The First CMOS-Integrated Voltage-Controlled MRAM with 0.7ns Switching Time

H. Suhail[1†], H. He[1†], J. Yang[1], Q. Shu[1], C. -Y. Wang[2], S. -Y. Yang[2], Y. -C. Hsin[2],
C. -Y. Shih[2], H. -H. Lee[2], D. Wu[1], A. Lee[1], J. -H. Wei[2], P. Gupta[1], K. L. Wang[1], S. Pamarti[1]

[1]University of California, Los Angeles, CA, USA, email: harissuhail@ucla.edu.
[2]Industrial Technology Research Institute, Hsinchu, Taiwan
†These authors contributed equally to this work

Abstract—Spin-Transfer-Torque (STT) Magnetic Random Access Memory (MRAM) has shown good density and compatibility with advanced nodes, but its write operation is slow and power-hungry. Voltage-Controlled (VC) Magnetic Tunneling Junction (MTJ) is a new magnetic storage device that shows significantly faster write speed and lower power, but no prior work has shown an embedded VC-MRAM array. In this paper, we present the first CMOS-integrated VC-MRAM. The VC-MRAM shows an ultra-fast 700ps switching time using 1.8V write voltage. The switching time has good uniformity, and 92% switching probability is achieved across the array. Reliability of $>10^{11}$ write cycles and read time of 8.5ns are demonstrated.

I. INTRODUCTION

Magnetic Random Access Memory (MRAM) has shown increasing popularity among mobile devices and micro-controllers as a replacement for last-level cache or Flash memory. Spin-Transfer-Torque (STT) MRAM provides dense and non-volatile storage solution, but its slow and power-hungry write operation makes it less advantageous than SRAM, despite its nonvolatility [1]. Its high switching current also limits its density due to the sizing requirement of the access transistor. Voltage-Controlled (VC)-MRAM, also referred to as Magneto-Electric RAM (MeRAM), is a promising candidate to drastically improve the write performance and array density [2,3], Fig.1. The voltage-based writing mechanism and high resistance of VC-MTJ (larger than 10x of STT-MTJ) allows the access transistor to be minimal size. The Voltage-Controlled Magnetic anisotropy (VCMA) effect at the interface of free and barrier layer allows the voltage to modulate the perpendicular field: H_{PMA} in Fig.1. The free layer's magnetization will precess under the torque from in-plane field and switches to the opposite state in <1ns. The free layer becomes stable at t_1 when voltage is removed. If the voltage pulse is long enough, the free layer's precession will be weakened by the damping effect and align with in-plane field at t_2. Previous works have demonstrated device properties of the VC-MTJ [3,4] and its application in true random number generator [5], but the devices are not integrated with CMOS. To demonstrate the feasibility of VC-MRAM as a high-performance embedded memory, we make the following contributions in this work: (1) Integrate VC-MRAM with CMOS technology through a foundry-and-lab collaboration; (2) Demonstrate an ultra-fast switching time of 0.7ns with 1.8V write voltage in a 8x8 memory array; (3) Study the variation of VC-MTJ's probabilistic switching behavior in an integrated array and demonstrate 92% switching probability under an un-calibrated and uniform pulse duration; (4) Demonstrate VC-MRAM array read time of 8.5ns and reliability of $>10^{11}$ cycles. The integration requires processing over the full wafer and the 180nm process is chosen for demonstration purposes due to its low cost.

II. INTEGRATION AND VC-MTJ DEVICE PROPERTY

The VC-MRAM film layer stack, Ta / Mo / CoFeB (free layer) / MgO / CoFeB / W / Co / Ru / [Co/Pt]$_n$ / Capping, was deposited by sputtering on 8 inch 0.18um CMOS wafer at room temperature. The 1 nm Mo insertion improves thermal stress stability and makes MTJ compatible with BEOL processing [6]. The deposited film was annealed at 360°C for 20 minutes. The critical MTJ pattern with diameter of 100 nm was defined by E-beam lithography. The integration flow is shown in Fig. 2. A closeup TEM image of the fabricated MTJ with clean sidewall is shown in Fig. 3(a), and Fig. 3(b) has a wider view showing the integrated MTJ together with the CMOS access transistor.

To characterize the VCMA coefficient, the stack with the same free layer but an in-plane reference layer was grown on the CMOS wafer. Effective perpendicular magnetic anisotropy (PMA) energy as a function of electric field [7] is shown in Fig. 3(d). VCMA coefficient is extracted by the slope of the plot as ~ 48 fJ/Vm, which is comparable with previous study [3-6]. Note that the same stack grown on Si/SiO$_2$ wafer shows higher VCMA coefficient ~ 65 fJ/Vm. The difference suggests high sensitivity of VCMA effect to the material interface, which is the major challenge of VCMA optimization. Recent studies achieve VCMA >100 fJ/Vm [8,9] by interfacial engineering on Si/SiO$_2$ wafer with CMOS compatible condition, however, dedicated process optimization on CMOS wafer might still be essential.

Individual devices on the same CMOS wafer are tested by probe station, giving tunneling magnetoresistance (TMR) ~ 120% and parallel resistance (RP) ~ 100kΩ. Note that the RP corresponds to resistance area product (RA) 800 Ωµm^2, ~50x higher than STT, which significantly reduces the write current. The switching tests were done by the standard set-up [2], showing writing error rate (WER) < 10^{-3} at 1.8 V.

III. INTEGRATED VC-MRAM CIRCUIT DESIGN

The MRAM circuit is designed for regular memory operations as well as direct measurements of the MTJs in the array to characterize array level device yield. The TEM image of Fig. 3(b) shows the bit-cell structure. The VC-MTJ device is built on top of VIA56. The source line (SL) of the

memory column connects to the fixed layer (top layer) of the MTJ through the CMOS VIA56, MTJ bottom electrode (BE), top electrode contact (TECT) and top electrode (TE). The bit line (BL) of the column connects to the free layer (bottom layer) of the MTJ device through VIA56 and BE. On the chip, the SL is on M1, the BL is on M5, and the word-line (WL) is on M4.

Fig. 4 shows a simplified block diagram of the memory array. The read is done through current sensing with added programmability and test modes to characterize the device. The MTJ state is determined by comparing it to a reference current that is the average of the MTJ current in the AP and P state. The MTJ read current is generated by grounding the SL (MTJ fixed layer) and applying a read voltage to the BL (MTJ free layer). For measurement of hysteresis curves, the current of each MTJ in the array can be mirrored off the chip.

For VC-MTJ write, standard write circuits used in prior STT-MRAMs cannot be employed. Instead, a sub-1ns unipolar write pulse needs to be applied to the VC-MTJ's fixed layer relative to the free layer. The write circuit has been designed to achieve a sharp and narrow pulse with a sub-100ps rise time and a programmable pulse width from 600ps to 5ns with a resolution of 100ps to allow testing of device behavior under different write conditions. To measure the pulse width, delay matched 'pulse start' and 'pulse stop' signals are sent off chip to an oscilloscope. To get high voltage drive capability, native thick-oxide MOS device was used as the access transistor. The word-line drivers were designed with deep-nwell thick-oxide inverters to provide the necessary negative voltage to turn off the native MOS device.

IV. CHIP MEASUREMENT RESULTS

The integrated VC-MRAM die is shown before and after MTJ processing in the micrographs of Fig. 5(a) and (b). The die was wire-bonded directly to a test PCB (Fig. 5(c)) and was tested with a variable magnetic field generator (Fig. 5(d)). We tested and characterized both the array level performance and direct device measurements.

A. Direct Device Measurements: The histogram of RP, RAP and TMR of 105 devices from two different dies is plotted in Fig. 6. A median cell TMR of 106% is achieved, which is close to the device TMR without access transistor and shows that the high MTJ resistance leads to lower TMR degradation. To determine retention, the thermal stability factor (Δ) was measured. The histogram of the stability factor and the offset perpendicular magnetic field (H_{off}) are shown in Fig. 7. Variation of Δ and H_{off} reduces the thermal stability, yield and switching probability, which needs to be improved further. Fig. 8 shows the magnetic hysteresis curves of 16 devices measured directly from devices in the array at polar angle $\theta \sim 60°$. Despite variations, there is a window where all devices can exist in both P and AP state. The external magnetic field is set accordingly to test the array performance.

B. Array Performance Results: Reliable switching is essential for memory writes. To test the switching, write pulses of different durations and voltages were applied to all devices and the MTJ state was read back through the memory access circuitry. The average switching probability was calculated for two different dies with 256 write attempts per device. Fig. 9 shows the average switching probability of devices from the two dies after removing outliers with switching probability of less than 80% at a pulse-width of 700ps (leaving 70 devices). Because PMA decreases with higher write voltage, the precession period also decreases. However, that higher voltage only changes the oscillation frequency but does not significantly change the peak probability, suggesting that 1.8V is sufficient for VCMA switching. For memories, it is not practical to calibrate the pulse width for each device, so it is important to achieve good device uniformity and select the best pulse width for the entire array. A distribution of the best pulse width (where switching probability is maximum for each device) for over 100 devices from two different dies is shown in Fig. 10. Most devices achieve their best switching probability at around 700ps. To test the array's endurance, 1.8V, 700ps write pulses at a rate of 1MHz are applied to one row of 8 devices, and the resistance of each device is measured intermittently. As shown in Fig. 11, the resistance does not show any sign of drift over 10^{11} write cycles, indicating excellent endurance. Fig. 12 shows the read shmoo plot. This was measured by forcing the MTJ state to AP/P with the external magnetic field and reading the expected state through the memory interface. Different write patterns were written to and read back from the MTJ array, as shown in Fig. 13. Non-volatility was confirmed by power cycling the chip in between write and read. A summary of the device performance is shown in the table of Fig. 14.

V. CONCLUSION

This paper demonstrated the first integration of the VC-MTJ device with CMOS and explored its array level performance. It achieved a 0.7ns switching time and demonstrated a faster and lower energy write than the STT-MRAM technology. While integration was done in 0.18um CMOS to keep wafer costs low, it can be extended to more advanced technology nodes. Through this demonstration, we showed the potential of VC-MRAM / MeRAM as an upcoming high density and energy efficient embedded memory. To further improve the switching probability, it is essential to reduce variation and increase thermal stability.

REFERENCES

[1] D. Edelstein *et al*., *IEEE International Electron Devices Meeting*.2022.
[2] S. Kanai, *et al. Appl. Phys. Lett* 101.12 (2012).
[3] C. Grezes *et al*., *Appl. Phys. Lett.* 2016.
[4] Y.C.Wu, *et al. IEEE Symposium on VLSI Technology*, 2020.
[5] J.Yang, *et al., European Solid-State Device Research Conference,* 2021.
[6] X.Li, *et al., Appl. Phys. Lett.* 107.14 (2015).
[7] T.Maruyama, *et al., Nature nanotechnology* (2009): 158-161.
[8] R.Carpenter, *et al., IEEE International Electron Devices Meeting*, 2021.
[9] Y.Shao et al. *Communications Materials* 3.1 (2022): 87.
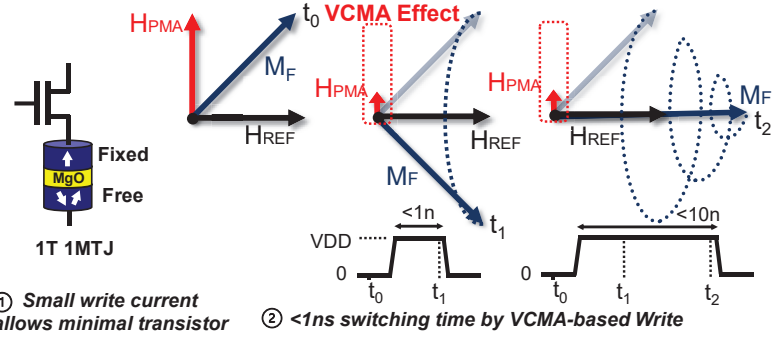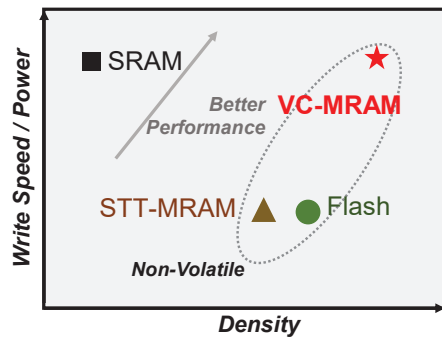[10] G. Hu et al. *IEEE International Electron Devices Meeting*. 2021.

Fig.1 Comparison between VC-MRAM and other memory technologies (left); 1T-1MTJ Cell (middle); Write operation: free layer (M_F) in parallel state at t_0; voltage modulates the H_{PMA} (by VCMA effect) and free layer precesses to the anti-parallel state at t_1; Free layer's precession is weakened and aligns with in-plane field at t_2.
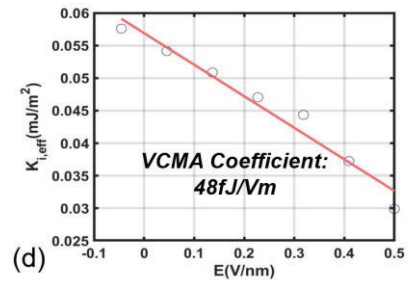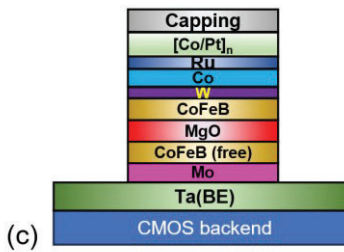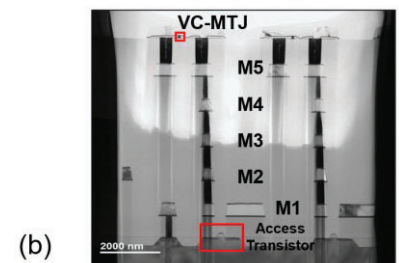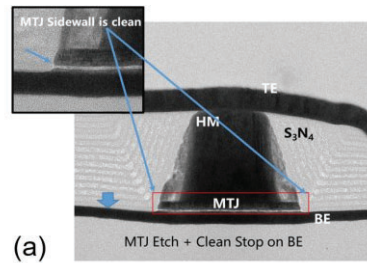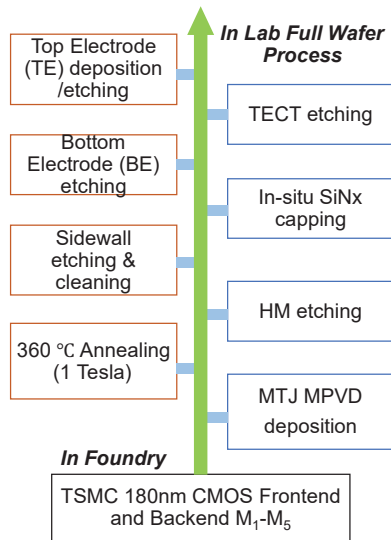

Fig.2 Integration flow: CMOS front end and M1-M5 backend fabricated in foundry then the whole wafer is processd in lab for MTJ film deposition and etching.

Fig. 3. (a)TEM image of the VC-MTJ and clean side walls; (b) CMOS backend TEM image of 1T-1MTJ cell; (c) material stack of the VC-MTJ; (d) Interfacial PMA vs electricl field plot of the VC-MTJ and the VCMA coefficient is extracted by the slope of the curve as 48fJ/Vm.
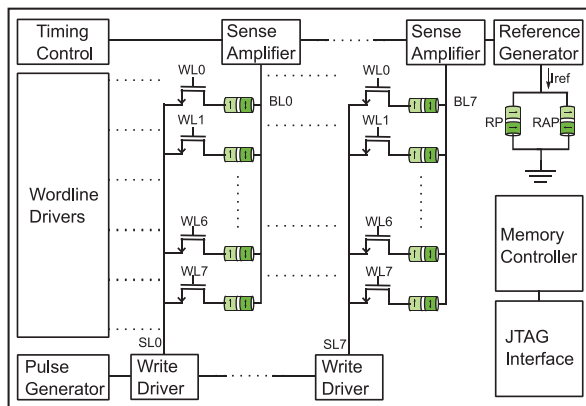

Fig. 4. Simplified block diagram of the memory system. The MTJs are arranged in an 8x8 pattern, and each of the 8 columns has a write driver and a read sense amplifier. All 8 sense amplifiers share a common reference. The timing control generates the control signals that transition faster than the digital clock. The chip communicates with a JTAG interface.
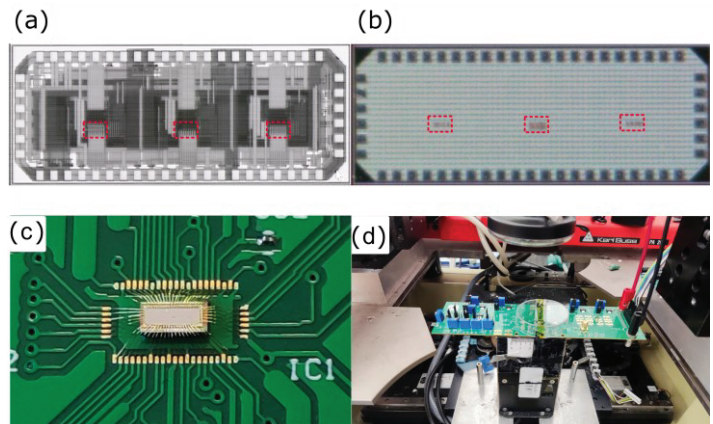

Fig. 5. (a) Chip micrograph before MTJ processing, (b) Chip micrograph after MTJ processing highlighting position of the MTJ arrays on chip, (c) memory chip wirebonded on the test PCB, and (d) Test PCB on top of a magnetic field generator.
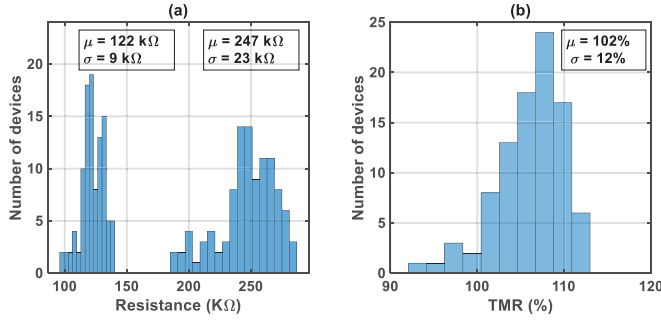
Fig. 6. (a) Measured distriubtion of RP and RAP, (b) Measured TMR distribution at a bias voltage of 100mV for two different dies. A median cell TMR of 106% is achieved.
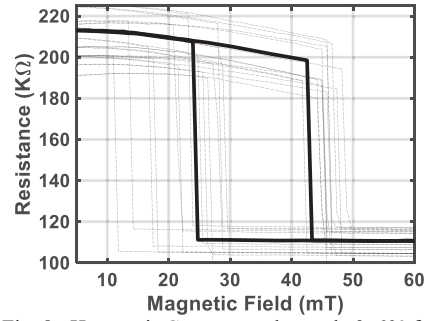

Fig.7. Measured distribution of (a) thermal stability factor delta, (b) offset magnetic field. Data are collected from one die.

Fig. 8. Hysteresis Curves at polar angle θ~60° for 16 devices from one die at a read voltage of 300 mV. A typical curve is highlighted.

Fig. 9. Measured average switching probability (with 1-sigma variation shaded) vs pulse width for devices from two different dies at a write voltage of (a) 1.8V and (b) 2.2V at an externally applied magnetic field of 37.5mT at θ~60°.
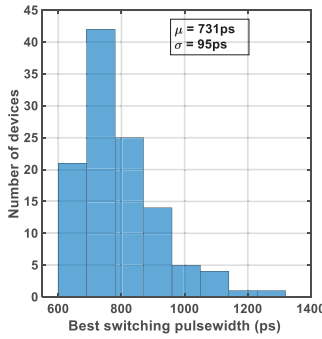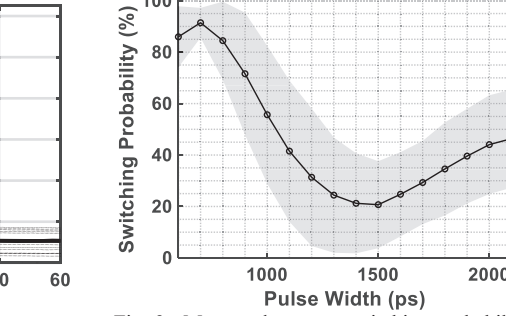

Fig. 10. Measured distriubtion of optimal pulse width for 113 devices from two different dies. Most devices show best switching at ~ 700ps.
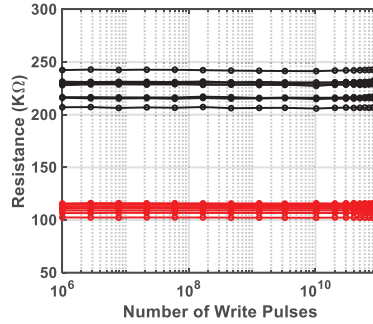

Fig. 11. Measured RP and RAP from 8 devices during application of 1.8V write pulses with a pulse width of 700 ps. RP or RAP does not change after 10^{11} write pulses, indicating good endurance.
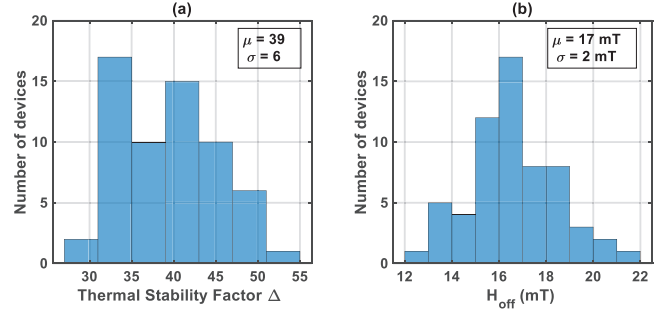
		Read Access Time (ns)											
Vread (mV)	VDD(V)	7.5	8	8.5	9	9.5	10	10.5	11	11.5	12	12.5	13
	2	P	P	P	P	P	P	P	P	P	P	P	P
	1.9	F	P	P	P	P	P	P	P	P	P	P	P
300	1.8	F	F	P	P	P	P	P	P	P	P	P	P
	1.7	F	F	F	P	P	P	P	P	P	P	P	P
	1.6	F	F	F	P	P	P	P	P	P	P	P	P
	2	F	F	F	F	P	P	P	P	P	P	P	P
	1.9	F	F	F	F	P	P	P	P	P	P	P	P
200	1.8	F	F	F	F	P	P	P	P	P	P	P	P
	1.7	F	F	F	F	F	P	P	P	P	P	P	P
	1.6	F	F	F	F	F	F	P	P	P	P	P	P
	2	F	F	F	F	F	F	F	F	F	P	P	P
	1.9	F	F	F	F	F	F	F	F	F	P	P	P
100	1.8	F	F	F	F	F	F	F	F	F	P	P	P
	1.7	F	F	F	F	F	F	F	F	F	F	P	P
	1.6	F	F	F	F	F	F	F	F	F	F	P	P

Fig. 12. Measured Read Shmoo Plot at different read voltages and VDD. A read time of 8.5ns is achieved at 1.8V VDD and 300mV read voltage.
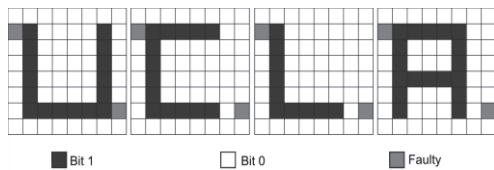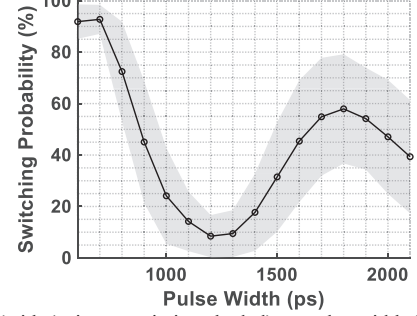

Bit 1   Bit 0   Faulty

Fig. 13. Bitmap of the 8x8 MTJ array read back from the chip after 4 different writes, consisting of the characters "UCLA". The chip is allowed multiple write attempts to write outlier devices which may have low switching probability. Two bits show faults – a stuck-at-1 fault due to incorrect device RP and a retention fault due to small hysteresis window. A bit 1 corresponds to RAP state and bit 0 corresponds to RP state.

Device Type	VCMA			STT
	This Work	VLSI 2020 [4]	IEDM 2021 [8]	IEDM 2021 [10]
Integration Level	CMOS Integrated	No Integration	No Integration	CMOS Integrated
MTJ Diameter (nm)	100	<75	100	38
RA(Ω*μm²)	800	Unknown	Unknown	~10
TMR Ratio	120%	246%	180%	120%
VCMA (fJ/Vm)	48	35	> 100	NA
Thermal Stability Factor, Δ	39	54	> 40	Unknown
Write Voltage(V)	1.8	Unknown	Unknown	0.7
Write Electric field (V/nm)	~1.1	1.4	Unknown	NA
Write time (ns)	0.7	0.9	0.6	3
Write energy (fJ/bit)	15	20	Unknown	~200
Endurance	>10^{11}	>10^{10}	>10^{8}	Unknown

Fig. 14. Summary of device performance achieved in this work. This work is the first demonstration of CMOS integrated VCMA showing memory operation. Compared to STT, this work demonstrates 4.3x better write speed and 13x better device write energy.