Understanding and Leveraging the Spatial Variation in Read Disturbance Vulnerability of Real DRAM Chips

A. Giray Yaşlıkçız Onur Mutlu
PhD Candidate Advisor

ETHzürich SAFARI
Executive Summary

Motivation: Read Disturbance is a worsening DRAM reliability/security problem

Problem: Existing mitigation mechanisms suffer from significant performance and energy overheads, limited effectiveness, or prohibitively high cost

Goal:
- To understand the spatial variation in read disturbance across DRAM rows
- To leverage this understanding to improve the existing mitigation mechanisms

Experimental study: 136 DDR4 DRAM chips from three major vendors
- Characterize all rows in a bank and a bank from each bank group
- A large variation in the necessary activation count to induce the first bitflip
- No strong correlation between a row’s spatial features and its vulnerability

SVÄRD: Dynamically adapts the aggressiveness of a RowHammer defense
- Implemented in either the DRAM chip or the memory controller
- Reduces the performance overheads by 2.4x, 1.6x, 2.7%, and 3.0x for BlockHammer, PARA, Hydra, and RRS

SAFARI
DRAM Organization
DRAM Operations

1. **ACTIVATE (ACT):** Fetch the row’s content into the **row buffer**

2. **Column Access (RD/WR):** Read/Write the target column and drive to I/O

3. **PRECHARGE (PRE):** Prepare the array for a new ACTIVATE
Repeatedly **opening** and **closing** a DRAM row or keeping a DRAM row **open** causes **bitflips** in nearby cells.
DRAM Read Disturbance and Increasing DRAM Chip Density

**DRAM read disturbance worsens as DRAM chip density increases**

- More than 10X reduction

**Preventive actions need to be performed more aggressively as DRAM chip density increases**
Our Goal

To understand the spatial variation in read disturbance across DRAM rows

To leverage this understanding to improve the existing mitigation mechanisms
DRAM Testing Infrastructure

FPGA-based SoftMC (Xilinx Virtex UltraScale+ XCU200)

Xilinx Alveo U200 FPGA Board (programmed with SoftMC*)

DRAM Module with Heaters

Fine-grained control over DRAM commands, timing parameters (±1.5ns), and temperature (±0.1°C)

Key Takeaways

DRAM read disturbance vulnerability significantly varies across DRAM rows.

No strong correlation is observed between a row’s spatial features & read disturbance vulnerability.
Distribution of $\text{HC}_{\text{first}}$ across DRAM Rows

Minimum Hammer Count to Induce the First Bitflip ($\text{HC}_{\text{first}}$)

- Weakest rows experience bitflips at hammer counts of $4\text{K}$
- Many rows do not experience bitflips at hammer counts below $24\text{K}$
Weakest rows experience bitflips at hammer counts of 4K or 8K

Many rows do not experience bitflips at hammer counts below 24K

Different rows may experience read disturbance bitflips at very different hammer counts
Effect of RowPress on the HC\textsubscript{\textit{first}} Distribution

RowPress reduces the mean of the distribution with increased t\textsubscript{AggOn}

RowPress can reduce the variation within the HC\textsubscript{\textit{first}} distribution
Effect of RowPress on the HC_{first} Distribution

There is a **significant variation** across rows under the effect of **RowPress**

DRAM read disturbance vulnerability **significantly varies** across DRAM rows **under the effect of RowPress**
Correlation Analysis

Correlation between a DRAM row’s read disturbance vulnerability

• bank address bits
• subarray address bits
• row address bits
• row’s distance to local row buffer

A small fraction of DRAM chips (28.5%) contain spatial features that provide >0.7 F1-score for predicting HC$_{\text{first}}$.

No strong correlation is observed between a row’s spatial features & read disturbance vulnerability.
Svärd: Spatial Variation Aware Read Disturbance Mitigation

- **Dynamically adapts** the aggressiveness of a mitigation mechanism based on the victim row’s **vulnerability level**

- Classifies DRAM rows into **several vulnerability-level bins**

- Maintains **a few bits** (e.g., four bits) for each DRAM row **within**
  - the memory controller in an SRAM array
  - the parity bits in DRAM
  - the DRAM row itself (e.g., 4 bits for each 8KB capacity)

- Svärd is implemented **nearby the mitigation mechanism**
  - the memory controller
  - or the DRAM chip
Performance Evaluation

• Cycle-level simulations using **Ramulator 2.0** [Luo+, TCAD 2023]

**System Configuration:**

<table>
<thead>
<tr>
<th>Component</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor</td>
<td>3.2 GHz, 8 core, 4-wide issue, 128-entry instr. window</td>
</tr>
<tr>
<td>Last-Level Cache</td>
<td>64-byte cache line, 8-way set-associative, 8 MB</td>
</tr>
<tr>
<td>Memory Scheduler</td>
<td>FR-FCFS</td>
</tr>
<tr>
<td>Address Mapping</td>
<td>Minimalistic Open Pages</td>
</tr>
<tr>
<td>Main Memory</td>
<td>DDR4, 4 bank group, 4 banks per bank group (16 banks per rank)</td>
</tr>
</tbody>
</table>

**Workloads:** 50 different **8-core** multiprogrammed workloads from **SPEC CPU2006, SPEC CPU2017, TPC, MediaBench, and YCSB** benchmark suites

• Paired with **BlockHammer, PARA, Hydra, and RRS**

**HC\textsuperscript{first}**: \{4K, 2K, 1K, 512, 256, 128, 64\} hammers

The minimum hammer count needed to induce the first bitflip

SAFARI
Implications on Future Solutions

Reduction in Performance Overhead

Weighted Speedup (Norm. to No Mitigation Baseline)

Minimum Hammer Count to Induce the First Bitflip
Implications on Future Solutions

Minimum Hammer Count to Induce the First Bitflip

Svärd reduces the performance overhead of mitigation mechanisms
Conclusion

**Experimental study:** 136 DDR4 DRAM chips from 3 major vendors

- A **large variation** in the necessary activation count to induce the first bitflip
- No strong correlation between a row’s **spatial features** and its **vulnerability** to read disturbance

**SVÄRD:** Dynamically adapts the aggressiveness of mitigations

- Implemented in either the DRAM chip or the memory controller
- Reduces the performance overheads by **2.4x, 2.7%, 1.6x, and 3.0x** for BlockHammer, Hydra, PARA, and RRS

**Future Work:**

- A deeper understanding is needed to account for irregularities in row and column addresses across chips
- Finding correlations is essential to reduce the hardware cost
- Reducing also the hardware cost of defenses

*SAFARI*
Understanding and Leveraging the Spatial Variation in Read Disturbance Vulnerability of Real DRAM Chips

A. Giray Yağlıkçı
PhD Candidate

Onur Mutlu
Advisor

ETH Zürich

SAFARI
Understanding and Leveraging the Spatial Variation in Read Disturbance Vulnerability of Real DRAM Chips

A. Giray Yağlıkçı
PhD Candidate

Onur Mutlu
Advisor
Variation in Bit Error Rate

Relative Location of DRAM Row (0 and 1 are the two edges of a DRAM bank)
$\text{HC}_{\text{first}}$ Across Rows

![Graphs showing the relationship between $\text{HC}_{\text{first}}$ and the relative location of DRAM rows for three manufacturers (H, M, S).](image)
# Tested DRAM Chips

<table>
<thead>
<tr>
<th>Mfr.</th>
<th>DIMMs</th>
<th>Chips</th>
<th>Density</th>
<th>Die Rev.</th>
<th>Org.</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mfr. H</td>
<td>1</td>
<td>8</td>
<td>16Gb</td>
<td>A</td>
<td>x8</td>
<td>2051</td>
</tr>
<tr>
<td>(SK Hynix)</td>
<td>3</td>
<td>8</td>
<td>16Gb</td>
<td>C</td>
<td>x8</td>
<td>2048</td>
</tr>
<tr>
<td>Mfr. M</td>
<td>1</td>
<td>16</td>
<td>8Gb</td>
<td>B</td>
<td>x4</td>
<td>N/A</td>
</tr>
<tr>
<td>(Micron)</td>
<td>2</td>
<td>4</td>
<td>16Gb</td>
<td>B</td>
<td>x16</td>
<td>2126</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>16</td>
<td>16Gb</td>
<td>E</td>
<td>x4</td>
<td>2014</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>16Gb</td>
<td>E</td>
<td>x16</td>
<td>2046</td>
</tr>
<tr>
<td>Mfr. S</td>
<td>2</td>
<td>8</td>
<td>8Gb</td>
<td>B</td>
<td>x8</td>
<td>2053</td>
</tr>
<tr>
<td>(Samsung)</td>
<td>1</td>
<td>16</td>
<td>8Gb</td>
<td>C</td>
<td>x4</td>
<td>2135</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>8</td>
<td>8Gb</td>
<td>D</td>
<td>x8</td>
<td>2110</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>8</td>
<td>4Gb</td>
<td>F</td>
<td>x8</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Figure 3: Overview of Hydra. Hydra splits tracking into two parts (a) aggregated tracking using GCT (b) per-row tracking using RCT (cached in RCC).

Understanding and Leveraging the Spatial Variation in Read Disturbance Vulnerability of Real DRAM Chips

BACKUP SLIDES

A. Giray Yağılıkçı
PhD Candidate

Onur Mutlu
Advisor