Dynamic Voltage Allocation with Quantized Voltage Levels and Simplified Channel Modeling

Haobo Wang, Nathan Wong, Richard Wesel
Electrical Engineering, UCLA

November 10, 2015
Mutual Information using Fixed Voltage Allocation Example

3020 P/E Cycles
Outline

- Dynamic Voltage Allocation
  - Channel Model
    - Channel Parameter Estimation
- Dynamic Voltage Allocation with Simplified Channel Modeling
- Dynamic Voltage Allocation with Quantized Voltage Levels
Dynamic Voltage Allocation

- Fixed threshold voltage allocation provides unnecessary margin at the beginning of Flash memory’s lifetime, causing accelerated wear-out.
Dynamic Voltage Allocation

• Fixed threshold voltage allocation provides unnecessary margin at the beginning of Flash memory’s lifetime, causing accelerated wear-out.

• Dynamic Voltage Allocation can reduce unnecessary wear-out, and thus increase lifetime by using lower threshold voltages for early writes.
Dynamic Voltage Allocation

• Fixed threshold voltage allocation provides unnecessary margin at the beginning of Flash memory’s lifetime, causing accelerated wear-out.

• Dynamic Voltage Allocation can reduce unnecessary wear-out, and thus increase lifetime by using lower threshold voltages for early writes.

• The threshold voltages will be periodically increased using a single scaling factor to combat channel degradation.
Dynamic Voltage Allocation

- Fixed threshold voltage allocation provides unnecessary margin at the beginning of Flash memory’s lifetime, causing accelerated wear-out.

- Dynamic Voltage Allocation can reduce unnecessary wear-out, and thus increase lifetime by using lower threshold voltages for early writes.

- The threshold voltages will be periodically increased using a single scaling factor to combat channel degradation.

- The target of the anti-degradation process is to maintain certain amount of mutual information as long as possible by increasing the mutual information to a point higher than the minimum mutual information limit.
Dynamic Voltage Allocation Scaling Factor Example

![Graph showing the comparison between fixed voltage allocation and dynamic voltage allocation. The graph plots Scaling Factor on the y-axis against P/E Cycles on the x-axis. The fixed voltage allocation is represented by a blue horizontal line, while the dynamic voltage allocation is represented by a red line that increases significantly after a P/E cycle of 4000.]
Dynamic Voltage Allocation Scaling Factor Example

![Graph showing dynamic voltage allocation scaling factor example with different scaling factors and P/E cycles. The graph illustrates the comparison between fixed voltage allocation and dynamic voltage allocation, highlighting the varying scaling factor over P/E cycles.]
Dynamic Voltage Allocation Scaling Example

Voltage Probability Density

Before DVA

P/E = 1000
1.936 bits

P/E = 3000
1.935 bits

P/E = 4800
1.929 bits
Dynamic Voltage Allocation Scaling Example

- P/E = 1000
  - 1.936 bits

- P/E = 3000
  - 1.935 bits

- P/E = 4800
  - 1.929 bits
Dynamic Voltage Allocation Scaling Example

- **P/E = 1000**
  - 1.936 bits $\rightarrow$ 1.94 bits

- **P/E = 3000**
  - 1.935 bits $\rightarrow$ 1.94 bits

- **P/E = 4800**
  - 1.929 bits $\rightarrow$ 1.94 bits
Dynamic Voltage Allocation v.s. Fixed Voltage Allocation Example

- Mutual Information with DVA
- Mutual Information with Fixed Allocation
- Mutual Information Target

3020 P/E Cycles
2164 P/E Cycles
71.7% Increase
DVA using Histogram-based Channel Estimation

- Histogram Measurement
- Parameter Estimation
- Dynamic Voltage Allocation
- Channel Model Assumption
We model the NAND flash memory cell data storage process as:

\[ y = x + \sum_i n_i \]

- \( y \): sensed threshold voltage
- \( x \): intended threshold voltage
- \( n_i \): ith noise component
Noise Components

• Possible noise components are

  • $n_{pe}$ : programming error
  • $n_p$ : programming noise
  • $n_w$ : wear-out noise
  • $n_{c2c}$ : cell-to-cell interference
  • $n_r$ : retention noise
Two Models

• Model 1

\[ y = x + n_p + n_w + n_r \]

• Model 2

\[ y = x + n_p + n_w + n_r + n_{c2c} + n_{pe} \]
Cell-to-cell Interference ($n_{c2c}$)

- Modern Flash memory cells are densely placed in a two or three dimensional array, this causes coupling between cells during write operations.

- The noise component models the coupling effect with coupling coefficients. They are classified in three categories:

  \[ \gamma_x, \quad \gamma_y, \quad \gamma_{xy}. \]
In order to reduce the area of a Flash chip, Flash memory write process programs a wordline in two batches, even cells and odd cells. In our research, we assume even cells in each wordline are written first by default.

\[
V_{n_{2c, odd}} = \gamma_{x,right} \times V_{x,right} + \gamma_{x,left} \times V_{x,left} + \gamma_{y} \times V_{y}
\]

\[
V_{n_{2c, even}} = \gamma_{x,right} \times V_{x,right} + \gamma_{x,left} \times V_{x,left} + \gamma_{xy,upper_left} \times V_{xy,upper_left} + \gamma_{xy,upper_right} \times V_{xy,upper_right}
\]
Replace the number of P/E cycles with accumulated voltage

• Channel degradation is usually modeled as a function of the number of program/erase (P/E) cycles.
Replace the number of P/E cycles with accumulated voltage

• Channel degradation is usually modeled as a function of the number of program/erase (P/E) cycles.

• The volume of charge passing through dielectrics actually causes the degradation.
Replace the number of P/E cycles with accumulated voltage

• Channel degradation is usually modeled as a function of the number of program/erase (P/E) cycles.

• The volume of charge passing through dielectrics actually causes the degradation.

• The use of the number of P/E cycles is approximately correct when the volume of charge passing through the dielectrics is the same for each P/E cycle.
Channel degradation is usually modeled as a function of the number of program/erase (P/E) cycles.

The volume of charge passing through dielectrics actually causes the degradation.

The use of the number of P/E cycles is approximately correct when the volume of charge passing through the dielectrics is the same for each P/E cycle.

We use a more precise metric named accumulated voltage $V_{acc}$ to directly characterize the volume of charge that has passed since the first write.
Accumulated Voltage

\[ V_{acc} = \sum_{j=1}^{N} (V_p^{(j)} - V_e) \]

- \( V_{acc} \): accumulated voltage over \( N \) P/E cycles,
- \( V_p^{(j)} \): programmed threshold voltage of the \( j \)th P/E cycle,
- \( V_e \): threshold voltage of the erased state.

The normalized accumulated voltage is \( V_{acc} / V_{max} \), where \( V_{max} \) is the maximum of \( V_p - V_e \).

When using fixed voltage levels, \( V_{acc} / V_{max} \approx \text{PE Cycle count} \).
Channel Parameter Estimation

- Channel parameter estimation workflow:
Channel Parameter Estimation

- Channel parameter estimation workflow:

  - Histogram Measurement
  - Channel Model Assumption
  - Levenberg-Marquardt algorithm
  - Least Squares Algorithm
  - Estimated Parameters
Channel Parameter Estimation

- Channel parameter estimation workflow:

  - Histogram Measurement
    - 9-read equal probability histogram

  - Least Squares Algorithm
    - Levenberg-Marquardt algorithm

  - Estimated Parameters

  - Channel Model Assumption
Histogram Measurement

Parameter Estimation

Dynamic Voltage Allocation

Histogram Measurement

![Graph 1](image1.png)

![Graph 2](image2.png)
Ground Truth: [0.0099, 0.3500, 0.0500, 0.0617, -0.5882]
**Parameter Estimation**

Ground Truth: \([0.0099, 0.3500, 0.0500, 0.0617, -0.5882]\)

Estimation Error: \(10^{-4} \times [0.0101, 0.0214, -0.1774, 0.0405, -0.0044]\)
Voltage Levels Adapted to Degraded Channel

Histogram Measurement

Parameter Estimation

Dynamic Voltage Allocation
Monte Carlo Simulation Result (Model 1)

Mutual Information (bit)

- Mutual Information with DVA (Monte Carlo)
- Mutual Information with Fixed Allocation
- Mutual Information Target

3020 P/E Cycles
2157 P/E Cycles
71.4% Increase
• For complex channels, it is not practical to calculate all the analytical expressions needed to estimate the channel.

• In some situations, there is no detailed knowledge about the actual channel.

• The simplest channel assumption is Gaussian distribution.
Monte Carlo Simulation Result for Model 1 (Gaussian Assumption)

- Mutual Information (bit)

- Mutual Information with DVA (Matching Assumption)
- Mutual Information with DVA (Gaussian Assumption)
- Mutual Information with Fixed Allocation
- Mutual Information Target

- 3020 P/E Cycles
- 2240 P/E Cycles
- 74.2% Increase
Dynamic Voltage Allocation for Model 2

- Even and odd cells perform differently.

- The first approach is to optimize for the worst case performance represented by the even-cell channel.
Theoretical Simulation for Model 2 with Ideal Channel Knowledge

- Odd-cell channel is over-optimized.
Dynamic Voltage Allocation for Model 2

- Even and odd cells perform differently.

- The first approach is to optimize for the worst case performance represented by the even cell channel.

- The second approach is to equalize the performance of even and odd-cell channels.
  - The write order is switched every 100 P/E cycles.
  - To avoid over-optimization, even and odd-cell channel are optimized separately.
Theoretical Simulation for Model 2 with Ideal Channel Knowledge

Mutual Information (bit)

- Mutual Information with DVA (Even Cell)
- Mutual Information with DVA (Odd Cell)
- Mutual Information with Fixed Allocation (Even Cell)
- Mutual Information with Fixed Allocation (Odd Cell)
- Mutual Information Target

P/E Cycle

2460 P/E Cycles
1892 P/E Cycles
76.9% Increase
Theoretical Simulation for Model 2 with Ideal Channel Knowledge

- Mutual Information with DVA (Even Cell)
- Mutual Information with DVA (Odd Cell)
- Mutual Information with Fixed Allocation (Even Cell)
- Mutual Information with Fixed Allocation (Odd Cell)
- Mutual Information Target

Mutual Information (bit)

P/E Cycle

2460 P/E Cycles

1892 P/E Cycles

76.9% Increase
Dynamic Voltage Allocation for Model 2

• Even and odd cells perform differently.

• The first approach is to optimize for the worst case performance represented by the even cell channel.

• The second approach is to equalize the performance of even and odd-cell channels.
  • The write order is switched every 100 P/E cycles.
  • To avoid over-optimization, even and odd-cell channel are optimized separately.
Monte Carlo Simulation Result for Model 2 (Gaussian Assumption)

- Mutual Information with DVA (Even Cell)
- Mutual Information with DVA (Odd Cell)
- Mutual Information with Fixed Allocation (Even Cell)
- Mutual Information with Fixed Allocation (Odd Cell)
- Mutual Information Target

Graph:
- X-axis: P/E Cycle
- Y-axis: Mutual Information (bit)
- Key points:
  - 2460 P/E Cycles
  - 2149 P/E Cycles
  - 87.4% Increase
Dynamic Voltage Allocation places intended threshold voltages with infinite precision.

Binning strategy places read reference voltages with infinite precision.

In reality, the intended threshold voltages and read reference voltages can only be placed at certain locations across the entire voltage range.

The number of possible locations is limited.
Monte Carlo Simulation Result for Model 2 (Gaussian Assumption, 256-level Quantization)

- **Mutual Information with DVA (Even Cell)**
- **Mutual Information with DVA (Odd Cell)**
- **Mutual Information with Fixed Allocation (Even Cell)**
- **Mutual Information with Fixed Allocation (Odd Cell)**
- **Mutual Information Target**

- No quantization: 87.4% Increase
- 2460 P/E Cycles
- 2105 P/E Cycles
- 85.6% Increase
Monte Carlo Simulation Result for Model 2 (Gaussian Assumption, 128-level Quantization)

- Mutual Information with DVA (Even Cell)
- Mutual Information with DVA (Odd Cell)
- Mutual Information with Fixed Allocation (Even Cell)
- Mutual Information with Fixed Allocation (Odd Cell)
- Mutual Information Target

2460 P/E Cycles
2140 P/E Cycles
87.0% Increase
Monte Carlo Simulation Result for Model 2 (Gaussian Assumption, 64-level Quantization)
Read Threshold Quantization v.s. Write Threshold Quantization

- Read threshold quantization affects channel estimation accuracy.
- Write threshold quantization affects dynamic voltage allocation accuracy.
Monte Carlo Simulation Result for Model 2 (Gaussian Assumption, 64-level Quantization, Read Threshold Quantization Only)
Monte Carlo Simulation Result for Model 2 (Gaussian Assumption, 64-level Quantization, Write Threshold Quantization Only)
Read Threshold Quantization v.s. Write Threshold Quantization

- Read threshold quantization affects channel estimation accuracy.
- Write threshold quantization affects dynamic voltage allocation accuracy.
- Read threshold quantization requires high accuracy.
Conclusion

- Dynamic Voltage Allocation can extend Flash lifetime even with practical constraints on model knowledge and quantization.

- The Gaussian distribution is a good approximation for Flash memory read channel for the channel estimation and dynamic voltage allocation.

- Quantization does not significantly affect Dynamic Voltage Allocation’s performance when minimal accuracy is assured for example by having 128 levels.

- The read threshold quantization requires higher precision than the dynamic voltage allocation.