Reducing Data Transfer Latency of NAND Flash Memory with Soft-Decision Sensing

Guiqiang Dong, Yuelin Zou, and Tong Zhang
Electrical, Computer and Systems Engineering Department
Rensselaer Polytechnic Institute
NAND Flash Memory

Hot Topic

CPU, Main memory, Hard disk

Bit cost reduction

100nm

10nm

2007 2008 2009 2010 2011 2012 2013 2014

Bit cost reduction
NAND Flash Memory

Error Rate  Endurance  Retention

Hard disk drive

Increasingly noisy media

Tb/in²

Areal Density Gb/in²

Two well-quoted features of LDPC codes

- Excellent error-correcting performance
- Parallel decoding for high-speed implementation
LDPC for NAND Flash

- Error floor of LDPC codes: At least partially solved!
- Low-cost, high-speed LDPC codec implementation: Completely solved!
- Memory read latency overhead
**NAND Flash Memory Read**

Accessed in the unit of page (e.g., 8kB)

- **Program**: Modulate transistor threshold voltage
- **Erase**: Modulate transistor threshold voltage

Threshold voltage: 11, 01, 00, 10
NAND Flash Memory Read

? Memory read latency overhead

LDPC Code

Controller

NAND Flash

NAND Flash

NAND Flash

Longer memory sensing latency

Longer data transfer latency

 Longer NAND flash memory read latency
Reducing Memory Read Latency

- Progressive soft-decision sensing

  1. Hard-decision memory sensing
  2. LDPC code decoding
  3. Success?
     - Yes
     - No
        - Highest sensing precision?
           - Yes
           - No
  4. Higher precision memory sensing

Success | Failure
Reducing Memory Read Latency

- Non-uniform quantization memory sensing
Reducing Memory Read Latency

? Memory read latency overhead

Controller
(LDPC decoder)

NAND Flash

NAND Flash

NAND Flash

Longer memory sensing latency

Longer data transfer latency

Lossless compression

Progressive soft-decision sensing
Non-uniform quantization memory sensing
Entropy Coding

Memory cell threshold distribution

Non-uniform quantization memory sensing

<table>
<thead>
<tr>
<th>Probability</th>
<th>0.238</th>
<th>0.0275</th>
<th>0.224</th>
<th>0.0283</th>
<th>0.223</th>
<th>0.0292</th>
<th>0.23</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed-length coding</td>
<td>111</td>
<td>110</td>
<td>100</td>
<td>000</td>
<td>001</td>
<td>011</td>
<td>010</td>
</tr>
<tr>
<td>Entropy coding</td>
<td>01</td>
<td>00000</td>
<td>10</td>
<td>00001</td>
<td>001</td>
<td>0001</td>
<td>11</td>
</tr>
</tbody>
</table>

Threshold voltage
One Step Further

- Progressive soft-decision sensing

1\textsuperscript{st} step: 4-level hard-decision sensing

2\textsuperscript{nd} step: 7-level soft-decision sensing

1\textsuperscript{st} step sensing results

Reduce 2\textsuperscript{nd} step sensing result coding overhead
Zoned Entropy Coding

Progressive soft-decision sensing $\Rightarrow$ zoned entropy coding

Four entropy coding zones

Threshold voltage
1\textsuperscript{st} step: 4-level hard-decision sensing

2\textsuperscript{nd} step: 7-level soft-decision sensing

\begin{itemize}
\item Memory Erase: \((\mu_e, \sigma_e)\)
\item Ideal Programming: \(\Delta V\)
\item Distorted by RTN: \(\lambda_r\)
\item Distorted by Cell-to-Cell Interference: \(\gamma\)
\item Retention noise: \((\mu_d, \sigma_d)\)
\end{itemize}
Evaluations

Probabilities and codewords of 2nd-step soft-decision sensing results.

<table>
<thead>
<tr>
<th>Level index</th>
<th>Probability</th>
<th>Fixed-length coding</th>
<th>Entropy coding</th>
<th>Zoned Entropy Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.2397</td>
<td>111</td>
<td>01</td>
<td>1</td>
</tr>
<tr>
<td>II</td>
<td>0.0255</td>
<td>110</td>
<td>00000</td>
<td>00</td>
</tr>
<tr>
<td>III</td>
<td>0.2255</td>
<td>100</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>IV</td>
<td>0.0263</td>
<td>000</td>
<td>00001</td>
<td>01</td>
</tr>
<tr>
<td>V</td>
<td>0.2245</td>
<td>001</td>
<td>001</td>
<td>1</td>
</tr>
<tr>
<td>VI</td>
<td>0.027</td>
<td>011</td>
<td>0001</td>
<td>00</td>
</tr>
<tr>
<td>VII</td>
<td>0.2315</td>
<td>010</td>
<td>11</td>
<td>1</td>
</tr>
</tbody>
</table>

20.4% reduction of transfer latency

64.8% reduction of transfer latency
Conclusion

- Error floor of LDPC codes  At least partially solved!
- Low cost, high-speed LDPC codec implementation  Completely solved!
- Memory read latency overhead
  - Progressive memory soft-sensing
  - Non-uniform sensing quantization
  - Two very simple compression schemes

THANK YOU!