World Journal of Innovation and Modern Technology, Vol. 7, Issue 5, (Oct) 2024
ISSN 2682-5910
DOI: 10.53469/wjimt.2024.07(05).06

A Better Sense Amplifier Improves the Resilience in Compute-In-Memory and Row Hammer

You Zhang, Tianyuan Xu, Yihan Wang

University of Michigan, Ann Arbor, Michigan, USA
{youzh, tyxu, wangyhhh}@umich.edu

Abstract: ***As machine learning (ML) workloads continue to scale, the demand for higher-density DRAM (Dynamic Random-Access Memory) and SRAM (Static Random-Access Memory) is intensifying, and also driving interest in compute-in-memory (CIM) architectures as a promising approach to enhancing computational efficiency. However, increasing DRAM cell density introduces significant challenges, particularly in the reliable identification of charge sharing between cells and bit lines. This challenge heightens the susceptibility of DRAM to row hammer attacks, where unintended data corruption occurs due to repeated access to adjacent rows. Furthermore, CIM architectures necessitate more precise sensing of bit lines to ensure accurate computation within memory. In light of these challenges, our study proposes an investigation into the potential benefits of utilizing an offset compensation sense amplifier (OCSA). The OCSA is designed to address the accuracy limitations posed by high-density DRAM in CIM applications. By enhancing the precision of bit line sensing, the OCSA may mitigate the vulnerability to row hammer attacks and improve the overall reliability and performance of CIM architectures. This study will explore the effectiveness of the OCSA in maintaining data integrity and computational accuracy within high-density DRAM environments, offering insights into its applicability for future ML workloads.***

1. INTRODUCTION

As machine learning (ML) workloads continue to grow in complexity and scale, the disparity between memory and computation power, often referred to as the "memory wall," is becoming increasingly problematic. While computational power continues to advance, the ability of memory systems, particularly DRAM, to keep pace is limited by several factors. This growing gap has heightened the demand for denser DRAM technologies, which are essential for supporting the memory-intensive nature of modern ML applications. Concurrently, there is a rising interest in compute-in-memory (CIM) architectures as well, which aim to alleviate data movement bottlenecks by integrating computation directly within the memory arrays [1], [2].

However, the pursuit of denser DRAM cells and in-memory computations introduce significant challenges. As DRAM cells become smaller and more densely packed, the charge sharing between the cell and bit lines decreases, making it increasingly difficult to accurately identify and retrieve stored data. This reduction in sensing margin exacerbates the vulnerability of DRAM to row hammer attacks, a phenomenon where repeated access to a row of memory can induce bit flips in adjacent rows, compromising data integrity. Furthermore, the shift towards compute-in-memory architectures imposes even greater demands on the precision of bit line sensing. The compute-in-memory mechanisms require bit cells in two bitlines to be open for read, causing the voltage difference to be even smaller compared to single bitcell read, necessitating more accurate and reliable sensing mechanisms.

In response to these challenges, we propose an investigation into the benefits of utilizing Offset Compensation Sense Amplifiers (OCSA) as a means to enhance sensing accuracy in DRAM. This study establishes four dedicated models to investigate the voltage change to victim in row hammer, sensing voltage requirements in the logical compute-in-memory, original sense amplifiers and OCSA. We utilized the bisection method to discover the sensing margin of sense amplifiers, and relate the sensing margin of sense amplifiers to the voltage requirements for row hammer model as well as compute-in-memory model [4].

Our preliminary studies indicate that by implementing OCSA, it is possible to reduce the voltage requirements for DRAM cells within voltage difference models. This not only improves the energy efficiency of the memory but also broadens the sensing range of the OCSA, thereby increasing the tolerance for voltage variations. Such variations could arise from the increased vulnerability to row hammer attacks or the added complexity of compute-in-memory operations. Through this research, we aim to demonstrate that OCSA can provide a robust solution to the dual challenges of memory reliability and CIM accuracy, potentially paving the way for more resilient and efficient memory architectures in the era of advanced ML workloads [5], [6], [8].

2. BACKGROUNDS

2.1 DRAM Cell, Charge Sharing and Sensing

Dynamic Random Access Memory (DRAM) cells are fundamental building blocks of DRAM arrays, known for their high density and small physical size. Each DRAM cell consists of two primary components: a transistor and a capacitor [9-10]. The capacitor is responsible for storing the data in the form of electrical charges. When a write operation is performed on the DRAM cell, the capacitor is either charged or discharged, corresponding to the binary values of 1 or 0, respectively.

However, the inherent nature of DRAM technology presents several challenges. One of the most significant issues is the gradual leakage of charge from the capacitor over time. Whether the capacitor is fully charged or completely discharged, it inevitably loses charge due to leakage currents. This leakage process means that without periodic refresh operations, the data stored in the DRAM cell would eventually degrade, leading to potential data loss. This necessity for continuous refreshing is a fundamental characteristic of DRAM, distinguishing it from other types of memory such as Static RAM (SRAM) [11-13].

During a read operation in a DRAM array, the bitline associated with the target cell is pre-charged to half of the supply voltage, typically denoted as V_{dd} / 2. The transistor in the DRAM cell is then activated, allowing the capacitor within the cell to share its charge with the bitline. Given the small capacitance of the DRAM cell relative to the large capacitance of the bitline, even a fully charged capacitor can only cause a slight change in the bitline voltage. This challenge is further compounded by the fact that, due to leakage, the capacitor might not hold a perfect binary value (a full 1 or a full 0) at the time of reading. As in the following charge sharing equation, the V_{BL}' will be eventually used for the comparison with reference voltage.

$$\mathrm{V_{BL}}' * (\mathrm{C_{BL}} + \mathrm{C_{cell}}) = \mathrm{V_{BL}} * \mathrm{C_{BL}} + \mathrm{V_{Cell}} * \mathrm{C_{Cell}} \quad (1)$$

To accurately determine the data stored in the DRAM cell, a sense amplifier is employed. The sense amplifier compares the voltage of the bitline with a reference line to detect the minute voltage difference and infer the correct stored value. This process is critical not only for standard memory operations but also for compute-in-memory (CIM) architectures, where the sense amplifier may play a role in interpreting computation results directly within the memory array.

The accuracy of this sensing process is heavily dependent on the "margin of readability," which refers to the extent to which the small voltage difference can be reliably detected. A high-quality sense amplifier is crucial in this context, as it provides a larger margin of readability, thereby enhancing the reliability of the memory system. In contrast, a less efficient sense amplifier might struggle to distinguish between the small voltage differences, leading to errors in data retrieval or computation. Consequently, improving sense amplifier design is vital for ensuring the correct operation of DRAM, especially as the demand for higher density and integration with computational functions increases. MPAAGN combines meta-path-guided neighbor selection, attention-based aggregation, and sampling strategies from GraphSAGE to enhance graph neural network performance, which can be utilized in other domains requiring efficient and scalable graph-based learning, such as social network analysis and recommendation systems [14].

2.2 Sense Amplifiers

Sense amplifiers are crucial components in various types of memory systems, including both Dynamic Random-Access Memory (DRAM) and Static Random-Access Memory (SRAM). In this study, we focus on voltage sense amplifiers, which are specifically designed to detect small voltage differences between two input lines and amplify these differences to drive the outputs to fully defined logic levels—either 0 or 1. The effectiveness of these amplifiers is central to the performance of memory systems, as their primary role is to ensure accurate data reading by distinguishing between closely spaced voltage levels [3].

A significant challenge in practical applications is that sense amplifiers in real-world manufacturing are rarely ideal. Non-idealities such as variations in transistor characteristics, mismatches in component values, and parasitic effects can impair the accuracy of the sense amplifiers. These imperfections impact the sensing margin, which is the critical range within which the sense amplifiers can reliably detect and distinguish between different voltage levels. The intrinsic performance limitations of the sense amplifiers fundamentally determine this sensing margin, influencing the overall reliability of data read operations.

Voltage sense amplifiers typically incorporate a pair of back-to-back inverters to form a feedback loop. This configuration allows them to drive bit lines with large capacitances effectively. During the sensing process, both the reading bit line and the reference bit line are initially pre-charged to a voltage level of Vdd/2. This pre-charging step is crucial for establishing a stable baseline. Following this, the transistor associated with the DRAM cell is activated, initiating charge sharing between the cell and the bit lines.

Once charge sharing has occurred, the sense amplifiers are connected to the bit lines. At this stage, they begin the sensing operation by comparing the voltages on the reading and reference bit lines. The feedback loop within the sense amplifier amplifies the voltage difference, causing the reading bit line to be driven to one logic level (0 or 1), while the reference bit line is driven to the opposite logic level. This differential amplification ensures that even small differences in voltage are accurately detected and converted into distinct binary values.

In summary, the design and performance of voltage sense amplifiers are pivotal in memory systems, influencing their ability to accurately read data. The non-ideal characteristics of real-world sense amplifiers impact the sensing margin, affecting the reliability of memory read operations. Understanding and addressing these challenges is essential for optimizing the performance of both DRAM and SRAM memory systems.

2.3 Row Hammer Attack

Row hammer is a critical vulnerability in modern DRAM systems, arising from the shrinking size of memory cells as DRAM technology advances. This phenomenon occurs when repeated access to a particular "aggressor" row in a DRAM module induces electrical disturbances that can cause charge leakage in adjacent "victim" rows, leading to unintended bit flips. As DRAM cells become more densely packed, the isolation between adjacent rows weakens, increasing the susceptibility to row hammer attacks. Since the DRAM cell itself becomes smaller, the charges stored in a single cell become less, which introduces further vulnerability to the aggressor's attack [6].

The implications of row hammer are significant, as it can lead to data corruption and potential security breaches. Although various mitigation strategies, such as increasing refresh rates and implementing error-correcting codes, have been proposed, these solutions often introduce trade-offs in power consumption, performance, and complexity. Understanding and addressing row hammer is essential for the reliability and security of DRAM, particularly in contexts where data integrity is paramount.

2.4 SRAM and Compute-In-Memory

Static Random-Access Memory (SRAM) is a type of volatile memory renowned for its exceptional speed and reliability. Unlike DRAM, which requires periodic refresh cycles to prevent data loss, SRAM maintains data integrity through a different mechanism. SRAM cells use a stable bistable latching configuration—typically built with six or eight transistors—that continuously holds data as long as power is supplied. This design eliminates the need for refresh operations, resulting in significantly faster data access times compared to DRAM.

Despite these advantages, SRAM generally consumes more power than DRAM. This higher power consumption is a trade-off for its rapid access speeds, which are crucial for applications demanding high performance, such as cache memory in processors and various embedded systems. The design of SRAM cells, which involves multiple transistors, leads to larger cell sizes compared to the more compact DRAM cells. While this results in a lower memory density, it ensures stable and reliable data storage, which is critical for high-speed data operations.

However, SRAM faces its own set of challenges. One such challenge is related to the ability of the small SRAM cells to drive large capacitive bit lines effectively. The capacitance of the bit lines can impact the speed and accuracy of data retrieval, particularly when the memory cells are required to handle significant data loads.

In the realm of compute-in-memory systems, this study investigates a specific mechanism involving logical computation directly within SRAM arrays. In this CIM approach, two lines of SRAM cells are simultaneously activated during each computational operation. The computational process involves performing logic operations within the memory array itself. Once the computation is completed, sense amplifiers are used to detect and measure the results of these operations. The output from the sense amplifiers is then routed to logic gates, where it can be used to perform more complex logical operations or further processing.

This integration of computation within the memory array introduces several challenges, particularly concerning the sensing margin—the range within which the sense amplifiers can accurately detect the computed values. Since both SRAM cells involved in a given operation are active during computation, there is potential for interference and racing effects between the driving capabilities of each cell. These racing effects occur when the driving strength of the cells influences each other's performance, potentially leading to inaccuracies in the sensed results. Managing these challenges is essential to maintaining the performance benefits of SRAM-based CIM systems, ensuring that the integration of computational tasks does not compromise the speed and reliability of data access.

3. MODELS

We used an open source library with CMRF8SF, a 120nm CD technology, and set up the models in Cadence Virtuoso. Simulations are done with SPICE and Python. In the model set-ups, we utilize the divide and conquer strategy by establishing separate models for rowhammer attack, compute in memory logical computation, sense amplifier and offset compensation sense amplifier. We used the rowhammer and CIM model to demonstrate the effect of voltage change on the bitline. And then we used the sense amplifier models with DC offset on the bitline to link to rowhammer model and CIM model. By synthesizing the results, we were able to draw comprehensive conclusions.

3.1 DRAM Cell and SRAM CELL

In our analysis, we employ a standard transistor and capacitor model for DRAM. In this model, each memory cell consists of a single transistor and a capacitor. The transistor functions as a switch that controls access to the capacitor, which stores the charge representing the data bit. When a word line is activated, it turns on the transistor, allowing charge to either flow into or out of the capacitor depending on the data being written or read. The bit lines, on the other hand, are used to transfer data to and from the memory cells during read and write operations. The combination of these components and their interactions determines the behavior and performance of the DRAM.

For SRAM, we use a typical 6-transistor (6T) model to represent each memory cell. The 6T SRAM cell configuration is well-known for its stable and reliable performance. It comprises six transistors arranged in a flip-flop configuration. This arrangement includes two cross-coupled inverters that form a feedback loop, which maintains the state of the cell as long as power is supplied. Additionally, four access transistors control the connection between the SRAM cell and the bit lines during read and write operations. The 6T SRAM cell's design allows for rapid access times and robust data storage, making it suitable for high-speed applications such as cache memory. By using this 6T model, we can accurately analyze and simulate the behavior of SRAM cells, including their performance characteristics and interactions with other components in the memory system [7].

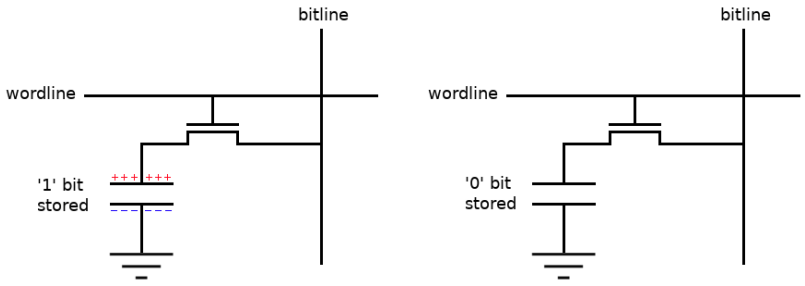

Figure 1: A typical DRAM simulation model setup

3.2 Compute-In-Memory Sub-Array

In our CIM model, we employ the 6T SRAM cells to construct a sophisticated subarray, which represents the top level of our memory architecture. The SRAM subarray is equipped with several crucial circuit components to facilitate both data handling and computational operations. Pre-charge circuits are employed to establish the initial voltage levels on the bit lines and other signal lines before initiating any read or write operations. This pre-charging ensures that all lines start from a known state, which is essential for accurate subsequent operations.

Two 16-bit SRAM word lines are used. In typical read or write operations, one single SRAM word line will be turned on by its transistors. And during the logical operations, two word lines will be enabled at the same time to compete. The bit line and inverted bit line carry the data to and from the SRAM cells. The bit line typically represents the data being stored or retrieved, while the inverted bit line holds the complementary data. These lines work together to facilitate differential sensing of the data, which enhances accuracy during read operations.

The reference line serves as a benchmark for comparison, providing a stable reference voltage against which the voltages on the bit lines can be measured. This comparison is crucial for the sense amplifiers, which are integrated into the model to detect and amplify small voltage differences between the bit lines and the reference line. These amplifiers convert minute voltage variations into distinct binary values, enabling precise data reading.

Additionally, an XOR (exclusive OR) gate is incorporated into the subarray to perform logical operations directly within the memory. This gate allows for computations such as addition or bitwise XOR to be executed within the SRAM array itself. By integrating computation capabilities within the memory, the model reduces the need for data transfers between the memory and the processor, thus enhancing overall system efficiency.

Moreover, in reality, the bit lines or inverted bit lines will be connected to thousands of cells. Therefore, to correctly model the capacitive bit lines, we have added the capacitors to the bit lines of the array and connected them to ground.

The purpose of the investigation on the sub-array is to demonstrate the higher voltage margin required for CIM systems.

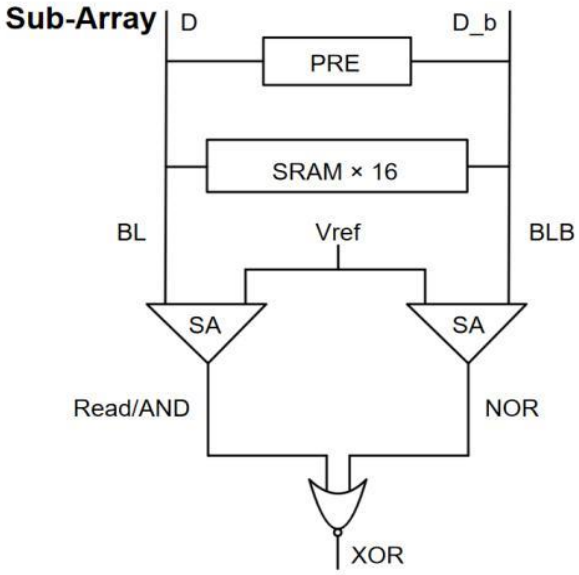

Figure 2: Sub-Array Model for the Compute In Memory

3.3 Row Hammer Attack Model

In row hammer attacks, it's important to relate to the physical and electrical characteristics of memory cells as they are implemented in real-world manufacturing processes. Memory cells in modern DRAM systems typically have metallic wires that connect various components of the memory array. The area between the transistors and capacitors within these cells is commonly filled with silicon dioxide, which serves as an insulator.

This arrangement naturally results in the formation of unintended capacitors between the metal wires of adjacent memory cells.

$$C = \varepsilon_{ox} * A \div D \quad (2)$$

As in the above capacitance equation, these parasitic capacitors are created due to the presence of the silicon dioxide dielectric material, which separates the metal interconnects. The silicon dioxide layer, while primarily acting as an insulator, also has a capacitance associated with it. This capacitance can influence the behavior of

adjacent cells, particularly in the presence of high-frequency or rapidly changing signals, as in Figure 3.

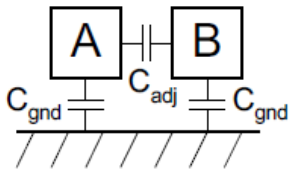

Figure 3: Thoracic Modelling of Crosstalk Between Wires

In a row hammer attack, the attacker rapidly activates and deactivates rows of memory cells in a specific pattern. This intense and repetitive activity generates significant electrical activity and switching noise. Due to the capacitive coupling between the metal wires and the silicon dioxide dielectric, this noise can induce voltage changes in adjacent cells, even those that are not directly activated by the attack. This phenomenon, known as crosstalk coupling, occurs because the rapid switching of signals in one row of memory cells can inadvertently affect the voltage levels in nearby rows.

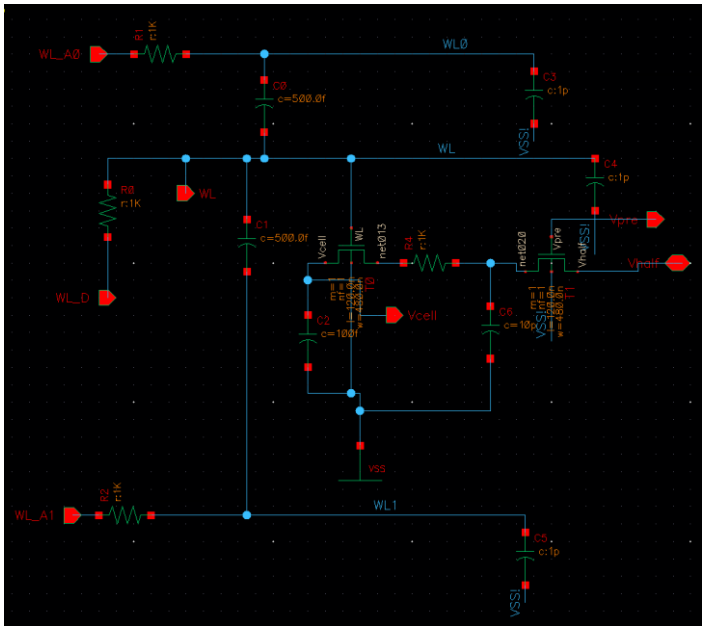

Figure 4: Row Hammer Coupling Model

The crosstalk coupling introduces noise into the victim cells, which can manifest as voltage fluctuations. These fluctuations can lead to unintended data corruption or bit flips in the victim cells, compromising the integrity of the stored data. The extent of the noise and its impact on the victim cells depend on several factors, including the proximity of the victim cells to the aggressor cells, the intensity of the row hammering activity, and the specific electrical characteristics of the memory array.

Therefore, we propose the above model to test the cell voltage change by time of attack. The purpose of this model is to evaluate the margin requirement for a DRAM cell in attack.

3.4 Original Sense Amplifier and Offset Model

We have implemented both standard sense amplifiers and offset compensation sense amplifiers, incorporating a direct DC voltage offset at the bitline. This offset is strategically used to simulate the voltage variations caused by a row hammer attack, which impacts the voltage on both the memory cell and the bitline [3].

A sense procedure is composed of precharge, charge sharing and sensing. In an offset compensation sense amplifier, there is one more procedure between precharge and charge sharing called compensation. In an offset compensation sense amplifier, there will be four more NMOS to choose the connection between the bitline of that node itself or the bitline of the opposite node. By connecting the gate of the NMOS of one of the inverters to its own node, when there is a voltage offset on the gate of the NMOS, there will be a discharge. And if there is a voltage offset on the gate of the NMOS, there will be a difference in the speed of discharge; thus, the NMOS with

higher voltage on its gate (its bitline as well) will have a higher discharge rate and the voltage on its bitline will be less.

The normal sense amplifier is composed of two back to back inverters. With both bitline and bitline bar precharged to half Vdd, the inverters will enter an unstable state, where a slight change of voltage could kick a feedback loop to rapidly drive both lines to their identified voltages, either Vdd or Vss.

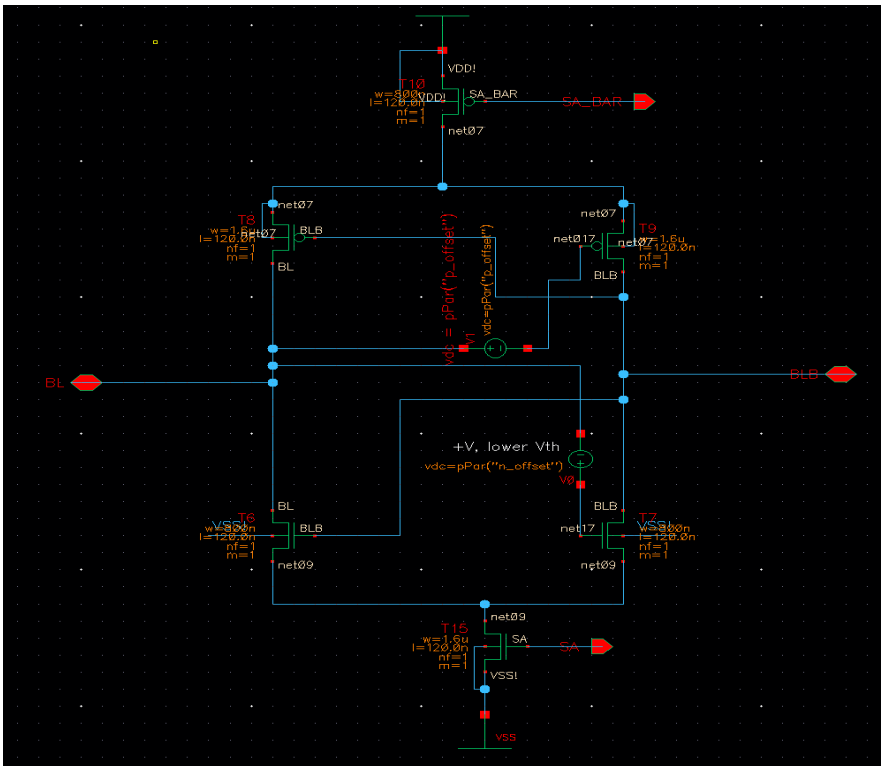

Figure 5: Normal Sense Amplifier

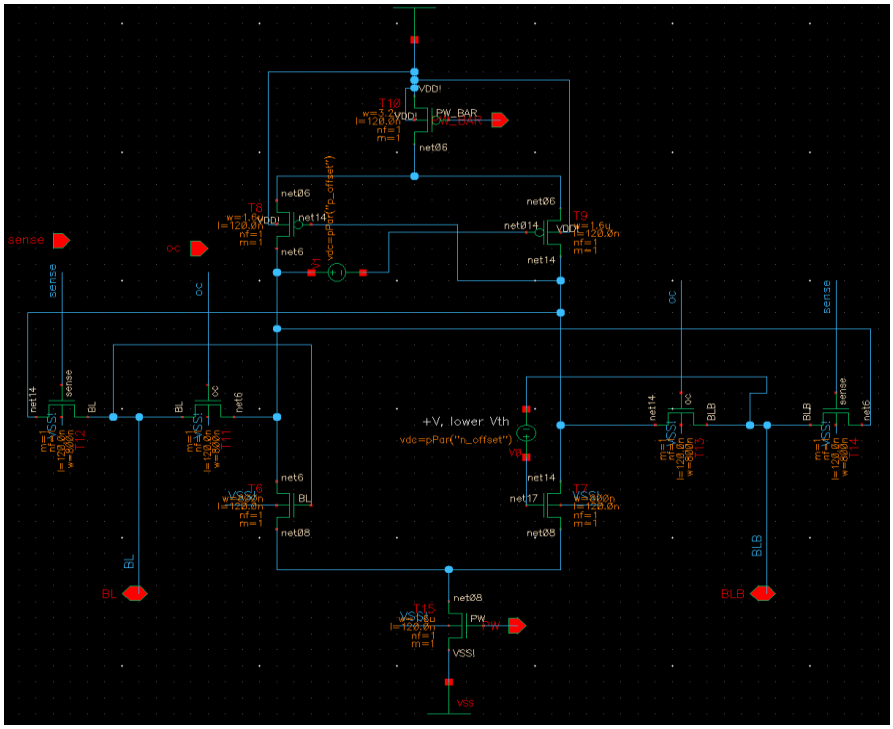

Figure 6: Offset Compensation Sense Amplifier

4. RESULTS

4.1 Voltage Change in Cell Induced by Row Hammer

By utilizing the model of the row hammer attack, we have achieved the following result. As in the result, we can conclude that, when the cell is precharged to Vdd, the attack will induce a voltage leak and eventually decrease the charges on the node, causing the voltage to decrease in the charge sharing. Moreover, if the node is precharged to Vss, the effect of the attack will be less, however, it will also cause the charge to be injected into the cell, and increase the voltage during the voltage sharing process. In both examples, the row hammer attack will change the voltages in the cells to be closer to half Vdd; and such change will cause the accuracy of sensign to be less and requires a more accurate sense amplifier with better margin.

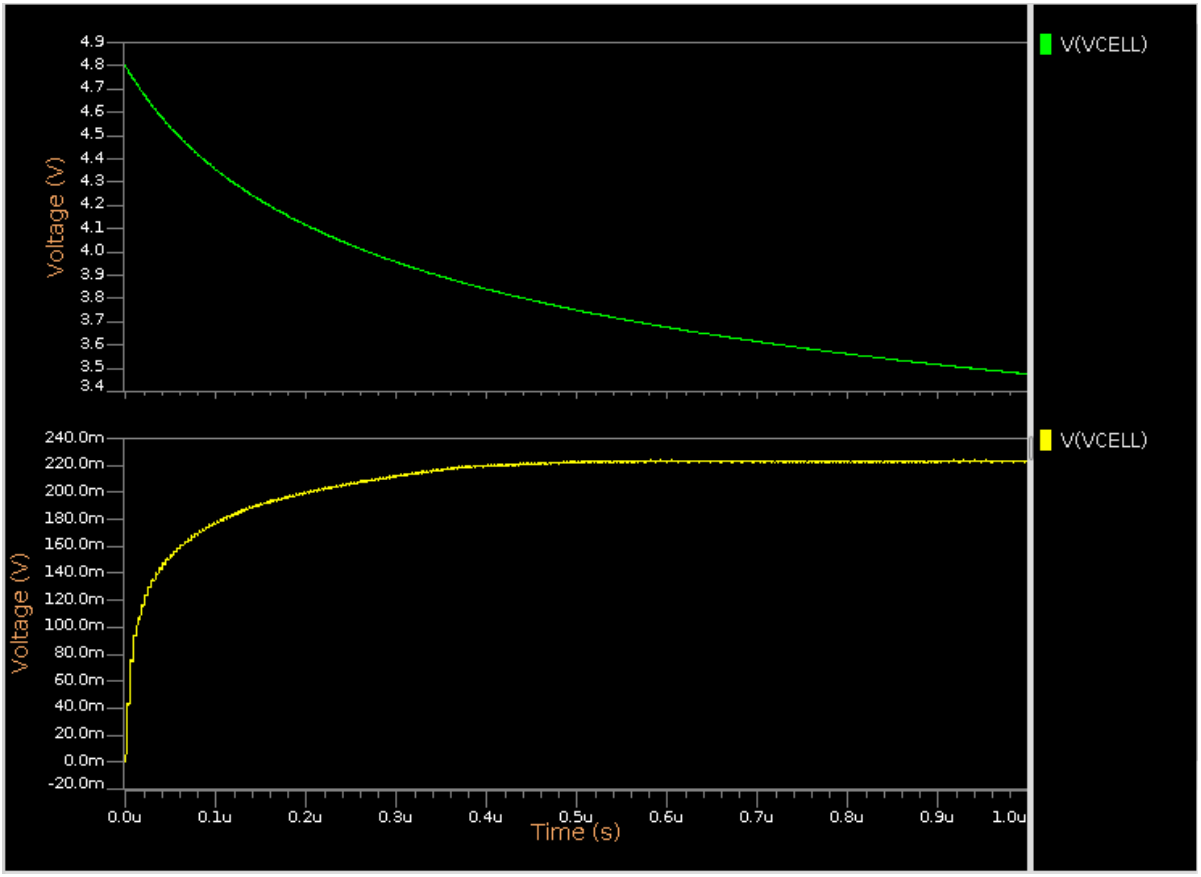

Figure 7: Results for Row Hammer Attack Simulation

4.2 Sensing Operation in Compute Cache

In the result of CIM computation, we can see an interference of both SRAMs' nodes, net 04 and 05. Such interference will not happen in normal SRAM and indicates that the SRAMs' data during CIM computation will be more easily corrupted. Moreover, in this simulation, we can see the SRAM unable to fully drive its own bar back to full Vdd or Vss before the SA kicks in. Therefore, a SA with better sensing to reduce the impact from the fabrication variation would be crucial to the sensing of CIM computation.

Figure 8: CIM Computation Waveform (See Appendix)

4.3 Sense Amplifier Operations

The results from a single test of the sense amplifier demonstrate that the standard sense amplifier performs effectively under baseline conditions, successfully sensing the expected signals without issue. However, when we introduced an increased voltage difference, designed to simulate the effects of a row hammer attack or the challenging conditions arising from logical computations in CIM operations—the performance of the standard sense amplifier degraded significantly. This heightened voltage difference, which mirrors the disturbances in the bitline and memory cell voltage during such scenarios, led to a sensing failure. The standard sense amplifier was unable to accurately detect the correct signal under these stressed conditions.

As in the graph, the blue line is the wordline, when the wordline opens, the charge sharing stats, and the cell will drive the bitline to a voltage higher than the reference bitline (BLB). After the charge sharing, the sense amplifier fires and starts driving the bitline to Vdd and BLB to Vss.

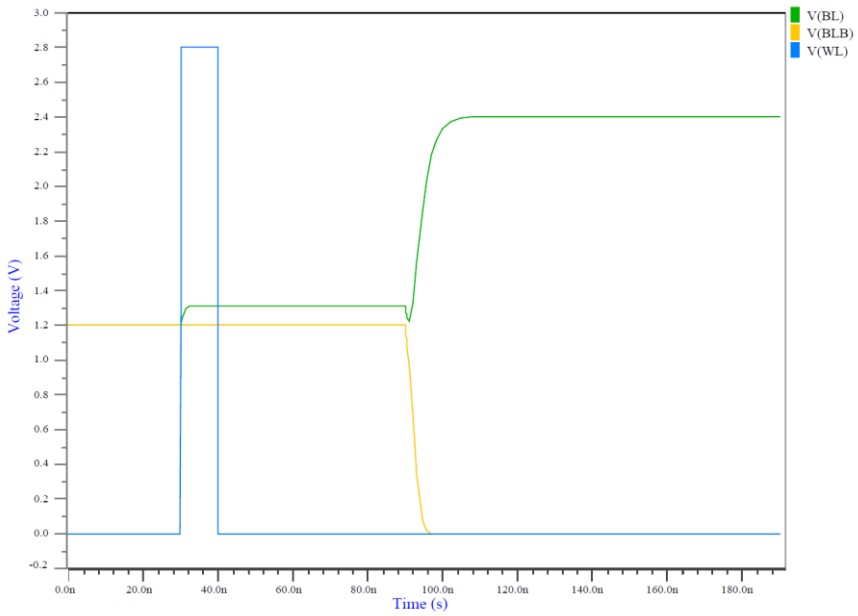

Figure 9: Sensing Success of Normal Sense Amplifier

With a positive offset on the gate of NMOS of the bitline, the sense amplifier will mistakenly think the BLB has a higher voltage. And this will finally cause a sense failure.

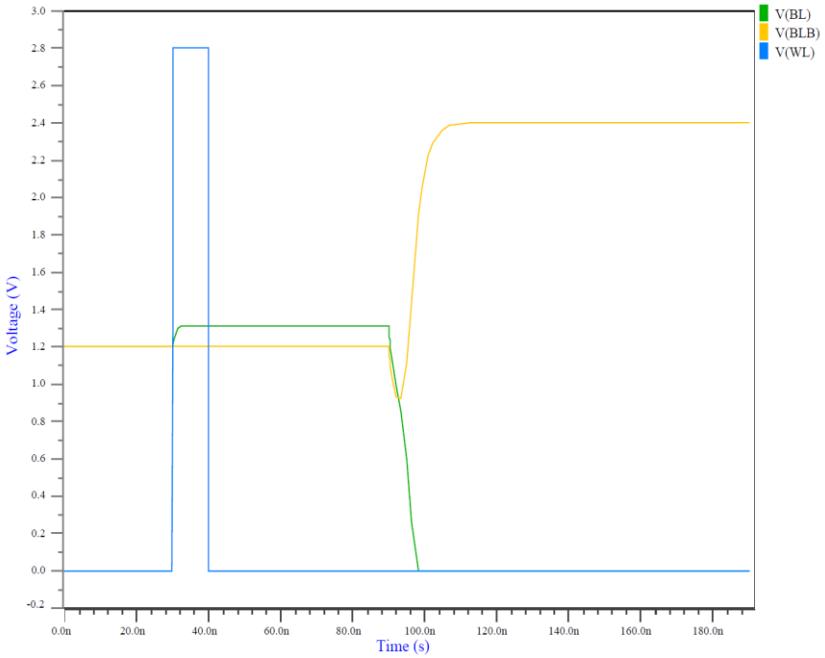

Figure 10: Sensing Failure of Normal SA with Increased Offset

With the offset compensation sense amplifier, the charge sharing after the precharge and compensation. In the compensation stage, because there is a DC offset to mimic a mistaken higher voltage on the BLB, there will be a voltage droop. When the gate is connected to each NMOS's each bitline, the voltage droops are induced and there is a voltage difference between the BLB and BL after compensation. Because the DC offset is positive to indicate a higher voltage on BLB, after the compensation, the voltage decrease is expected to decrease. Such result demonstrated the better sensing margin of the offset compensation sense amplifier.

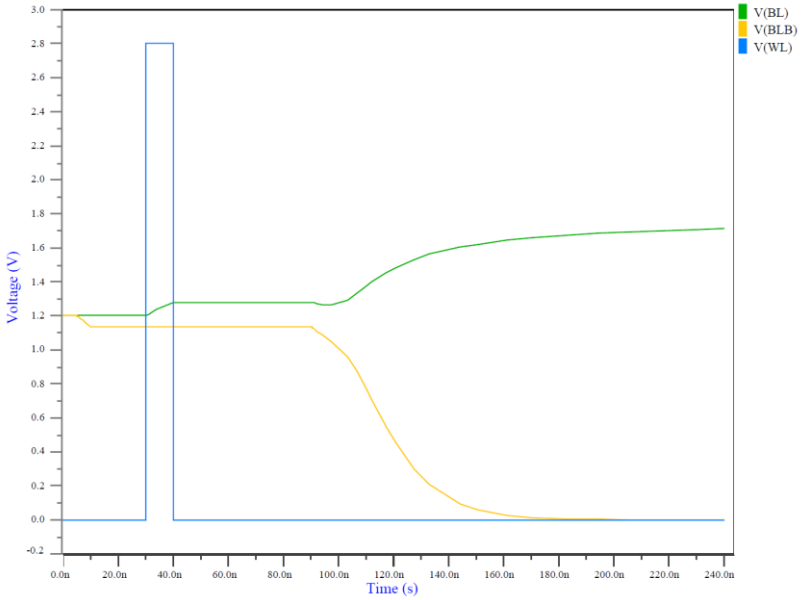

Figure 11: OCSA Sensing

5. CONCLUSION

In conclusion, our study provides several critical insights into the challenges of reliable sensing in memory operations, particularly in the presence of adverse conditions such as row hammer attacks and CIM operations.

The row hammer attack, a well-known vulnerability in DRAM, was shown to create a substantial voltage leak within the memory cell. This leakage results in a diminished sensing margin, making it increasingly difficult for traditional sense amplifiers to accurately detect the correct data. Such a scenario highlights the importance of developing more advanced sensing technologies, particularly offset sense amplifiers, which can offer an improved sensing margin by compensating for the induced voltage variance.

Additionally, our simulations of CIM operations revealed another layer of complexity. In these scenarios, we observed that the inherent racing conditions between two SRAM cells during logical computations can cause significant glitches in the output. These glitches further reduce the sensing margin, making accurate detection even more challenging. The racing conditions and the associated noise in CIM operations demand a more robust sensing approach to ensure the integrity of the computational results.

The OCSA was particularly effective in addressing these challenges. In our experiments, the OCSA consistently demonstrated a significantly better sensing margin compared to standard sense amplifiers. This improved performance was evident under both the stress conditions of a row hammer attack and the complex, noisy environment of CIM operations. The ability of the OCSA to maintain a stable and accurate sensing margin under such conditions suggests that it is an optimal solution for scenarios where voltage fluctuations and glitches are prevalent.

Therefore, we can conclusively state that the OCSA represents a superior choice for applications that are vulnerable to row hammer attacks and the challenges posed by compute-in-memory operations. Its enhanced sensing capabilities make it an essential tool for ensuring data integrity and reliable performance in modern memory systems, where the demands for both security and computational efficiency are continually increasing.

REFERENCES

[1] Shaizeen Aga, Supreet Jeloka, Arun Subramaniyan, Satish Narayanasamy, David Blaauw, and Reetuparna Das. 2017. Compute Caches. In 2017 IEEE International Symposium on High Performance Computer Architecture (HPCA). 481–492. https://doi.org/10.1109/HPCA.2017.21

[2] Y. Zhang, K. Yang, Y. Wang, P. Yang and X. Liu, "Speculative ECC and LCIM Enabled NUMA Device Core," 2023 3rd International Symposium on Computer Technology and Information Science (ISCTIS), Chengdu, China, 2023, pp. 624-631, doi: 10.1109/ISCTIS58954.2023.10213102.

[3] Liu, Li; Tai, Dele; Qiang, Bin; Guan, Lijun; Peng, Chunyu; Dai, Chenghu; Hao, Licai; Lu, Wenjuan; Lin, Zhiting; Zhao, Qiang; Wu, Xiulong, “Flip Point Offset-Compensation Sense Amplifier With Sensing-Margin-Enhancement for Dynamic Random-Access Memory,” IEEE transactions on circuits and systems. II, Express briefs, vol. 71, no. 4. New York: IEEE, pp. 1759–1763, 2024.

[4] Y. Jiang, H. Zhu, D. Sullivan, X. Guo, X. Zhang, and Y. Jin, “Quantifying rowhammer vulnerability for DRAM security,” 2021 58th ACM/IEEE Design Automation Conference (DAC), 2021.

[5] Taehui Na; Byungkyu Song; Jung Pill Kim; Kang, Seung H.; Seong-Ook Jung, “Offset-Canceling Current-Sampling Sense Amplifier for Resistive Nonvolatile Memory in 65 nm CMOS,” IEEE journal of solid-state circuits, vol. 52, no. 2. IEEE, pp. 496–504, 2017.

[6] Yang, Thomas; Lin, Xi-Wei, “Trap-Assisted DRAM Row Hammer Effect,” IEEE electron device letters, vol. 40, no. 3. New York: IEEE, pp. 391–394, 2019.

[7] Kwon, Semyung; Yun, Ilgu, “Asymmetric and Double-Layered Gate-All- Around Structures of 1T-DRAM for Sensing Margin and Retention Improvement,” IEEE transactions on electron devices, vol. 71, no. 6. IEEE, pp. 3627–3632, 2024.

[8] Wang, Qipeng; Xu, Mengwei; Jin, Chao; Dong, Xinran; Yuan, Jinliang; Jin, Xin; Huang, Gang; Liu, Yunxin; Liu, Xuanzhe, “Melon: breaking the memory wall for resource-efficient on-device machine learning.” New York, NY, USA: ACM, pp. 450–463, 2022.

[9] Xu, S., Wu, P., & Chen, Y. (2024). Interfacial thermal conductance in 2D WS2/MoSe2 and MoS2/MoSe2 lateral heterostructures. Computational Materials Science, 245, 113282.

[10] Wu, P., Iquebal, A. S., & Ankit, K. (2023). Emulating microstructural evolution during spinodal decomposition using a tensor decomposed convolutional and recurrent neural network. Computational Materials Science, 224, 112187.

[11] Chen, H., Yang, Y., & Shao, C. (2021). Multi-task learning for data-efficient spatiotemporal modeling of tool surface progression in ultrasonic metal welding. Journal of Manufacturing Systems, 58, 306-315.

[12] Cao, Y., Cao, P., Chen, H., Kochendorfer, K. M., Trotter, A. B., Galanter, W. L., ... & Iyer, R. K. (2022). Predicting ICU admissions for hospitalized COVID-19 patients with a factor graph-based model. In

Multimodal AI in healthcare: A paradigm shift in health intelligence (pp. 245-256). Cham: Springer International Publishing.

[13] Z. Ren, "A Novel Feature Fusion-Based and Complex Contextual Model for Smoking Detection," 2024 6th International Conference on Communications, Information System and Computer Engineering (CISCE), Guangzhou, China, 2024, pp. 1181-1185, doi: 10.1109/CISCE62493.2024.10653351.

[14] Wu, Z. (2024). MPGAAN: Effective and Efficient Heterogeneous Information Network Classification. Journal of Computer Science and Technology Studies, 6(4), 08-16.

Appendix

1. CIM Computation Result