Improving LDPC Performance Via Asymmetric Sensing Level Placement on Flash Memory

Qiao Li §, Liang Shi §, Chun Jason Xue※
Qingfeng Zhuge §, and Edwin H.-M. Sha §

§ College of Computer Science, Chongqing University
※Department of Computer Science, City University of Hong Kong
Outline

- Background
- Problem and Challenge
- Design
- Experiments and Evaluations
- Conclusions
Flash Memory Deployment

- NAND flash memory is widely applied, from USB to data centers.
- NAND flash memory is developed in various aspects.
- Various flash products have been used around us.
Challenges of State-of-the-art Flash

- With the development, from SLC to 3D TLC, the reliability is degraded.
- Stronger ECC has been a requirement, from BCH to the LDPC.
**LDPC Codes in Flash**

- Low-Density Parity Code (LDPC) is applied in Flash for strong ECC capability
- The decoding strength of LDPC depends on the accuracy of input information

Flash read: Step1. Sensing  Step2. Transfer

To decode data with high RBER, long read latency is needed.

Read performance is degraded using LDPC to guarantee reliability.
LDPC Codes in Flash

Low RBER → Less Sensing Levels → Low Read Latency

High RBER → More Sensing Levels → High Read Latency
Our Goal

- Reduce LDPC sensing levels
  - As well as the information bits
- While maintaining the same error correction capability for high RBER.
- Achieved by exploiting flash reliability characteristics.
Different types of error sources impact flash reliability.
Flash Reliability Characteristics

- **Inter-State Asymmetric Errors**
  - States $S_2$ and $S_3$ show the highest error rate.
  - States $S_0$ and $S_1$ show the lowest error rate.

- **Intra-State Asymmetric Errors**
  - The voltage states mainly shift to the left with long retention time.
  - The voltage states mainly shift to the right with great P/E cycles.
Basic Idea

- Place more sensing levels in the region with high RBER
- Place less sensing levels in the region with low RBER

![Graph showing log likelihood ratio (LLR) and sensed threshold voltage](image)

\[
LLR(y_i) = \log \left( \frac{P(x_i = 0 \mid y_i)}{P(x_i = 1 \mid y_i)} \right)
\]
Sensing Level Placement

- Inter-state Asymmetry aware.

Unaware

Aware

Hard-decision level

Soft-decision level

Sensing Level Placement

Unaware

Aware

Error-free region

Error region
Intra-state Asymmetry aware.

- Hard-decision level
- Soft-decision level

Left-shift unaware

Left-shift aware

Retention error

Right-shift unaware

Right-shift aware

Read/Program disturb
Implementation

- Inter-state Asymmetry aware.
  - The region between $S_2$ and $S_3$ takes the priority for sensing level adding.

![Diagram](image)

<table>
<thead>
<tr>
<th>Steps</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_0$</td>
<td>1</td>
<td></td>
<td>2</td>
<td></td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_1$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_2$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Implementation

- Intra-state Asymmetry aware.
  - Left-shift aware: left side takes priority for sensing level adding
  - Right-shift aware: right side takes priority for sensing level adding

<table>
<thead>
<tr>
<th>Steps</th>
<th>P/E Cycles &gt; $T_{PE}$</th>
<th>P/E Cycles ≤ $T_{PE}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1  2  3  4  5  6</td>
<td>1  2  3  4  5  6</td>
</tr>
<tr>
<td>$N^l_k$</td>
<td>1</td>
<td>2  3</td>
</tr>
<tr>
<td>$N^r_k$</td>
<td>1  2</td>
<td>3</td>
</tr>
</tbody>
</table>
Experiments Setup

- **Sensing level**
  - Calculate number of sensing level bases on bit error rates [2].
  - Bit error rates are computed from the widely used flash memory error model [12].

- **Read performance**
  - Simulator: Disksim [15]
  - 8 channels, 8 chips per channel and 4 planes per chip
  - Default FTL, page mapping, garbage collection and wear leveling
  - 6 workloads from MSR [16]
Experiments Results

- Sensing level comparison between the proposed asymmetry aware and traditional asymmetry unaware.

<table>
<thead>
<tr>
<th>Symmetric</th>
<th>Left-shift aware (P/E = 5K)</th>
<th>Right-shift aware (Retention = 1 day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_{total}$</td>
<td>bits</td>
<td>level placement</td>
</tr>
<tr>
<td>21</td>
<td>5</td>
<td>(3,3)(3,3)(3,3)</td>
</tr>
<tr>
<td>18</td>
<td>5</td>
<td>(3,2)(3,2)(3,2)</td>
</tr>
<tr>
<td>15</td>
<td>4</td>
<td>(2,2)(2,2)(2,2)</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>(2,1)(2,1)(2,1)</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>(1,1)(1,1)(1,1)</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>(1,0)(1,0)(1,0)</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>(0,0)(0,0)(0,0)</td>
</tr>
</tbody>
</table>

Traditional Symmetric  | Inter- and Left Shift Aware  | Inter- and Right Shift Aware

- Both sensing level and information bits reduction
- More reduction when there are more sensing levels.
Experiments Results

- Reduced read latency comparison

- The latency reduction increases with retention time.

- The reduction is degraded when retention time is 6 months, because only sensing level is reduced without information bits reduction.
Conclusions

- We studied the read performance reduction, which caused by adoption of LDPC for high RBER.
- We presented two reliability characteristics in flash memory.
- We proposed asymmetric sensing level placement approach based on the characteristics.
- Results show that the proposed approach achieves significant performance improvement with no overhead.
Questions!

Thanks!