

# A Novel Embedded 1T-1R SOT-MRAM Macro Achieving 32% Area Reduction Compared with The Traditional 2T-1R SOT Cell

Weiliang Huang<sup>1,†</sup>, Jinhao Li<sup>1,†</sup>, Yimo Du<sup>1,†</sup>, Hongchao Zhang<sup>2</sup>, Chengyuan Sun<sup>2</sup>, Hong-xi Liu<sup>2,\*</sup>, Hui Jin<sup>1</sup>, Xuan Li<sup>2</sup>, Gefei Wang<sup>2</sup>, Kaihua Cao<sup>1</sup>, Zhaohao Wang<sup>1</sup>, He Zhang<sup>1,\*</sup>, Weisheng Zhao<sup>1,\*</sup>

<sup>1</sup>Beihang University, Beijing, China <sup>2</sup>Truth Memory Corporation, Beijing, China

<sup>†</sup>authors contribute equally \*Email: hongxi\_liu@tmc-bj.cn, {zhanghe, weisheng.zhao}@buaa.edu.cn

**Abstract**—Emerging memory technologies have attracted significant research attention owing to their non-volatile characteristics and exceptional silicon integration compatibility. Among these, Spin-Transfer Torque Magnetic Random-Access Memory (STT-MRAM) has emerged as a promising candidate for edge-side applications due to its superior read/write performance and high storage density. However, as a two-terminal device, STT-MRAM faces critical challenges, particularly unintended data writes triggered by intrinsic read-disturbance vulnerabilities. To overcome this challenge, we introduces the first implementation of a 32Kb embedded 1T-1R Spin-Orbit Torque MRAM (SOT-MRAM) cell and array architecture, achieving 32% area reduction compared with traditional 2T-1R SOT cell structure for high-speed scenarios like Last-Level Cache (LLC). Furthermore, the macro incorporates an innovative row/column grouping architecture to mitigate crosstalk in three-terminal devices, significantly enhancing operational reliability through systematic interference rejection at both row and column levels. A mid-point readout scheme is proposed to achieve reliable readout for high-precision read requirements in less than 47.5ns @ (-40, 125)°C at Monte-Carlo (MC) and 5 MOS corners simulation.

**Keywords**—SOT-MRAM, high density, 1T-1R

## I. INTRODUCTION

With the continuous scaling down of CMOS process nodes, volatile memory media such as SRAM and DRAM exhibit exponentially increasing leakage currents at advanced process nodes, resulting in prohibitive static power dissipation. This technological shift has driven substantial research interest in emerging Non-Volatile Memory (NVM) technologies, including MRAM, Resistive RAM (ReRAM), and Phase-Change RAM (PCRAM) have attracted considerable academic attention recently, which demonstrate potential for zero standby power consumption and high-speed cache applications.

STT-MRAM has garnered significant attention as a promising non-volatile memory technology and has achieved commercial viability with successful deployment in practical applications [1-3]. However, as a two-terminal memory architecture, its shared read/write pathways introduce inherent risks of inadvertent data corruption during access operations. Furthermore, the technology faces critical limitations including substantial write latency (commonly referred to as incubation delay) and reliability concerns under high-frequency operations,

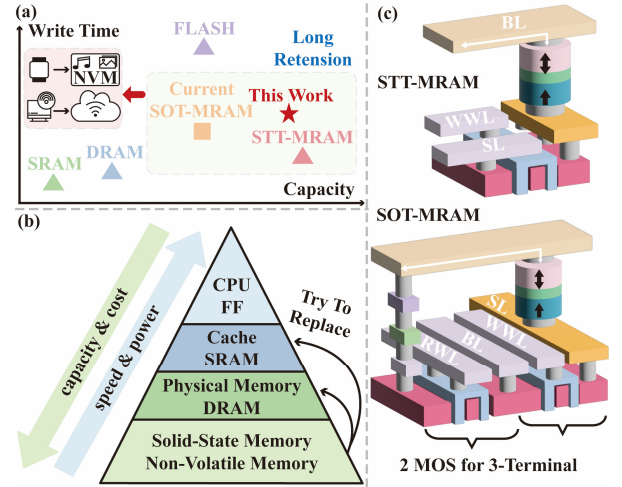


Fig. 1. (a) Comparison between different memory types and main applications for SOT-MRAM; (b) Memory level and opportunity; (c) Layout of 1T-1R STT-MRAM and 2T-1R SOT-MRAM

which fundamentally restrict its applicability in high-performance computing architectures requiring nanosecond-scale access times, such as L1/L2 cache memory subsystems [4].

SOT-MRAM is a three-terminal device with independent read and write paths, as shown in Fig. 1, which eliminates the impact of read operations on data retention. Recently, SOT-MRAM has been regarded as a promising candidate for its subnanosecond switching speed and  $>10^{12}$  cycles endurance, which can be applied in various high-speed and high-reliability scenarios such as CPU caches and Block RAMs (BRAMs) in field-programmable gate arrays (FPGAs) [5]. Both academic research groups and industrial leaders have successfully demonstrated CMOS-compatible SOT-MRAM integration schemes with competitive performance metrics.[6-8]. However, SOT-MRAM requires integrating MOS in both read and write paths as a 3-terminal device, resulting in poor area efficiency compared with 1T-1R STT-MRAM cell, as shown in Fig. 1(c). This technical difficulty urgently needs to be addressed, considering the strict area requirements of LLC and edge-side application scenarios.

In this paper, we propose a 32Kb SOT-MRAM array based on novel 1T-1R SOT cells proposed for the first time. Compared

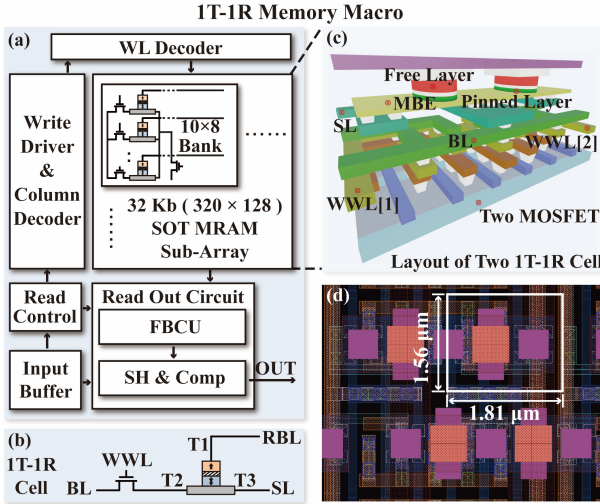


Fig. 2. (a) Block diagram of the proposed 1T-1R Memory Macro; (b) The proposed 1T-1R cell; (c) 3D view of two 1T-1R Layout; (d) Four 1T-1R cell layout with 110-nm CMOS design rule.

with traditional 2T-1R SOT-MRAM array, this work achieve 32% area reduction. To address critical challenges in three-terminal device implementation, we present two key innovations: 1) a Row-Column Grouping (RCG) architecture for crosstalk mitigation and 2) a Mid-Point Clamp Readout (MPCR) scheme for leakage current suppression, which pass Monte Carlo (MC) simulation at 5 MOS corners in less than 47.5ns read access time. 1T-1R SOT-MRAM cell and array structure proposed in this work significantly contribute to resolving SOT-MRAM's dilemma of area efficiency for 2T-1R structure.

## II. 1T1R SOT-MRAM MACRO AND RCG SCHEME

Fig. 2(a) shows the proposed 32Kb SOT-MRAM macro, which consists of a  $320 \times 128$  1T-1R SOT-MRAM array with RCG scheme, MPCR module, read control unit, write driving units and row & column decoder. The RCG scheme divides the array into  $10 \times 8$  sub-arrays, in which two rows are used as complementary reference units and eight rows store binary data.

Design Technology Co-optimization (DTCO) is used on the layout design of the 1T-1R cell for higher area efficiency, as shown in Fig. 2(b). Fig. 2(c) and 2(d) show two dimensions of the layout design of adjacent two or four 1T-1R storage cells. For the SOT process used in this work, the write and group-selection transistors are customized designed and one side of the heavy metal layer of two adjacent SOT devices are reused. By eliminating the vias and selection MOS between the top electrode T1 and metal layer in the proposed 1T-1R cell structure, higher area efficiency is achieved, with a cell area of  $233.3 f^2$  in the 180nm process. Compared with 2T-1R SOT-MRAM cell with a cell area of  $332.8 f^2$  in the same technology node, the proposed 1T-1R SOT-MRAM cell achieves a 32% efficiency improvement

SOT-MRAM requires integration of additional transistors at both write and read paths to prevent row and column leakage and parasitic interference caused by the array, which introduces

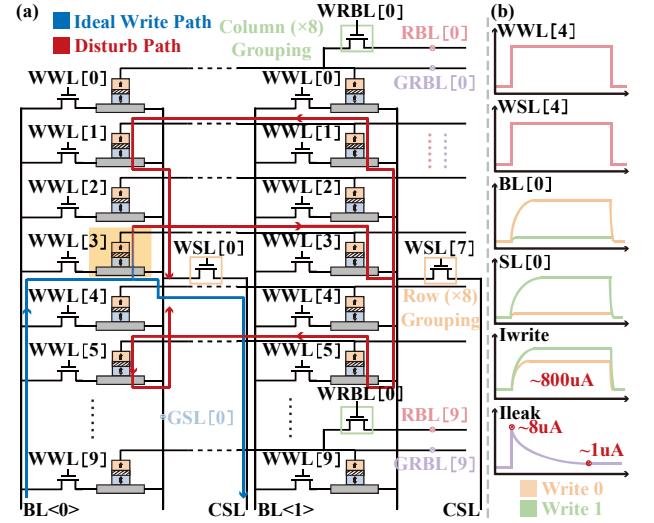


Fig. 3. (a) The proposed Row and Column Grouping (RCG) scheme and write disturb path analysis; (b) Write timing waveforms.

extra overhead compared to STT-MRAM or other two-terminal memory devices. To overcome this limitation, we propose a RCG strategy as shown in Fig. 3(a). The 1T-1R SOT-MRAM array is partitioned into  $10 \times 8$  sub-arrays to mitigate interference from inactive three-terminal memory cells on target units. Within each sub-array, SL terminals of cells in 10 rows per column are interconnected to form a Group Source Line (GSL). Each GSL is regulated by the Write Select Line (WSL) signal for activation control, effectively isolating the sub-array from other rows. The GSL is connected to the Common SL (CSL) terminals through MOS switches, with CSL multiplexing implemented every 8 columns to enhance storage density. Top electrodes of SOT cells in 8 columns along the same row are linked via a Group Read Bit Line (GRBL), controlled by WRBL transistors and signals to isolate the sub-array from other columns. This architecture confines leakage current and crosstalk from the 1T-1R cell's top electrode terminals within the  $10 \times 8$  sub-array boundaries, significantly improving signal integrity while maintaining design compactness.

As shown in Fig. 3(b), the highlighted cell is used to illustrate the write process of the 1T-1R array with proposed RCG scheme. the WWL[3] and WSL[0] generated by the row & column decoder control the writing transistor at BL[0] and GSL[0] to turn on, while WRBL remains off. BL[0] and CSL are driven and form a write current greater than 800uA. Furthermore, high-resistance SOT process is adopted to suppress the leakage current to acceptable 8uA, compared to the over 100uA leakage observed in conventional 1T-1R SOT-MRAM arrays with standard resistance parameters.

## III. MPCR TECHNIQUE FOR CROSSTALK SUPPRESSION

SOT-MRAM typically employs a top-electrode readout scheme, where the read voltage is applied to the T1 terminal (top electrode) to measure the resultant sensing current. However, in 1T-1R array architectures, applying read voltage to the target row's RBL introduces a critical design challenge: the formation

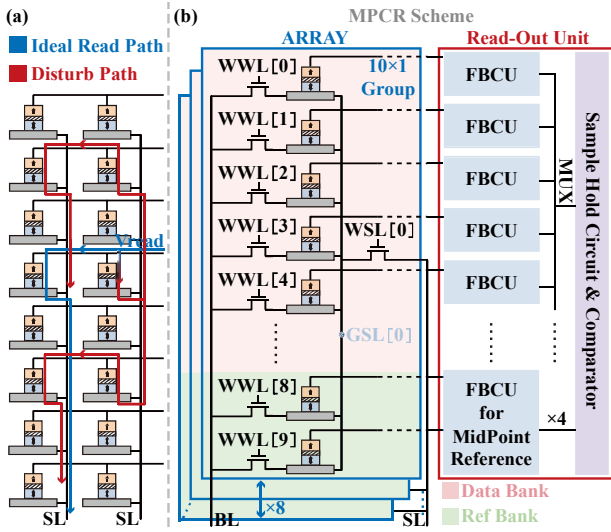


Fig. 4. (a) Read disturb path analysis; (b) The block diagram of the proposed Mid-Point Clamp Readout (MPCR) Scheme.

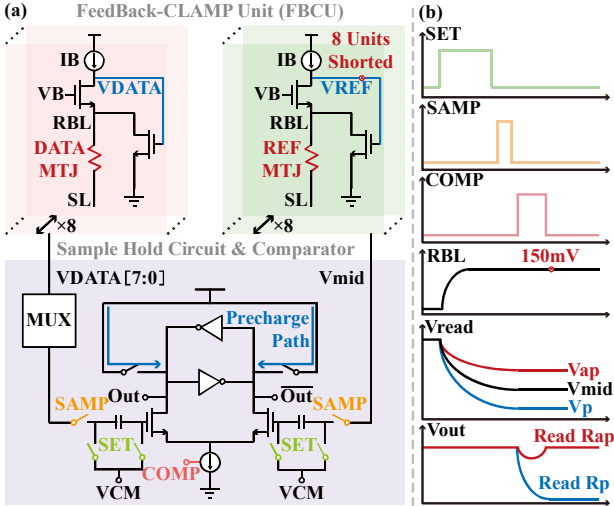


Fig. 5. (a) The schematic diagram of the FeedBack-CLAMP Unit (FBCU) and voltage boosting comparator; (b) Timing sequence waveforms of the MPCR scheme.

of unintended discharge paths through RBLs of inactive rows or columns. This parasitic current leakage originates from the unterminated third terminal inherent in three-terminal memory cells, which creates parallel conduction paths during read operations, ultimately compromising sensing accuracy, as shown in Fig. 4(a). Although The RCG strategy we proposed suppress it to some extent, it still cannot meet the high-precision read requirements, especially for the low switching ratio of MRAM.

Here, we introduce the MPCR technique, as shown in Fig. 4(b). The RBL terminals in each sub-array are clamped by the same set of ten Feedback-Clamp Units (FBCU), which prevents discharge paths from non-target cells. The 320 rows are divided into 4 pages, with each section consisting of eight (8+2) rows. Each page corresponds to one set of FBCUs. As shown in Fig. 5(a),  $V_{data}$  generated by the Data Bank is selected through a

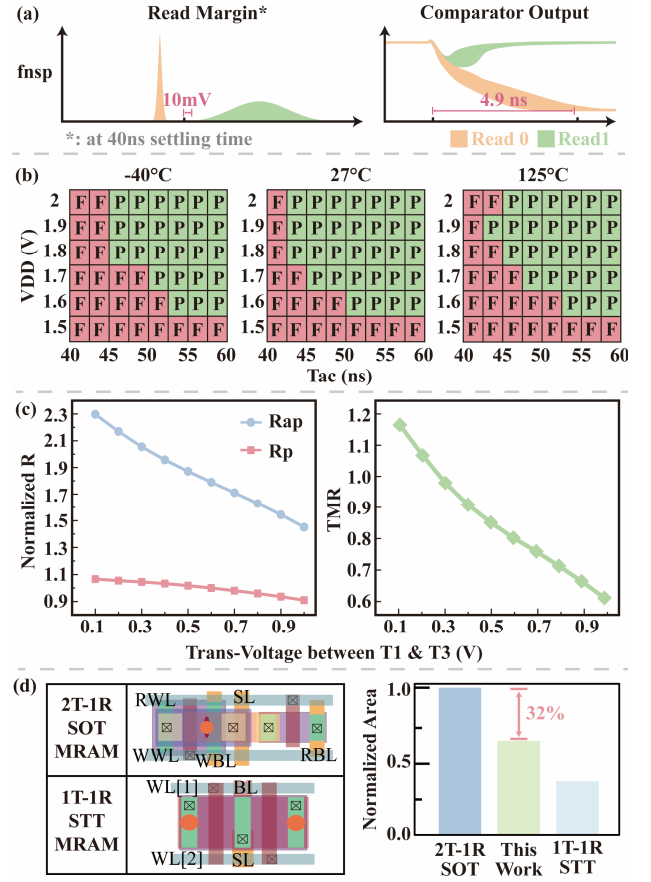


Fig. 6 (a) Worst-case MC simulations for MPCR scheme; (b) The read shmoos plots in -40°C, 27°C and 125°C; (c) Relationship between  $R_p/R_{ap}$ , TMR and Trans-Voltage between T1 and T3. (d) Layout of 2T-1R SOT-MRAM cell and 1T-1R STT-MRAM cell and normalized area comparison.

MUX based on the address. The gate terminals of the FBCUs in the Ref Bank's two rows are shorted with the corresponding positions of the other three activated Ref Banks in the other three sub-arrays.  $V_{mid}$  is generated by averaging the current from eight reference cells in four sub-arrays. This implementation of  $V_{mid}$  suppresses the impact of process variations caused by individual set of sub-arrays or FBCU, which ensures correlation with voltage fluctuations and variations in the Data Bank, improving the reliability of the read operation.

For the detailed readout logic, once  $V_{data}$  and  $V_{mid}$  are settled, they are input to the capacitor, both sides of which are preset to  $V_{cm}$ . The clock signals generated by clock generator in read control unit switch as shown in Fig. 5(b), the comparator's input will be boosted to

$$V_g = V_{cm} + \frac{V_{in}}{1 - C_g/C_s} \quad (1)$$

Thus, the capacitors at the gate of input transistors with an appropriate capacity enhance the input voltage for high read speed, especially for specific cases with low input voltage. Furthermore, the sampling switch connected to the FBCU is disabled during comparison, which prevents the kickback noise from interfering with consecutive readouts.



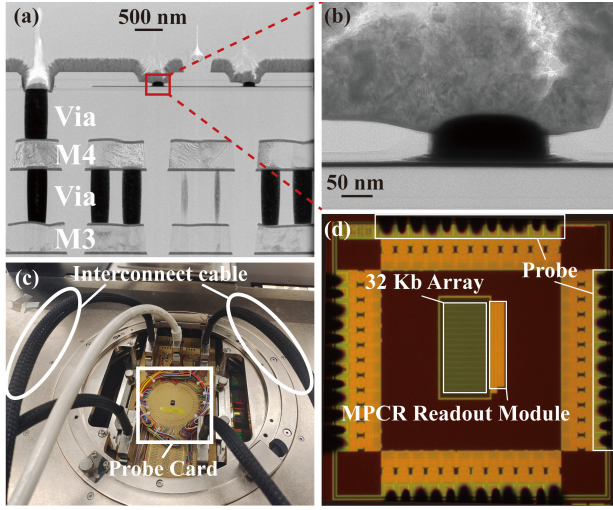


Fig. 7 (a) TEM image of a single cell in the chip; (b) TEM image of SOT MTJs and via; (c) The photo of the test platform; (d) Top view of the 32Kb 1T-1R SOT-MRAM macro under optical microscope.

We performed 2000 MC simulations considering  $3\sigma$  process variations across five distinct MOS process corners applied to both the readout module and SOT device, with worst-case TMR of 50%, using a 47.5ns clock (42.5ns setup + 5ns compare). The results in Fig. 6(a) shows that under different simulation conditions, the minimum read margin remains above 10 mV, showcasing exceptional read reliability.

#### IV. RESULTS ANALYSES

Fig. 6(b) shows the access time ( $T_{ac}$ ) shmoo plots for the proposed MPCR scheme. In the temperature range of  $-40^\circ\text{C}$  to  $150^\circ\text{C}$ , the reading speed of 47.5ns  $T_{ac}$  is also feasible. Fig. 6(c) demonstrates the variations in  $R_p/R_{ap}$  and TMR with respect to the read voltage. Low read clamp voltages of 150mV are used to ensure high TMR more than 105% for reliable readout. Fig. 6(d) shows the layout of several common structures. The proposed 1T-1R structure achieves 32% area reduction compared with the traditional 2T-1R structure. Fig. 7(a) and 7(b) show a Transmission Electron Microscope (TEM) image of a single cell in the array fabricated on the CMOS wafer. The MTJ is patterned with dimensions of  $250\text{ nm} \times 700\text{ nm}$  in an ellipse shape. Fig. 7(c) presents the top view photograph of SOT MRAM cell fabricated on a 300 mm wafer. Comparison with other state-of-the-art works is shown in Table I.

#### V. CONCLUSION

This work addresses the critical challenges of area inefficiency and read-disturbance reliability in SOT-MRAM for high-speed edge-side applications. By proposing a novel 32Kb 1T-1R SOT-MRAM array—the first of its kind—we achieve a 32% area reduction compared to conventional 2T-1R SOT-MRAM architectures. The RCG strategy effectively mitigates crosstalk and leakage in three-terminal devices, while the MPCR scheme ensures reliable read within 47.5ns across extreme temperatures and process variations.

TABLE I. Comparison with state-of-the-art SOT-MRAM macros

	This Work	JSSC'21[9]	IEDM'24[10]	IEDM'24[11]
Memory Type	SOT	SOT	SOT	SOT
Process Node (nm)	180	55	180	NA
Cell Structure	1T-1R	2T-1R	2T-1R	2D-1R
Cell Size ( $\text{f}^2$ )	233.3	1725.6	NA	NA
Capacity	32 Kb	32 Kb	128 Kb	NA
Read Time (ns)	47.5	11.1	5	NA
Read Energy (pJ/bit)	4.3	NA	NA	NA
Write Time (ns)	< 20	16.7	15	10
Write Energy (pJ/bit)	< 32.4	NA	NA	NA
Supply (V)	1.8	1.2	NA	NA
SOT Node (nm)	$250 \times 700$	$88 \times 315$	$260 \times 720$	NA
Thermal Stability	> 60	> 70	> 72	> 54.1
Endurance	> $1\text{e}12$	NA	> $1\text{e}10$	> $1\text{e}10$

#### ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (No. 62301019), the Key R&D Program of Shandong Province, China (No. 2024CXGC010112).

#### REFERENCES

- [1] B. Dieny et al., "Opportunities and challenges for spintronics in the microelectronics industry," *Nature Electronics*, vol. 3, no. 8, pp. 446–459, 2020.
- [2] H. Honjo et al., "25 nm iPMA-type Hexa-MTJ with solder reflow capability and endurance  $> 10^7$  for eFlash-type MRAM," in 2022 International Electron Devices Meeting (IEDM). IEEE, 2022, pp. 10–3.
- [3] O. Golonzka et al., "MRAM as Embedded Non-Volatile Memory Solution for 22FFL FinFET Technology," in 2018 IEEE International Electron Devices Meeting (IEDM). IEEE, 2018, pp. 18–1.
- [4] J. Alzate et al., "2 Mb array-level demonstration of STT-MRAM process and performance towards L4 cache applications," in 2019 IEEE International Electron Devices Meeting (IEDM). IEEE, 2019, pp. 2–4.
- [5] D. Suzuki and T. Hanyu, "Design of a low-power nonvolatile flip-flop using three-terminal magnetic-tunnel-junction-based self-terminated mechanism," *Japanese Journal of Applied Physics*, vol. 56, no. 4S, p. 04CN06, 2017.
- [6] A. Lu et al., "High-speed emerging memories for AI hardware accelerators," *Nature Reviews Electrical Engineering*, vol. 1, no. 1, pp. 24–34, 2024.
- [7] S. Van Beek et al., "Scaling the SOT track – A path towards maximizing efficiency in SOT-MRAM," in 2023 International Electron Devices Meeting (IEDM). IEEE, 2023, pp. 1–4.
- [8] M. Song et al., "High RA Dual-MTJ SOT-MRAM devices for High Speed (10ns) Compute-in-Memory Applications," in 2023 International Electron Devices Meeting (IEDM). IEEE, 2023, pp. 1–4.
- [9] M. Natsui et al., "Dual-Port SOT-MRAM Achieving 90-MHz Read and 60-MHz Write Operations under Field-Assistance-Free Condition," *IEEE Journal of Solid-State Circuits*, vol. 56, no. 4, pp. 1116–1128, 2020.
- [10] C. Jiang et al., "Demonstration of 128 Kb SOT-MRAM chip with 5 ns write and 15 ns read speed, high endurance over  $10^{10}$  and low ECC-on bit error rate," in 2024 IEEE International Electron Devices Meeting (IEDM). IEEE, 2024, in press.
- [11] C. Yang et al., "Dual-Function Unipolar Top-pSOT-MRAM for All-Spin Probabilistic Computing with Ultra-Dense Coupling and Adaptive Temporal Coding," in 2024 IEEE International Electron Devices Meeting (IEDM). IEEE, 2024, in press.