Using Machine Learning to Enhance Flash Endurance & Latency

Cloud Zeng
LiteOn/Storage/NVM Lab
Challenges in Error Recovery

- **During SSD Service Time:**
  - P/E Cycle, Data Retention, Read Disturb
  - Critical R/W Condition (Temperature)

- **Decoding Strategy:**
  - Read Retry
  - Soft Decoding
  - ...

- **Challenges:**
  - Keep High Reliability & Low Latency under Variant Operation Condition
**Status Prediction**

- **Input Parameters:**
  - Some factors will affect NAND Flash Status. *(P/E Cycle, Retention Time, Read Count, Temperature…)*
  - Some information from NAND Flash are also collected as Input Parameters *(Program/Erase Time, …)*

- **Status Prediction:**
  - Our target is to predict NAND Flash Status *(Ex: Optimal Read Level, Error Recovery Flow …..)* by Input Parameters.

- **Input:**
  - P/E Cycle
  - Read Count
  - Retention Time
  - …

- **Output:**
  - Optimal Read Level
  - Error Recovery Flow
  - …
Optimal Decoding Parameters

Example: Binary LDPC - MSA

1. Read Level for **Hard Bit**
   - Minimize Error Bits

2. Read Level for **Soft Bit & LLR Value**
   - Maximize the Decoding Capability

- **Optimal Read Level & LLR Prediction**
  - Maximize Decoding Capability. Extend the Endurance
  - Vary with **Operation Condition** (P/E Cycle, Retention Time, Read Count, Temperature...)

Flash Memory SUMMIT
Error Recovery Flow - Prediction

1. Default Read Level with Hard Decoding
2. Optimal HD Read Level with Hard Decoding
3. Optimal HD/SD Read Level with Soft Decoding

Ref: NAND Flash Status Prediction, FMS 2016
Visual Illustration - Error Recovery Flow

Probability Density (for NAND Error Bits)
Frame Error Rate (for Decoder)

Begin of Life

End of Life

Hard Decoding Capability

P/E Cycle
Data Retention
Read Disturb
Critical R/W Condition (Temperature)....

Decoding Fail

Error Bits Count/Chunk Size

Read Level

Vth
Visual Illustration – Read Retry

- Hard Decoding Capability
- Read Retry
- Default Read
- Decoding Fail
- Optimal Read Level
- How to quickly Find/Predict Optimal Read Level ?!
- End of Life
- Default Read Level
- Error Bits Count/Chunk Size
Visual Illustration – Soft Decode

- Hard Decoding Capability
- Optimal Read Level
- Soft Decoding Capability
- Decoding Fail
- End of Life
- Soft Read Level

How to quickly Find/Predict Optimal Soft Read Level & LLR ?!
1. **Default Read Level with Hard Decoding**

2. **Optimal Read Level with Hard Decoding**

3. **Optimal Read Level with Soft Decoding**

**Hard Decoding Capability**

**Soft Decoding Capability**

Recover the Data as soon as possible

- **Hard Read Level**
- **Decode Flow** \{ 1, 2, 3 ... n \}
- **Soft Read Level**
- **Priority Arrangement** (Fixed vs Dynamic)

**Error Bits Count/Chunk Size**
Parameter Optimization with ML

- A Smart Error Recovery Scheme is developed by Machine Learning
- This Scheme can be applied to variant operation condition (combination of \{PE, DR, RD, Temperature\})
- This Scheme can extend the endurance and reduce the latency

<table>
<thead>
<tr>
<th>Category</th>
<th>Item</th>
<th>Description</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>P/E</td>
<td>Cycle</td>
<td>0, 1000, ...~</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td>(Random)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dwell</td>
<td>(Random)</td>
<td></td>
</tr>
<tr>
<td>Test Item</td>
<td>Data Retention</td>
<td>0, 1, ...~ (Days)</td>
<td>Room Temperature</td>
</tr>
<tr>
<td></td>
<td>Data Retention</td>
<td>0, 1, ...~ (Days)</td>
<td>High Temperature</td>
</tr>
<tr>
<td></td>
<td>Read Disturb</td>
<td>0, 1000, ...~</td>
<td></td>
</tr>
</tbody>
</table>
Our Error Recovery Scheme use ML to find Optimal Read Level for variant operation conditions (combination of \{PE, DR, RD, Temperature\})

- 5x Extension for Baking Time & 2x Extension for P/E Count
Endurance with Hard/Soft Decoding

Decoding Coverage/Endurance Comparison

- Proposed Error Recovery Scheme with only Hard Decode is still better than Traditional Read Retry + Soft Decode in Decoding Coverage

Optimal Read Level with Soft Decoding

Hard Read Level
Soft Read Level
LLR Value
Throughput/IOPS Comparison

- **Proposed Error Recovery Scheme** always has less read latency compared with **Traditional Error Recovery Scheme**.

- **9% vs 47%**

- **9% vs 57%**

- **Decode Flow { 2, 3 ... n }**
  - **Priority Arrangement**

![Graph showing Throughput/IOPS Comparison and Decoding Coverage/Endurance Comparison](image-url)
What if we can predict Future Status of NAND Flash/SSD?!
Read Performance Drop can be further reduced with **Future Status Prediction**

- 9% vs 57% vs 81%
THANK YOU!

Any questions?

Come by LITEON Booth# 621 for Live Demo!
Learn about Machine Learning & TSV Technology
Get a chance to win a special prize