13.3 A 7Mb STT-MRAM in 22FFL FinFET Technology with 4ns Read Sensing Time at 0.9V Using Write-Verify-Write Scheme and Offset-Cancellation Sensing Technique

Liqiong Wei, Juan G. Alzate, Umut Arslan, Justin Brockman,
Nilanjan Das, Kevin Fischer, Tahir Ghani, Oleg Golonzka, Patrick Hentges,
Rawshan Jahan, Pulkit Jain, Blake Lin, Mesut Meterelliyoz,
Jim O'Donnell, Conor Puls, Pedro Quintero, Tanaya Sahu,
Meenakshi Sekhar, Ajay Vangapaty, Chris Wiegand, Fatih Hamzaoglu

Intel, Hillsboro, OR

STT-MRAM has been emerging as a very-promising high-density embedded non-volatile memory (eNVM) [1, 2]. Embedded Flash memory has been the leading eNVM technology, but STT-MRAM has been developed as a better solution for continuing scaling, speed and cost. This paper presents a write-verify-write (WvW) scheme and a programmable offset cancellation sensing technique that achieves a high-yield, high-performance and high-endurance 7Mb STT-MRAM arrays in a 22FFL FinFET technology [3]. The developed technology supports a wide range of operating temperatures between -40 - 105°C. Compared to prior-art [4,5], the two-stage current-sensing technique with a die-by-die tuning of a thin-film precision resistor that is used as a reference can significantly improve the sensing margin during verify and read operations. Read disturb for reference cells is eliminated as there is no MTJ in the reference path.

Figure 13.3.1 illustrates the organization of the 144KB STT-MRAM block. Each block has 4 subarrays sharing one charge pump circuit and control logic. The charge pump circuit can support a WL over-drive voltage during the write operation and an under-drive voltage at temperatures larger than 85°C. The 36KB subarray features 258b/BL and 288b/WL in a butterfly-array configuration. Subarray density, including the ECC bits, is 10.6Mb/mm^2. Each logical IO uses 32:1 column interleaving, with write drivers located on both sides of the 256-row column. The offset-cancellation sense amplifier with a programmable reference resistor (a thin-film precision resistor) is in the center of the column IO to balance the sensing margin of the two array sectors. Each 0.0486um^2 1T1MTJ bitcell is 216×225nm^2, with 2-polysilicon WLs, one Metal 4 BL, and one Metal 1 SL. To improve the WL slew rate, multiple straps are inserted to connect the two polysilicon wordlines to Metal 5. The cross-sectional TEM of the MTJ array embedded between Metal 2 and Metal 4 of a 9-Metal CMOS process is shown in Fig. 13.3.1. The tunnel-magneto-resistance ratio (TMR) of the MTJs is 180% at 25°C, with a target device critical dimension (CD) between 60-80nm [6].

Figure 13.3.2 shows the details of the write circuit using a programmable WvW scheme. WvW control signals are generated in mid-logic, where the write-to-verify time, the number of write pulses, the write pulse width, the write settings, and total write time can be programmed. The write current is determined by the write settings. The center write driver is located in the column IO, which has both strong- and weak-write settings, while the edge write driver only has strong settings to save area.

Figure 13.3.3(a) shows the relative write current versus the available 16 settings, and the maximum current at different WL locations. Writing from both sides of the array can significantly reduce the voltage drop due to interconnect resistance; therefore, the write current is observed to be independent of the row address. Figure 13.3.3(b) illustrates the WvW timing diagram: during verify, the read D_{OUT} is compared with the D_{IN} to be written, upon match the verify passes and the corresponding write driver is disabled; otherwise, it is enabled along with a boosted WL. In the example presented in Fig. 13.3.3(b), the first verify has a mismatch result; hence, the first write is attempted with a lower write current, but in the example, the cell is not switched. As a consequence, the 2nd verify fails and the circuit attempts a 2nd and stronger write pulse. Due to the stronger write current, the cell in this example is switched from a high-resistance state (AP-state) to a low-resistance state (P-state); this results in passing the 3rd verify and the write driver being disabled for the 3rd write.

Figure 13.3.4(a) shows the read-sensing circuit, with the offset-cancellation technique and its read-timing diagram (see Figure 13.3.4(b)). First, BL and SL are equalized and pre-discharged to 0V, and the sense amplifier output is charged to V_{CC}. Then, the WL is driven high for one cycle. When the temperature is higher than 85°C, WL underdrive is enabled to minimize the impact of unselected bitcell leakage to the current sensing margin. The first phase is evaluation, where PH1 is ON and PH2 is OFF. SL remains at 0V and BL is clamped low enough to minimize read disturb. C_{G0} and C_{G1} are gate capacitors. Voltage for C_{G1} is determined by the bitcell current and voltage for C_{G0} is given by the current of the reference resistor (R_{REF}). If the bitcell is in a high-resistance state (AP-state) and the cell current is smaller than reference resistor, $V_{CG1} > V_{CG0}$. The second phase is for amplification, where PH2 is ON and PH1 is OFF. Since the AP-state bitcell's current is weaker and the PMOS load is stronger due to a lower V_{CG0}, V_{OUT} goes high. Given that the same pair of PMOS load and NMOS clamp transistors are used during the two phases, the mismatch between data and reference branches, due to process variation, is cancelled and the sensing margin is significantly improved. Finally, SAE turns ON and the outputs of the final sense amplifier are rectified to supply levels. The reference resistor is made using a thin-film precision resistor, which is less sensitive to process variation. Die-by-die programming is utilized to optimize the reference resistance for maximum yield. A smaller address space is searched, with coarse and fine tunings at hot and cold temperatures, to find the optimal reference resistance values, within a limited test time. The bit error rate (BER) versus the deviation from the picked optimum R_{REF} setting for a given die is shown in Figure 13.3.4(c); showing that the selected R_{REF} setting is optimal and there is enough sense margin for both hot and cold temperatures.

Combining the read and write design techniques described above, the yield for MB-size arrays is significantly enhanced. Figure 13.3.5(a) shows that the resulting yield rate can be >99.997%, across temperature, at a supply voltage of 1.1V; this meets the ECC design target BER. The voltage-read sensing time (t_{SEN}) shmoo for 7Mb STT-MRAM arrays tested at various temperatures is shown in Figure 13.3.5(b). The design achieves a 4ns read sensing time at 0.9V. The array is also tolerant of a wide supply power range, down to 0.6V with an 8ns read sensing time.

Figure 13.3.6(a) provides the BER after 10^6 cycles of write with and without the WvW design technique, as well as the read disturb rate at various clamp voltages (V_{CLAMP}) (see Fig. 13.3.6(b)). No endurance driven errors are detected using the WvW design technique and no read disturb is observed after 10^{12}-read cycles at the optimal clamping voltages.

Figure 13.3.7 shows the die photograph of the test chip with eight 7Mb perpendicular-STT-MRAM arrays (56Mb/die total). The chip is fabricated in a 22FFL FinFET technology. The chip also implements an SRAM array, a PLL, an eFuse, a BIST and DDR IOs, showing that the integration of STT-MRAM has no impact on the functionality or yield of these blocks.

References:
[1] Y. Lee, et al., "Embedded STT-MRAM in 28-nm FDSOI Logic Process for Industrial MCU/IoT Application," *VLSI*, pp. 181-182, 2018.
[2] Y. J. Song, et al., "Highly Functional and Reliable 8Mb STT-MRAM Embedded in 28nm Logic," *IEDM Tech. Dig.*, pp. 663-666, 2016.
[3] B. Sell, et al., "22FFL: A High Performance and Ultra Low Power FinFET Technology for Mobile and RF Applications," *IEDM Tech. Dig.*, pp. 686-688, 2017.
[4] Y. Shih, et al., "Logic Process Compatible 40nm 16Mb, Embedded Perpendicular MRAM with Hybrid-Resistance Reference, sub-uA Sensing Resolution, and 17.5ns Read Access Time," *VLSI*, pp. 79-80, 2018.
[5] Q. Dong, et al., "A 1Mb 28nm STT-MRAM with 2.8ns Read Access Time at 1.2V VDD Using Single Cap Offset Cancelled Sense Amplifier and In-situ Self-Write-Termination," *ISSCC*, pp. 480-481, 2018.
[6] O. Golonzka, et al., "MRAM as Embedded Non-Volatile Memory Solution for 22FFL FinFET Technology," *IEDM Tech. Dig.*, 2018.
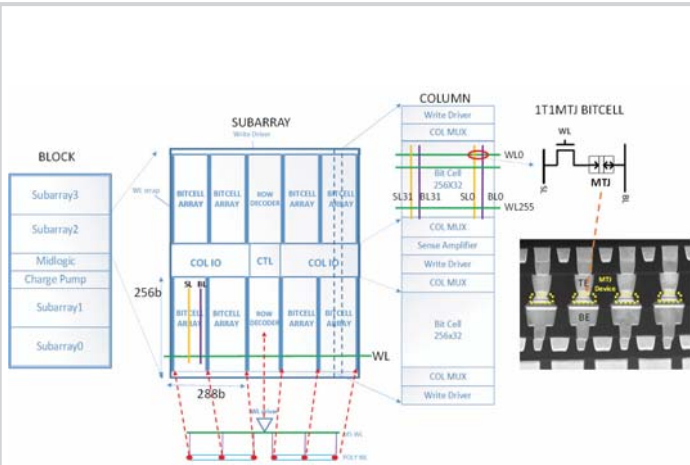
Figure 13.3.1: STT-MRAM block, subarray, column and bitcell diagrams with TEM of MTJ array.
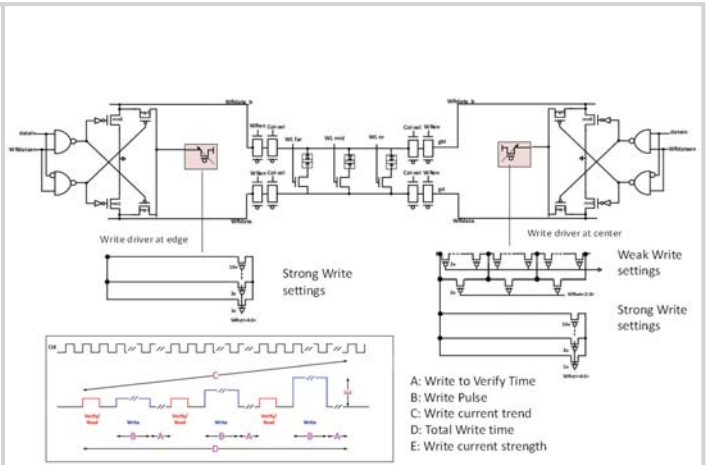

Figure 13.3.2: Write circuit with programmable write verify write scheme.
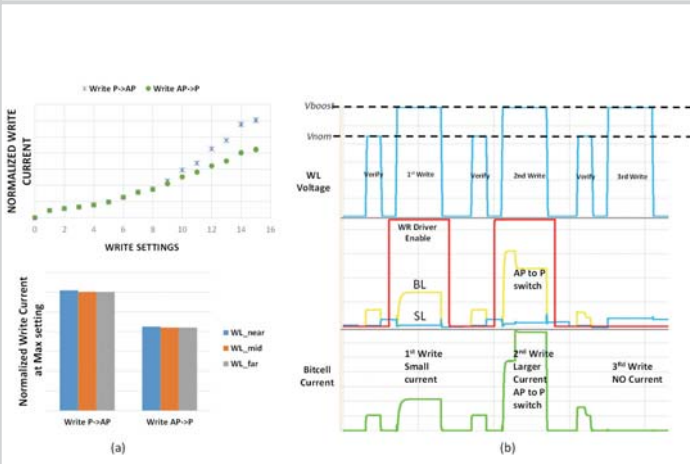

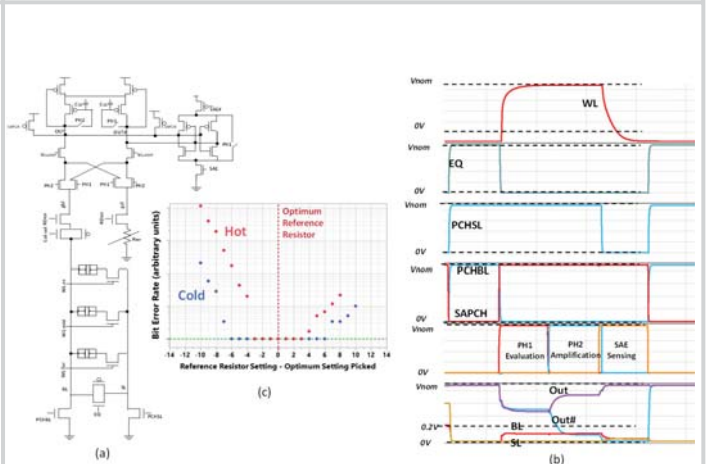Figure 13.3.3: Write current at different settings and WLs, and WvW timing diagram.

Figure 13.3.4: Read Sensing Circuit and Read Timing Diagram with BER at the deviation of optimal R_{REF} setting picked for a given Die.
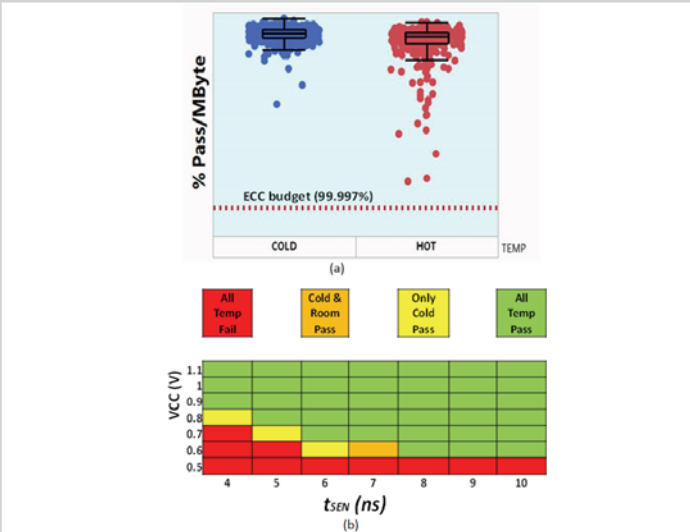

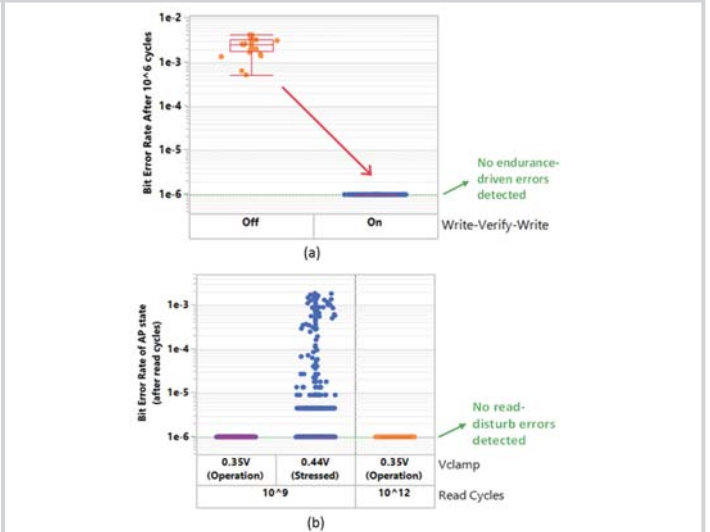Figure 13.3.5: Bit Yield at 1.1V for Hot and Cold and Read Shmoo plot for 7Mb Array with ECC enabled.

Figure 13.3.6: Write Endurance with and without WvW, and Read disturb at different clamp voltages.

13

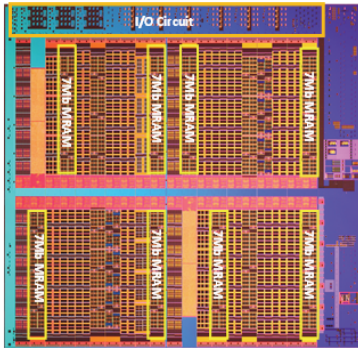| Technology	22FFL FinFET Technology
Memory	Perpendicular STT-MRAM
TMR	>180%@25C
Cell type	1T1MTJ
Cell size	$0.0486 \mu m^2$
Capacity	7Mb
Subarray Density (Incl. ECC bits)	10.6 Mbits/mm^2
Read Sense Time	4ns@0.9V, 8ns@0.6V
Bit Yield	> 99.997%
Retention	200C 10 years
Write Endurance	>1E06
READ Disturb	>1E12
Temp Range	-40C to 105C

Figure 13.3.7: Die Photo and Summary Table.