Artificial Neural Network Coupled LDPC ECC for 3D-NAND Flash Memories

Toshiki Nakamura and Ken Takeuchi
Chuo University, Japan
Outline

- Introduction
- Proposed Artificial Neural Network Coupled (ANN) LDPC ECC (ANN-LDPC ECC)
  - Training of ANN
  - Case 1-3 (1 hidden layer ANN)
  - Case 4 and 5 (2 hidden layer ANN)
- Conclusion
Outline

- Introduction
- Proposed Artificial Neural Network Coupled (ANN) LDPC ECC (ANN-LDPC ECC)
  - Training of ANN
  - Case 1-3 (1 hidden layer ANN)
  - Case 4 and 5 (2 hidden layer ANN)
- Conclusion
3D-TLC NAND Flash Memory

- TLC NAND flash memory stores 3 bits/cell
- Data is stored by controlling amount of electrons

V<sub>TH</sub> state:
- Erase A
- B: 1 0 0 0
- C: 1 1 0 0
- D: 1 1 1 0
- E: 1 1 1 1
- F: 0 0 0 0
- G: 0 0 0 0

V<sub>TH</sub> distribution:

3D-NAND flash cell array:
- WL: Word-line
- BL: Bit-Line
- SGD M
- SGD 0
- SGS

Page type:
- WL 0
- WL 1
- WL 2
- WL N
Low-Density Parity-Check (LDPC) ECC

- LDPC ECC corrects errors gradually by repeating decoding
- Log-likelihood ratio (LLR) is required for LDPC decoding

Log-likelihood ratio (LLR)

- LDPC decode
- Parity check ok?
  - Yes: Data output
  - No: Max iteration?
    - Yes: Read fail
    - No: Iteration++

Charge-trap 3D-TLC NAND flash, Write/erase cycles ($N_{WE}$) = 500, Data-retention time = 3 days

@85degC
Relation between LLR and BER

- LLR represents reliability of each bit data
- Error-correcting capability depends on accuracy of LLR
- LLR is calculated based on bit-error rate (BER)

LLR (1) = \( \ln \{ \frac{BER}{1-BER} \} \)

LLR (0) = \( \ln \{ \frac{1-BER}{BER} \} \)

Read bit: 0 0 ? 0 ?? 1 ?? 1 1

BER: Low High Low

LLR: 12 ... 5 ... -5 ... -12

Large positive value

Large negative value

Large positive value

Large negative value
Problem of Conventional LDPC ECC

- BER varies complicatedly in 3D-NAND flash
  Conventional LDPC ECC cannot predict BER precisely

Reliability parameters of 3D-NAND flash

Static information
- $V_{TH}$ state
- Page type (Lower/Middle/Upper)
- Neighboring cell data
- WL number

Dynamic information
- Data-retention time
- Write/erase cycles
Word-line Variations

- Complicated inter word-line variations of errors exist [1]

Charge-trap 3D-TLC NAND flash

@85degC

$N_{W/E} = 500$

Data-retention time = 2.8 days

[1] K. Mizoguchi et al., IMW, pp. 119-122, May 2017
Lateral Charge Migration

- Charge loss is caused by lateral charge migration [2]

In 3D structure, charge-trap layer is connected to neighboring cell

Floating-Gate (FG) Cell

- Inter floating-gate capacitive coupling noise cause errors [3]

Interstitial dielectric

FG-FG interference

Channel

Floating-gate 3D-TLC NAND flash, $N_{W/E} = 300$, Data-retention time = 2.8days

@85degC

Measured BER of “F”-state at WL (n) (a.u.)

V_TH state of WL (n+1) V_TH state of WL (n-1)

Concept of Proposed ANN-LDPC

Proposed Artificial Neural Network Coupled (ANN) LDPC ECC (ANN-LDPC ECC)

Proposed ANN-LDPC adaptively and automatically correct errors [4]

Outline

- Introduction
- Proposed Artificial Neural Network Coupled (ANN) LDPC ECC (ANN-LDPC ECC)
  - Training of ANN
  - Case 1-3 (1 hidden layer ANN)
  - Case 4 and 5 (2 hidden layer ANN)
- Conclusion
Training of ANN

Training flowchart

1. **Input parameter**
2. **Calculate output**
3. **Calculate loss function**
4. **Update weights**

- Training of ANN is performed before product shipment
- Calculated synaptic weights are pre-recorded in proposed storage controller
Training of ANN (Step 1 and 2)

**Step 1**: Input parameter $x$  
- Random data (4MBytes)
- Separate 200 batches  
- Next epoch

**Step 2**: Calculate output $y$  
- $y_n = h_2 \{ \sum w^{(2)} h_1(\sum w^{(1)} x) \}$

- ReLU (hidden layer)  
  $$h_1(x) = \begin{cases} 
  x & (x \geq 0) \\
  0 & (x < 0) 
  \end{cases}$$

- Softmax (output layer)  
  $$h_2(x_n) = \frac{\exp(x_n)}{\exp(x_1) + \exp(x_2)}$$

Activation function
Training of ANN (Step 3 and 4)

Step 3: Calculate loss function $E$

Loss function
- Softmax cross entropy
  \[
  E = -\{L_1 \ln(y_1) + L_2 \ln(y_2)\}
  \]

Ex. 1: Bad training
\[
\begin{align*}
  y_1 &= 0.9, \quad L_1 = 0 \\
  y_2 &= 0.1, \quad L_2 = 1 \\
  E &= -\ln(0.1) = 2.3
\end{align*}
\]

Ex. 2: Good training
\[
\begin{align*}
  y_1 &= 0.1, \quad L_1 = 0 \\
  y_2 &= 0.9, \quad L_2 = 1 \\
  E &= -\ln(0.9) = 0.1
\end{align*}
\]
Outline

- Introduction
- Proposed Artificial Neural Network Coupled (ANN) LDPC ECC (ANN-LDPC ECC)
  - Training of ANN
  - Case 1-3 (1 hidden layer ANN)
  - Case 4 and 5 (2 hidden layer ANN)
- Conclusion
Case 1-3 (1 Hidden Layer ANN)

- Case 1-3 consider only static information
- Value of input parameters is 0 or 1

### 1 hidden layer ANN

<table>
<thead>
<tr>
<th>Input parameters</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{TH}$ state</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Page type</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Neighboring cell data</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>WL number</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Read offset level</td>
<td>◯</td>
<td>◯</td>
<td>○</td>
</tr>
</tbody>
</table>

0 or 1

Value of input parameters is 0 or 1

Predicted BER

$\begin{align*}
x_1 & \quad \cdots \quad x_k \\
a_1 & \quad \cdots \quad a_{4k} \\
y_1 & \\
y_2 & \\
\end{align*}$
Example: Input Parameters of Case 1

Input parameters

<table>
<thead>
<tr>
<th>$V_{TH}$ state</th>
<th>Page type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Erase</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>U</td>
</tr>
<tr>
<td></td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>L</td>
</tr>
<tr>
<td>$x_1$</td>
<td>0</td>
</tr>
<tr>
<td>$x_2$</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$x_8$</td>
<td>0</td>
</tr>
<tr>
<td>$x_9$</td>
<td>0</td>
</tr>
<tr>
<td>$x_{10}$</td>
<td>0</td>
</tr>
<tr>
<td>$x_{11}$</td>
<td>1</td>
</tr>
</tbody>
</table>

Ex. $V_{TH}$ state: “Erase”  $\Rightarrow x_1 = 1$
Page type: “Lower”  $\Rightarrow x_{11} = 1$

Upper 1WL

TLC NAND flash memory

Read 1WL

$V_{TH}$ state:

Upper 1 0 1 1
Middle 0 0 0 1
Lower 0 0 1 0

Erase

Predicted BER

x_1 : 1
x_2 : 0
x_3 : 0
x_{11} : 1
Predicted BER of Case 2

- Case 2 reproduces lateral charge migration successfully
  - Charge-trap, $N_{W/E} = 500$, Data-retention time = 2.8 days, @85degC

- More precise BER is obtained to consider analog $V_{TH}$

![Graph showing BER of different cases](image-url)
Predicted BER of Case 3

- In Case 3, analog $V_{TH}$ value within each $V_{TH}$ state is included.

Charge-trap, $N_{W/E} = 500$, Data-retention time = 18 days

- BER is most precisely predicted with extra reads.
Case 1-3 Decoding Result

- Acceptable data-retention time increases by 76-times

Charge-trap, $N_{W/E} = 500$, @85degC
Code rate = 9/10, Max iteration = 30

- Case 1
- Case 2
- Case 3

<table>
<thead>
<tr>
<th>Input parameters</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{TH}$ state</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Page type</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Neighboring cell data</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>WL number</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Read offset level</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
Outline

- Introduction
- Proposed Artificial Neural Network Coupled (ANN) LDPC ECC (ANN-LDPC ECC)
  - Training of ANN
  - Case 1-3 (1 hidden layer ANN)
  - Case 4 and 5 (2 hidden layer ANN)
- Conclusion
Case 4 and 5 consider both static and dynamic information

- Case 4 and 5 (2 Hidden Layer ANN)

<table>
<thead>
<tr>
<th>Input parameters</th>
<th>Case 4</th>
<th>Case 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{TH}$ state</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Page type</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Neighboring cell data</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>WL number</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Data-retention time</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Write/erase cycles</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
Predicted BER of Case 4

- Predicted BER fits well with measured BER at various data-retention time and endurance

Charge-trap, @85degC

![Graph showing BER vs data-retention time for different $N_{W/E}$ values]

Measured BER (answer data):
- $N_{W/E} = 300$
- $N_{W/E} = 500$

Predicted BER:
- $N_{W/E} = 300$
- $N_{W/E} = 500$

<table>
<thead>
<tr>
<th>Case 4 Input parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{TH}$ state</td>
</tr>
<tr>
<td>Page type</td>
</tr>
<tr>
<td>Data-retention time</td>
</tr>
<tr>
<td>Write/erase cycles</td>
</tr>
</tbody>
</table>
Case 4 and 5 Decoding Result

- Acceptable data-retention time increases by 8-times

<table>
<thead>
<tr>
<th>Input parameters</th>
<th>Case 4</th>
<th>Case 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{TH}$ state</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Page type</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Neighboring cell data</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>WL number</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Data-retention time</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Write/erase cycles</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
Case 4 and 5 Decoding Result

- ANN corrects errors in both charge-trap and floating-gate 3D-NAND

Floating-gate, $N_{\text{W/E}} = 300$, @85degC, Code rate = 9/10, Max iteration = 30

- ANN corrects errors in both charge-trap and floating-gate 3D-NAND

<table>
<thead>
<tr>
<th>Input Parameters</th>
<th>Case 4</th>
<th>Case 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{\text{TH}}$ state</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Page type</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Neighboring cell data</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>WL number</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Data-retention time</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Write/erase cycles</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
Outline

- Introduction
- Proposed Artificial Neural Network Coupled (ANN) LDPC ECC (ANN-LDPC ECC)
  - Training of ANN
  - Case 1-3 (1 hidden layer ANN)
  - Case 4 and 5 (2 hidden layer ANN)
- Conclusion
## Conclusion

<table>
<thead>
<tr>
<th>Charge-trap 3D-TLC NAND flash</th>
<th>$N_{W/E} = 500$</th>
<th>$N_{W/E} = 400$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case 1</strong></td>
<td>7V&lt;sub&gt;REF&lt;/sub&gt; sensing (1WL)</td>
<td>21V&lt;sub&gt;REF&lt;/sub&gt; sensing (3WL)</td>
</tr>
<tr>
<td><strong>Case 2</strong></td>
<td>21V&lt;sub&gt;REF&lt;/sub&gt; sensing (3WL)</td>
<td>77V&lt;sub&gt;REF&lt;/sub&gt; sensing (3WL)</td>
</tr>
<tr>
<td><strong>Case 3</strong></td>
<td>7V&lt;sub&gt;REF&lt;/sub&gt; sensing (1WL)</td>
<td>21V&lt;sub&gt;REF&lt;/sub&gt; sensing (3WL)</td>
</tr>
<tr>
<td><strong>Case 4</strong></td>
<td>7V&lt;sub&gt;REF&lt;/sub&gt; sensing (1WL)</td>
<td>21V&lt;sub&gt;REF&lt;/sub&gt; sensing (3WL)</td>
</tr>
<tr>
<td><strong>Case 5</strong></td>
<td>7V&lt;sub&gt;REF&lt;/sub&gt; sensing (1WL)</td>
<td>21V&lt;sub&gt;REF&lt;/sub&gt; sensing (3WL)</td>
</tr>
</tbody>
</table>

- **Read operation**
  - 7V<sub>REF</sub> sensing (1WL)
  - 21V<sub>REF</sub> sensing (3WL)
  - 77V<sub>REF</sub> sensing (3WL)
  - 7V<sub>REF</sub> sensing (1WL)
  - 21V<sub>REF</sub> sensing (3WL)

- **Acceptable data-retention time compared with BCH ECC**
  - 2.1x
  - 5.3x
  - Over 76x
  - 4.6x
  - 8.0x

- **Weight table size**
  - 22MB
  - 928MB
  - 1.1GB
  - 14KB
  - 490KB

- Case 5 achieves high reliability with small weight tables
- Case 3 is applied if reliability is seriously degraded
Thank you for your attention

This work was supported by JST CREST Grant Number JPMJCR1532, Japan