CHAPTER-4

Area-Efficient Auto-Write-Terminate Circuit for NV Latch and Logic-In-Memory Applications

The part of the work is adopted from-

J. Rajpoot and S. Verma, "Area-Efficient Auto-Write-Terminate Circuit for NV Latch and Logic-In-Memory Applications," ***IEEE Transactions on Circuits and Systems II: Express Briefs***, vol. 70, no. 7, pp. 2630-2634, July 2023, **doi: 10.1109/TCSII.2023.3242989**.

Abstract:

Spin transfer torque (STT) based magnetic tunnel junction (MTJ) device is a commercially appealing option for non-volatile latches and flip-flops. In this work, an STT-MTJ based non-volatile latch with auto-write-terminate (AWT) feature is proposed. The proposed latch has a simple structure, better stability, and higher speed, and is easy to integrate with CMOS logic styles. Also, no additional write driver circuit is needed, resulting in a smaller footprint while writing into the MTJs. Therefore, the proposed latch consumes less power than existing latches while using fewer transistors for a write operation and logic implementation. The proposed AWT circuit continuously monitors the write operation, prevents redundant MTJ writing, and removes excessive write current flow, thereby saving write energy. The AWT circuitry can save the total write energy by ~83 % when compared to the conventional write circuit. Also, the proposed circuitry for auto-write termination uses 75 % fewer transistors compared to previously proposed circuits.

4.1 Introduction

Magnetic tunnel junctions (MTJ) are being extensively explored for magnetoresistive RAMs (MRAMs), hybrid logic circuits, non-volatile computing, and logic-in-memory (LiM) applications due to the non-volatility of MTJ devices and CMOS compatibility. In MTJ-based logic-in-memory (LiM) architecture, MTJs are mounted over the CMOS devices, which shortens the distance between memory and logic blocks. This leads to a higher speed of data transfer and a decrease in power consumption. Furthermore, the non-volatile nature of MTJ-based computing architecture eradicates the need for extra power required for retaining data during standby mode in non-volatile computation [122], [186], [187], [188].

Amirany *et al*. [170] proposed a fully non-volatile and low-power full adder circuit by using MTJs, however, this technique faces the challenge of high area and power consumption as a separate write circuit is needed for writing/storing data in MTJs. Another non-volatile data path architecture, illustrated in **Figure 4.1**, stores data in the non-volatile (NV)-latch and accesses it as needed during logic implementation, which can be used to save power during standby mode in computation. NV-latches are essential entities in both LiM and novel NV computing architectures. Different NV latches have been proposed in recent years envisaged to fulfill the requirements of LiM and NV computing architectures. The NV-latch proposed in [171] is based on the separated latch and sensing circuit (SLS) structure that can consume a large area and energy due to separated sensing and writing circuits. Also, high sensing delay can be caused by the sensing operation being performed in two stages. The NV-latch proposed in [172] requires the use of a separate write circuit, which also increases area and energy consumption. While the NV-latch in [173] has an inbuilt read and write circuit, using the same path for reading and writing of the MTJ can increase read disturbance. Furthermore, existing NV-latches do not have a write energy-saving mechanism, and their design is unsuitable for integrating with an auto-write termination (AWT) technique. Therefore, a large amount of energy can be wasted during the write operation.

Figure 4.1 Non-volatile computing using MTJs

Both LiM implementation and non-volatile computation (shown in **Figure 4.1**) involve MTJs and associated write circuits. The wastage of write energy is a significant challenge that MTJ-based LiM and non-volatile architectures face. This wastage of write energy is due to the current flow even after the MTJ state has switched. The time required for writing into STT-MTJ varies significantly as the switching of MTJ is stochastic due to thermal fluctuations [152]. To ensure reliable writing, the write enable time for STT-MTJ is generally kept much longer than the average write time, around five times the average time [153]. It causes a waste of energy as the write current continues to flow even after the MTJ has switched and slows down the write operation. Thus, considering the challenges mentioned above of the existing latch and write circuits, this article mainly focuses on area-efficient auto-write-termination circuits for STT-MTJ based novel non-volatile latch, which are simple, area-efficient, and consume less power.

4.2 Proposed Non-Volatile Latch

The proposed NV latch has two MTJs connected symmetrically, and NMOS pass transistors (PT) connected to the clock (clk) for accessing both the MTJs, as shown in **Figure 4.2**. A cross-coupled inverter (CCI) is also used to store the read data at appropriate logic levels. Both MTJs will be in the opposite magnetization state, one in the P configuration and the other in the AP configuration. MTJs in P (AP) configuration represent logic HIGH(LOW) and vice versa. The read/write operation of the proposed latch is as follows:

Read operation: At $clk = 1$, transistors MP1, MN1, and MN2 get turned ON. When MTJ1 (MTJ2) are placed in a P (AP) arrangement, they have LOW (HIGH) resistance, which causes a LOW (HIGH) voltage drop across MTJ1 (MTJ2). This results in HIGH (LOW) voltage at the drain terminal of MN1 (MN2). PT1 (PT2) passes the output from node A (B) to $Q(Q_b)$, the input (output) of the CCI. The regenerative action of the CCI leads to a HIGH (LOW) voltage at $Q(Q_b)$. When the clk is back to LOW, all the NMOS transistors controlled by the clk are turned OFF, and the CCI is isolated, retaining the values of $Q = HIGH$ and $Q_b = LOW$ as long as the CCI has a power source.

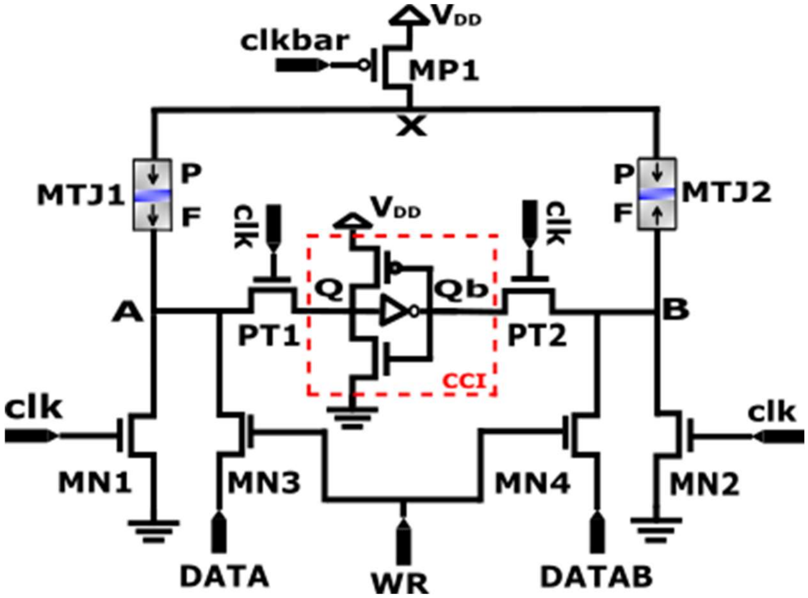


Figure 4.2 Proposed non-volatile latch circuit with inbuilt read and write circuit

Write operation: For writing/switching the MTJs, the WR goes HIGH at $clk = 0$, which turns ON the MN3 and MN4, and then by applying the appropriate input data logic at the node $DATA$ and $DATAB$, we can change the state of the MTJs from P (AP) to AP (P). If node $DATA$ is HIGH, the current flows from MN3 to MN4 through MTJs and node X, which changes the state of MTJ1 (MTJ2) from AP (P) to P (AP). Similarly, when node $DATA$ is LOW, current will flow from MN4 to MN3 via MTJs and node X, which changes the state of MTJ1(MTJ2) from P (AP) to AP (P).

4.3 Existing and Proposed AWT Approaches

Several ways to deal with the issue of energy wastage with a write termination scheme have been reported [152], [153], [154], [189], [190], [191], [192], [193], [194], [195]. A continuous write detection scheme proposed by Bi *et al.* [189] is used to detect the state of MTJ in cache memory. Zheng *et al.* [152] proposed a scheme to monitor both the MTJ states, but the use of two delay flip-flops incurs a large area overhead. Suzuki *et al.* [153] put forward a simple and attractive method that uses a multiplexer and XNOR gate to detect write completion. However, there is extra power consumption because the write completion detector keeps working even after the write enable generator is deactivated. In [189], [192], a periodic read operation is used to identify write completion. The termination can be ineffective sometimes since there is an inevitable time lag between the actual time of switching and its detection. Gupta *et al*. [194] show an improvement over the scheme proposed in [153], but the area overhead is still very large. In [195], authors propose a self-write-terminated driver for the hybrid STT-MTJ/CMOS circuits based on a logic-in-memory (LIM) structure. Here, the area overhead can be high due to the use of XOR logic, a MUX, and inverters for signal amplification. Almost all the mentioned schemes incur large hardware cost due to the usage of XOR logic and MUXes for data comparison in addition to buffers for signal amplification. Therefore, we propose a simple and cost-efficient write termination scheme for a novel non-volatile latch circuit presented in section 4.2. Here, a carefully designed inverter (INV1 in **Figure 4.3**) and a dynamic inverter (see **Figure 4.3**) automatically stop the write operation as soon as the intended data is written into the MTJ pair. Herein, two CMOS inverters, a dynamic inverter, and an AND gate are used to generate a write control signal. Some access transistors are also used to disable the circuit during read/sense, as shown in **Figure 4.3**. INV1 and INV2 sense and amplify the voltage at the node X of the NV-latch, as shown in **Figure 4.2** and **Figure 4.3**.

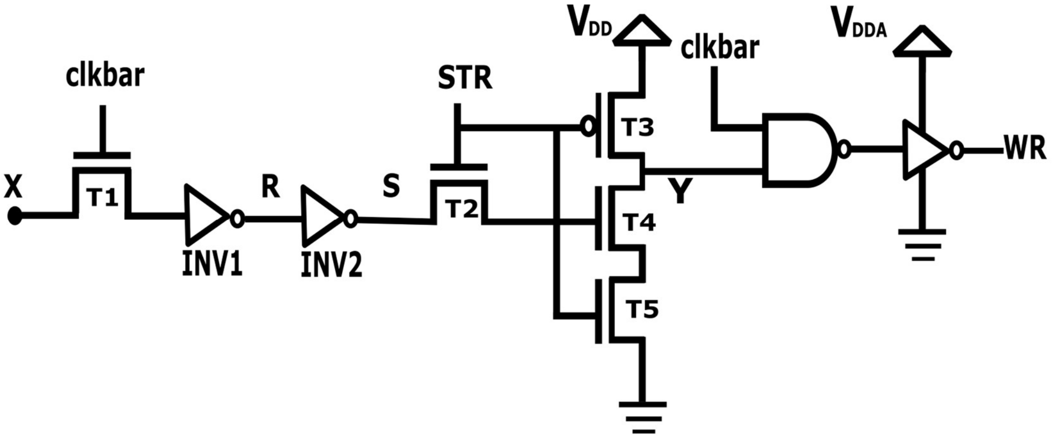


Figure 4.3 Proposed auto-write termination circuit for the non-volatile latch

The circuit is designed such that the LOW (HIGH) voltage drops at node X is less (more) than the switching threshold of the inverter INV1. Since the inverter INV1 amplifies the voltage at node X, the output (R) will be almost HIGH (LOW) while node S will be LOW (HIGH). The output S of the INV2 is fed to the dynamic inverter, which generates the write control signal. The write current can pass through the MTJs only when the transistors MN3 and MN4 (see **Figure 4.2**) are ON, which requires WR to be HIGH. To initialize the write operation, $clkbar$ is set to HIGH, and node Y is precharged to $\sim V_{DD}$ by setting $STR = LOW$ for a short duration of time, after which STR is set to HIGH to detect write completion. Since node Y is precharged to $\sim V_{DD}$, WR becomes HIGH such that the transistors MN3 and MN4 are turned ON, thereby initializing the write operation, as shown in **Figure 4.2** and **Figure**

4.3. Once the write operation is complete, the voltage at node X becomes HIGH due to the voltage divider effect. Once the voltage at node X is HIGH, the input to the INV2 (R) becomes LOW, and the output of the INV2 (S) becomes HIGH. Since STR is HIGH after the initial precharge, node Y gets discharged, and eventually, WR becomes LOW. Further, this turns OFF transistors MN3 and MN4, thereby terminating the write operation by cutting off the write current through the MTJs. The different cases are demonstrated in **Figure 4.4**, along with various input signals, changes in node voltages, and MTJ magnetization. The voltage at node X remains set to HIGH if the data being written and the stored data are the same or if the write operations are completed. For example, if MTJ1 (MTJ2) is in AP (P) and we want to write AP (P) to MTJ1 (MTJ2). In that case, $DATA = 0$ and $DATAB = 1$. The direction of current would be from node $DATAB$ to node $DATA$ via MTJs. So, the voltage drops at node X would become HIGH from the beginning since MTJ1 is in the AP state, and the write operation terminates immediately. **Table 4.1** further summarizes the cases discussed and other possibilities.

Table 4.1 Summary of Data Write Operation using AWT

Case	DATA	DATAB	MTJ1	MTJ2	Action
(a)	1	0	AP	P	TWC
(b)	0	1	AP	P	TI
(c)	1	0	P	AP	TI
(d)	0	1	P	AP	TWC

**TI: Terminate Immediately, **TWC: Terminate on Write Completion

4.4 Simulation And Results

This section discusses and demonstrates the circuit implementation of the proposed latch and AWT using the post-layout simulations that have been carried out in NGSPICE and Magic

IC Layout. Herein, an open-source industry standard skywater 130 nm process design kit (PDK) [196] has been used for the simulations along with the STT-MTJ SPICE model from [172]. The simulation parameters are given in **Table 4.2**. The supply voltage (V_{DD}) used for the simulations is 1.2 V, whereas a higher supply voltage (V_{ADD}) of 1.8 V is used in the write driver used to generate the write control signal WR. The simulation is performed without load using rise and fall time 100 ps for 100 ns.

Table 4.2 MTJ Model and Transistor Parameters and Their Variation

Parameters		**Parameters Description**	**Value**	**± σ**
MTJ	$l_x \times l_y$	Free Layer Dimensions	45×45 nm^2	6%
	t_F	Free Layer Thickness	1.48 nm	3%
	M_{S0}	Saturation Mag. at 0K	1300 emu/cm^3	--
	K_u	Crystal Anisotropy const.	621600 emu/cm^3	--
	P_0	Polarization Factor at 0K	1	3%
	T_0	Temperature	300 K	--
	γ	Gyromagnetic Ratio	$1.76*10^7$ $Oe^{-1}s^{-1}$	--
	α	Gilbert Damping Coeff.	0.04	--
	H_k	Uniaxial Field Strength	100 Oe	--
	RA_0	Resistance-Area Product	4 $\Omega\mu m^2$	6%
	R_P	Parallel Resistance	1.95 kΩ	--
	R_{AP}	Anti-parallel Resistance	3.70 kΩ	--
Transistors	Channel Length		150 nm	--
	Channel width	MP1	2000 nm	--
		MN3, MN4	3000 nm	--
		All PMOS and T4, T5	900 nm	--
		All NMOS and T3	450 nm	--

*Minimum channel length in the skywater 130 nm PDK is 150 nm
*σ = variation

To verify the operation of the proposed non-volatile latch with AWT, we performed the simulation in four modes as demonstrated in **Figure 4.4**, *i.e.*, normal, backup (write and termination), sleep (power down), and restore mode. In normal mode, the clock signal (clk) is set to ground, thereby turning OFF the pass transistors (PT1 and PT2) and isolating the CCI from the MTJs. Backup mode is performed to store data in the MTJs corresponding to input data at the node $DATA$ and $DATAB$. The power supply voltage is cut-off after writing data to MTJs in sleep mode. Restore operation is performed to retrieve data from the MTJs to the latch at nodes Q and Q_b after power is turned ON.

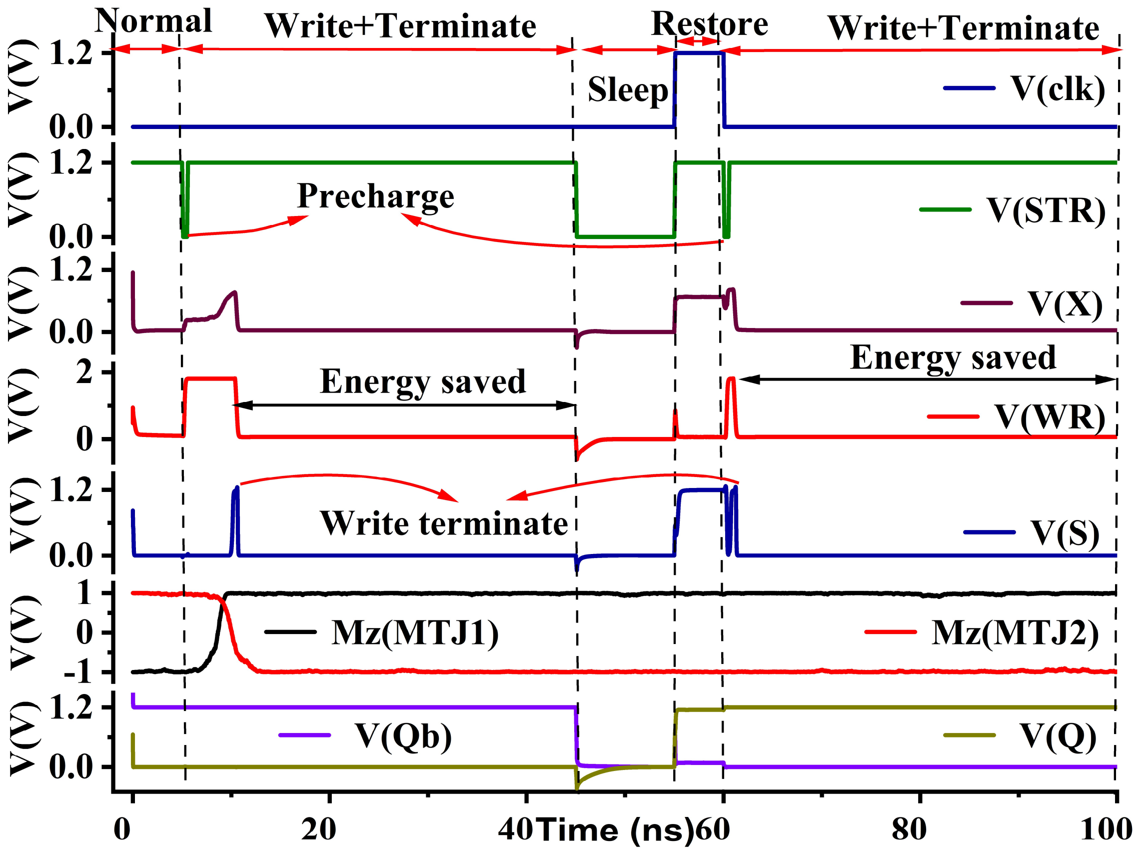


Figure 4.4 Timing diagram demonstrating all possible operations performed by integrating AWT with the proposed latch. The first write operation performed for data to be stored is not the same as stored data. This corresponds to case (a) in **Table 4.1**. The second write operation for data to be stored is the same as stored data to show immediate write termination corresponding to case (c), as given in **Table 4.1**.

Backup mode (5ns-45ns): Here, initially, MTJ1 (MTJ2) is in the AP (P) state, *i.e.*, case (a), as given in **Table 4.1**. $DATA$ is constantly held at HIGH to switch MTJ1 (MTJ2) from AP (P) to P (AP) state. Backup mode begins with clock (clk) being set to LOW and initiated by setting $STR = LOW$ for a small duration (400 ps) during which WR gets precharged to $\sim V_{DD}$. The voltages at nodes R and S change according to the change in voltage at node X during the write operation. The inverter INV1 amplifies the voltage difference created at node X due to the voltage divider effect, which results in a higher sense margin and more reliable write termination operation. A high voltage drops at node X marks the completion of the write operation. Once the write operation is complete, the WR discharges, thereby terminating the current flow through the MTJs and start saving write energy, as shown in **Figure 4.4**. MTJ1 (MTJ2) will now be in P(AP) state such that $Q_b(Q)$ must be LOW (HIGH) during the restore operation. The second write operation is performed between 60 ns and 100 ns when the data to be stored in MTJs is the same as input data, where the WR signal immediately discharges just after STR becomes HIGH. Hence, terminating the write operation immediately saves a significant amount of write energy.

Restore mode (55ns-60ns)*:* It is performed at $clk = HIGH$. The timing diagram shown in **Figure 4.4** demonstrates the restore operation performed during a time interval of 55 ns to 60 ns after sleep mode. The stored data in the MTJ1 (MTJ2) is in a P (AP) state, MTJ1 is in a low resistance state, and MTJ2 is in a high resistance state such that the voltage drop at node A is higher than at the node B. The HIGH (LOW) voltage drop at node $A(B)$ is greater (less) than the switching threshold of the inverters in the CCI, which causes $Q = HIGH$ and $Q_b = LOW$.

4.5 Performance Analysis of the AWT Circuit

In this section, we analyze the performance of the proposed AWT in terms of power consumed, speed, area occupied, and reliability. The proposed AWT circuit for the latch is able to significantly cut-off wastage of write energy. The amount of time/energy needed to switch STT-MTJ from P to AP is much higher than from AP to P state. Moreover, the write pulse

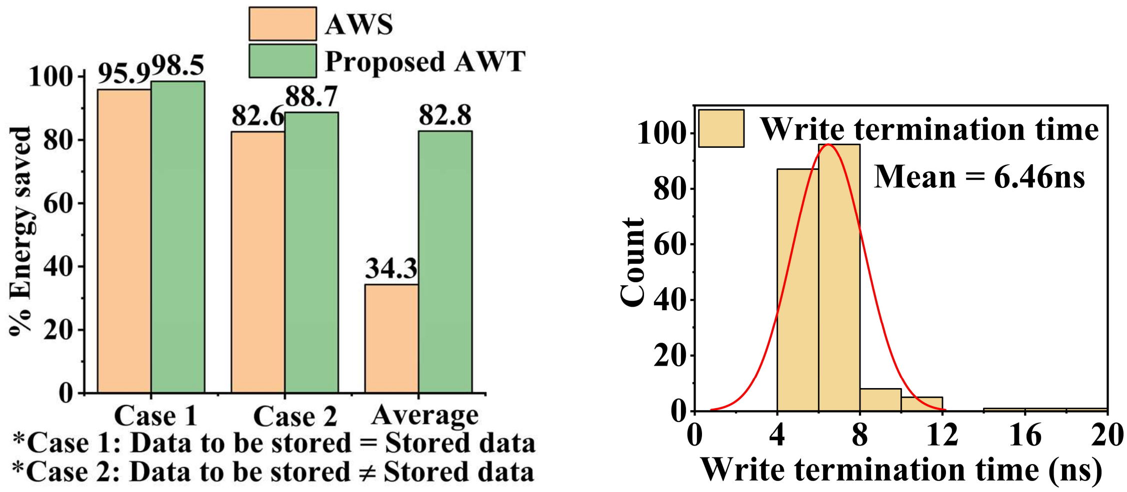
(40 ns) duration is kept sufficiently higher, *i.e.*, five times the average write time as in [153], to ensure reliable writing. We analyze the proposed and existing AWT circuits [195] by calculating the amount of write time the AWT is able to save by cutting off the flow of current. The proposed AWT circuit is able to considerably reduce the transistor count as compared to previous schemes for self-write termination. It is worth noting that the proposed AWT circuit can be integrated with any CCI-based latch, like the NV SR latch [154] or the PCSA-based latch [142].

Table 4.3 gives the performance overview of the proposed AWT circuit as compared to the conventional write circuit and existing write termination scheme referred to as the auto-write-stop (AWS) for LiM architecture [195]. The reduction in total write energy per bit achieved for both bit '1' and bit '0' is 82.8 % from the conventional circuits and 73.9 % from the AWS [163]. Furthermore, a significant amount of energy is saved when the data to be stored is the same as the data already stored, *i.e.*, 98.5 % from the conventional write and 63.4 % from the existing AWS write circuits [195]. However, the control circuitry consumes 80.7 % less energy than the AWS scheme. Additionally, the write control circuitry in the AWT uses 52 fewer transistors than the AWS [195] scheme. Furthermore, a comparison of the percentage of energy saved with the existing write-stop circuit [195] in different cases is shown in **Figure 4.5(a)**. The charging of node Y to generate a write control signal is also fast since the charging does not depend on any propagated signal. The discharging of node Y and, hence, the write termination is quick since the circuit continuously monitors the voltage at node X. As a result, the proposed AWT is fast, compact in size, and energy efficient.

Table 4.3 Performance Comparison of Auto-Write Terminate Circuits

Particulars		**Conventional write circuit**	**AWS** [195]	**Proposed AWT**
Self-write termination		No	Yes	Yes
Monitoring process		No	Yes	Yes
Total number of transistors		20	69	17
Control circuitry energy/bit (fJ) ^{P}		1406.2	4176.1	802.6
Write energy/bit (fJ)	bit '1' or '0'	6163.2	1069.8	697.6
	IN=Bit stored in MTJs	6163.2	248.4	90.72
	Average ^{Q}	6163.2	796	495.1
Total write energy/bit (fJ) $^{P+Q}$		7569.4	4972.1	1297.7
Energy Saved	Data to be stored = Stored data (%)	0	95.9	98.5
	Data to be stored ≠ Stored data (%)	0	82.6	88.7
Total energy saved (Avg. energy/bit) (%)		0	34.3	82.8

The AWT circuit is also reliable and has been verified for different values of TMR and RA products of the MTJs. We performed 200 Monte-Carlo simulations for the write operation Under a 3 % variation in free layer dimensions, initial polarization, and *RA* products of the MTJ. Monte-Carlo simulation results show that the AWT circuit terminates 100 % of the writing process successfully, with an average write termination time of 6.46 ns, as shown in **Figure 4.5**(b).



(a) (b)

Figure 4.5 (a) Comparison of energy saved (%) with existing auto write stop (AWS) circuit from [195] in different write cases. (b) Variation of write termination time using AWT with 3 % variation in MTJ parameters. The simulations are performed for case (a), corresponding to **Table 4.1**.

Table 4.4 Comparison of Non-volatile Latch with existing NV latch

Particulars	[172]	[173]	**This work**
Device count	7T+2MTJ	11T+2MTJ	11T+2MTJ
Write circuit required	Yes	No	No
Write termination	No	No	Yes
MTJ stress time	85	85	12.3
Sensing current (µA)	43.5	84	48.85
Sensing delay (ps)	230	850	90.5
Write current (µA)	87.5	92	90
Write delay (ns)	40	40	6.46
Redundant write delay (ns)	40	40	0.84
Write energy/bit(pJ)	6.3	4.42	0.697
Redundant write energy/bit (pJ)	6.3	4.42	0.091
D-Q delay (ps)	190	240	204.5
Total power (µW)	142.08	174.81	39.76
PDP (fJ)	26.9	41.9	8.13
Layout area (μm^2)	77.34	61.13	57.24
Area-delay product($\mu m^2 \times ns$)	14.69	14.67	11.70
Read failures	0	16	0

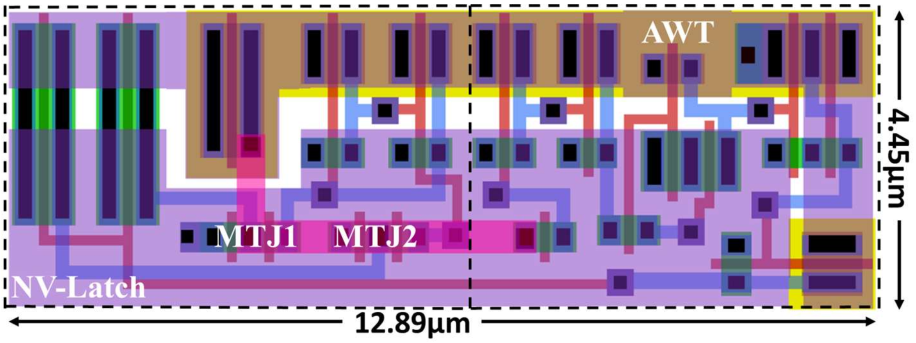
*All are average values for reading or writing logic 0 and logic 1

4.6 Advantages of Proposed Non-volatile Latch

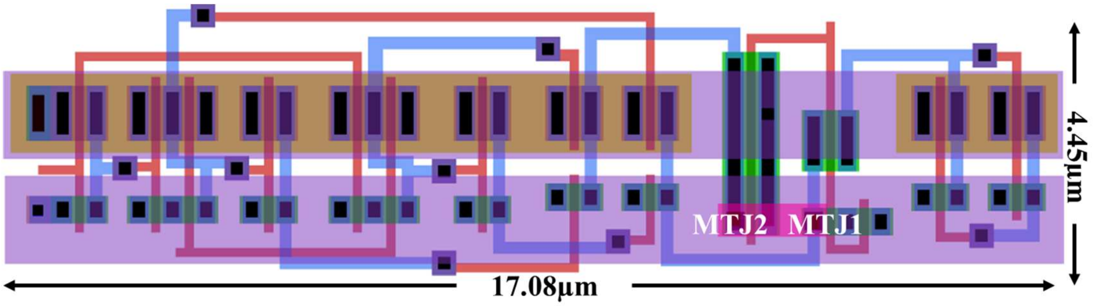
Table 4.4 summarizes the comparison of the proposed non-volatile latch with the recent work in [172] and [173]. The performance comparison of the proposed latch is done in terms of die area, sensing delay, write delay, and write energy consumption. Herein, we have implemented our proposed work in an advanced 130 nm technology node by using an open-source skywater 130 nm process design kit (PDK) [196] and compared our results with the recent works [172], [173]. For a fair comparison, we have implemented the existing latch design from [172] and [173] using 130 nm technology in our simulation setup. The proposed non-volatile latch offers 2.54x, 6.2x, 47.6x, 9x, 70x, 3.58x, 3.30x improvement with latch [172] and 9.4x, 6.2x, 47.6x, 6.4x, 48.5x, 4.4x, 5.2x, improvement with latch NVFF-II [173] in sensing delay, write delay, redundant write delay, write energy/bit, redundant write energy/bit, total power, and PDP respectively. Further, the smaller layout area (with write circuit or AWT) of 57.24 µm^2 highlights the main advantages of the proposed latch by eliminating the need for write circuits while maintaining a separate read and write path of MTJs. Further, the proposed design is suitable for integrating the auto write termination technique easily. Hence, low-power consumption, less area overhead, and elimination of read disturbance can be achieved. The integration of AWT not only cuts down the wastage of write energy but alleviates the endurance issues for the MTJ. The flow of unnecessary write current into MTJ can break down the dielectric barrier which leads to endurance issues.

4.7 Layout Implementation of Proposed Latch and Existing Latches

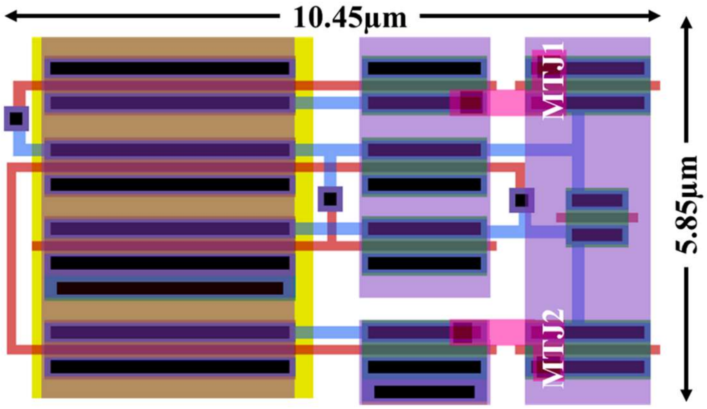
The layout implementation of the proposed latch with AWT and existing latches are demonstrated in **Figure 4.6(a)** and **Figure 4.6(b)-(c)**, respectively. The MTJs are placed between metals 3 and 4, as in [197]. The proposed latch with AWT has only 26 transistors with an area of 57.24 µm^2 with a significant improvement in overall footprint area. **Figure 4.6(b)** shows the layout of the latch from [172] with a conventional write circuit that has 35 % more area, *i.e.,* 77.34 µm^2. Herein, skywater 130 nm technology is used for the layout implementation. The layout implementation of the latch proposed in [173] has an area of 61.13 µm^2, as shown in **Figure 4.6(c)**. It is worth noting that both the implementation in [172] and [173] do not use write termination.



(a)



(b)



(c)

Figure 4.6 Layout implementation of (a) proposed NV-latch with AWT, (b) NV-latch with write circuit in [172], and (c) NVFF-II [173] with inbuilt write circuit.

4.8 Conclusion

To prevent the redundant MTJ writing and excessive current flow in the novel NV-latch during the write operation, we presented novel latch and AWT circuits, which made our implementation power and area efficient for LiM and NV computing applications. Herein, only 17 transistors were used for writing and controlling the write operation. The proposed AWT saved approximately 77 % of total writing power and our implementation was 26 % more area efficient.