Optimized Degree Distributions for Binary and Non-Binary LDPC Codes in Flash Memory

ISITA 2014

Kasra Vakilinia
Contributors: Dariush Divsalar, and Richard D. Wesel

Communication Systems Laboratory
University of California, Los Angeles

October 27, 2014
Introduction to NAND Flash Memory
Outline

1. Introduction to NAND Flash Memory
2. Different Number of Reads and Equivalent Communication Channels
Outline

1. Introduction to NAND Flash Memory
2. Different Number of Reads and Equivalent Communication Channels
3. Read Voltages and Mutual Information Maximization
Outline

1. Introduction to NAND Flash Memory
2. Different Number of Reads and Equivalent Communication Channels
3. Read Voltages and Mutual Information Maximization
4. LDPC Code Design and Threshold Calculation
Outline

1. Introduction to NAND Flash Memory
2. Different Number of Reads and Equivalent Communication Channels
3. Read Voltages and Mutual Information Maximization
4. LDPC Code Design and Threshold Calculation
5. Conclusions
NAND Flash Memory

Word line

Control Gate
Oxide Layer
Floating Gate
Oxide Layer
Substrate
P-well

Sense-amp Comparator
Line

Drain
Source
N+
N+

Erase Voltage
To store information, add a specified amount of charge to the floating gate.
To store information, add a specified amount of charge to the floating gate.
To read information, apply a specified word-line voltage to the control gate.
To read information, apply a specified word-line voltage to the control gate. The sense-amp comparator then provides a single bit of information (whether the transistor is “on”, i.e. the drain current is above a specified threshold). The threshold voltage is the lowest voltage at which the transistor turns on.
Variations in Threshold Voltage

- The threshold voltage is proportional to the amount of charge in the floating gate.
Variations in Threshold Voltage

- The threshold voltage is proportional to the amount of charge in the floating gate.

- The actual charge level written to the floating gate can vary with:
  Overcharge in the write operation
  Leakage in the retention period
  Interference from nearby cells
Variations in Threshold Voltage

- The threshold voltage is proportional to the amount of charge in the floating gate.

- The actual charge level written to the floating gate can vary with:
  - Overcharge in the write operation
  - Leakage in the retention period
  - Interference from nearby cells

- So, there is a *distribution* associated with the threshold voltage.
One read system is equivalent to binary symmetric channel (BSC).
Two Reads

With two symmetric reads at $q_1$ and $-q_1$, the equivalent channel has an erasure region.
Three Reads

Additional reads result in additional outputs and more complicated equivalent communication channel.
For \( n \) reads, the channel has \( n+1 \) outputs.
Mutual Information Maximization for One Read

Find the read voltages to maximize the mutual information.

\[ maximize \ I(X; Y) = H(Y) - H(Y|X) \]

For the 1-read system, the optimized read is at zero.
Mutual Information Maximization for Three Reads

Find the threshold voltages to maximize the mutual information.

\[ \text{maximize } I(X; Y) = H(Y) - H(Y|X) \]

For 2,3 reads, \( q_1 \) is the solution to \( \frac{dl}{dq_1} = 0 \).
For the 5 reads, the Max MI is the solution of $\frac{dl}{dq_1} = 0$ and $\frac{dl}{dq_2} = 0$. 
For the 5 reads, the Max MI is the solution of $\frac{dl}{dq_1} = 0$ and $\frac{dl}{dq_2} = 0$.

If $q_2 \geq q_1$, MI is quasi-concave in $q_1$ for a fixed value of $q_2$ and vice-versa.
Fully optimized single-read channel \[ \times \times 0 \times \times \]: $MI = 0.560$

We can sacrifice some $MI$ at three reads to optimize the $MI$ with five reads.

Fully optimized single-read channel \[ \times \times 0 \times \times \]: $MI = 0.560$

Progressive 3-read channel \[ \times -0.4 0 0.4 \times \]: $MI = 0.643$

Fully optimized for 5-read channel \[-0.9 -0.4 0 0.4 0.9\]: $MI = 0.668$

We can sacrifice some $MI$ at five reads to optimize the $MI$ at three reads.

Fully optimized single-read channel \[ \times \times 0 \times \times \]: $MI = 0.560$

Fully optimized for 3-read channel \[ \times -0.61 0 0.61 \times \]: $MI = 0.652$

Progressive 5-read channel \[-0.61 -0.28 0 0.28 0.61\]: $MI = 0.663$
Read Voltage Optimization

- Fully optimized single-read channel $[ x \ x \ 0 \ x \ x ]$: $MI = 0.560$
- Fully optimized for 3-read channel $[ x \ -0.61 \ 0 \ 0.61 \ x ]$: $MI = 0.652$
Read Voltage Optimization

- Fully optimized single-read channel $[\times \times 0 \times \times ]$: $MI = 0.560$
- Fully optimized for 3-read channel $[\times -0.61 0 0.61 \times ]$: $MI = 0.652$
- Fully optimized for 5-read channel $[-0.9 -0.4 0 0.4 0.9]$: $MI = 0.668$
Read Voltage Optimization

- Fully optimized single-read channel $[ \times \times 0 \times \times ]$: $MI = 0.560$
- Fully optimized for 3-read channel $[ \times -0.61 0 0.61 \times ]$: $MI = 0.652$
- Fully optimized for 5-read channel $[-0.9 -0.4 0 0.4 0.9]$: $MI = 0.668$

- We can sacrifice some MI at three reads to optimize the MI with five reads.
Read Voltage Optimization

- Fully optimized single-read channel $[\times \times 0 \times \times ]$: $MI = 0.560$
- Fully optimized for 3-read channel $[\times -0.61 0 0.61 \times ]$: $MI = 0.652$
- Fully optimized for 5-read channel $[-0.9 -0.4 0 0.4 0.9]$: $MI = 0.668$

We can sacrifice some MI at three reads to optimize the MI with five reads.

- Fully optimized single-read channel $[\times \times 0 \times \times ]$: $MI = 0.560$
Read Voltage Optimization

- Fully optimized single-read channel $[\times \times 0 \times \times ]$: $MI = 0.560$
- Fully optimized for 3-read channel $[\times -0.61 0 0.61 \times ]$: $MI = 0.652$
- Fully optimized for 5-read channel $[-0.9 -0.4 0 0.4 0.9]$: $MI = 0.668$

We can sacrifice some MI at three reads to optimize the MI with five reads.

- Fully optimized single-read channel $[\times \times 0 \times \times ]$: $MI = 0.560$
- Progressive 3-read channel $[\times -0.4 0 0.4 \times ]$: $MI = 0.643$
Read Voltage Optimization

- Fully optimized single-read channel \([ x \ x \ 0 \ x \ x ]\): \(MI = 0.560\)
- Fully optimized for 3-read channel \([ x \ -0.61 \ 0 \ 0.61 \ x ]\): \(MI = 0.652\)
- Fully optimized for 5-read channel \([-0.9 \ -0.4 \ 0 \ 0.4 \ 0.9]\): \(MI = 0.668\)

We can sacrifice some MI at three reads to optimize the MI with five reads.

- Fully optimized single-read channel \([ x \ x \ 0 \ x \ x ]\): \(MI = 0.560\)
- Progressive 3-read channel \([ x \ -0.4 \ 0 \ 0.4 \ x ]\): \(MI = 0.643\)
- Fully optimized for 5-read channel \([-0.9 \ -0.4 \ 0 \ 0.4 \ 0.9]\): \(MI = 0.668\)
Read Voltage Optimization

- Fully optimized single-read channel \([ x \ x \ 0 \ x \ x ]\): \(MI = 0.560\)
- Fully optimized for 3-read channel \([ x \ -0.61 \ 0 \ 0.61 \ x ]\): \(MI = 0.652\)
- Fully optimized for 5-read channel \([-0.9 \ -0.4 \ 0 \ 0.4 \ 0.9]\): \(MI = 0.668\)

We can sacrifice some MI at three reads to optimize the MI with five reads.

- Fully optimized single-read channel \([ x \ x \ 0 \ x \ x ]\): \(MI = 0.560\)
- Progressive 3-read channel \([ x \ -0.4 \ 0 \ 0.4 \ x ]\): \(MI = 0.643\)
- Fully optimized for 5-read channel \([-0.9 \ -0.4 \ 0 \ 0.4 \ 0.9]\): \(MI = 0.668\)

We can sacrifice some MI at five reads to optimize the MI at three reads.
Read Voltage Optimization

- Fully optimized single-read channel $[\times \times 0 \times \times ]$: $MI = 0.560$
- Fully optimized for 3-read channel $[\times -0.61 0 0.61 \times ]$: $MI = 0.652$
- Fully optimized for 5-read channel $[-0.9 -0.4 0 0.4 0.9]$: $MI = 0.668$

We can sacrifice some MI at three reads to optimize the MI with five reads.

- Fully optimized single-read channel $[\times \times 0 \times \times ]$: $MI = 0.560$
- Progressive 3-read channel $[\times -0.4 0 0.4 \times ]$: $MI = 0.643$
- Fully optimized for 5-read channel $[-0.9 -0.4 0 0.4 0.9]$: $MI = 0.668$

We can sacrifice some MI at five reads to optimize the MI at three reads.

- Fully optimized single-read channel $[\times \times 0 \times \times ]$: $MI = 0.560$
Read Voltage Optimization

- Fully optimized single-read channel \([ x \ x \ 0 \ x \ x ]\): \(MI = 0.560\)
- Fully optimized for 3-read channel \([ x \ -0.61 \ 0 \ 0.61 \ x ]\): \(MI = 0.652\)
- Fully optimized for 5-read channel \([-0.9 \ -0.4 \ 0 \ 0.4 \ 0.9]\): \(MI = 0.668\)

We can sacrifice some MI at three reads to optimize the MI with five reads.

- Fully optimized single-read channel \([ x \ x \ 0 \ x \ x ]\): \(MI = 0.560\)
- Progressive 3-read channel \([ x \ -0.4 \ 0 \ 0.4 \ x ]\): \(MI = 0.643\)
- Fully optimized for 5-read channel \([-0.9 \ -0.4 \ 0 \ 0.4 \ 0.9]\): \(MI = 0.668\)

We can sacrifice some MI at five reads to optimize the MI at three reads.

- Fully optimized single-read channel \([ x \ x \ 0 \ x \ x ]\): \(MI = 0.560\)
- Fully optimized for 3-read channel \([ x \ -0.61 \ 0 \ 0.61 \ x ]\): \(MI = 0.652\)
Read Voltage Optimization

- Fully optimized single-read channel \([ x \ x \ 0 \ x \ x ]\): \(MI = 0.560\)
- Fully optimized for 3-read channel \([ x \ -0.61 \ 0 \ 0.61 \ x ]\): \(MI = 0.652\)
- Fully optimized for 5-read channel \([-0.9 \ -0.4 \ 0 \ 0.4 \ 0.9]\): \(MI = 0.668\)

We can sacrifice some MI at three reads to optimize the MI with five reads.

- Fully optimized single-read channel \([ x \ x \ 0 \ x \ x ]\): \(MI = 0.560\)
- Progressive 3-read channel \([ x \ -0.4 \ 0 \ 0.4 \ x ]\): \(MI = 0.643\)
- Fully optimized for 5-read channel \([-0.9 \ -0.4 \ 0 \ 0.4 \ 0.9]\): \(MI = 0.668\)

We can sacrifice some MI at five reads to optimize the MI at three reads.

- Fully optimized single-read channel \([ x \ x \ 0 \ x \ x ]\): \(MI = 0.560\)
- Fully optimized for 3-read channel \([ x \ -0.61 \ 0 \ 0.61 \ x ]\): \(MI = 0.652\)
- Progressive 5-read channel \([-0.61 \ -0.28 \ 0 \ 0.28 \ 0.61]\): \(MI = 0.663\)
Concluding Remarks on MMI

<table>
<thead>
<tr>
<th>Method</th>
<th>MI</th>
<th>Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimized single read</td>
<td>0.560</td>
<td>[ \times \times 0 \times \times ]</td>
</tr>
<tr>
<td>Optimized three reads</td>
<td>0.652</td>
<td>[ \times -0.61 0 0.61 \times ]</td>
</tr>
<tr>
<td>Optimized five reads</td>
<td>0.668</td>
<td>[-0.9 -0.4 0 0.4 0.9]</td>
</tr>
<tr>
<td>Progressive three reads</td>
<td>0.643</td>
<td>[ \times -0.4 0 0.4 \times ]</td>
</tr>
<tr>
<td>Progressive five reads</td>
<td>0.663</td>
<td>[-0.61 -0.28 0 0.28 0.61]</td>
</tr>
</tbody>
</table>

The MI differences are small, but large enough to have noticeable effects on frame error rate (according to [1]).

Optimizing $q_1$ for three-read performance maximizes the probability of decoding successfully after three reads.

Optimizing $q_1$ for the five-read scenario will produce the lowest probability of losing a page.

In order to design LDPC codes that are matched to the Flash read channel, we use EXIT functions with the Gaussian approximation and Reciprocal Channel Approximation (faster alternative to Density Evolution) to compute LDPC decoding thresholds for different number of reads.
For a fixed degree distribution and $q_1$, RCA-EXIT analysis can determine the word-line voltages that minimize $E_b/N_0$ threshold for multiple-read Flash channels.
For a fixed degree distribution, there is a $q_1$ that has the absolute $E_b/N_0$ threshold for multiple-read Flash channels.
For each $Eb/N_0$ there is a $q_1$ that maximizes the MI.
The threshold $Eb/N_0$ achieved with $q_1$ chosen to maximize MI is only 1% away from the best possible $Eb/N_0^*$ achieved with a density-evolution optimized $q_1$. 
The optimized degree distribution is found by calculating the threshold at $q_{MMI}$ and further optimizing the $q_1$. 

Kasra Vakilinia (UCLA)
Table: Thresholds for optimized degree distributions on SLC flash with 1, 3, and 5 reads as well as soft information.

<table>
<thead>
<tr>
<th>Target Threshold</th>
<th>1 read</th>
<th>3 reads</th>
<th>5 reads</th>
<th>Soft</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimized for 1 read</td>
<td>4.752</td>
<td>3.728</td>
<td>3.542</td>
<td>3.398</td>
</tr>
<tr>
<td>Optimized for 3 reads</td>
<td>4.923</td>
<td>3.640</td>
<td>3.441</td>
<td>3.295</td>
</tr>
<tr>
<td>Optimized for 5 reads</td>
<td>4.926</td>
<td>3.649</td>
<td>3.437</td>
<td>3.288</td>
</tr>
<tr>
<td>Optimized for Soft</td>
<td>4.926</td>
<td>3.662</td>
<td>3.443</td>
<td>3.275</td>
</tr>
<tr>
<td>Shannon-Limit</td>
<td>4.400</td>
<td>3.495</td>
<td>3.328</td>
<td>3.198</td>
</tr>
</tbody>
</table>

Different codes have different thresholds for different levels of quantization. A code designed for one level of quantization is not the best for another level.
Table: Gap of thresholds for optimized degree distributions on SLC flash with 1, 3, and 5 reads as well as soft information.

<table>
<thead>
<tr>
<th>Target threshold gap</th>
<th>1 read Th. gap</th>
<th>3 reads Th. gap</th>
<th>5 reads Th. gap</th>
<th>Soft Th. gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimized for 1 read</td>
<td>0</td>
<td>0.088</td>
<td>0.105</td>
<td>0.123</td>
</tr>
<tr>
<td>Optimized for 3 reads</td>
<td>0.171</td>
<td>0</td>
<td>0.004</td>
<td>0.020</td>
</tr>
<tr>
<td>Optimized for 5 reads</td>
<td>0.174</td>
<td>0.009</td>
<td>0</td>
<td>0.013</td>
</tr>
<tr>
<td>Optimized for Soft</td>
<td>0.1740</td>
<td>0.022</td>
<td>0.006</td>
<td>0</td>
</tr>
</tbody>
</table>

The code optimized for a single read performs well across multiple reads. There is reason to consider using the code designed for a single read.
The threshold of two regular binary LDPC codes vs. Shannon-limit for different number of reads.
The threshold of regular NB-LDPC code for different number of reads.
The threshold of an optimized binary LDPC code for different number of reads.
Huge gain can be achieved by using a rate-0.45 NB-LDPC code on MLC with the same number of reads and for the same spectral efficiency of 0.9.
Huge gain can be achieved by using a rate-0.45 NB-LDPC code on MLC with the same number of reads and for the same spectral efficiency of 0.9.
<table>
<thead>
<tr>
<th># Reads</th>
<th>Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLC</td>
<td></td>
</tr>
<tr>
<td>1 read</td>
<td>$\lambda_2$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_3$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_7$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_8$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_{27}$</td>
</tr>
<tr>
<td></td>
<td>$\rho_{61}$</td>
</tr>
<tr>
<td>SLC</td>
<td></td>
</tr>
<tr>
<td>3 reads</td>
<td>$\lambda_2$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_3$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_6$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_7$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_{26}$</td>
</tr>
<tr>
<td></td>
<td>$\rho_{56}$</td>
</tr>
<tr>
<td>SLC</td>
<td></td>
</tr>
<tr>
<td>5 reads</td>
<td>$\lambda_2$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_3$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_5$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_8$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_{25}$</td>
</tr>
<tr>
<td></td>
<td>$\rho_{56}$</td>
</tr>
<tr>
<td>SLC</td>
<td></td>
</tr>
<tr>
<td>soft</td>
<td>$\lambda_2$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_3$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_6$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_8$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_{26}$</td>
</tr>
<tr>
<td></td>
<td>$\rho_{56}$</td>
</tr>
<tr>
<td>MLC</td>
<td></td>
</tr>
<tr>
<td>3 reads</td>
<td>$\lambda_2$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_3$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_4$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_7$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_8$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_{11}$</td>
</tr>
<tr>
<td></td>
<td>$\rho_5$</td>
</tr>
<tr>
<td></td>
<td>$\rho_6$</td>
</tr>
<tr>
<td></td>
<td>$\rho_8$</td>
</tr>
<tr>
<td></td>
<td>$\rho_{22}$</td>
</tr>
<tr>
<td></td>
<td>$\rho_{22}$</td>
</tr>
<tr>
<td>MLC</td>
<td></td>
</tr>
<tr>
<td>soft</td>
<td>$\lambda_2$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_3$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_4$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_7$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_9$</td>
</tr>
<tr>
<td></td>
<td>$\lambda_{11}$</td>
</tr>
<tr>
<td></td>
<td>$\rho_5$</td>
</tr>
<tr>
<td></td>
<td>$\rho_6$</td>
</tr>
<tr>
<td></td>
<td>$\rho_8$</td>
</tr>
<tr>
<td></td>
<td>$\rho_9$</td>
</tr>
<tr>
<td></td>
<td>$\rho_{22}$</td>
</tr>
</tbody>
</table>
Conclusions

- $q_{MMI}$ is the solution to $\frac{dl}{dq_1} = 0$ (and $\frac{dl}{dq_2} = 0$) and can be computed quickly.
Conclusions

- $q_{MMI}$ is the solution to $\frac{dl}{dq_1} = 0$ (and $\frac{dl}{dq_2} = 0$) and can be computed quickly.

- Optimized Degree Distributions can be found by calculating the threshold at $q_{MMI}$ and further optimizing the $q$ to get the lowest $E_b/N_0$ thresholds.
Conclusions

- $q_{\text{MMI}}$ is the solution to $\frac{dl}{dq_1} = 0$ (and $\frac{dl}{dq_2} = 0$) and can be computed quickly.

- Optimized Degree Distributions can be found by calculating the threshold at $q_{\text{MMI}}$ and further optimizing the $q$ to get the lowest $E_b/N_0$ thresholds.

- Even though a code designed for one level of quantization is not the best for another level of quantization, the code optimized for a single read performs well across multiple reads.
Conclusions

- $q_{MMI}$ is the solution to $\frac{dl}{dq_1} = 0$ (and \(\frac{dl}{dq_2} = 0\)) and can be computed quickly.

- Optimized Degree Distributions can be found by calculating the threshold at $q_{MMI}$ and further optimizing the $q$ to get the lowest $E_b/N_0$ thresholds.

- Even though a code designed for one level of quantization is not the best for another level of quantization, the code optimized for a single read performs well across multiple reads.

- Using a low-rate LDPC code on MLC for the same spectral efficiency as a high-rate code on SLC can provide huge gain in terms of the thresholds.