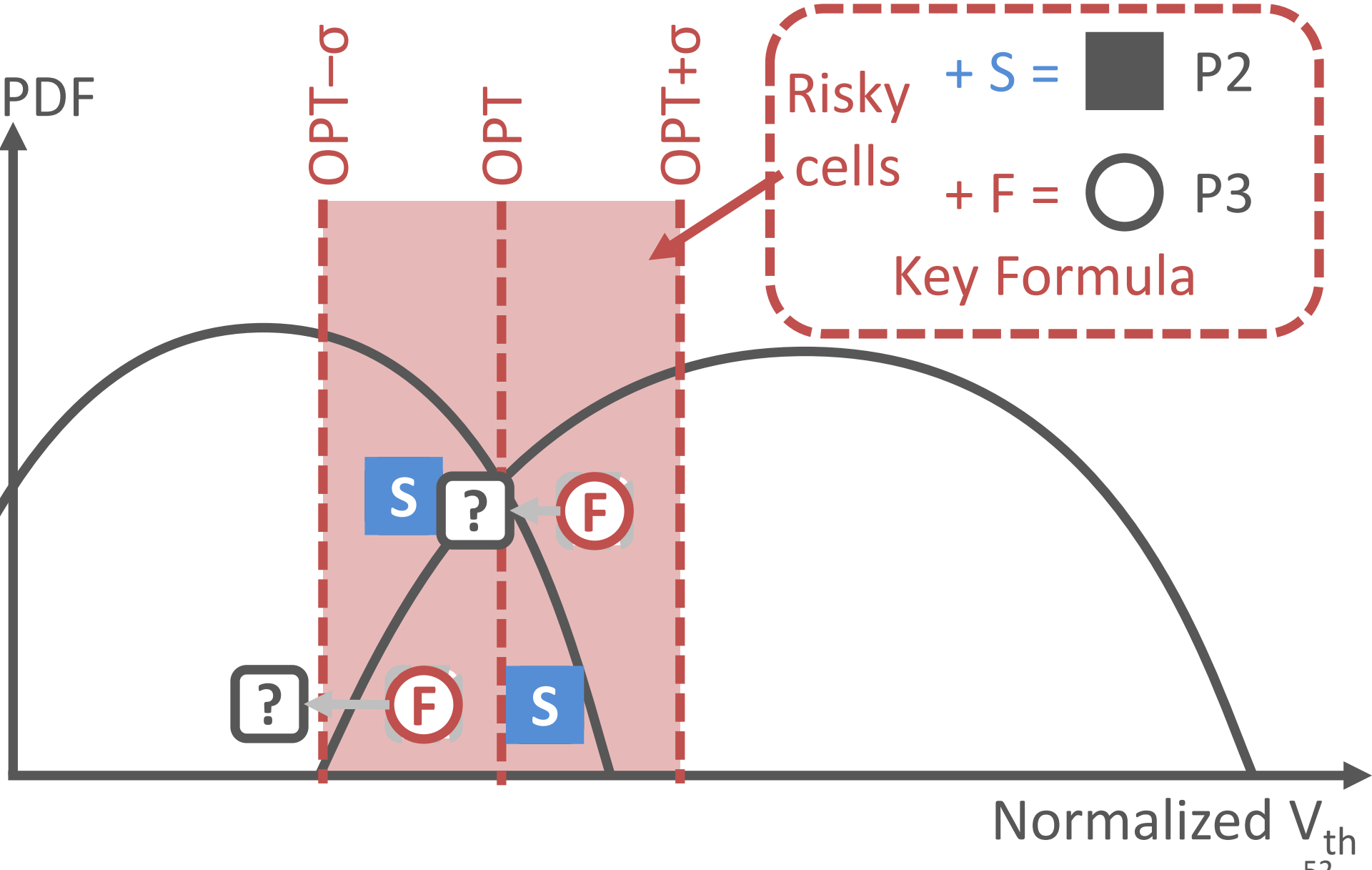
1. Identify Risky Cells



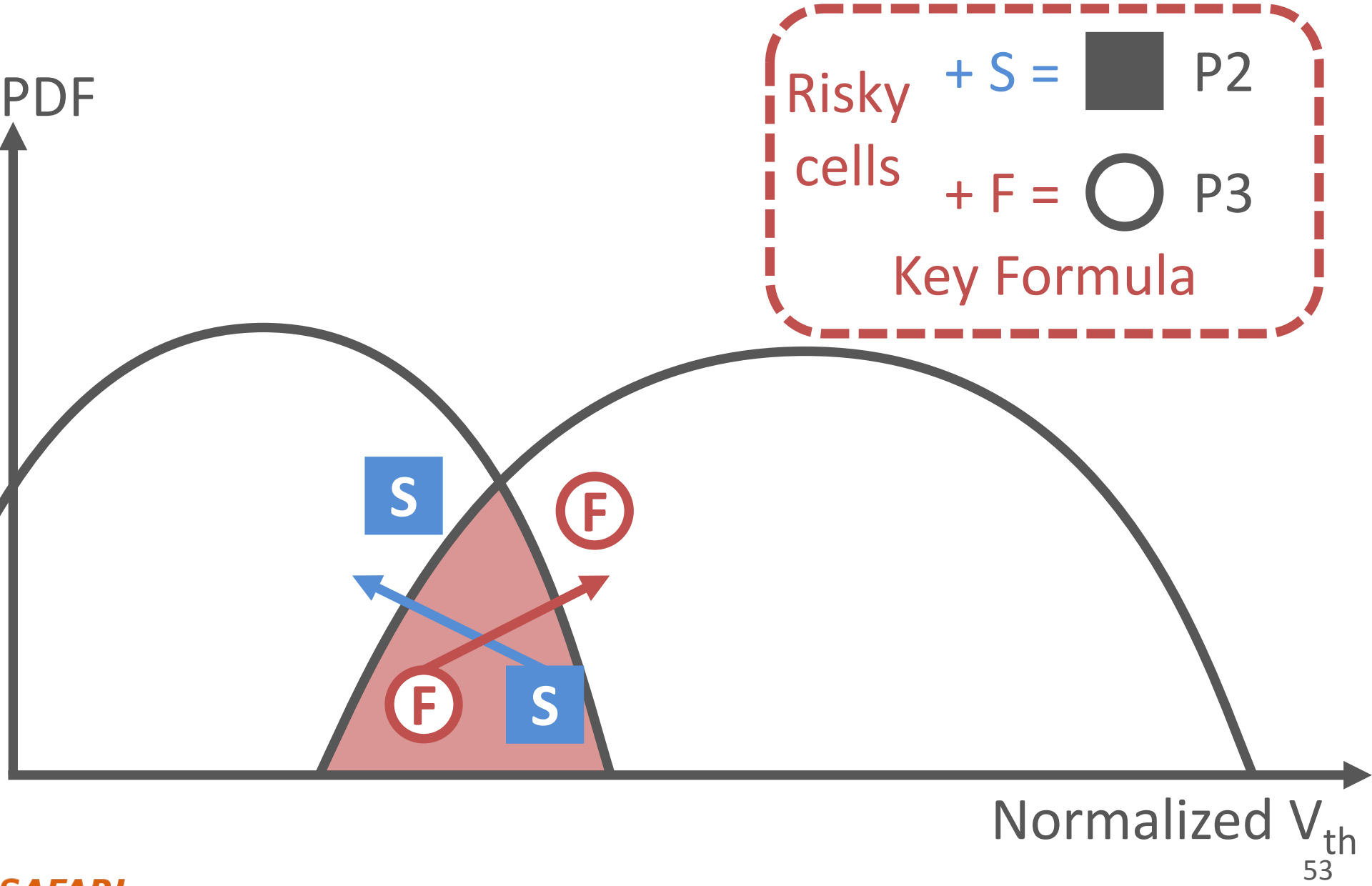
2. Identifying Fast- vs. Slow-Leaking Cells



2. Identifying Fast- vs. Slow-Leaking Cells

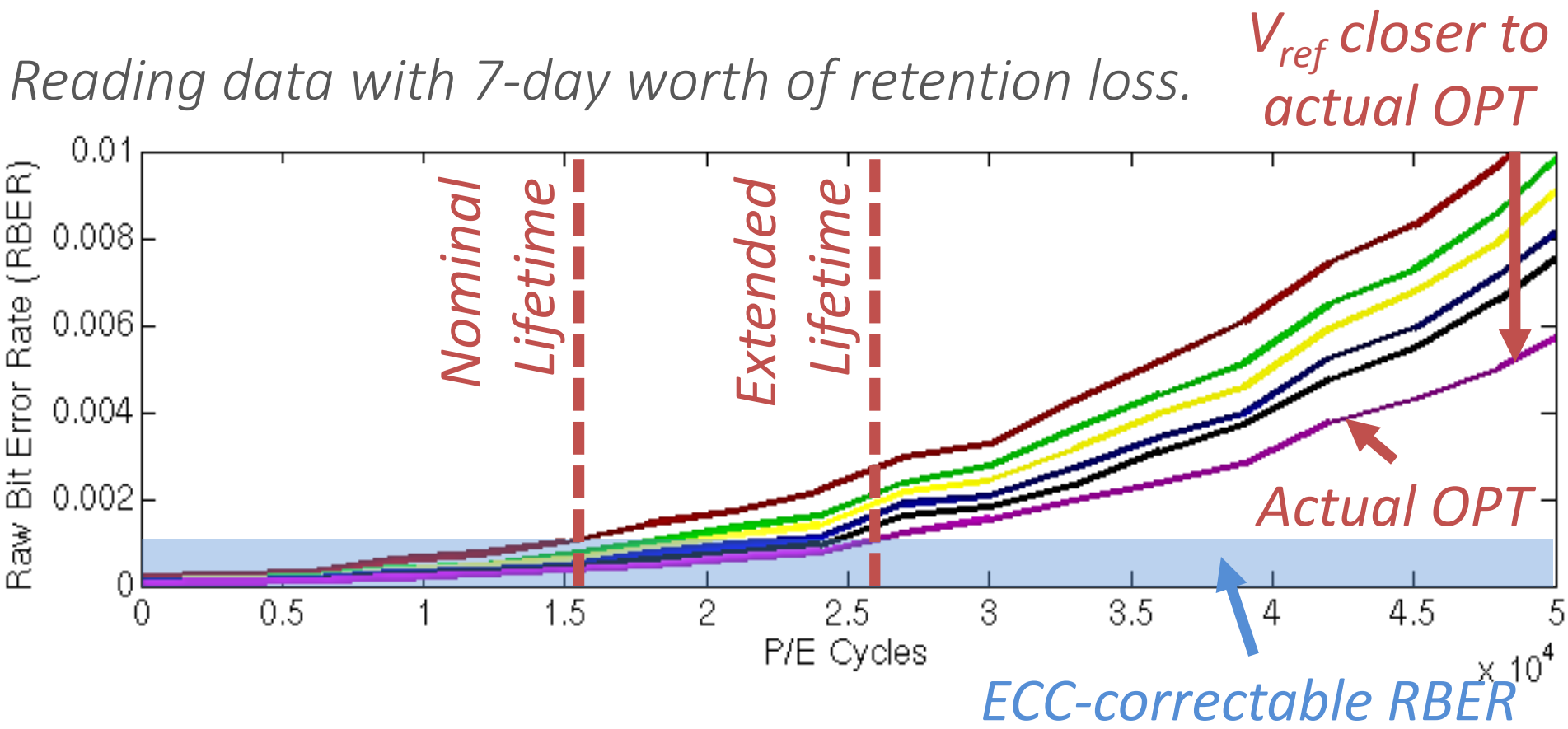


3. Guess Original States



3. RBER and P/E Cycle Lifetime

Reading data with 7-day worth of retention loss.



Finding: Using actual OPT achieves the longest lifetime

Characterization Summary

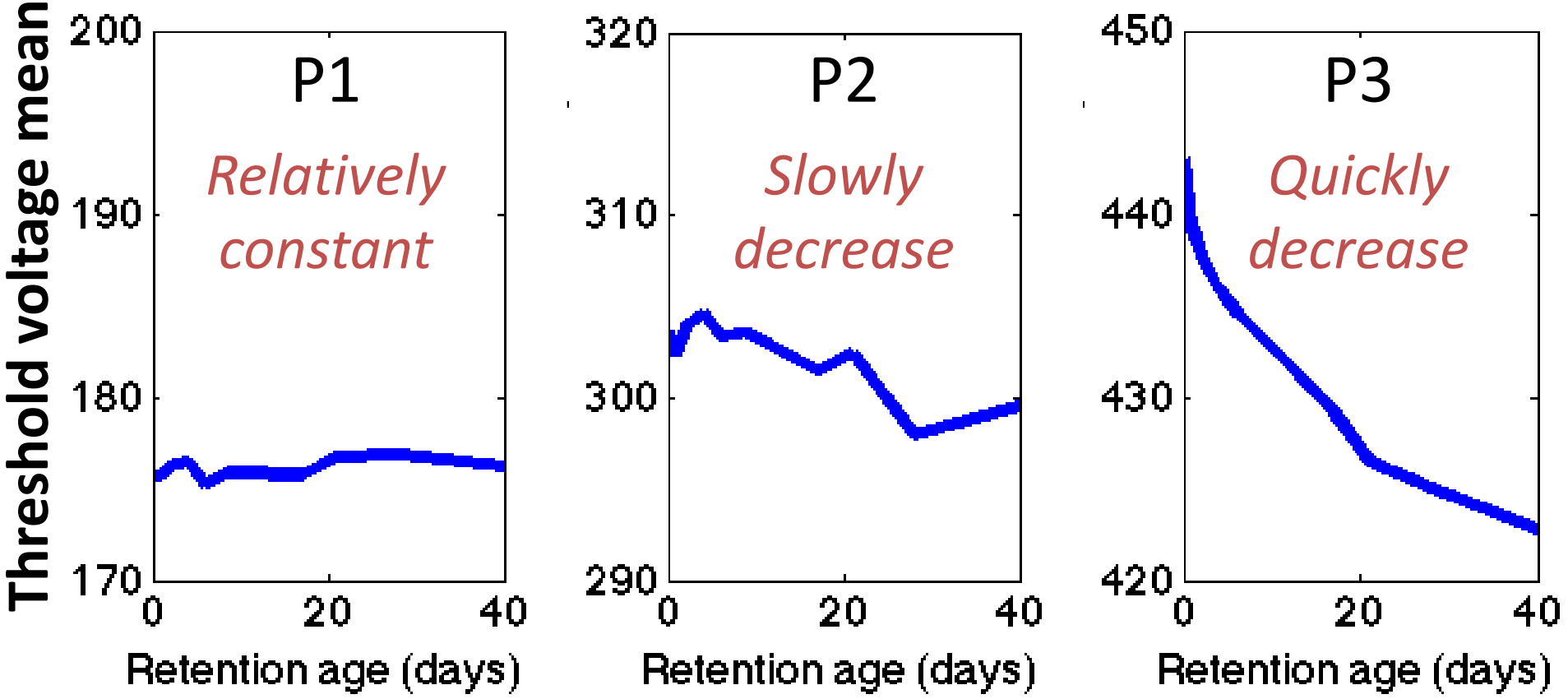
*Due to **retention loss***

- **Cell's threshold voltage** (V_{th}) decreases over time
- **Optimal read reference voltage** (OPT) decreases over time

*Using the **actual OPT** for reading*

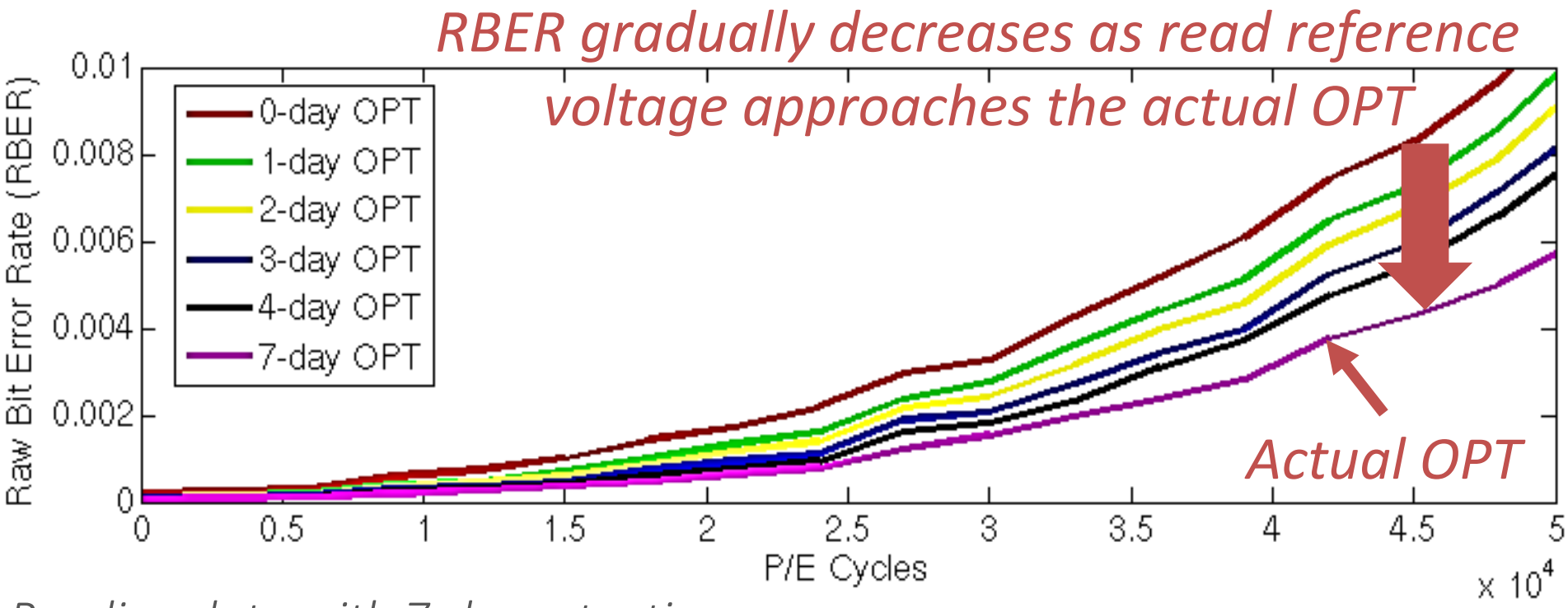
- Achieves the longest **lifetime**

Threshold Voltage (V_{th}) Mean



Finding: V_{th} shifts faster in higher voltage states

Raw Bit Error Rate (RBER)



Reading data with 7-day retention age.

Finding: The actual OPT achieves the lowest RBER

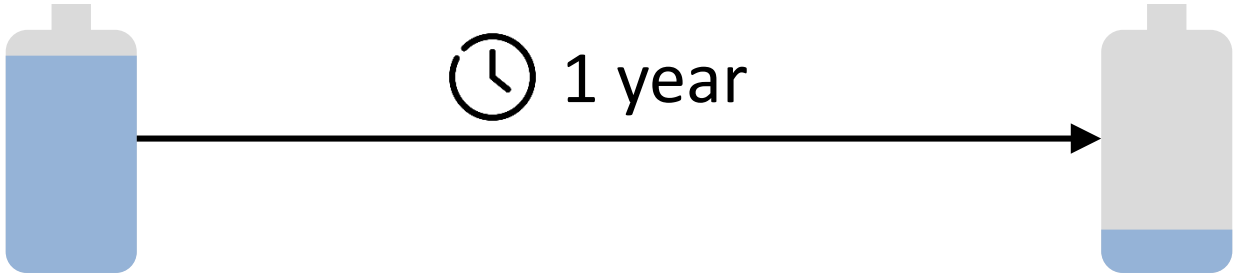
Online Pre-Optimization Algorithm

- ***Periodically*** *learn and record* ***OPT for page 255*** *as* ***per-block*** *starting read reference voltage (V_0)*
  - Page 255 has the shortest retention age
  - Other pages within the block have longer retention age and retention age will increase over time

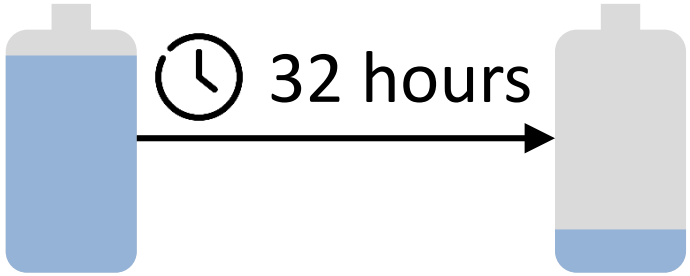
- *Step 1: Read with V_{ref} = old V_0, record RBER*
- *Step 2: Decrease V_{ref} = V_{ref} – ΔV^* compare RBER*
- *Step 3: Increase V_{ref} = V_{ref} + ΔV compare RBER*
- *Step 4: Record new V_0 = V_{ref} | minimal RBER*

$^*\Delta V$ is the smallest step size for changing read reference voltage.

Arrhenius Law



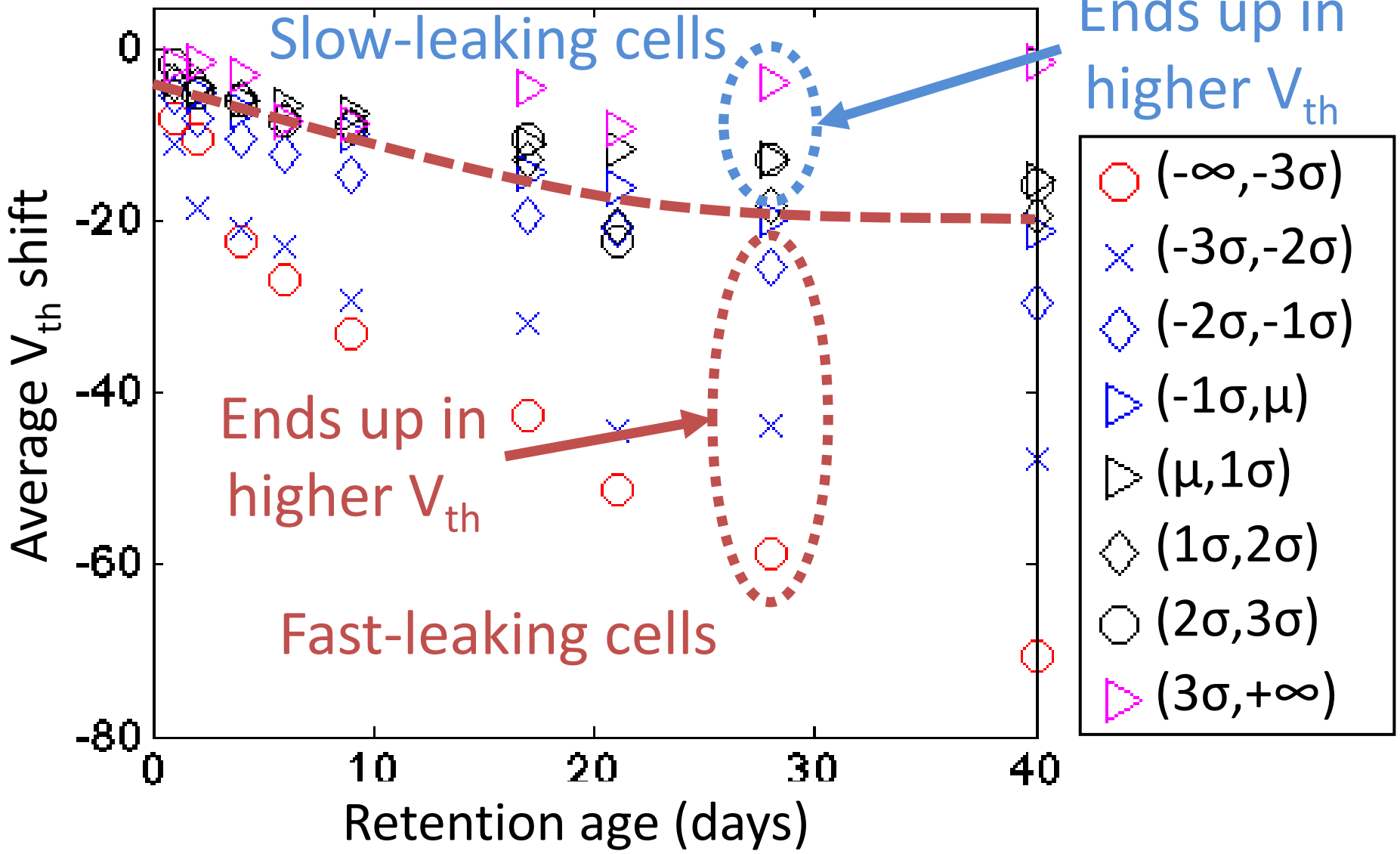
Room temperature (20°C)



High temperature (70°C)

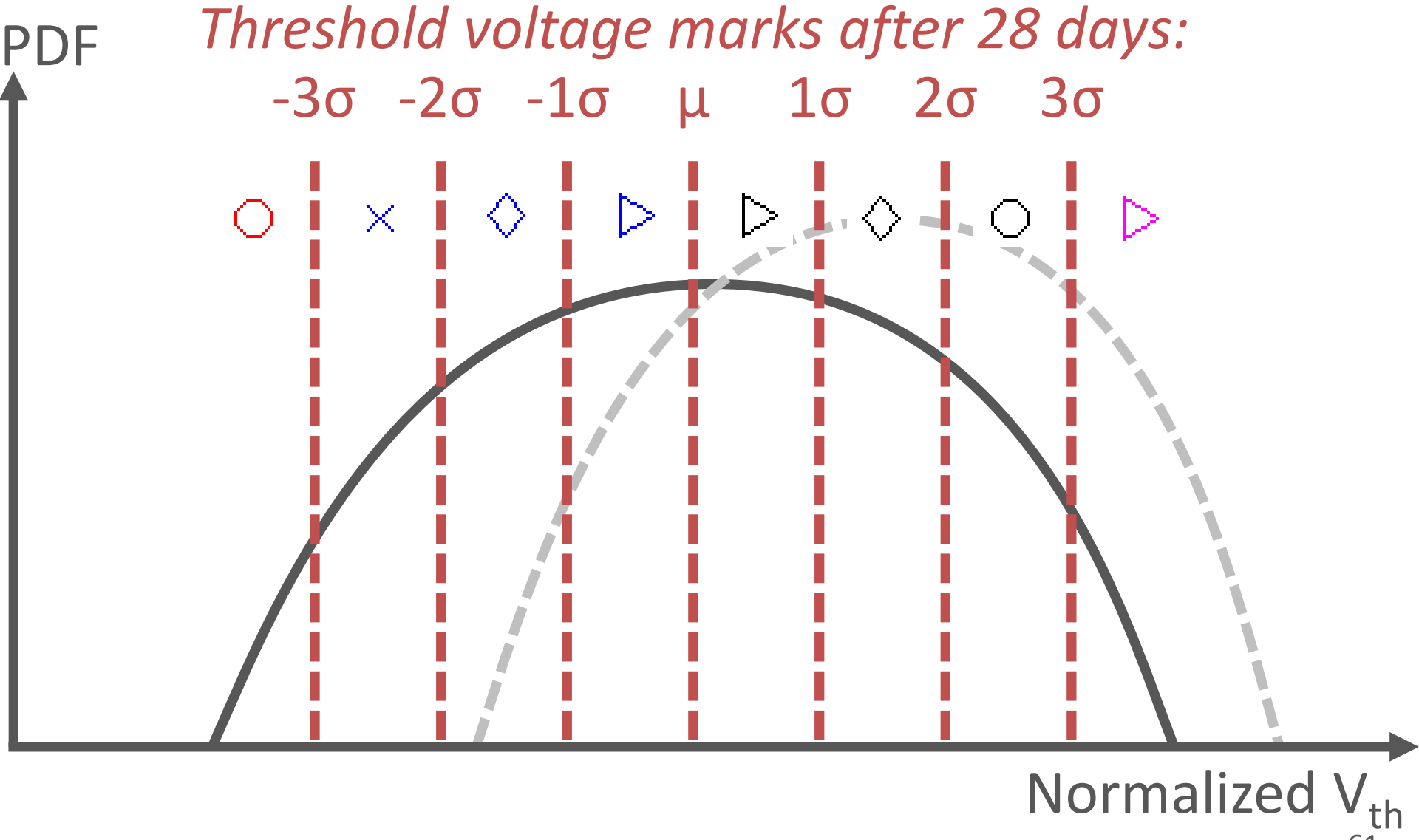
High temperature accelerates retention loss

Fast- and Slow-Leaking Cells

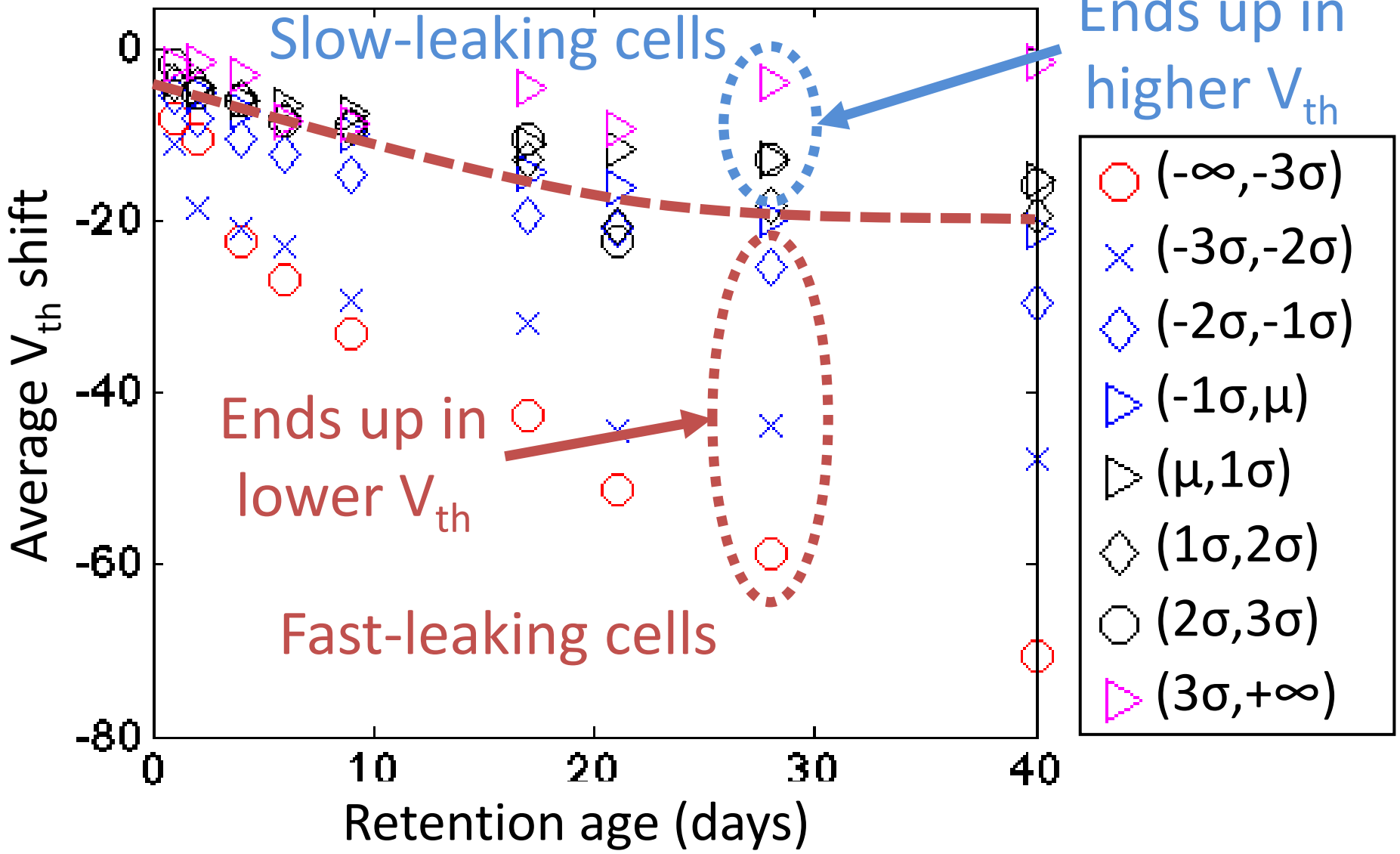


*Similar trends are found in P2 state, as shown in the paper.

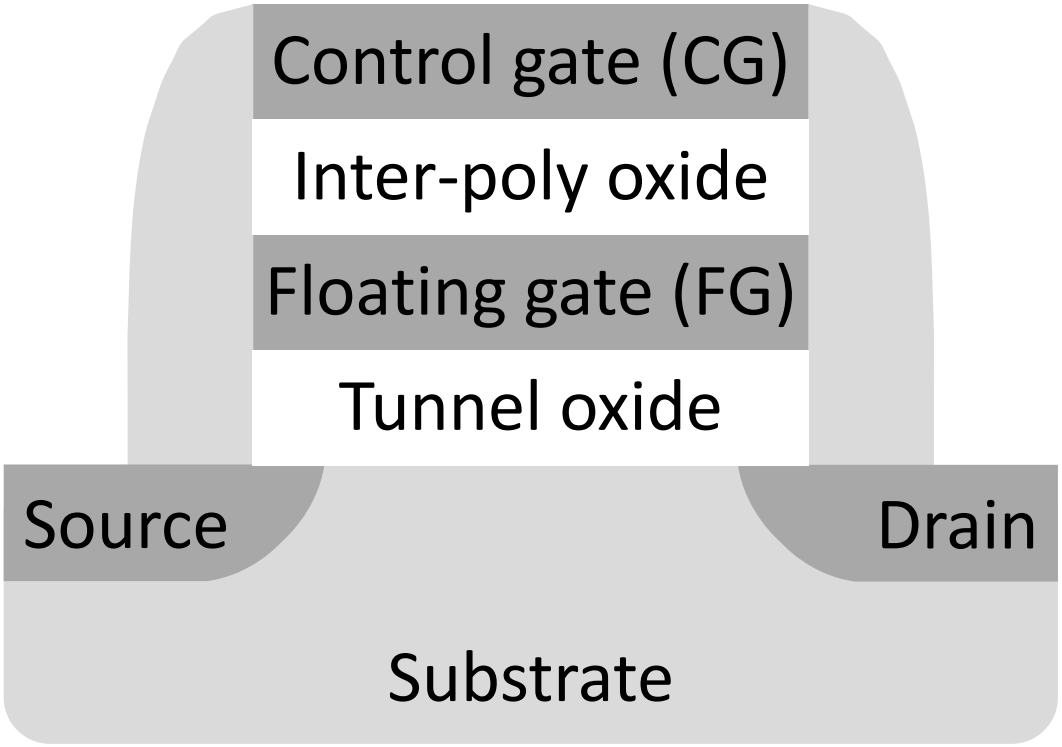
Fast- and Slow-Leaking Cells

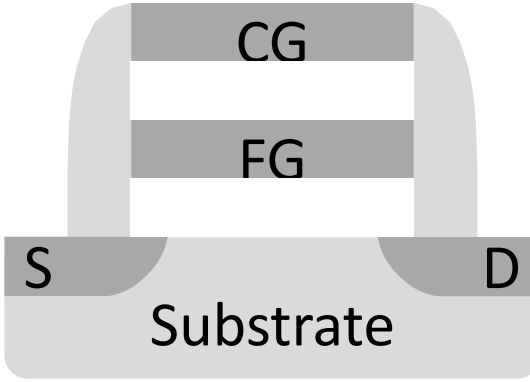


Fast- and Slow-Leaking Cells



*Similar trends are found in P2 state, as shown in the paper.


Control gate (CG)
Inter-poly oxide
Floating gate (FG)
Tunnel oxide
Source
Drain
Substrate


CG
FG
S
D
Substrate