

Rubix: Reducing the Overhead of Secure Rowhammer Mitigations via Randomized Line-to-Row Mapping

Anish Saxena

anish.saxena@cc.gatech.edu
Georgia Institute of Technology
Atlanta, USA

Saurav Mathur

smathur44@gatech.edu
Georgia Institute of Technology
Atlanta, USA

Moinuddin Qureshi

moin@gatech.edu
Georgia Institute of Technology
Atlanta, USA

Abstract

Modern systems mitigate Rowhammer using *victim refresh*, which refreshes neighbours of an aggressor row when it encounters a specified number of activations. Unfortunately, complex attack patterns like Half-Double break victim-refresh, rendering current systems vulnerable. Instead, recently proposed secure Rowhammer mitigations perform mitigative action on the aggressor rather than the victims. Such schemes employ mitigative actions such as *row-migration* or *access-control* and include AQUA, SRS, and Blockhammer. While these schemes incur only modest slowdowns at Rowhammer thresholds of few thousand, they incur prohibitive slowdowns (15%-600%) for lower thresholds that are likely in the near future. The goal of our paper is to make secure Rowhammer mitigations practical at such low thresholds.

Our paper provides the key insights that benign application encounter thousands of *hot rows* (receiving more activations than the threshold) due to the memory mapping, which places spatially proximate lines in the same row to maximize row-buffer hitrate. Unfortunately, this causes row to receive activations for many frequently used lines. We propose *Rubix*, which breaks the spatial correlation in the line-to-row mapping by using an encrypted address to access the memory, reducing the likelihood of hot rows by 2 to 3 orders of magnitude. To aid row-buffer hits, *Rubix* randomizes a group of 1-4 lines. We also propose *Rubix-D*, which dynamically changes the line-to-row mapping. *Rubix-D* minimizes hot-rows and makes it much harder for an adversary to learn the spatial neighbourhood of a row. *Rubix* reduces the slowdown of AQUA (from 15% to 1%), SRS (from 60% to 2%), and Blockhammer (from 600% to 3%) while incurring a storage of less than 1 Kilobyte.

CCS Concepts: • Security and privacy → Systems security; Hardware security implementation.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

ASPLOS '24, April 27-May 1, 2024, La Jolla, CA, USA

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0385-0/24/04

<https://doi.org/10.1145/3620665.3640404>

Keywords: DRAM, Rowhammer, Memory Mapping

ACM Reference Format:

Anish Saxena, Saurav Mathur, and Moinuddin Qureshi. 2024. Rubix: Reducing the Overhead of Secure Rowhammer Mitigations via Randomized Line-to-Row Mapping. In *29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24)*, April 27-May 1, 2024, La Jolla, CA, USA. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3620665.3640404>

1 Introduction

Rowhammer is a data-disturbance error where frequently activating a row induces bit flips in nearby rows [24]. Rowhammer is a severe security threat and has been used to leak confidential data and escalate privilege [3, 6–9, 13, 27, 27, 44, 48]. Rowhammer worsens with higher memory density. The number of activations required to induce bit-flips, termed as the *Rowhammer Threshold* (T_{RH}), has plummeted from 139K (DDR3) in 2014 to just 4.8K in 2020 (LPDDR4), as shown in Figure 1 (a). The threshold is expected to reduce even further, and if the current trend continues (30X reduction in 6 years), we can expect T_{RH} of about 100 over the next decade. Solutions that protect against Rowhammer must be viable not just at the current threshold, but also at future thresholds.

Rowhammer defenses typically incorporate a *tracking mechanism* to count row activations and a *mitigative action* to perform when the activation count reaches the threshold. The most popular form of mitigative action is *victim refresh*, which simply refreshes the nearby *victim rows* when the *aggressor row* reaches the specified threshold of activations. Victim refresh has been deployed in commercial systems (e.g. DDR4, DDR5) in the form of Target Row Refresh (TRR) [7]. However, the drastic reduction of T_{RH} poses two problems. First, due to severe area limitation in DRAM (9% area required for per-row tracking [11]), TRR is unable to identify all aggressors, even in DDR5 [11]. In fact, two recent whitepapers from JEDEC [15, 16] mention that “in-DRAM mitigations cannot eliminate all forms of Rowhammer attacks”. Second, even if tracking is perfect, the act of victim-refresh can itself be used to induce bit-flips. As shown in Figure 1 (b), Half-Double [2] leverages victim-refresh to cause bit-flips at a distance-of-2 from the aggressor row, thereby breaking all defenses relying on victim-refresh. Thus, current systems remain vulnerable to Rowhammer. In this paper, we focus on mitigations resilient to complex patterns.

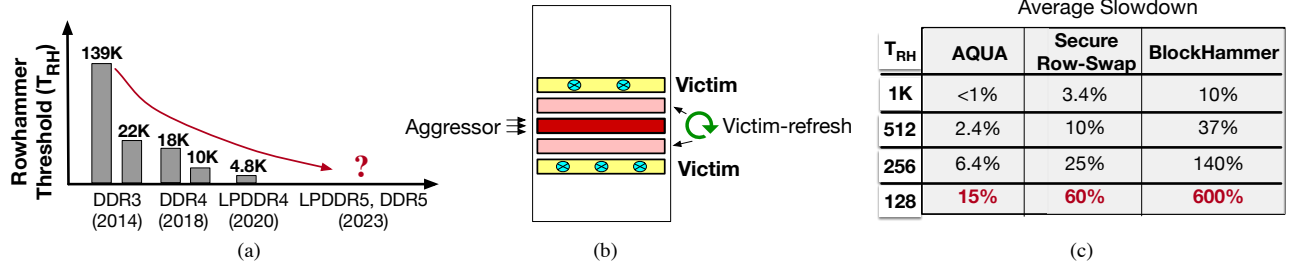


Figure 1. (a) Trend of Rowhammer threshold (30x lower in 6 years). (b) Half-Double breaks victim-refresh. (c) Secure Rowhammer mitigations that are resilient against Half-Double, incur impractical slowdown at low thresholds (T_{RH} of 128).

Recent studies [43, 53, 54] propose such resilient mitigations that are aggressor-focused instead of being victim-focused (such as victim-refresh). AQUA [43] and SRS [53] migrate the aggressor row to another row, breaking the spatial correlation between the aggressor and victim. Blockhammer [54] limits the number of activations to any row to less than T_{RH} , preventing large number of activations to a row (critical for complex attacks). While these schemes invoke high-overhead mitigating actions that take several microseconds or more, at current T_{RH} only a small number of rows require any mitigation, and these schemes incur a modest slowdown. Unfortunately, as T_{RH} reduces, many more rows reach the threshold, requiring mitigations, which causes significantly higher overheads. Figure 1 (c) shows the average slowdown of AQUA, SRS, and Blockhammer as the threshold reduces from 1K to 128. At a threshold of 128, AQUA suffers a slowdown of 15%, SRS of 60%, and Blockhammer of 600%, rendering them impractical at lower thresholds.

The goal of our paper is to make secure Rowhammer mitigations practical even at a low threshold (128), as such thresholds can occur if the trend holds for the next decade.

High slowdown occurs due to the dramatic increase in number of rows that receive more than T_{RH} activations in 64ms, which we define as *hot rows*. In our evaluations, we observe only about 200 hot-rows, on average, with 512 or more activations, but 9500 hot-rows with 64 or more activations (45X more). Reducing the number of hot-rows would reduce the slowdown stemming from secure mitigations.

We make the key observation that hot rows are primarily caused by the memory mapping function, which determines the set of lines co-residing within the same row. The memory-mapping in modern processors places lines with spatial proximity in the same row to maximize row-buffer hits. For example, Intel Coffee Lake [49] mapping places the entire 4KB page within the same row and Intel Skylake [49] round-robins the lines of each 4KB page between rows of two banks. Thus, 32-64 lines of each 4KB page co-reside within the same row, and if the page is heavily accessed, these lines would contribute to the aggregate activation count of the row. While each line incurs only a few activations, the sum of activations due to all the lines makes the row a hot-row.

Typical workloads access only a small fraction of the memory within 64ms. In our evaluations, less than 5% of rows are touched within 64ms. Thus, spreading activations from hot-rows to the entire memory would greatly reduce hot-rows. With this insight, we propose *Rubix*, a memory mapping that breaks spatial correlation of lines to rows by using an encrypted address to access memory. We present two flavors of Rubix: Static (*Rubix-S*) and Dynamic (*Rubix-D*).

Rubix-S uses the low-latency programmable bit-width K-Cipher [26] for address-space randomization, which is kept in the memory controller. On a memory access, the memory controller encrypts the line-address, accessing the memory with the encrypted line address. Encryption randomizes the line-to-row mapping, so the lines co-resident in the same row have no spatial correlation. This significantly reduces the likelihood of heavily accessed lines getting placed in the same row, virtually eliminating all the hot-rows. As a result, mitigations are invoked much less, reducing slowdown. To preserve row buffer hit rate, Rubix-S encrypts a *gang* of 1-4 contiguous lines, as line-level address encryption would result in virtually zero row buffer hits. Our evaluations show that at T_{RH} of 128, Rubix-S reduces the slowdown of AQUA (from 15% to 1%), SRS (from 60% to 3%), and Blockhammer (from 600% to 3%) while requiring just 16 bytes of storage, thereby making it practical to deploy secure mitigations.

With Rubix-S, the group of lines that co-reside in the row are randomized, however, this group remains unchanged throughout the system uptime. To this end, Rubix-D provides dynamic randomization of line-to-row mapping without needing a programmable cipher, by using an xor operation with a randomly generated key. The mapping changes gradually from the current-key to the next-key. Rubix-D remaps vertically (gangs in the same position of different rows) instead of horizontally (gangs within the row), which not only reduces hot-rows, but also makes it much harder for an adversary to determine set of spatially contiguous rows, a critical step in launching a targeted Rowhammer attack. Our evaluations show that at T_{RH} of 128, Rubix-D reduces the slowdown of AQUA (from 15% to 1.5%), SRS (from 60% to 2%), and Blockhammer (from 600% to 3%) while incurring a storage overhead of less than 1 KB.

Overall, our paper makes the following contributions:

1. To the best of our knowledge, this is the first paper to analyze the impact of memory (line-to-row) mapping on the efficacy of Rowhammer mitigations.
2. We demonstrate that the line-to-row mapping is the primary reason for hot-rows in benign workloads.
3. We propose Rubix-S, which breaks the spatial correlation in line-to-row mapping by accessing the memory with an encrypted address (with gangs of 1-4 lines).
4. We propose Rubix-D that provides dynamic randomization without needing a programmable cipher and makes it harder to identify spatially proximate rows.

2 Background and Motivation

2.1 Threat Model

We assume an unprivileged attacker that can run code on the system vulnerable to Rowhammer. The attacker can run a process under user privilege and exploit Rowhammer to flip bits in critical data structures (such as page-table) or in the data of another program. We assume the Rowhammer bit-flip occurs at the victim location when any row in memory incurs more activations than T_{RH} within the refresh interval of 64ms. Thus, the attack is successful if no mitigation is issued when a row has encountered more than T_{RH} activations.

2.2 Background on DRAM

Modern DRAM-based memory is organized into several banks, each of which is a two-dimensional array of DRAM cells, organized as rows and columns. Each bank caches the most recently opened row in a *row buffer*. Data is accessed by bringing it into the row buffer. To access data in another row, the bank clears the row buffer, followed by activation of the given row. DRAM cells leak charge and require periodic refresh operations (at 64ms). T_{RC} determines the time between consecutive activations for a given bank and is about 45ns.

2.3 Memory Mapping

The memory-mapping function routes a given line address to a particular bank and row, determining the set of lines that co-reside in a row [10, 49]. It also affects row-buffer hit-rate and performance. Memory systems place spatially proximate lines in the same row and we consider two mappings used in Intel systems. While not exhaustive, we note that most deployed mappings use similar xor-based hashing [49].

Coffee Lake Mapping: This mapping places consecutive 128 lines within the same row buffer. So, two consecutive 4KB pages would be resident in the same row. It uses a xor-based hashed mapping for bank selection.

Skylake Mapping: This mapping alternatively places a pair of lines between two banks (selected using xor). So, for a 4KB page, lines 0,1,4,5 ... 60, 61 reside in a row of one bank,

and lines 2,3,6,7 ... 62,63 are in row of the other bank. This mapping causes 32 lines from a 4KB page to reside in a row, with contents of four consecutive pages in the same row.

2.4 Rowhammer

Rowhammer is a data-disturbance error [17, 24] where activating a row frequently induces bit-flips in nearby rows. The *Rowhammer Threshold* (T_{RH}) denotes minimum activations required on a row to induce bit-flips with any access pattern. Rowhammer is a severe security threat as the attacker can flip bits in the page table and take over the system. When Rowhammer was characterized in 2014, T_{RH} was 139K with single-sided attack, whereas it reduced by 30x to 4.8K [19] in 2020 (with double-sided attack). As established by a decade of threshold characterization [20], ultra-low thresholds will be reached by the next decade and as memory gets denser, more nearby rows experience aggressor activations [28].

We can avoid ultra-low thresholds if DRAM organization changes fundamentally or DRAM vendors mitigate Rowhammer. Unfortunately, neither option has materialized, as stated by JEDEC [15, 16] and recent industry papers [11, 23]. Moreover, as systems remain deployed for several years, Rowhammer defenses must scale to future T_{RH} .

Hardware-based defenses for Rowhammer have two parts: activation-tracker and mitigating-action. Several studies [4, 22, 29, 36, 37, 46, 47, 56] have looked at storage-efficient trackers. Comparatively, mitigating action is less well studied. Most modern systems simply rely on victim-refresh, which is vulnerable to address-correlation attacks [25]. Thus, modern systems continue to be vulnerable to Rowhammer attacks.

2.5 Secure Rowhammer Mitigation

Performing mitigative action on aggressor row, instead of the victim, prevents address-correlation attacks. Figure 2 shows three such recently proposed *aggressor-focused mitigation*.

Blockhammer [54] controls the access rate of frequently accessed rows, such that no row incurs more than T_{RH} activations within 64ms, by delaying accesses for an appropriate time. As the adversary cannot perform an overwhelmingly large number of activations (required for Half-Double) on a single row, it prevents complex attacks.

AQUA [43] migrates the aggressor row when it receives $T_{RH}/2$ activations (halving of threshold due to tracker reset), to a *quarantine-region* in memory. AQUA breaks the spatial connection between aggressor and victim, limiting the time for an attacker to craft complex attacks.

Secure Row-Swap (SRS) [53] swaps the aggressor row, once it has received $T_{RH}/3$ activations (reduction due to birthday-paradox attacks), with another randomly selected row in memory. Like AQUA, SRS breaks the spatial connection between the aggressor and the victim.

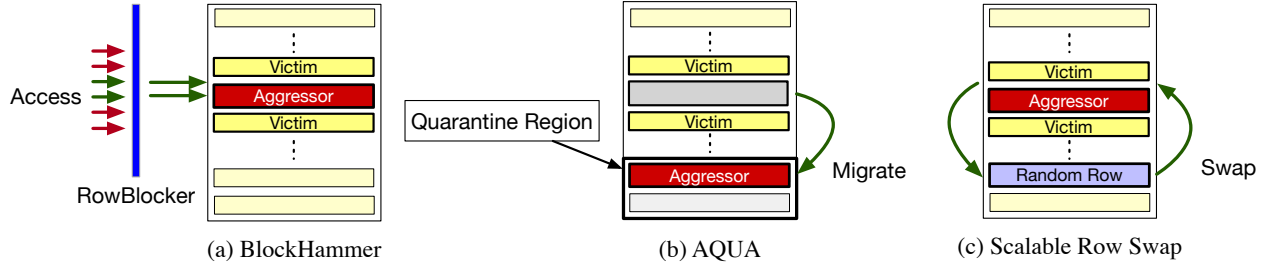


Figure 2. Secure Rowhammer Mitigations: (a) BlockHammer controls rate of accesses to each row. (b) AQUA quarantines aggressor rows in a dedicated region. (c) Scalable Row-Swap (SRS) swaps the aggressor row with a random row.

2.6 Scalability Problem of Secure Mitigations

Secure Rowhammer mitigations (such as Blockhammer, AQUA, and SRS) incur significantly more overhead than victim-refresh. While performing two victim-refresh activation takes less than 100 nanoseconds, these schemes incur significantly more latency. For example, row migration in AQUA and SRS ties up the memory bus for several microseconds, during which the channel cannot service any requests. The problem is even worse for Blockhammer, as rate control can delay accesses by several tens to hundreds of microseconds.

These schemes are designed for the current thresholds of few thousands, where very few rows reach the threshold in benign workloads and require mitigation. However, at lower thresholds, many more rows reach the threshold, which causes more frequent high-overhead mitigations, causing drastic slowdown. Figure 3 shows the performance of AQUA, SRS, and Blockhammer as thresholds ranging from 1K to 128, for the Coffee Lake and Skylake memory mappings, normalized to Coffee Lake-based baseline.

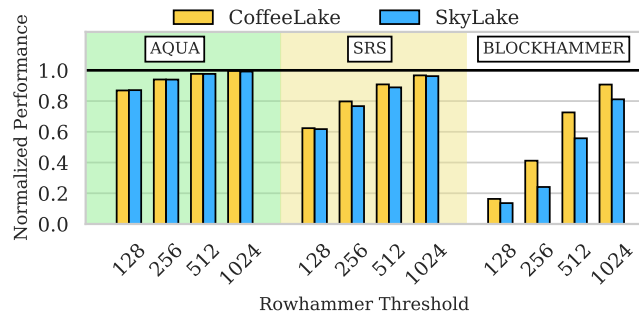


Figure 3. Normalized performance of AQUA, SRS, Blockhammer at varying T_{RH} . At T_{RH} of 128, all schemes incur significant slowdowns.

At the threshold of 1K, AQUA and SRS have negligible slowdown, whereas Blockhammer suffers 10% (Coffee Lake) to 25% (Skylake). However, at T_{RH} of 128, all schemes incur significant overheads. AQUA and SRS incur 15% and 60% slowdown, respectively, and Blockhammer has 500% to 600% slowdown (note that normalized IPC of 0.2 implies 5x slowdown), making secure mitigations impractical for adoption.

2.7 Goal of Our Paper

The goal of our paper is to make secure Rowhammer mitigations viable at low thresholds (T_{RH} of 128), which are likely to be present in the near future. We want to accomplish this without incurring significant hardware overheads. We develop a general framework that can be used by current and also by future secure Rowhammer mitigations.

3 Evaluation Methodology

3.1 System Configuration

We use the Gem5 [31] simulator to perform multi-core simulations in Syscall Emulation (SE) mode with an out-of-order core and DDR4 memory model. We use DDR4 2400MT/s based on Micron MT40A2G4 [12]. Table 1 shows our baseline system configuration. We use the open-adaptive memory page policy which keeps the row open for maximum of 16 accesses before closing it. Moreover, we use first-ready FCFS (FR-FCFS) scheduling policy to prioritize row hits and minimize unnecessary activations. We use the Coffee Lake mapping as our baseline. AQUA and SRS use the Misra-Gries [35] tracker and for Blockhammer, we use an idealized SRAM tracker with one counter per row in memory. Due to tracker state reset, we use a tracker threshold of $T_{RH}/2$.

Table 1. Baseline System Configuration

Out-of-Order Cores	4 cores, 8 wide at 3GHz
Last Level Cache (Shared)	8MB, 16-Way, 64B lines
Memory size	16 GB – DDR4 2400MT/s
t_{RCD} - t_{CL} - t_{RP} - t_{RC}	14.2-14.2-14.2-45 ns
Rows x Banks x Ranks x Channels	128K×16×1×1
Size of row	8KB

3.2 Workloads

We evaluate with 18 SPEC2017 [1] *rate* workloads and 16 mixed workloads (each with four random SPEC2017 workloads). We fast-forward 25 billion instructions and simulate for 250 million instructions. Table 2 shows the Misses Per 1K Instructions (MPKI) and the average number of unique rows touched, and "hot-rows" with 64 or more activations (ACT-64+) and with 512 or more activations (ACT-512+).

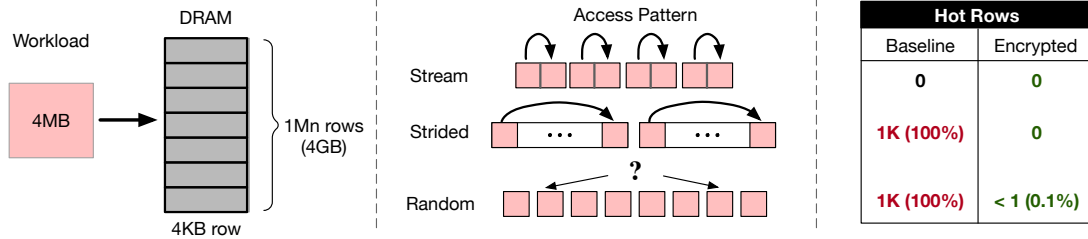


Figure 4. Illustration: Understanding the impact of memory-mapping in generating hot-rows (a) System configuration (b) Workloads (c) Number of hot-rows for 4MB footprint (and 4KB rows). Under baseline mapping, both stride-64 and random have 1K hot rows (100%), however, with an encrypted address virtually all the hot-rows are eliminated.

Table 2. Workloads Characteristics: MPKI, Unique Rows Touched (within 64ms), and Hot-Rows (within 64ms).

Workload	MPKI (LLC)	Unique Rows Activated	Total number of "Hot-Rows"	
			ACT-64+	ACT-512+
blender	12.78	8.8K	347K	2.9K
lbm	20.87	29.4K	70.3K	0
gcc	6.12	10.4K	21.8K	384
cactuBSSN	2.57	5.2K	12.2K	0
mcf	5.81	4.9K	10.5K	425
roms	3.33	27.9K	6.6K	9
perlbench	0.71	11.4K	1.7K	0
xz	0.40	10.8K	496	0
nab	0.53	4.4K	189	0
namd	0.37	3.4K	105	0
imagick	0.13	1.1K	89	0
bwaves	0.21	1.7K	20	0
wrf	0.02	702	20	0
exchange2	0.01	122	14	0
deepsjeng	0.25	68.1K	12	0
povray	0.01	390	8	0
parest	0.10	2.4K	3	0
leela	0.02	879	0	0
Average	3.01	10.7K	9528	206

4 A Case for Randomized Memory

The reason secure Rowhammer mitigations incur significant overheads at low thresholds is because more rows reach the threshold number of activations (we refer to such rows as *hot-rows*). We identify the root cause of hot-rows to be the memory mapping function that determines the line-to-row mapping. In this section, we first present this insight, then our workload characterization, then our solution *Rubix*, and results for slowdown and mitigation.

4.1 Dependence of "Hot Rows" on Mapping

We illustrate the dependence of hot-rows on line-to-row mapping using a simple model, as shown in Figure 4 (a). The processor accesses a memory system containing one bank. The memory system is 4GB and contains 1 million rows of 4KB each. We use sequential mapping that places the 4KB page within the same row.

We consider three kernels as shown in Figure 4 (b): *stream*, *stride-64*, and *random*, with a 4MB footprint. Each kernel

incurs 1 million memory accesses within 64ms. We deem a row to be a hot-row if it has at least 64 activations. We analyze the number of hot-rows for these kernels.

For the *stream* kernel, the first access causes an activation, and subsequent 63 accesses get a row-buffer hit. Therefore, a million memory accesses cause a total of only 15.6K activations, which get spread over the 1K rows, with a uniform activation rate of about 16 activations per row, with no hot-rows. The *stride-64* kernel has a stride of 64 lines and each access goes to a different page. When all pages are exhausted, the stride continues with the next line on the page. As each memory access causes an activation, this kernel incurs 1 million activations, spread equally over 1K pages, and each row gets 1K activations. Thus, all the 1K rows are hot-rows. The *random* kernel accesses a random line in memory. The likelihood of a row buffer hit is negligibly small, so the 1 million accesses cause 1 million activations, spread over 1K rows. The average activations per row are 1000 (standard deviation of 32), with more than 99% of the rows having more than 900 activations. Thus, we deem all the 1K rows to be hot-rows. The results are summarized in Figure 4 (c).

The conventional mapping of placing sequential lines in the same row buffer causes hot-rows for both the *stride* pattern and the *random* pattern. We have 64 lines that cause activation of the same row in memory, thus compounding the total number of activations incurred by the given row.

Consider an alternative mapping that uses an encrypted line-address to access the memory system. There are 64K lines in a 4MB footprint. These 64K lines would be spread over the 1 million rows in memory. We estimate (using binomial distribution) that 61.5K rows have exactly 1 line from the kernel mapped to them, 1.9K rows with 2 lines, and 40 rows with 3 lines (no row with 4 or more lines). For both *stream* and *stride*, each line gets accessed 16 times. So, we have 61.5K rows with 16 activations, 1.9K rows with 32 activations, and 40 rows with 48 activations. Thus, no row is deemed a hot-row. For *random*, we estimate the expected number of hot-rows to be 0.4, so less than 1 row will be deemed a hot-row. Thus, randomizing the line-to-row mapping eliminates the hot rows of all three kernels.

4.2 Characterizing Lines in Hot-Rows

For our baseline system, we examine how many lines (out of the 128 lines) of the row contribute to making the row a hot-row. For each row that reaches 64 activations, we measure the number of lines in the row that encountered at least 1 activation. Table 3 shows the percentage of hot-rows that had 1-8 lines, 8-16 lines, 32-64 lines, and 64-128 lines (and the average) contributing to the row activation counts.

Table 3. Number of lines that add to activation counts of hot-rows (data for workloads with 100+ hot-rows).

Workload	Number of Activating Lines in a Hot-Row			
	1-32	32-64	64-128	Average
blender	2%	98%	0	60
lbm	0	100%	0	58
gcc	1%	99%	0	60
cactuBSSN	0	100%	0	63
mcf	0	100%	0	52
roms	3%	97%	0	51
perlbench	7.3%	92%	0	47
xz	0	100%	0	57
nab	0	100%	0	58
namd	0	100%	0	54
Average	2%	98%	0	56

We observe that for 98% of hot-row activations come from at-least 32 lines in the row. On average, 56 out of 128 lines incur at least one-activations within the hot-row. This validates our hypothesis that hot-rows occur because many lines of the row contribute to the activation counts. Thus, the line-to-row mapping which decides which set of lines co-reside within the same row is the main reason for the occurrence of hot-rows.

4.3 Rubix: Randomized Line-to-Row Mapping

Rubix breaks the spatial correlation of lines to row by using an encrypted address to access memory. Figure 5 shows an overview of the static version of Rubix, called Rubix-S. Consider the access pattern where requests for four consecutive lines A, B, C, D are set to memory. In conventional mapping, these four lines will co-reside within the same row. However, with encryption, these lines get scattered to different rows.

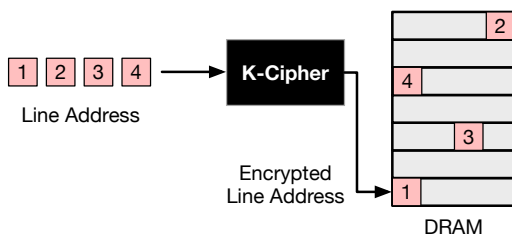


Figure 5. Overview of Rubix-S: breaking spatial correlation in line-to-row mapping with encrypted line-address.

Rubix-S uses K-Cipher [26], a low-latency programmable bit-width cipher, for address-space randomization. K-Cipher is kept in the memory controller and incurs a latency of 3 cycles (with 10nm process technology [26]). On a memory access, it encrypts the line-address which is used to access the memory. As we have 16GB memory, we use a 28-bit cipher. Encryption randomizes the line-to-row mapping, breaking the spatial correlation between lines co-residing in the row.

The exact line-to-row mapping depends on the 96-bit key of the K-Cipher. The key is set to a random value (based on PRNG) at boot time. As each system will have a different key, the memory mapping for each system will be different.

4.4 Recouping Row-Buffer Hits via Gangs

While line-address encryption virtually eliminates hot-rows, it degrades the row-buffer hit-rate to approximately zero. Rubix minimizes hot-rows while still retaining some row-buffer hits, by encrypting a gang of 2-4 contiguous lines. Figure 6 shows Rubix-S with gang-level randomization.

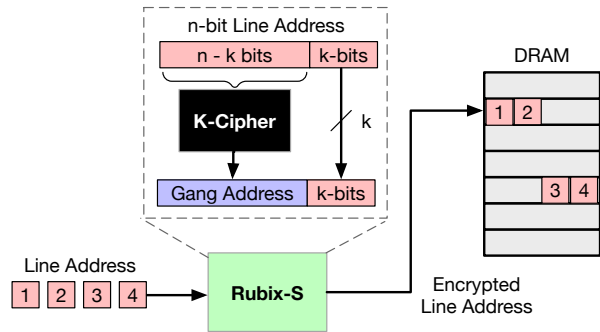


Figure 6. Rubix-S: Using gang-address encryption to balance both row-buffer hits and reduced hot-rows .

Instead of encrypting the entire n-bit line-address, Rubix-S skips the k least significant bits of the line-address and only encrypts the gang-address, which is the remaining (n-k) bits. The encrypted gang-address is concatenated with the unmodified k-bits, and this line-address is used to access the memory. Thus, lines within a gang that co-reside in a row provide temporal locality to aid row-buffer hits. For example, in Figure 6, lines 1 and 2 co-reside in the same row. Note that with k-bits, we would have a gang-size of 2^k lines and if k is set to zero, this design degenerates into Rubix-S with line-address encryption. We denote Rubix-S with a gang-size of X lines as Rubix-S (GSX). The size of the cipher is adjusted per gang size, so Rubix-S (GS4) uses a 26-bit K-cipher.

4.5 Results: Impact on Mitigations

We observe that Rubix-S (GS1, line-level) eliminates all hot-rows for our workloads. Figure 7 shows the number of hot-rows for the baseline system with Coffee Lake mapping, Skylake mapping, and Rubix-S (GS4). Rubix-S eliminates hot-rows for all but six workloads. On average, Coffee Lake

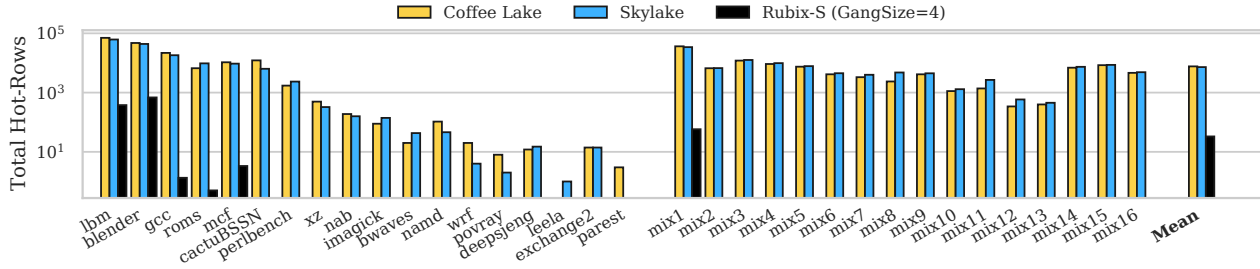


Figure 7. Number of hot-rows (activations of 64 or more) with Intel mappings and Rubix-S with Gang-Size of 4 (GS4). Mean implies arithmetic mean. While baselines have more than 7K hot rows on average, Rubix-S (GS4) reduces it by 220x to 33.

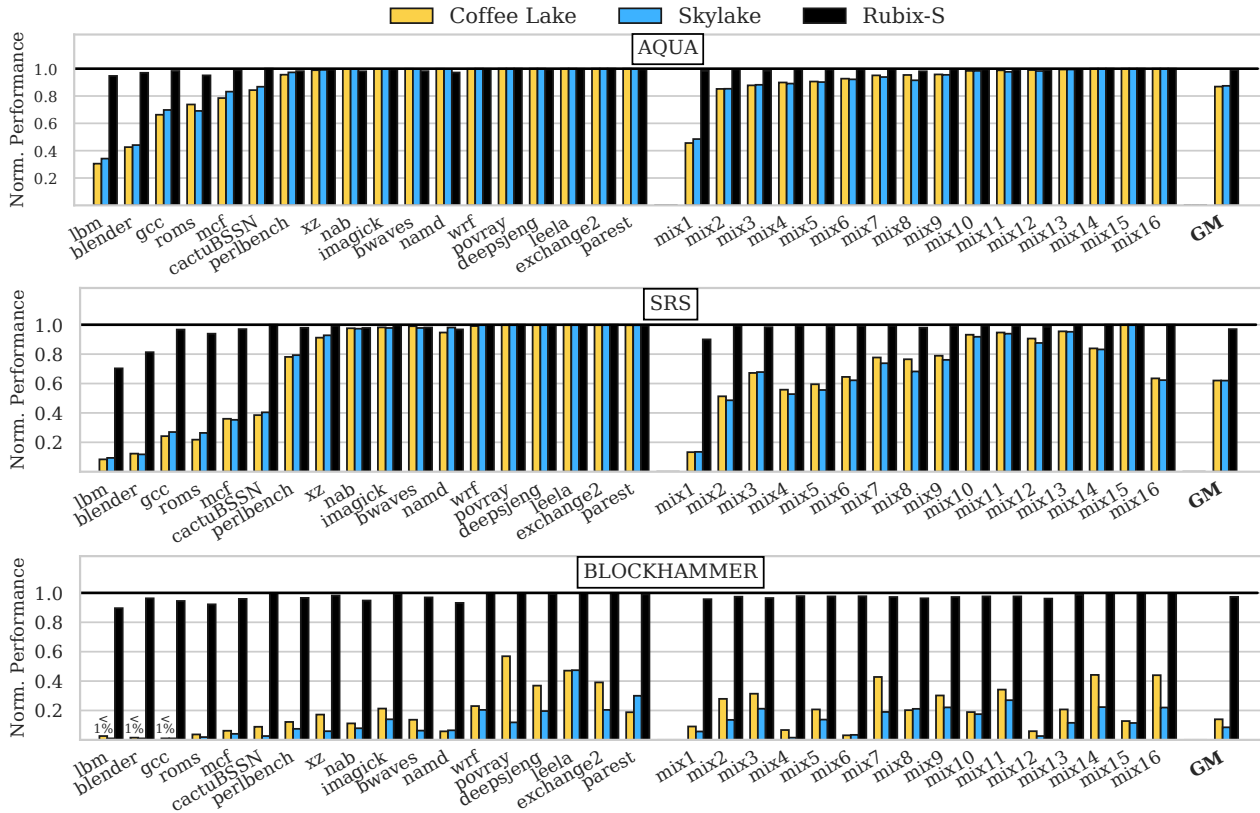


Figure 8. Performance of secure mitigations at T_{RH} of 128 for Intel mappings and Rubix-S, normalized to an unprotected Coffee Lake baseline. Rubix-S uses GS4 for AQUA and SRS, and GS1 for Blockhammer, and reduces the average slowdown to 1.1%, 3.1%, and 2.9%, respectively (down from 15%, 60%, and 600%), making them viable at ultra-low thresholds.

and Skylake mappings have 7.6K and 7.2K hot-rows respectively, whereas Rubix-S (GS4) reduces it by 220x to only 33. Line-to-row mapping primarily determines hot-rows, and our design significantly reduces hot-rows. Mitigations are invoked much less, greatly reducing performance overheads.

4.6 Results: Impact on Performance

Figure 8 shows performance of AQUA, SRS, and Blockhammer with Intel Coffee Lake, Skylake, and Rubix-S mappings. Performance is normalized to an unprotected Coffee Lake

baseline. Intel mappings incur unacceptable slowdown with secure mitigations. We compare Rubix-S with Coffee Lake mapping which performs slightly better than Skylake. Coffee Lake incurs a significant average slowdown of 15% for AQUA, while Rubix-S reduces it to a negligible 1% (for gang-size 4). SRS and BlockHammer are impractical with baseline policies, incurring 60% and 600% average slowdown, respectively. Rubix-S not only enables SRS and BlockHammer with a negligible average slowdown of 3.1% (GS 4) and 2.8% (GS 1), respectively, it retains application-level performance with

a worst-case slowdown of 42% for lbm with SRS and just 11% for BlockHammer, 28X and 350X improvement.

Overall, Rubix-S makes secure mitigations viable at ultra-low T_{RH} of 128 with just 2-3% overhead. While we do not change access scheduling and DRAM page policies, fine-tuning them would likely reduce the overheads even further.

4.7 Sensitivity: Varying Gang-Size

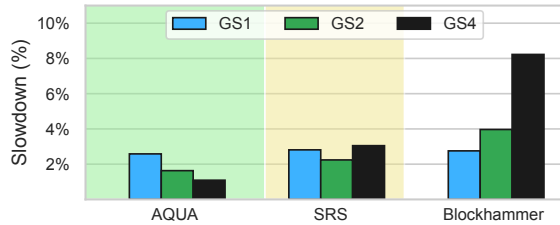


Figure 9. Performance of Rubix-S with Gang-Size of 1-4.

Gang-size (GS) balances row-buffer hits and reduction in hot-rows. With larger GS, row-buffer hit rate increases along with hot-rows and mitigation overheads. Figure 9 shows the performance of secure mitigations with Rubix-S as GS is varied from 1 to 4. Due to high mitigation overhead, Blockhammer works best with GS1 which eliminates hot-rows. AQUA has lower overhead mitigation, so GS4 works best which retains row-buffer hits. For SRS, GS2 offers the best balance between row-buffer hits and minimizing hot-rows. Thus, the best GS size depends on the scheme and the mitigation overhead and Rubix-S provides the flexible trade-off of minimizing hot-rows while retaining row-buffer locality.

4.8 Results: Impact on Row-Buffer Hits

A key effect of small gang-size is decreased row-buffer hit rate. The baseline Coffee Lake and Skylake policies provide an average row-buffer hit rate of 55% and 63%, respectively. Rubix-S shows a gradual increase in row-buffer hit-rate from 0 with GS1, to 19% at GS2 to 31% at GS4, with up-to 2.7X more activations for GS1. Thus, GS2 and GS4 recoup some of the row-buffer hits. The overall system performance depends not only on row-buffer hits but also on mitigation overheads.

4.9 Results: Storage and Power Overheads

Rubix requires negligible power for the K-Cipher and address mapping logic. We use Micron’s power calculator [34] to compute DRAM power, the primary overhead due to lower row-buffer hit rate. Rubix-S increases the DRAM power by 120mW at a gang-size of 4 (a 4.3% increase), and by 300mW at gang-size of 1 (10.6% increase), due to a lower row-buffer hit rate than baseline that result in additional activations. The power consumption of Rubix-S with secure mitigations remains within 10% of the baseline, because of virtually eliminating mitigations, unlike existing memory mappings, which incur prohibitive energy overheads.

4.10 Security Analysis of Rubix-S

The security of Rubix-S stems from the security of the underlying mitigation schemes (SRS, AQUA, Blockhammer). The security guarantees of these schemes are not dependent on using a specific memory-mapping. Rubix-S remains secure because we simply change the memory mapping.

4.10.1 Defining T_{RH} . We define T_{RH} as the minimum number of activations to **at least** one row within 64ms which causes a bit flip via any attack pattern (single-sided, double-sided, Half-Double[25], or a future attack pattern). So to ensure security of our solution, our only assumption is:

A successful Rowhammer attack requires activating **at least** one row more than T_{RH} times within a refresh period.

4.10.2 Security of SRS, AQUA, and BlockHammer. SRS and AQUA rely on row migration to guarantee that now row receives more than T_{RH} activations within a 64ms window. SRS does so by randomization, guaranteeing that even under continuous attacks for several years, the likelihood of randomly finding migrated rows is negligibly small. With AQUA, a row that receives T_{RH} activations is moved to a quarantine region, and by design, it guarantees that no physical row will ever receive more than T_{RH} activations. BlockHammer controls the activation rates to a physical row such that no row ever receives more than T_{RH} activations. These schemes rely on accurate tracking of row counts, and we use Misra-Gries tracker for SRS and AQUA, and one-counter-per-row for BlockHammer, which provide guaranteed tracking. The security guarantees of SRS, AQUA, Blockhammer are applicable for all access patterns (including Half Double) and all possible memory mapping (the mapping of lines to rows).

4.10.3 Proving Security of Rubix-S Using Lemmas. Rubix-S reduces performance overheads while retaining the security guarantees of the underlying SRS, AQUA, and Blockhammer schemes. The underlying schemes (SRS, AQUA, and Blockhammer) are secure against all access patterns (including Half Double), and these guarantees work for any memory mapping. Using the randomized memory mapping of Rubix-S retains these guarantees, as we show using lemmas.

Lemma-1: *The security guarantee of SRS, AQUA, and Blockhammer is not dependent on memory mapping, so these designs are secure for all memory mappings.*

Lemma-2: *Rubix-S is a memory mapping which randomizes the line-to-row mapping.*

From Lemma-1 and Lemma-2, it follows that secure mitigations continue to be secure with Rubix-S. For example, Half Double requires that an aggressor row be activated about 100x more times than T_{RH} . As no row is activated T_{RH} times in secure mitigations, their security is unaffected by Rubix-S.

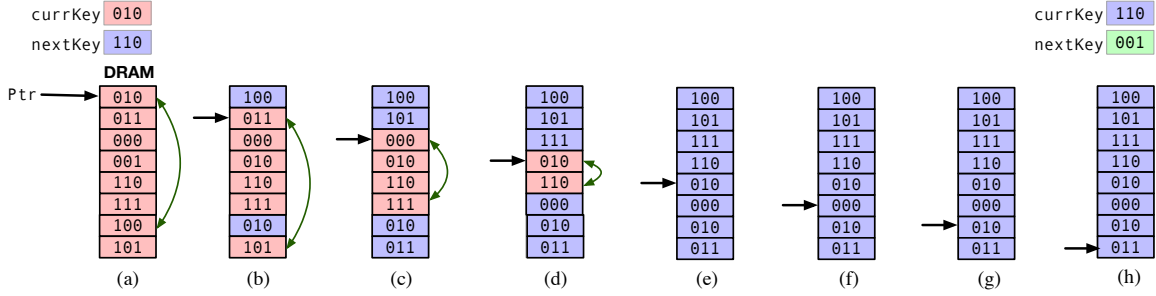


Figure 10. An example of dynamically changing xor-based mapping. The effective address is the line-addressed xor-ed with a key. The dynamic remapping algorithm gradually remaps all the lines from a currKey (010) to the nextKey (110).

5 Rubix-D: Dynamic Randomization

With Rubix-S, lines co-residing in the row are randomized. However, this mapping remains unchanged throughout the system uptime. We propose *Rubix-D*, an alternative approach that not only randomizes the line-to-row mapping but changes this mapping dynamically at system runtime. Rubix-D reduces hot-rows and makes it much harder to determine rows that are spatially contiguous to each other (a critical step in a targeted Rowhammer attack).

We adapt the ideas presented in seminal works on dynamic memory remapping [40, 45] to suit our constraints and objectives. Rubix-D performs xor with a randomly generated key to randomize [45], and this mapping is gradually changed from a given key to new key. In this section, we first provide an example of dynamically changing xor mapping, then present Rubix-D, and finally the results.

5.1 Overview of Xor-Based Remapping

Figure 10 provides an example of the xor-based dynamic remapping for a memory containing 8 lines (000-111). The system contains a pointer (Ptr) to aid with remapping and two sets of keys currKey and nextKey. The effective line-address is computed as the xor operation with one of the keys. At the start of the epoch, all lines use the currKey whereas by the end of the epoch, all lines use the nextKey. We perform remapping every 100 accesses. Figure 10 (a) shows the mapping at the start of the epoch with all lines located at their original address xor-ed with the currKey (010). After 100 accesses, the first remapping is invoked, so the physical location 000 (pointed by the Ptr) is swapped with the destination 110 (Ptr xor-ed with the nextKey). Ptr is incremented to 001.

The next three remappings also result in swaps (Figure 10 (b), (c), and (d)) and the pointer is incremented. For the next four remapping episodes, swapping is skipped as the Ptr points to an already remapped line. After 8 episodes, all lines use the mapping with nextKey (Figure 10(h)). Next, the currKey becomes currkey xor-ed with nextKey, and the nextKey is initialized to a new value obtained using a hardware-based PRNG. The Ptr is reset to 000, indicating a new epoch.

We translate line-address to physical-address in two steps:

- (1) Translate line-address L to $L' = (L \text{ xor } \text{currKey})$.
- (2) Perform two checks: First, is $L' < \text{Ptr}$? and Second, is $(L' \text{ xor } \text{nextKey}) < \text{Ptr}$?. If either is yes, $L' = (L' \text{ xor } \text{nextKey})$.

The memory access is routed to location L' . The simple xor and check operations are performed within one cycle. Thus, xor-based dynamic remapping randomizes line-to-row addresses with negligible SRAM (three registers – currKey, nextKey, and Ptr) and latency (one cycle). For properties and proof of xor-based randomization, please refer to [45].

5.2 Pitfall of Xor at Randomizing Line-to-Row

While xor-based mapping dynamically randomizes memory addresses, we cannot directly apply it in our context, due to the linear mapping of xor. For example, if there are 128 lines co-residing in a row, then after an xor with a random key, these 128 lines still co-reside in one row (at another location). As all the top $(n-7)$ bits of the lines that get mapped to the same row are identical, an xor with the $(n-7)$ bits in the key results in the same remapped value. Reordering of lines within the destination row, unfortunately, does not reduce the likelihood of it becoming a hot-row. Instead, our proposal *Rubix-D* reorganizes the xor-based mapping to dynamically randomize the group of lines that co-reside in a row.

5.3 Overview of Rubix-D

Figure 11 shows an overview of Rubix-D. We randomize gangs vertically (across rows but for same gang-in-row). For G gangs in a row, we provision G sets of remapping circuits (currKey, nextKey, and Ptr). As each gang in the row uses a different key, gangs co-residing in the same row in the baseline are scattered to different rows in memory, breaking the spatial correlation between gang mapping to a row.

In Figure 11, the memory has 4 gangs in a row (colored red, yellow, etc.). The same-colored gangs across all rows form a *vertical-group* (*v-group*). Each v-group is allocated a pair of keys (curr and next) and a pointer. The line-address is split into three parts: the least significant k bits identify the line-in-gang, next p bits identify the gang-in-row, and remaining $n - p - k$ bits identify the row-address.

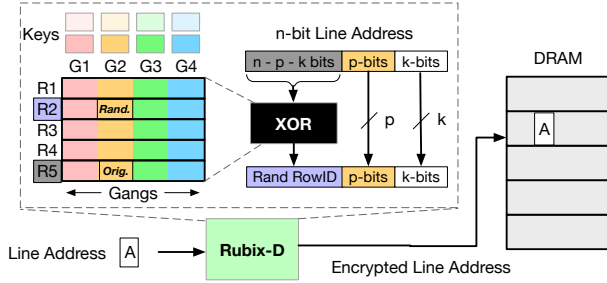


Figure 11. Overview of Rubix-D: gangs within a vertical group (G1, G2, etc.) are routed to random rows in memory.

Rubix-D keeps the $k+p$ bits of the line address unchanged, randomizing only the bits for (global) row address. The p bits identify the v -group and its keys and pointer translate the row-address to the remapped-row-address. The remapped-row-address is concatenated with the $k+p$ bits to form the remapped-line-address, which is used to access the memory. With a 28-bit line address, Rubix-D with gang-size of 4 uses 2 bits to identify line-in-gang, the next 5 bits for gang-in-row, and remaining 21 bits for global row address. With less than 8 bytes for each pair of keys and ptr, we need total SRAM of 512 bytes (for 32 v -groups).

5.4 Remapping Rate and Remapping Period

Remapping-Rate (RR) determines the frequency of remapping. We set RR to occur with 1% probability on each activation (thereby avoiding ACT counters for v -groups). V -gangs with more activations are remapped more frequently. During remap, the gangs pointed by the Ptr of the v -group are swapped with their destination (based on nextKey). At GS4, the memory controller streams 4 lines from source and destination rows and swaps them (open-row-X, read-DataX, open-row-Y, read-DataY, write-DataX-to-Y, open-row-X, write-DataY-to-X). Swapping incurs 3 ACTs, 8 CAS reads and 8 CAS writes, consuming bandwidth and energy. As half of the remap operations are skipped (Figure 10 (e)-(h)), at an RR of 1%, the average overhead is low at 1.5% extra activations.

Remapping Period (RP) is the time to remap the v -group. With RR=1% and two million rows in memory, a v -group has a remap-period of about 200 million activations. We can reduce the remapping-period by dividing such that every Nth row of the v -group to forms a v -segment. Each v -segment has its own set of keys and pointer. With $N=32$, the remapping-period of the v -segment is 6.25 million activations; however, this requires 16 KB SRAM overhead for metadata.

5.5 Security Analysis of Rubix-D

Even though Rubix-D remaps dynamically, it is not a standalone mitigation for Rowhammer, as an adversary can use Flush+Reload [55] to cause bit-flips. Thus, Rubix-D must always be used with a Rowhammer mitigation scheme. Rubix-D’s security stems from the underlying mitigation (AQUA/

SRS/ Blockhammer). As the security of these schemes is not dependent on line-to-row mapping, Rubix-D retains their security (please see Section 4.10). Thus, per Lemma-1 and the fact that Rubix-D is simply a memory mapping, the overall design (with AQUA,SRS, Blockhammer) of Rubix-D is secure against all access patterns, including Half-Double.

5.6 Impact of Rubix-D on Future Attacks

Complex attacks, such as Half-Double and BLASTER [28], attack multiple rows and identifying spatially contiguous rows is critical for success [5]. Once inferred, the mapping remains constant in Rubix-S until system reboot, whereas with Rubix-D the neighbor information gets changed within a few seconds due to remapping. Thus, with Rubix-D, not only do we get security for known attacks, it makes orchestrating future complex pattern attacks much harder.

5.7 Results: Storage and Power Overheads

Rubix-D needs 8-byte metadata (currKey, nextKey, Ptr) for each v -group, so 512 bytes for gang-size of 4 lines. For segmented Rubix-D, the storage overhead is proportional to the number of segments (e.g., 16KB SRAM for 32 segments). DRAM power, computed using Micron’s power calculator [34], increases by 130mW at GS4 (4.2% more than baseline), 180mW at GS2 (5.8% increase), and 320mW at GS1 (10.9% increase). Note that with baseline mappings, secure mitigations would incur significant energy overheads.

5.8 Results: Impact on Mitigations

Rubix-D reduces hot-rows within 64ms as shown in Figure 12, which plots hot-rows for conventional policies, Rubix-S, and Rubix-D (as GS is varied). The baseline policies each have more than 7K hot-rows. Rubix with GS1 eliminates hot-rows which GS2 incurs a negligible number of hot-rows, which increase to few tens with GS4. The reduction in hot-rows makes secure mitigations viable at T_{RH} of 128.

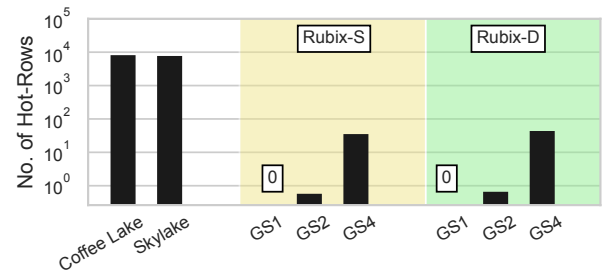


Figure 12. Hot-rows in baseline and Rubix (atleast 100x less).

5.9 Results: Impact on Performance

We evaluate Rubix-D with Remapping-Rate of 1% without any segments as they do not impact performance (they affect the Remapping-Period and storage overheads). Figure 13

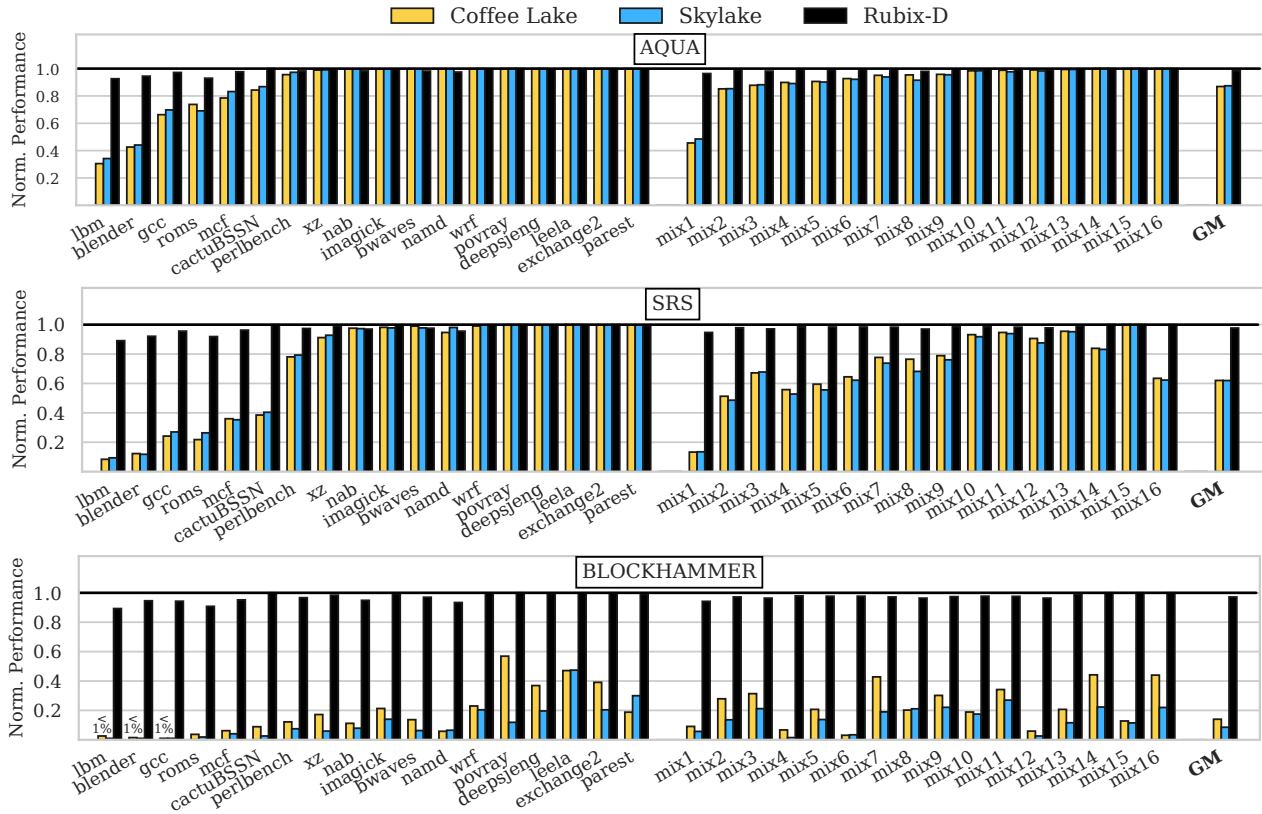


Figure 13. Performance of secure mitigations at T_{RH} of 128 with Intel mappings and Rubix-D, normalized to unprotected Coffee Lake baseline. With GS4 for AQUA, GS2 for SRS, and GS1 for BlockHammer, Rubix-D incurs a low average slowdown of 1.5%, 2.3%, and 2.8%, respectively (down from 15%, 60%, and 600%).

shows the performance of Rubix-D compared to Intel mappings, normalized to an unprotected Coffee Lake baseline. Rubix-D incurs low overhead of just 1-3% on average at T_{RH} of 128. AQUA, SRS, and BlockHammer perform best at different gang-sizes. AQUA launches almost no mitigations and benefits from row buffer locality at GS4. SRS operates at a lower threshold of $\frac{T_{RH}}{3}$ and launches more mitigations, performing best at GS2 with negligible hot rows. BlockHammer has high mitigation overhead and works best minimal hot-rows at GS1. Rubix-D incurs worst-case slowdown of just 10%, compared to more than 100X in baseline (for BlockHammer). The remapping of Rubix-D also avoids getting stuck with an accidentally bad mapping, as the mapping gets changed over program execution.

5.10 Sensitivity: Mapping Overhead of Rubix

Table 4 shows the isolated slowdown of Rubix mappings without any mitigative action. Rubix incurs low overhead of 1%-3% due to lower row-buffer hit rate than baseline mapping. Rubix-D overheads are slightly higher than Rubix-S due to extra activations required for dynamic remapping. As randomization incurs a small performance cost while minimizing episodes of hot rows, it avoids expensive mitigations.

Table 4. Isolated overhead of Rubix without mitigative action

Slowdown of Rubix Mapping	Rubix-S	Rubix-D
Gang of 4 lines (GS4)	1%	1.3%
Gang of 2 lines (GS2)	1.6%	1.9%
Gang of 1 line (GS1)	2.6%	2.7%

5.11 Sensitivity: Higher Rowhammer Thresholds

Figure 14 shows the slowdown of secure mitigations with Rubix-S and Rubix-D at higher T_{RH} . Randomizing the line-to-row mapping practically eliminates hot-rows at higher thresholds, even at gang-size of 4 employed by Rubix at T_{RH} of 1K for all secure mitigations. Consequently, the performance overhead is a negligible 1.1% to 1.4% at T_{RH} of 1K.

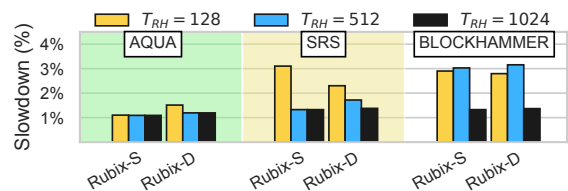


Figure 14. Slowdown of Rubix at Higher Thresholds. Rubix-GS4 incurs less than 2% slowdown at $T_{RH} = 1K$.

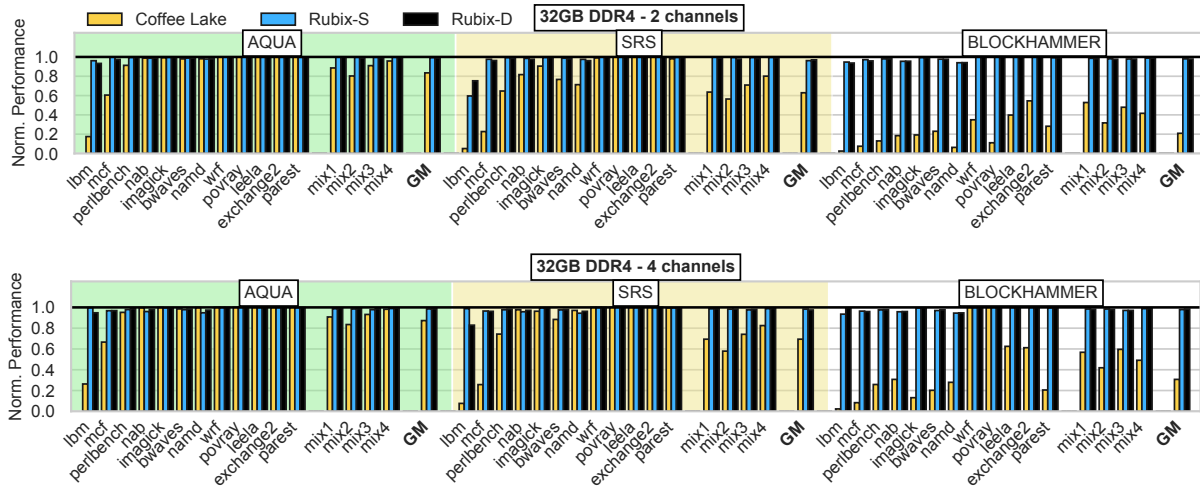


Figure 15. Normalized performance of secure mitigations with Intel and Rubix mappings for an 8-core multi-channel system. While Intel mappings incur impractical average overheads of 15%-380% (AQUA-BlockHammer), Rubix reduces it to 1%-4%.

5.12 Sensitivity: Scaled-up Multi-Channel Systems

We evaluate Intel Coffee Lake and Rubix mappings on a subset of workloads with 8-core simulations with 2 and 4 channels (32GB DDR4 memory and 16 MB LLC, other configuration same as Table 1). As Figure 15 shows, Intel’s mapping incurs impractical overheads of 15%, 45%, and 380% for AQUA, SRS, and BlockHammer (bottom graph), even though it stripes gangs of 4 lines across 4 channels, because contiguous lines end up in the same row in a strided pattern. Rubix breaks the spatial correlation of line-to-row, resulting in low overheads of just 1-3% (4% for 2-ch SRS with Rubix-S).

5.13 Sensitivity: Memory-Intensive Workloads

We evaluate Rubix and the baseline mappings with memory-intensive STREAM workloads [33] using 1 GiB arrays (LLC MPKI of more than 50). Figure 16 shows the performance of Rubix normalized to unprotected Coffeelake and Skylake mappings. Rubix eliminates hot-rows in all STREAM workloads. On average, Rubix incurs 2% to 5% slowdown compared to Coffeelake mapping (5% to 8% slowdown compared to Skylake mapping) due to lower row buffer hit rate (Rubix-D incurs more slowdown due to dynamic remapping). Overall, Rubix is low-cost even with memory-bound workloads.

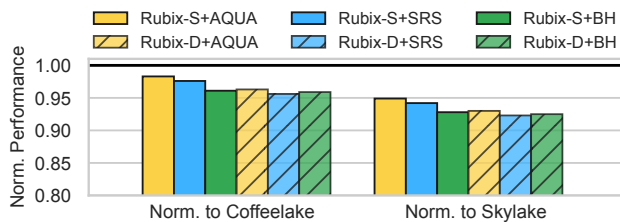


Figure 16. Rubix with secure mitigations incurs 2% to 8% average slowdown (geomean) for memory-intensive workloads compared to unprotected baseline memory mappings.

6 Discussion

In this section, we describe alternative designs that can reduce hot-rows without relying on a cipher or remapping.

6.1 Randomizing Line-to-Row without Cipher

Rubix-S breaks line-to-row spatial proximity via randomization. An alternative strategy to reduce hot rows is to use the most significant bits of the memory address for the gang-in-row, which strides gangs co-resident in a row. For example, 16GB memory and 32 gangs-per-row strides the gangs in the same row by 512MB. As lines that are much further away (512MB) from each other are unlikely to be accessed within a short time of each other, this mapping also reduces the line-to-row correlation without relying on a cipher. We also evaluated such a large-stride design and found that it has overheads similar to Rubix-S (1.8% to 3.8% slowdown with secure mitigations compared to unprotected Coffee Lake mapping). However, unlike Rubix-S, gang-level striding would not be robust against all access patterns, such as patterns with large strides, whereas, cipher-based randomization provides a principled solution for all patterns.

6.2 Randomizing Line-to-Row with Keyed XOR

Rubix-D assigns each gang-in-row to a separate remapping circuit to XOR-hash with its randomly generated key. If dynamic remapping is skipped, Rubix-D still retains static randomization while avoiding the performance and energy overheads of swapping gangs. In our evaluations, Rubix-D without dynamic remapping incurs an average slowdown of just 0.9%-2.6% with secure mitigations. The randomized mapping remains unchanged until system is rebooted (like Rubix-S). Note that static randomization virtually eliminates all hot-rows, and the additional benefit of dynamic randomization is to make targeted Rowhammer attacks difficult.

7 Related Works

7.1 Mapping of Memory Systems

Minimalist Open-Page (MOP) [18] balances performance and fairness by placing only four lines of a 4KB page in the same row. Unfortunately, as MOP round-robins across all banks, spatially proximate lines from consecutive pages co-reside in the same row, maintaining spatial correlation. We find hot-rows with MOP are similar to our baseline mapping. Figure 17 shows the normalized performance of secure mitigations with MOP, Rubix-S, and Rubix-D. We observe that MOP still suffers significant slowdowns, whereas Rubix virtually eliminates the hot-rows and the associated slowdown. Rather than hand-crafting a mapping, our work uses encryption for breaking the spatial correlation of lines.

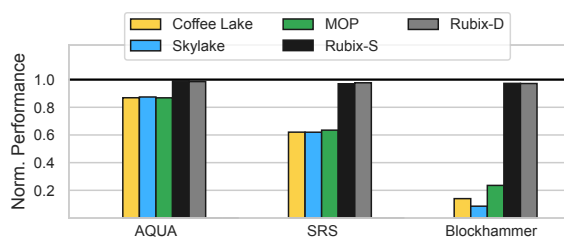


Figure 17. Performance of AQUA, SRS, and Blockhammer on MOP and Rubix. MOP suffers large slowdowns.

7.2 Randomization in Memory Systems

Randomization is a popular technique to improve the reliability and security of memory systems. For example, Start-Gap [40] and Security-Refresh [45] randomize mapping in non-volatile memories for wear-leveing. Cache randomization [30, 38, 39, 41, 50, 51] techniques randomize the line-to-set mapping to mitigate conflict-based cache attacks.

7.3 In-DRAM Rowhammer Mitigations

DRAM modules contain *Target Row Refresh (TRR)*, which tracks aggressors and refreshes victims. Recent attacks [7, 14], break TRR by exploiting its insufficient tracking capability to capture all possible aggressor rows. Samsung’s DSAC [11] and SK Hynix’s PAT [23] improve TRR for DDR5, but due to severe area limitation in DRAM, still allow aggressors to escape detection. DSAC has an escape probability of 13.9% between two mitigations and PAT fails 6.9% of the time (compared to DDR4-TRR). Two recent whitepapers from JEDEC [15, 16] mention that the deployed “in-DRAM mitigations cannot eliminate all forms of Rowhammer attacks”.

Even if all aggressors are tracked accurately, victim-refresh is still not secure as it preserves spatial proximity between aggressor and victims, enabling attacks such as Half-Double. Note that increasing the victim refreshed to two on each side does not solve Half-Double, as rows distance-of-three away can now incur bit flips. Instead, our solution Rubix

makes secure Rowhammer mitigations, which are resilient to complex attacks, practical at ultra-low thresholds, as shown in Table 5. Rubix is a memory mapping and is compatible with any tracking and mitigation mechanism.

Table 5. Comparison of Rowhammer Mitigations

Mitigation	Security	Slowdown
in-DRAM TRR	Not Secure	< 1%
AQUA	Secure – Isolation	15%
SRS	Secure – Randomization	60%
BlockHammer	Secure – Rate Control	600%
Rubix with AQUA/SRS/ BlockHammer	Secure – underlying mitigation	1% to 3%

Rubix can also greatly reduce the overheads of existing mitigations, which rely on victim refresh, by eliminating the root cause of overheads – hot-rows, thereby requiring much reduced number of mitigative actions. Moreover, all our evaluated secure mitigations (AQUA, SRS, and BlockHammer) work with commodity DRAM and DDR protocol, while in-DRAM mechanisms like Rega [32] typically require changes to DDR protocol and DRAM architecture (and significant energy overheads to mitigate low thresholds). Thus, such solutions are orthogonal to our work.

7.4 Randomization to Mitigate Rowhammer

Recent row migration proposals [42, 43, 52, 53] mitigate Rowhammer by moving an aggressor row to another row in memory. However, such row-to-row randomization does not change the set of lines that co-reside in the row Likewise, randomized DRAM address remapping [21] retains the set of lines coresident in the same memory row. Thus, unlike our solution, these schemes do not reduce the hot-rows.

8 Conclusion

Rowhammer gets worse as thresholds drop and attacks develop complex patterns that defeat the commonly used victim-refresh. Mitigations resilient to complex attacks, like AQUA, SRS, and Blockhammer, suffer from drastic slowdown at low thresholds due to many *hot-rows*. We identify the line-to-row mapping as the root cause of hot-rows, as it places spatially correlated lines in same row. Our proposal, Rubix, breaks this spatial correlation by randomizing the line-to-row mapping, reducing the number of hot rows by more than 100x. Rubix reduces overheads of the prior schemes by 10-100x, making them viable for practical adoption.

Acknowledgments

We thank Salman Qazi (Google) for feedback on an earlier draft of our paper. We also thank our shephard, Prof. Alaa Alameldeen, and the anonymous reviewers of MICRO-2023 and ASPLOS-2024 for their comments and feedback. This work was supported in part by a gift from Intel.

References

- [1] [n. d.]. SPEC CPU2017 Benchmark Suite. In *Standard Performance Evaluation Corporation*. <http://www.spec.org/cpu2017/>
- [2] [n. d.]. "Half-Double": Next-Row-Over Assisted RowHammer.
- [3] Zelalem Birhanu Aweke, Salessawi Ferede Yitbarek, Rui Qiao, Reetuparna Das, Matthew Hicks, Yossi Oren, and Todd Austin. 2016. ANVIL: Software-based protection against next-generation rowhammer attacks. *ACM SIGPLAN Notices* 51, 4 (2016), 743–755.
- [4] Tanj Bennett, Stefan Saroiu, Alec Wolman, and Lucian Cojocar. 2021. Panopticon: A Complete In-DRAM Rowhammer Mitigation. In *Workshop on DRAM Security (DRAMSec)*.
- [5] Lucian Cojocar, Jeremie Kim, Minesh Patel, Lillian Tsai, Stefan Saroiu, Alec Wolman, and Onur Mutlu. 2020. Are we susceptible to rowhammer? an end-to-end methodology for cloud providers. In *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 712–728.
- [6] Lucian Cojocar, Kaveh Razavi, Cristiano Giuffrida, and Herbert Bos. 2019. Exploiting correcting codes: On the effectiveness of ecc memory against rowhammer attacks. In *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 55–71.
- [7] Pietro Frigo, Emanuele Vannacc, Hasan Hassan, Victor Van Der Veen, Onur Mutlu, Cristiano Giuffrida, Herbert Bos, and Kaveh Razavi. 2020. TRRepass: Exploiting the many sides of target row refresh. In *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 747–762.
- [8] Daniel Gruss, Moritz Lipp, Michael Schwarz, Daniel Genkin, Jonas Juffinger, Sioli O'Connell, Wolfgang Schoecl, and Yuval Yarom. 2018. Another flip in the wall of rowhammer defenses. In *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 245–261.
- [9] Daniel Gruss, Clémentine Maurice, and Stefan Mangard. 2016. Rowhammer.js: A remote software-induced fault attack in javascript. In *International conference on detection of intrusions and malware, and vulnerability assessment*. Springer, 300–321.
- [10] Martin Heckel and Florian Adamsky. 2023. Reverse-Engineering Bank Addressing Functions on AMD CPUs. (2023).
- [11] Seungki Hong, Dongha Kim, Jaehyung Lee, Reum Oh, Changsik Yoo, Sangjoon Hwang, and Jooyoung Lee. 2023. DSAC: Low-Cost Rowhammer Mitigation Using In-DRAM Stochastic and Approximate Counting Algorithm. arXiv:2302.03591 [cs.CR]
- [12] Micron Technology Inc. 2015. DDR4 SDRAM Datasheet (MT40A2G4). (2015). https://www.micron.com/-/media/client/global/documents/products/data-sheet/dram/ddr4/8gb_ddr4_sdram.pdf
- [13] Yeongjin Jang, Jaehyuk Lee, Sangho Lee, and Taesoo Kim. 2017. SGX-Bomb: Locking down the processor via Rowhammer attack. In *Proceedings of the 2nd Workshop on System Software for Trusted Execution*. 1–6.
- [14] Patrick Jattke, Victor van der Veen, Pietro Frigo, Stijn Gunter, and Kaveh Razavi. 2022. BLACKSMITH: Rowhammering in the Frequency Domain. In *43rd IEEE Symposium on Security and Privacy'22 (Oakland)*. https://comsec.ethz.ch/wp-content/files/blacksmith_sp22.pdf.
- [15] JEDEC. 2021. Near-Term DRAM Level Rowhammer Mitigation (JEP300-1). (2021).
- [16] JEDEC. 2021. System Level Rowhammer Mitigation (JEP301-1). (2021).
- [17] Wen Jiang, Gautam Khera, Roger Wood, Mason Williams, Neil Smith, and Yoshihiro Ikeda. 2003. Cross-track noise profile measurement for adjacent-track interference study and write-current optimization in perpendicular recording. *Journal of Applied Physics* 93, 10 (05 2003), 6754–6756. <https://doi.org/10.1063/1.1557716> arXiv:https://pubs.aip.org/aip/jap/article-pdf/93/10/6754/8061756/6754_1_online.pdf
- [18] Dimitris Kaseridis, Jeffrey Stuecheli, and Lizy Kurian John. 2011. Minimalist Open-Page: A DRAM Page-Mode Scheduling Policy for the Many-Core Era. In *Proceedings of the 44th Annual IEEE/ACM International Symposium on Microarchitecture (Porto Alegre, Brazil) (MICRO-44)*. Association for Computing Machinery, New York, NY, USA, 24–35. <https://doi.org/10.1145/2155620.2155624>
- [19] Jeremie S Kim, Minesh Patel, A Giray Yağlıkçı, Hasan Hassan, Roknoddin Azizi, Lois Orosa, and Onur Mutlu. 2020. Revisiting rowhammer: An experimental analysis of modern dram devices and mitigation techniques. In *2020 ACM/IEEE 47th ISCA*. IEEE, 638–651.
- [20] Jeremie S Kim, Minesh Patel, A Giray Yağlıkçı, Hasan Hassan, Roknoddin Azizi, Lois Orosa, and Onur Mutlu. 2020. Revisiting rowhammer: An experimental analysis of modern dram devices and mitigation techniques. In *ISCA*. IEEE, 638–651.
- [21] Moonsoo Kim, Jungwoo Choi, Hyun Kim, and Hyuk-Jae Lee. 2019. An effective DRAM address remapping for mitigating rowhammer errors. *IEEE Trans. Comput.* 68, 10 (2019), 1428–1441.
- [22] Michael Jaemin Kim, Jaehyun Park, Yeonhong Park, Wanju Doh, Namhoon Kim, Tae Jun Ham, Jae W Lee, and Jung Ho Ahn. 2021. Mithril: Cooperative Row Hammer Protection on Commodity DRAM Leveraging Managed Refresh. *arXiv preprint arXiv:2108.06703* (2021).
- [23] Woongrae Kim, Chulmoon Jung, Seongnyuh Yoo, Duckhwa Hong, Jeongjin Hwang, Jungmin Yoon, Ohyoung Jung, Joonwoo Choi, Sanga Hyun, Mankeun Kang, Sangho Lee, Dohong Kim, Sanghyun Ku, Donhyun Choi, Nogeun Joo, Sangwoo Yoon, Junseok Noh, Byeongyong Go, Cheolhoe Kim, Sunil Hwang, Mihyun Hwang, Seol-Min Yi, Hyungmin Kim, Sanghyuk Heo, Yeonsu Jang, Kyoungchul Jang, Shinho Chu, Yoonna Oh, Kwidong Kim, Junghyun Kim, Soohwan Kim, Jeongtae Hwang, Sangil Park, Junphyo Lee, Inchl Jeong, Joohwan Cho, and Jonghwan Kim. 2023. A 1.1V 16Gb DDR5 DRAM with Probabilistic-Aggressor Tracking, Refresh-Management Functionality, Per-Row Hammer Tracking, a Multi-Step Precharge, and Core-Bias Modulation for Security and Reliability Enhancement. In *2023 IEEE International Solid-State Circuits Conference (ISSCC)*. 1–3. <https://doi.org/10.1109/ISSCC42615.2023.10067805>
- [24] Yoongu Kim, Ross Daly, Jeremie Kim, Chris Fallin, Ji Hye Lee, Donghyuk Lee, Chris Wilkerson, Konrad Lai, and Onur Mutlu. 2014. Flipping bits in memory without accessing them: An experimental study of DRAM disturbance errors. *ISCA* (2014).
- [25] Andreas Kogler, Jonas Juffinger, Salman Qazi, Yoongu Kim, Moritz Lipp, Nicolas Boichat, Eric Shiu, Mattias Nissler, and Daniel Gruss. 2022. Half-Double: Hammering from the next row over. In *USENIX Security Symposium*.
- [26] Michael Kounavis, Sergej Deutsch, Santosh Ghosh, and David Durham. 2020. K-cipher: A low latency, bit length parameterizable cipher. In *2020 IEEE Symposium on Computers and Communications (ISCC)*. IEEE, 1–7.
- [27] Andrew Kwong, Daniel Genkin, Daniel Gruss, and Yuval Yarom. 2020. Rambleed: Reading bits in memory without accessing them. In *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, 695–711.
- [28] Zhenrong Lang, Patrick Jattke, Michele Marazzi, and Kaveh Razavi. 2023. BLASTER: Characterizing the Blast Radius of Rowhammer. In *3rd Workshop on DRAM Security (DRAMSec) co-located with ISCA 2023*. ETH Zurich.
- [29] Eojin Lee, Ingab Kang, Sukhan Lee, G Edward Suh, and Jung Ho Ahn. 2019. TWiCe: preventing row-hammering by exploiting time window counters. In *ISCA*.
- [30] Fangfei Liu, Hao Wu, Kenneth Mai, and Ruby B Lee. 2016. Newcache: Secure cache architecture thwarting cache side-channel attacks. *IEEE Micro* 36, 5 (2016), 8–16.
- [31] Jason Lowe-Power, Abdul Mutaal Ahmad, Ayaz Akram, Mohammad Alian, Rico Amslinger, Matteo Andreozzi, Adria Arnejach, Nils Asmussen, Srikant Bharadwaj, Gabe Black, Gedare Bloom, Bobby R. Bruce, Daniel Rodrigues Carvalho, Jerónimo Castrillón, Lizhong Chen, Nicolas Derumigny, Stephan Diestelhorst, Wendy Elsasser, Marjan Fariborz, Amin Farmahini Farahani, Pouya Fotouhi, Ryan Gambord, Jayneel Gandhi, Dibakar Gope, Thomas Grass, Bagus Hanindhito, Andreas Hansson, Swapnil Haria, Austin Harris, Timothy Hayes, Adrian

- Herrera, Matthew Horsnell, Syed Ali Raza Jafri, Radhika Jagtap, Hanhwi Jang, Reiley Jeyapaul, Timothy M. Jones, Matthias Jung, Subash Kannoth, Hamidreza Khaleghzadeh, Yuetsu Kodama, Tushar Krishna, Tommaso Marinelli, Christian Menard, Andrea Mondelli, Tiago Mück, Omar Naji, Krishnendra Nathella, Hoa Nguyen, Nikos Nikoleris, Lena E. Olson, Marc S. Orr, Binh Pham, Pablo Prieto, Trivikram Reddy, Alec Roelke, Mahyar Samani, Andreas Sandberg, Javier Setoain, Boris Shingarov, Matthew D. Sinclair, Tuan Ta, Rahul Thakur, Giacomo Travaglini, Michael Upton, Nilay Vaish, Ilias Vougioukas, Zhengrong Wang, Norbert Wehn, Christian Weis, David A. Wood, Hongil Yoon, and Éder F. Zulian. 2020. The gem5 simulator: Version 20.0+. *arXiv preprint arXiv:2007.03152* (2020).
- [32] Michele Marazzi, Flavien Solt, Patrick Jattke, Kubo Takashi, and Kaveh Razavi. 2023. REGA: Scalable Rowhammer Mitigation with Refresh-Generating Activations. In *44rd IEEE Symposium on Security and Privacy (SP 2023)*. IEEE.
- [33] John D. McCalpin. 1995. Memory Bandwidth and Machine Balance in Current High Performance Computers. *IEEE Computer Society Technical Committee on Computer Architecture (TCCA) Newsletter* (1995).
- [34] Micron Technology Inc. [n. d.]. System Power Calculators. ([n. d.]). <https://www.micron.com/support/tools-and-utilities/power-calc>.
- [35] Yeonhong Park, Woosuk Kwon, Eojin Lee, Tae Jun Ham, Jung Ho Ahn, and Jae W Lee. 2020. Graphene: Strong yet Lightweight Row Hammer Protection. In *MICRO*. IEEE, 1–13.
- [36] Yeonhong Park, Woosuk Kwon, Eojin Lee, Tae Jun Ham, Jung Ho Ahn, and Jae W. Lee. 2020. Graphene: Strong yet Lightweight Row Hammer Protection. In *2020 53rd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*. IEEE, Athens, Greece, 1–13. <https://doi.org/10.1109/MICRO50266.2020.00014>
- [37] Moinuddin Qureshi, Aditya Rohan, Gururaj Saileshwar, and Prashant J Nair. 2022. Hydra: enabling low-overhead mitigation of row-hammer at ultra-low thresholds via hybrid tracking. In *Proceedings of the 49th Annual International Symposium on Computer Architecture*. 699–710.
- [38] Moinuddin K Qureshi. 2018. CEASER: Mitigating conflict-based cache attacks via encrypted-address and remapping. In *2018 51st Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*. IEEE, 775–787.
- [39] Moinuddin K Qureshi. 2019. New attacks and defense for encrypted-address cache. In *Proceedings of the 46th International Symposium on Computer Architecture*. 360–371.
- [40] Moinuddin K Qureshi, John Karidis, Michele Franceschini, Vijayalakshmi Srinivasan, Luis Lastras, and Bulent Abali. 2009. Enhancing lifetime and security of PCM-based main memory with start-gap wear leveling. In *Proceedings of the 42nd annual IEEE/ACM international symposium on microarchitecture*. 14–23.
- [41] Gururaj Saileshwar and Moinuddin Qureshi. 2021. MIRAGE: Mitigating Conflict-Based Cache Attacks with a Practical Fully-Associative Design. In *30th USENIX Security Symposium (USENIX Security 21)*. 1379–1396.
- [42] Gururaj Saileshwar, Bolin Wang, Moinuddin Qureshi, and Prashant J. Nair. 2022. Randomized Row-Swap: Mitigating Row Hammer by Breaking Spatial Correlation between Aggressor and Victim Rows. In *Proceedings of the 27th ACM International Conference on Architectural Support for Programming Languages and Operating Systems (Lausanne, Switzerland) (ASPLOS '22)*. Association for Computing Machinery, New York, NY, USA, 1056–1069. <https://doi.org/10.1145/3503222.3507716>
- [43] Anish Saxena, Gururaj Saileshwar, Prashant J Nair, and Moinuddin Qureshi. 2022. Aqua: Scalable rowhammer mitigation by quarantining aggressor rows at runtime. In *2022 55th IEEE/ACM International Symposium on Microarchitecture (MICRO)*. IEEE, 108–123.
- [44] Mark Seaborn and Thomas Dullien. 2015. Exploiting the DRAM rowhammer bug to gain kernel privileges. *Black Hat* 15 (2015), 71.
- [45] Nak Hee Seong, Dong Hyuk Woo, and Hsien-Hsin S Lee. 2010. Security refresh: Prevent malicious wear-out and increase durability for phase-change memory with dynamically randomized address mapping. *ACM SIGARCH computer architecture news* 38, 3 (2010), 383–394.
- [46] Seyed Mohammad Seyedzadeh, Alex K Jones, and Rami Melhem. 2018. Mitigating wordline crosstalk using adaptive trees of counters. In *2018 ACM/IEEE 45th Annual International Symposium on Computer Architecture (ISCA)*. IEEE, 612–623.
- [47] Mungyu Son, Hyunsun Park, Junwhan Ahn, and Sungjoo Yoo. 2017. Making DRAM stronger against row hammering. In *Proceedings of the 54th Annual Design Automation Conference 2017*. 1–6.
- [48] Victor Van Der Veen, Yanick Fratantonio, Martina Lindorfer, Daniel Gruss, Clémentine Maurice, Giovanni Vigna, Herbert Bos, Kaveh Razavi, and Cristiano Giuffrida. 2016. Drammer: Deterministic rowhammer attacks on mobile platforms. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*. 1675–1689.
- [49] Minghua Wang, Zhi Zhang, Yueqiang Cheng, and Surya Nepal. 2020. Dramdig: A knowledge-assisted tool to uncover dram address mapping. In *2020 57th ACM/IEEE Design Automation Conference (DAC)*. IEEE, 1–6.
- [50] Zhenghong Wang and Ruby B Lee. 2007. New cache designs for thwarting software cache-based side channel attacks. In *Proceedings of the 34th annual international symposium on Computer architecture*. 494–505.
- [51] Mario Werner, Thomas Unterluggauer, Lukas Giner, Michael Schwarz, Daniel Gruss, and Stefan Mangard. 2019. ScatterCache: Thwarting Cache Attacks via Cache Set Randomization. In *USENIX Security Symposium*. 675–692.
- [52] Minbok Wi, Jaehyun Park, Seoyoung Ko, Michael Jaemin Kim, Nam Sung Kim, Eojin Lee, and Jung Ho Ahn. 2023. SHADOW: Preventing Row Hammer in DRAM with Intra-Subarray Row Shuffling. In *2023 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*. 333–346.
- [53] Jeonghyun Woo, Gururaj Saileshwar, and Prashant J Nair. 2023. Scalable and Secure Row-Swap: Efficient and Safe Row Hammer Mitigation in Memory Systems. In *2023 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*. IEEE, 374–389.
- [54] A Giray Yağlıkçı, Minesh Patel, Jeremie S Kim, Roknoddin Azizi, Ataberk Olgun, Lois Orosa, Hasan Hassan, Jisung Park, Konstantinos Kanellopoulos, Taha Shahroodi, et al. 2021. BlockHammer: Preventing RowHammer at Low Cost by Blacklisting Rapidly-Accessed DRAM Rows. In *2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*. IEEE, 345–358.
- [55] Yuval Yarom and Katrina Falkner. 2014. {FLUSH+ RELOAD}: A high resolution, low noise, l3 cache {Side-Channel} attack. In *23rd USENIX security symposium (USENIX security 14)*. 719–732.
- [56] Jung Min You and Joon-Sung Yang. 2019. MRLoc: Mitigating Row-hammering based on memory Locality. In *2019 56th ACM/IEEE Design Automation Conference (DAC)*. IEEE, 1–6.