Energy Efficient Data Encoding in DRAM channels exploiting Data Value Similarity

Hoseok Seol, Wongyu Shin, Jaemin Jang, Jungwhan Choi, Jinwoong Suh, Lee-Sup Kim

Department of Electrical Engineering
Outline

1. Introduction
2. BD-Encoding
3. Evaluation Results
4. Conclusion
Modern DRAM Interface

- **DRAM off-chip data bus** consumes significant energy.
- Data Bus Energy: Switching + Termination (dominant)
- Modern DRAMs introduce asymmetric termination.
  - **Pseudo Open Drain (POD)**: DDR4, GDDR4/5
  - **Low Voltage Swing Terminated Logic (LVSTL)**: LPDDR4
Hamming Weight & Interface Energy

- **Hamming Weight**: number of 1’s in a string of bits.

- Decreasing Hamming Weight reduces both the termination and switching energy.

- We propose novel *data encoding* to reduce data bus energy.

Ex) LVSTL interface

Data: “11101010”

- Hamming Weight: 5
- Switching Activity: 6

Data: “00000010”

- Hamming Weight: 1
- Switching Activity: 2
Bitwise Difference (BD) Encoding

- **Observation:** Similar data words are sent over the DRAM data bus.

- **Key Idea:** Transfer the bit-wise difference between a current data word and the most similar data words.

  ![Diagram of Bitwise Difference Encoding]

- **Energy Reduction:** 58.3% of termination and 45.3% of switching energy.
Outline

1. Introduction
2. BD-Encoding
3. Evaluation Results
4. Conclusion
Motivation

> Energy dissipated in DDR4 data bus:

Termination (14.1%) + Switching Activity (7%)
## Observation: Data Value Similarity

<table>
<thead>
<tr>
<th>Transfer</th>
<th>libquantum</th>
<th>mcf</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>38 ad b3 00 18 83 24 00</td>
<td>18 67 df aa aa 2a 00 00</td>
</tr>
<tr>
<td>2</td>
<td>58 ad b3 00 18 83 24 00</td>
<td>01 00 00 00 00 00 00 00</td>
</tr>
<tr>
<td>3</td>
<td>78 ad b3 00 18 83 24 00</td>
<td>98 53 b8 aa aa 2a 00 00</td>
</tr>
<tr>
<td>4</td>
<td>98 ad b3 00 18 83 24 00</td>
<td>08 63 b8 aa aa 2a 00 00</td>
</tr>
<tr>
<td>5</td>
<td>a8 ad b3 00 18 83 24 00</td>
<td>00 00 00 00 00 00 00 00</td>
</tr>
<tr>
<td>6</td>
<td>c8 ad b3 00 18 83 24 00</td>
<td>00 27 bd aa aa 2a 00 00</td>
</tr>
</tbody>
</table>

> Observation: Similar data words are sent over the DRAM data bus.
Observation: Data Value Similarity

- All the workloads in SPEC 2006 have Data Value Similarity.
- The probability of the similar data occurrence (with recent 64 data words) is 72% in SPEC 2006 workloads.
Bitwise Difference Coder

- Recent data is stored in both tables in Encoder / Decoder
- When transfer data, search the **most similar data word**.
- If similar data exists, transfer 1) **bitwise difference**, 2) index NO.
  If not, transfer the original data.
Example of BD-encoding

BD-encoding (xor data)
0 0 0 0 0 0 1 0

Data
11101010

XOR

Off Chip Data Bus

XOR

Data
11101010

W/O encoding
1 1 1 0 1 0 1 0

Recent Values
1 0 0 0 1 1 0 0 1 1 1 0 1 0 0 0

Recent Values
1 0 0 0 1 1 0 0 1 1 1 0 1 0 0 0

<table>
<thead>
<tr>
<th>W/O encoding</th>
<th>BD-encoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hamming Weight</td>
<td>5</td>
</tr>
<tr>
<td>Switching Activity</td>
<td>6</td>
</tr>
</tbody>
</table>
Hardware Overheads

- **Coder** (data table 64 entries)
  - Area: 0.044% of commodity DDR4
  - Latency: 2.3ns (Transmitter), 0.7ns (Receiver)
  - Energy: 7pJ (Transmitter), 2pJ (Receiver)
  - Designed by 65nm logic process

- **Index Line**
  - a single extra line per 8 data lines.
  - can be shared with DBI / DM pins in DDR4.
Outline

1. Introduction
2. BD-Encoding
3. Evaluation Results
4. Conclusion
Methodology

<table>
<thead>
<tr>
<th>Component</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>Gem5, X86, 3.3GHz</td>
</tr>
<tr>
<td>Caches</td>
<td>L1 I-cache : 32KB, 4way</td>
</tr>
<tr>
<td></td>
<td>L1 D-cache : 64KB, 4way</td>
</tr>
<tr>
<td></td>
<td>L2 cache : 2MB, 8way</td>
</tr>
<tr>
<td>DRAM</td>
<td>DDR4-2133, 8GB</td>
</tr>
<tr>
<td>Interface</td>
<td>Pseudo Open Drain (DDR4)</td>
</tr>
<tr>
<td></td>
<td>Termination Energy Calculation:</td>
</tr>
<tr>
<td></td>
<td>Micron DDR4 Power Calculator</td>
</tr>
<tr>
<td></td>
<td>Switching Energy Calculation:</td>
</tr>
<tr>
<td></td>
<td>$E = CV^2$</td>
</tr>
<tr>
<td></td>
<td>Channel capacitance: 15 [pF]</td>
</tr>
<tr>
<td>Workloads</td>
<td>SPEC CPU 2006</td>
</tr>
</tbody>
</table>
Comparison Points

➢ Data Bus Inversion [M.stan, TVLSI ‘95]
   ⇒ Transfer inverted data if the hamming weight of inverted one is smaller.
   ⇒ Adopted in the commodity DRAMs (GDDR4/5, DDR4, LPDDR4)

➢ Power Protocol [K.Basu, MICRO’02], Frequent Value Encoding
   [J.Yang, ISLPED’01]
   ⇒ Transfer the table index instead of data when current data is the same as data transferred recently.

➢ Variable Length Value Encoder [D.suresh, ICCD’05]
   ⇒ Transfer the table index instead of data when current data is partly matched with data transferred recently.
Hamming Weight Reduction

BD-Encoding decreases the hamming weights in all workloads (the least effect in bzip: 29%)

The results increase as the number of table entries increases (28-58% for 1-64 table entries)
Comparison to Prior Works

BD-encoding reduces 58.3% of the termination and 45.3% of the switching energy.

The probability for similar data occurrence is much higher than that for the same data ⇒ BD-encoding shows better results than Power Protocol and VALVE.

**< Energy Reduction Rate >**

<table>
<thead>
<tr>
<th></th>
<th>DBI</th>
<th>PP_64</th>
<th>VALVE_64</th>
<th>Proposed Work_64</th>
<th>DBI</th>
<th>PP_64</th>
<th>VALVE_64</th>
<th>Proposed Work_64</th>
</tr>
</thead>
<tbody>
<tr>
<td>Termination Energy</td>
<td>12.4</td>
<td>25.5</td>
<td>34.8</td>
<td>58.3</td>
<td>10.9</td>
<td>20.7</td>
<td>25.8</td>
<td>45.3</td>
</tr>
<tr>
<td>Switching Energy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DBI: Data Bus Inversion
PP: Power Protocol
VALVE: Variable Length Value Encoder
Interface Energy Reduction

BD-encoding reduces overall interface energy including coder hardware energy (24-47.6% for 1-64 entries)

Optimal number of entries exists (32ea) due to overhead of index line and coder hardware.
Outline

1. Introduction
2. BD-Encoding
3. Evaluation Results
4. Conclusion
**Conclusion**

- **Reducing hamming weight** decreases both the termination and switching energy.

- Data Value Similarity: **Similar data words** are sent over the DRAM data bus.

- Bitwise Different Encoding: Transfer the bit-wise difference between a **current data word** and the most similar data word recently transferred.

- Evaluation Results: Reduce **58.3% of termination** and **45.3% of switching energy**.