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
Characterize retention loss in real NAND chip

Optimize read performance for old data

Recover old data after failure
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 consistently slower than normal as slow as 30MB/s whereas newly-written files ones used in benchmarks, perform as fast as new – around 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.

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

PDF

Erased (11)  P1 (10)  P2 (00)  P3 (01)

\( V_a \)  \( V_b \)  \( V_c \)

Normalized \( V_{th} \)
Experimental Testing Platform


Cai et al., FPGA-based Solid-State Drive prototyping platform, FCCM 2011.
Finding: Cell’s threshold voltage decreases over time
Threshold voltage reduces over time

Old data

Less charge

PDF

New data

More charge

Normalized $V_{th}$

P1 (10)

P2 (00)

P3 (01)
First read attempt fails

Old data

PDF

Less charge

More charge

Normalized $V_{th}$

$V_b$

$V_c$

$P_1$ (10)

$P_2$ (00)

$P_3$ (01)

Raw bit errors $>$ ECC correctable errors
Read-retry

Old data

PDF

Normalized $V_{th}$

Increase read latency

Fewer raw bit errors

$V_b'$, $V_b$

$V_c'$, $V_c$

P1 (10)

P2 (00)

P3 (01)
Why is old data slower?

Retention loss

→ Leak charge over time
  → Generate retention errors
  → Require read-retry
  → Longer read latency
Characterize retention loss in real NAND chip

Optimize read performance for old data

Recover old data after failure
The ideal read voltage

Optimal read reference voltage \( \rightarrow \) minimal read latency

Minimal raw bit errors

OLD DATA

PDF

\( OPT_b \) \( \rightarrow \) OPT_c

P1 (10)

P2 (00)

P3 (01)
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 vs. ROR**
  - **Read-retry count:** Baseline, 70%; ROR, 40% (30% decrease)
  - **BCH decoding latency:** Baseline, 100%; ROR, 90% (10% decrease)
  - **Total read latency:** Baseline, 100%; ROR, 29% (71% decrease)
Retention optimized reading

Retention loss $\rightarrow$ longer read latency

Optimal read reference voltage (OPT)

$\rightarrow$ Shortest read latency

$\rightarrow$ Decreases gradually over time (retention)

$\rightarrow$ Learn OPT periodically

$\rightarrow$ Minimize read-retry & RBER

$\rightarrow$ Shorter read latency
Characterize retention loss in real NAND chip

Optimize read performance for old data

Recover old data after failure
Retention failure

Very old data

PDF

P1
(10)

OPT\(_b\)

P2
(00)

OPT\(_c\)

P3
(01)

Uncorrectable errors

Normalized \(V_{th}\)
Leakage speed variation

PDF

N-day retention

Low-leaking cell

Fast-leaking cell

N-day retention

Normalized $V_{th}$
A simplified example

Normalized $V_{th}$

PDF

P2

P3

S

F

S

F

S
Reading very old data

Very old data

Fast-leaking cells have lower $V_{th}$

Slow-leaking cells have higher $V_{th}$

PDF

Normalized $V_{th}$

Fast-leaking cells (F) have lower $V_{th}$, slow-leaking cells (S) have higher $V_{th}$.
“Risky” cells

PDF

Normalized $V_{th}$

Risky cells

$+ S = \text{P2}$

$+ F = \text{P3}$

Key Formula

Uncorrectable errors
Retention Failure Recovery (RFR)

**Key idea:** 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

Program with random data

Detect failure, backup data

Recover data

28 days

12 add’l. days
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, *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**

2. **Flash-based SSD prototyping and testing platform**
3. **Overall flash error analysis**

4. **Program and erase noise study**
5. Cell-to-cell interference characterization and tolerance


6. Read disturb noise study


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
Thank you!

Feel free to email me with any questions & feedback

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

- Attempt 1
- Attempt 2
- Attempt 3

Flash Read Latency  \[ \propto \text{Raw bit error} \]

\[ = \text{Constant} \]
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 \(\rightarrow\) 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°C)
- 0 to 50k P/E Cycles
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:
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

<table>
<thead>
<tr>
<th>Flash Read Techniques</th>
<th>Lifetime (P/E Cycle)</th>
<th>Performance (Read Latency)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed $V_{\text{ref}}$</td>
<td>✕</td>
<td>✓</td>
</tr>
<tr>
<td>Sweeping $V_{\text{ref}}$</td>
<td>✓ 64% ↑</td>
<td>✕</td>
</tr>
<tr>
<td>ROR</td>
<td>✓ 64% ↑</td>
<td>✓</td>
</tr>
</tbody>
</table>

- Nom. Life: 2.4% ↓
- Ext. Life: 70.4% ↓
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 $\rightarrow$ 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

- Performing as normal read
- $V_{\text{pred}}$ already close to actual OPT
- Decrease $V_{\text{ref}}$ if $V_{\text{pred}}$ fails, and retry

![Diagram showing PDF, OPT, $V_{\text{pred}}$, and Very close normalized $V_{\text{th}}$]
1. Identify Risky Cells

- **Key Formula**
  
  
  
  Risky cells  
  
  + S = P2  
  
  + F = P3  
  
  Normalized $V_{th}$
2. Identifying Fast- vs. Slow-Leaking Cells

- **Risky cells**
  - $+ S = \square$ P2
  - $+ F = \bigcirc$ P3

**Key Formula**

- Normalized $V_{th}$

- PDF

- OPT $- \sigma$
- OPT
- OPT $+ \sigma$
2. Identifying Fast- vs. Slow-Leaking Cells

Risky cells

Key Formula

PDF

Normalized $V_{th}$

$OPT - \sigma$

$OPT$

$OPT + \sigma$

$+ S = \Box$ P2

$+ F = \bigcirc$ P3
3. Guess Original States

Key Formula

PDF

Risky cells + S = P2

+ F = P3

Normalized $V_{th}$
3. RBER and P/E Cycle Lifetime

Finding: Using actual OPT achieves the longest lifetime

Reading data with 7-day worth of retention loss.

V_ref closer to actual OPT

ECC-correctable RBER

Finding: Using actual OPT achieves the longest lifetime
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

- **P1**: Relatively constant
- **P2**: Slowly decrease
- **P3**: Quickly decrease
Raw Bit Error Rate (RBER)

RBER gradually decreases as read reference voltage approaches the actual OPT.

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 \( V_{\text{ref}} = \text{old } V_0 \), record RBER

• **Step 2**: Decrease \( V_{\text{ref}} = V_{\text{ref}} - \Delta V^* \) compare RBER

• **Step 3**: Increase \( V_{\text{ref}} = V_{\text{ref}} + \Delta V \) compare RBER

• **Step 4**: Record new \( V_0 = V_{\text{ref}} \mid \text{minimal RBER} \)

\( \Delta V \) 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

Retention age (days)

Average $V_{th}$ shift

Slow-leaking cells

Fast-leaking cells

Ends up in higher $V_{th}$

Similar trends are found in P2 state, as shown in the paper.
Fast- and Slow-Leaking Cells

Threshold voltage marks after 28 days:
-3σ, -2σ, -1σ, μ, 1σ, 2σ, 3σ
Fast- and Slow-Leaking Cells

- Slow-leaking cells
  - Ends up in higher $V_{th}$

- Fast-leaking cells
  - Ends up in lower $V_{th}$

*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
SAFARI