Novel Error Recovery Architecture
Based on Machine Learning

Cloud Zeng
LITEON/Storage/NVM Lab
Error Recovery Flow

1. **Default Read Level with Hard Decoding**
2. **Retry/Optimal Read Level with Hard Decoding**
3. **Retry/Optimal Read Level with Soft Decoding**

**Hard Decoding Capability**

- **Probability Density (Error Bits)**
- **FER (Frame Error Rate)**

**Soft Decoding Capability**

1. **Recover the Data - Coverage**
2. **As soon as possible - Latency**

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

**Error Bits Count/Chunk Size**
Error Recovery Scheme with ML

- An Error Recovery Scheme is developed by Machine Learning
- This Scheme can be applied to variant operation condition (combination of {PE, DR, RD, Temperature, Cross-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/High Temperature</td>
</tr>
<tr>
<td></td>
<td>Read Disturb</td>
<td>0, 1000, … ~</td>
<td>High Temperature</td>
</tr>
<tr>
<td></td>
<td>Cross-Temperature</td>
<td>HT/LT Write – LT/HT Read</td>
<td></td>
</tr>
</tbody>
</table>
Endurance with Hard Decoding

- Our Error Recovery Scheme use ML to find Optimal Parameters for variant operation conditions (combination of \{PE, DR, RD, Temperature\})
- 5x Extension for Baking Time & 2x Extension for P/E Count
# Prediction Model - Optimal Read Level

## Example: Data Collection

<table>
<thead>
<tr>
<th>Data 1</th>
<th>Input Para 1</th>
<th>Input Para 2</th>
<th>Input Para 3</th>
<th>Input Para 4</th>
<th>Input Para 5</th>
<th>Input Para 6</th>
<th>Optimal HD Read Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1100</td>
<td>589</td>
<td>1794</td>
<td>6322</td>
<td>1000</td>
<td>1000</td>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data 2</th>
<th>Input Para 1</th>
<th>Input Para 2</th>
<th>Input Para 3</th>
<th>Input Para 4</th>
<th>Input Para 5</th>
<th>Input Para 6</th>
<th>Optimal HD Read Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>932</td>
<td>908</td>
<td>1503</td>
<td>7849</td>
<td>500</td>
<td>500</td>
<td>-5</td>
</tr>
</tbody>
</table>

| ...    | ...          | ...          | ...          | ...          | ...          | ...          | ...                   |

<table>
<thead>
<tr>
<th>Data N</th>
<th>Input Para 1</th>
<th>Input Para 2</th>
<th>Input Para 3</th>
<th>Input Para 4</th>
<th>Input Para 5</th>
<th>Input Para 6</th>
<th>Optimal HD Read Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>990</td>
<td>842</td>
<td>1894</td>
<td>5692</td>
<td>300</td>
<td>400</td>
<td>3</td>
</tr>
</tbody>
</table>

## Feature Selection

## What’s the Optimal HD Read Level after n Days/Weeks?

## Input Parameters:
- P/E Cycle, Retention Time, Read Count, Temperature, Dwell …
- Program/Erase Time, Histogram …. 

## Regression Problem:
- Ordinary Least Square (OLS) Regression
- Ridge Regression (Hoerl and Kennard, 1970)
- Other Regression Analysis can be used to solve this problem
• Proposed Error Recovery Scheme always has less read latency compared with Traditional Error Recovery Scheme
Optimized Read Retry Sequence

- **Optimal Read Level**
- **Retry 1**
- **Retry 2**
- **Default Read**
- **Retry/Optimal Read Level with Hard Decoding**
- **Hard Decoding Capability**

Error Bits Count/Chunk Size
Billions of ECC Chunks Info were collected over dice under different failure mode

<table>
<thead>
<tr>
<th>Die</th>
<th>Plane</th>
<th>BLK</th>
<th>WL</th>
<th>PageType</th>
<th>P/E Count</th>
<th>Baking Time</th>
<th>Optimal Read LV1</th>
<th>Optimal Read LV2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0</td>
<td>100</td>
<td>64</td>
<td>0</td>
<td>3000</td>
<td>24</td>
<td>+10</td>
<td>-6</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>101</td>
<td>78</td>
<td>0</td>
<td>4000</td>
<td>24</td>
<td>+7</td>
<td>+9</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>120</td>
<td>31</td>
<td>0</td>
<td>4000</td>
<td>36</td>
<td>+3</td>
<td>-12</td>
</tr>
</tbody>
</table>

...
How many Retry Tables are required to cover the following case?

- 60 sets → 100%
- 19 sets → 90%
- 173 sets → 100%
- 40 sets → 90%
- 57 sets → 100%
- 17 sets → 90%
Reduced Retry Table – Coverage

Find some indexes to reduce retry tables without Coverage Loss

PageType 0 - By Failure Mode

PageType 1 - By Failure Mode

PageType 2 - By Failure Mode
Reduced Retry Table – Latency

Change Default Read Level and the Priority of Retry Table dynamically

Ideal: monotonically decreasing

Read Latency Distribution

Probability

Read Count/Latency

1. Proposed Error Recovery Scheme with extra overhead
2. Proposed Error Recovery Scheme without overhead
3. Traditional Retry Table (Fixed Order)

Last line of defense → Prediction Model: Optimal Read Level/LLR
Throughput with Future Status Prediction

Read Performance Comparison (4K Rand IOPS)

- Traditional Read Retry: 9% vs 57%
- Proposed Error Recovery: 9% vs 57% vs 81%

Decoding Coverage/Endurance Comparison

- Proposed Error Recovery + Hard Decode
- Traditional Read Retry + Soft Decode
- Traditional Read Retry + Hard Decode
- Default Read + Hard Decode

- Read Performance Drop can be further reduced with Future Status Prediction
Difficulty For Future Status Prediction

Prediction Flow Triggered

Select Block/Page
Block_{sel} / Page_{sel}

Collect/Save Required Parameters

Apply Future Status Prediction Model
Ex. Retry/Fail Rate, Optimal Read Level .. after 1 week

P1 : Trigger Condition/Frequency ?

P2 : Which block/page(s) should be selected ?

P3 : Important/Required Parameters ?
P 1~3 : Machine Learning

P4 : Operation Condition in 1 week !!!
P 4~5 : Dynamically Adjusted Prediction Model

P5 : Power-Off !!!
Summary

- Current NAND Flash Endurance can be Greatly Extended
  - Optimal Parameters: Retry/Optimal Read Level, LLR
  - Powerful Recovery Flow: Soft Decode, Future Status Prediction...
  - The key point is ... QoS (Quality of Service)

- Error Recovery Scheme based on Machine Learning
  - Optimized Read Retry Sequence
  - Optimal Read Level, LLR Estimation/Prediction Model
  - Future Status Prediction Model

- New Error Recovery Architecture
  - Adjust Error Recovery Flow based on failure mode/operation condition
  - Dynamically Adjusted Estimation/Prediction Model
THANK YOU!
Any questions?

Come by LITE-ON Booth# 621 for Demos!

- Learn about Machine Learning & the latest SSD Technology
- Get a chance to win special prizes!