

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Area-Efficient Spin-Orbit Torque Magnetic Random-Access Memory

Karim Ali

Abstract

Spin-orbit torque magnetic random-access memory (SOT-MRAM) has shown promising potential to realize reliable, high-speed and energy-efficient on-chip memory. However, conventional SOT-MRAM requires two access transistors per cell. This limits the use of conventional SOT-MRAM in high-density memories. Thus, various architectures in the literature have been proposed to improve the area efficiency of the SOT-MRAM. In this chapter, these proposals are divided into two categories: non-diode-based SOT-MRAM and diode-based SOT-MRAM cells. The non-diode-based proposals may result in a 1-bit effective area saving up to 50% compared to the conventional SOT-MRAM, whereas the diode-based designs may result in 1-bit effective area-saving of up to 75%. However, the area saving may be accompanied by higher energy and reliability issue penalties. Therefore, here, the various proposals in the literature are presented, highlighting the pros and cons of each design. Moreover, the technology requirements to realize these proposals are discussed. Finally, the various designs are evaluated from both cell and system level perspectives.

Keywords: SOT-MRAM, MOM diode, SLC, MLC, MTJ

1. Introduction

In the era of the Internet of things (IoT), ultra-low power and energy-efficient computing become essential. The expected large number of the IoT nodes forces the limitation of their power budget [1]. In particular, for the majority of the IoT nodes that are powered by energy harvesters, their anticipated power budget can be as small as sub μ Watt [1]. Thus, employing nonvolatile (NV) memory in these nodes would aid in reducing their power consumption. This is because leakage power constitutes a significant percentage of the total power consumed due to the long idle durations. Hence, the memory can be power gated, thanks to its nonvolatility, and consequently eliminates its leakage contribution. In addition, IoT nodes powered by energy harvesters may suffer from multiple durations of power discontinuities as a result of using the unreliable power source. Therefore, having NV memory aids to continue the computation from where it's stopped without restarting every single operation once the power goes down [2], which aids in increasing the overall performance as well. Furthermore, the different modules on these nodes need to be area efficient. Area efficiency aids in decreasing the overall parasitic contribution,

which enhances the performance and energy efficiency. More importantly, the smaller the silicon area consumed by the IoT node, the lower the overall cost. The cost is an important metric for the success of a specific IoT design as it is targeted to have these IoT chips with prices as low as 50 cents. Thus, a combination of improved technologies and circuits is needed to achieve energy- and area-efficient designs.

Emerging devices such as a magnetic tunnel junction (MTJ) may be used to implement a nonvolatile memory, which would be called magnetic random-access memory (MRAM). An MTJ, highlighted in **Figure 1**, comprises two ferromagnetic (FM) layers separated by a tunneling oxide barrier. One FM layer has a pinned magnetization (i.e. pinned layer (PL)), while the other has a free magnetization (M_F) (i.e. free layer (FL)). Manipulating the direction of the FL to be either parallel (P) or anti-parallel (AP) to the PL magnetization direction determines the electrical resistance state of the MTJ to be either low resistance (R_P) or high resistance (R_{AP}), respectively. An MTJ offers different benefits like nonvolatility, programmability, high endurance, long state retention time and compatibility with CMOS fabrication flow. Consequently, the MRAM inherits the MTJ advantages, which makes MRAM a standout solution to implement a nonvolatile memory that is suitable to replace or co-exist with these leaky charge-based memories in both on-chip and off-chip memory.

The MRAM can be classified based on the writing technique employed to switch the MTJ resistance state. These techniques, such as field-induced magnetization reversal (FIM) [3], spin-transfer torque (STT) [4, 5], voltage-controlled magnetic anisotropy (VCMA) [6] and spin-orbit torque (SOT) [7–11], result in the different corresponding MRAM types like FIM MRAM, STT MRAM, VCMA MRAM, and SOT MRAM, respectively. FIM requires applying an external magnetic field to program the MTJ, which hinders its scalability and increases its energy consumption, while STT, VCMA, and SOT are known to be highly scalable [9]. STT programming has a common current path for both writing and reading the MTJ, whereby the write current (I_{write}) must flow through the MTJ directly. This results in higher writing voltage requirements, due to the MTJ high resistance, and device reliability degradation. Similarly, VCMA demands voltage application across the MTJ, which subjects it to reliability issues such as tunnel barrier breakdown. The FIM, STT, and VCMA technologies use the MTJ device independently in its two-terminal form, shown in **Figure 1(a)**. On the other hand, the SOT technology requires placing the MTJ over a heavy metal (HM) electrode from the FL side, which results in a three-terminal device as depicted in **Figure 1(b)**. In SOT technology, a charge current passes through the HM electrode that in return results in spin accumulation in the FL of the MTJ due to the spin-orbit interaction, which assists in the reversal of the FL magnetization to either P or AP state depending on the current direction relative to the FL easy axis. Thus, SOT programming solves

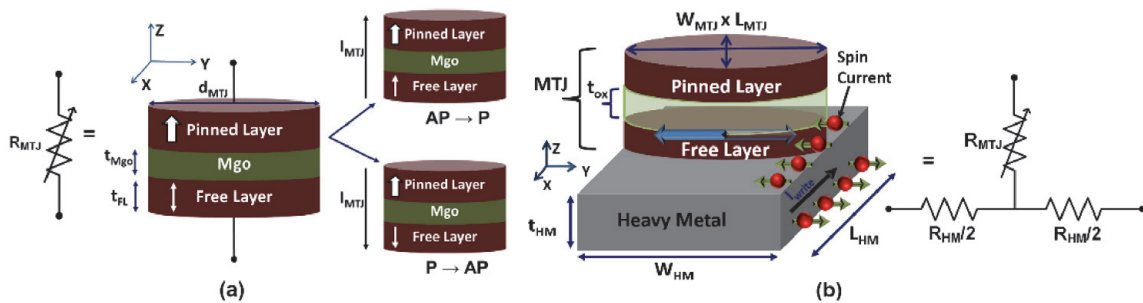


Figure 1.

(a) Two-terminal MTJ with two magnetization configurations reflects two resistance states depending on the write current direction if STT technology is used. (b) Three terminal SOT-MTJ.

the aforementioned issues [9]. Firstly, it features high energy efficiency due to its low critical current requirements and fast switching speed. In particular, due to the SOT-MRAM high switching speed (i.e. high performance) that can be down to 100 s of ps [12, 13], in addition to its nonvolatility and smaller cell area, it becomes more appealing to replace SRAM in cache memory applications. Secondly, the I_{write} flows through the low resistance heavy metal (HM) rather than the MTJ. This improves the device reliability, permits the usage of smaller write voltages, and increases the voltage headroom margin for the write transistor, which relaxes the current source design. Finally, the write and read paths are separated, which leads to a more optimized design as it permits satisfying the contradictory requirements for the access transistors sizing in both read and write modes. In the write mode, larger access transistor is required to supply higher I_{write} (at least larger than SOT-MTJ critical switching current), whereas in the read mode, smaller access transistor is required to reduce the read current to avoid read disturbances. Consequently, SOT-MRAM is considered as the most promising MRAM to realize reliable, high speed, and energy-efficient on-chip memory [14, 15].

However, the separation of the read and write paths comes with a penalty of requiring two access transistors per cell to fully isolate both paths of the nonselected cells in the conventional approach of implementing the SOT-MRAM, presented in **Figure 2(a)**. This increases the 1-bit effective area of the conventional SOT-MRAM, which limits its use in high-density memories. Thus, various SOT-MRAM cell designs in the literature have been proposed to improve the area efficiency of the SOT-MRAM. In this chapter, these proposals in the literature are going to be presented, highlighting the advantages and disadvantages of each design. The various proposals are divided into two main categories, which are diode-based and non-diode-based SOT-MRAM. Moreover, the technology requirements to realize these proposals are discussed. Finally, the proposals are evaluated from both cell and system level perspectives.

The chapter is organized as follows: Section 2 presented the nondiode-based SOT-MRAM proposals discussing their operation, pros, and cons. Similarly, the operation, pros, and cons of the various diode-based SOT-MRAMs are illustrated in Section 3. Section 4 evaluates the various diode and non-diode-based SOT-MRAM proposals from both cell and system-level perspectives. Finally, Section 5 provides the concluding remarks.

2. Nondiode-based SOT-MRAM

In this section, the various SOT-MRAM proposals in the literature that do not require employing a diode (or selector) in the cell are presented. Nondiode based SOT-MRAM cells only rely on access transistors for current flow control. Avoiding using diodes offer an advantage of a simpler fabrication process and may achieve higher energy efficiency as lower read voltages may be applied. The nondiode-based SOT-MRAMs in the literature include the conventional single-level cell (SLC) SOT-MRAM, as shown in **Figure 2(a)**, and two multi-level cell (MLC) proposals depicted in **Figure 2(c)** and **(d)**.

2.1 Conventional SLC SOT-MRAM

The conventional SOT-MRAM, depicted in **Figure 2(a)**, as discussed in various works [15, 14] requires two transistors to access one bit per cell. One of the transistors is used to control the write current flow (i.e. write Tx shown in **Figure 2(a)**) and the second transistor is required to control the read current flow (i.e. read

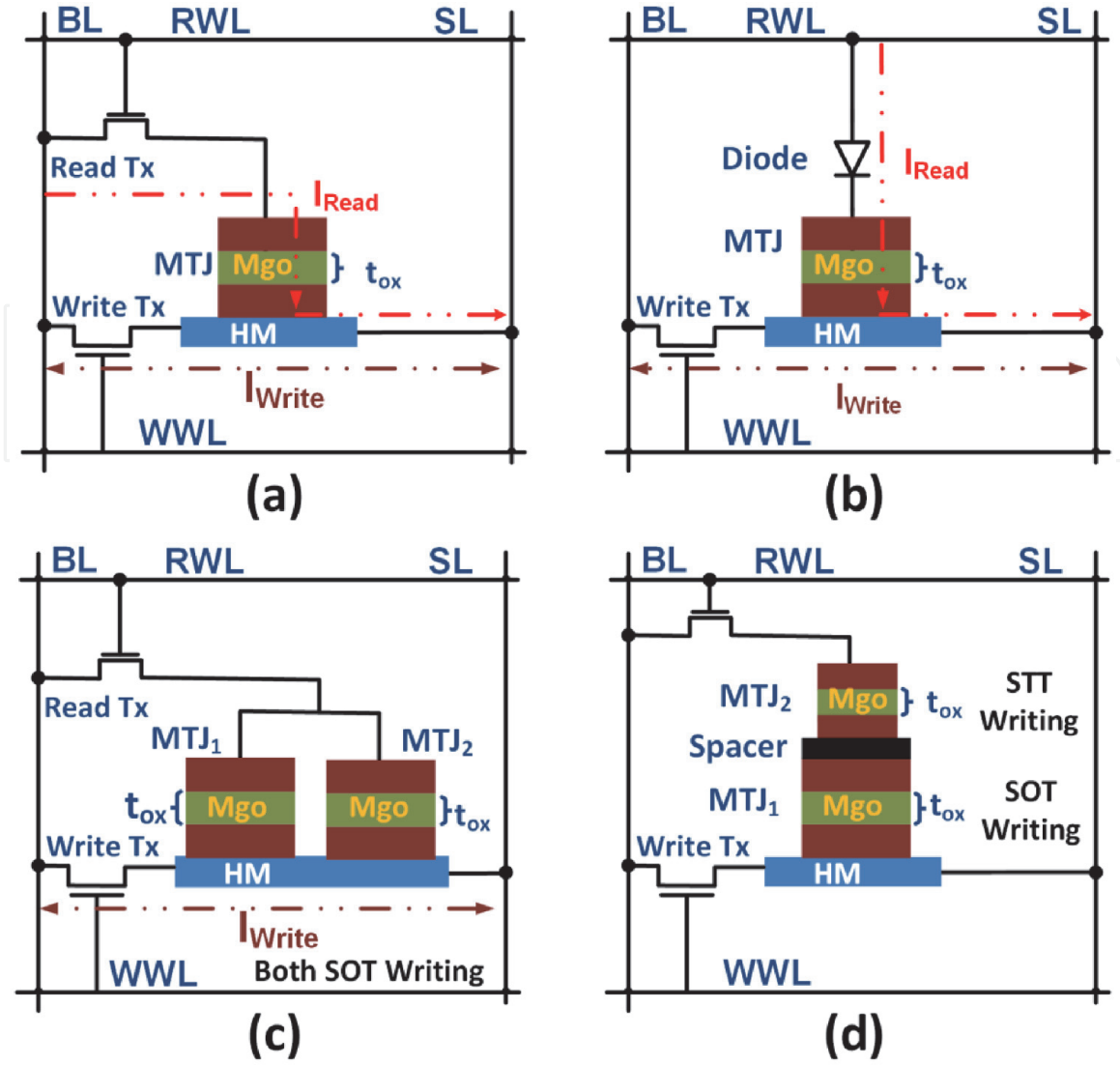


Figure 2. Various SOT-MRAM proposals in the literature. (a) Conventional SOT-MRAM [15], (b) Schottky diode-based SLC SOT-MRAM [14], (c) P-MLC [15], and (d) S-MLC [15].

transistor). The two transistors are essential to fully isolate the nonselected bits to avoid nonintentional read and write of these bits when not selected. Furthermore, the existence of a single MTJ per heavy metal (HM) electrode to be programmed for each cell write operation maintains the high energy efficiency of SOT technology. In addition, the single-level cell sensing scheme for this cell enhances the distinguishability, bit error rate (BER), and read energy consumption. Consequently, the conventional SLC SOT-MRAM design, shown in **Figure 2(a)**, represents the baseline of the energy consumption for the SOT-MRAM. However, the usage of two transistors to access a single bit increases the 1-bit effective area significantly, which is estimated to be $69F^2$ using the design rules in [16] that is used in the area estimation throughout this chapter. This limits the application of the conventional SOT-MRAM as it will not be suitable for high-density memory applications. Thus, further SOT-MRAM cells are proposed in the literature to reduce the 1-bit effective area with an expected penalty in the various performance and energy metrics of the SOT-MRAM as discussed in the following sections.

2.2 MLC SOT-MRAM

Kim et al. [15] proposed to employ a multi-level cell (MLC) SOT-MRAM instead of the single-level cell (SLC) SOT-MRAM to reduce the 1-bit effective area. In the

MLC memory, each cell comprises two bits (i.e. two MTJs exists per cell) instead of only 1-bit (1 MTJ) in the SLC. This means that the two MTJ resistance combinations should represent four different resistance states instead of only two resistance states in the SLC. In this MLC design, the two bits are accessed by two transistors (i.e. effectively one transistor per bit), which results in an approximately 50% smaller 1-bit effective area compared to conventional SLC SOT-MRAM. However, there is still a margin to improve the 1-bit effective area than their estimated value of $34.5F^2$, using the design rules in [16], as illustrated later in Section 3. Moreover, their two proposed MLC designs, which are known as parallel MLC (P-MLC), shown in **Figure 2(c)**, and series MLC (S-MLC), depicted in **Figure 2(d)**, do suffer from various drawbacks that result in higher energy consumption and reliability issues as discussed below.

2.2.1 Parallel multi-level cell SOT-MRAM (P-MLC SOT-MRAM)

P-MLC, depicted in **Figure 2(c)**, encloses two MTJs in-parallel and both placed side-by-side over a common HM electrode. The main advantage of this cell structure is that the two bits are accessed by two transistors, which results in approximately 50% reduction in the 1-bit effective area. As the two MTJs are placed on the HM, both of the MTJs may be programmed by SOT effect. Writing the two MTJs with identical bits ('00' or '11') can be done using only one write pulse with duration following the slower MTJ. However, to write independent bits on the two MTJs ('01' or '10'), the two MTJs must have different critical currents (I_c) (i.e. the MTJs with smaller I_c switches faster for the same supplied current amplitude). Thereafter, either time-dependent or current-dependent writing [17] can be adopted. In time-dependent writing, a constant write current flows through the HM electrode, where both MTJs are firstly programmed within pulse duration t_1 , followed by the programming of faster MTJ2 with time t_2 ($t_1 > t_2$ and t_2 is not long enough for MTJ1 to switch), whereas in current-dependent writing, both MTJs are firstly written with larger write current (I_{write1}). Subsequently, the faster MTJ2 is programmed with smaller write current (I_{write2}) ($I_{write1} > I_{write2}$ and I_{write2} is not large enough for MTJ1 to switch).

Releasing the two MTJs with different I_c can be achieved by having different MTJ free layer thickness, dimensions [17], and/or width of underlying electrodes [15]. Therefore, P-MLC manufacturing might be challenging due to this imposed nonuniformity in cell architecture. The imposed different current requirements may result in lower energy efficiency (i.e. additional energy penalty) as one of the MTJs should switch with larger current amplitude compared to the other MTJ (under equivalent switching time assumption), whereas in conventional SLC SOT-MRAM, all the SOT-MTJs consume the same energy (i.e. no enforced rule of using two SOT-MTJs with different I_c). However, this energy penalty can be reduced by having a smaller ΔI_c between the two SOT-MTJs, as depicted in **Figure 3**. The minimum ΔI_c (minimum energy penalty) depends on the tolerable switching probability (PSW) of MTJ2 (slower SOT-MTJ) while writing MTJ1, as the smaller the ΔI_c , the higher the PSW of MTJ2. It is important to note that in real MRAM chip, the ΔI_c should be large enough to accommodate for the different distributions of switching time, switching current, and error rates. This should be chosen carefully by considering the potential variability of critical current ($\sigma I_c / \mu I_c$), which can be as large as 10% [13].

The reading operation is done similarly to the conventional SOT-MRAM, where the sense current or voltage is used to identify the equivalent resistance state. However, it is required to differentiate between four different resistances states unlike the only two states that exist in the conventional SLC. In addition, the

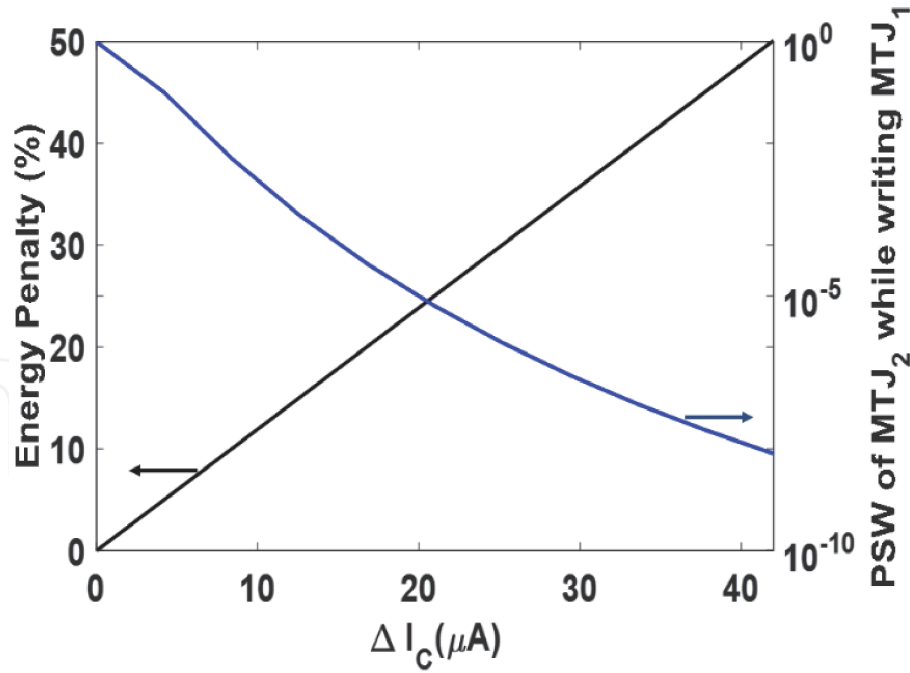


Figure 3.

Percentage of increase in the energy resulting from having two MTJs of different IC versus $\Delta I_c = I_{c2} - I_{c1}$ and the corresponding probability of switching (PSW) of slower MTJ2 while writing faster MTJ1 [17].

in-parallel configuration of the two MTJs during reading results in a reduced minimum difference between the various resistance states (and consequently reduced read margin compared to in-series configuration and SLC), as the equivalent resistance for two parallel resistances is always smaller than the smallest resistance. For instance, if the parallel configuration resistance (R_p) of MTJ1 (R_{p1}) = 7 k Ω , R_p of MTJ2 (R_{p2}) = 12 k Ω , anti-parallel configuration resistance (R_{ap}) of MTJ2 (R_{ap2}) = 22 k Ω , the equivalent resistance for in-parallel MTJs connecting for case1 (R_{p1} , R_{p2}) is $R_{tot1} = 4.4$ k Ω and for case2 (R_{p1} , R_{ap2}) is $R_{tot2} = 5.3$ k Ω . That results in a minimum resistance difference (ΔR_{min}) of 0.9 k Ω only, while for SLC or even in-series MTJ connection, a larger ΔR_{min} can be achieved. For instance, if the MTJ connected in-series instead of in parallel using the same MTJ resistance values, the R_{tot1} becomes equal to 19 k Ω and $R_{tot2} = 29$ k Ω , which results in ΔR_{min} of 10 k Ω . Consequently, the combination of both MLC scheme and in-parallel connectivity increases the expected BER.

2.2.2 Series multi-level cell SOT-MRAM (S-MLC SOT-MRAM)

S-MLC, shown in **Figure 2(d)**, consists of two in-series MTJs placed over an electrode made of heavy metal. The first MTJ (MTJ1) in contact with the heavy metal electrode can be programmed by SOT effect, whilst the second MTJ (MTJ2) (stacked over MTJ1) must be programmed by conventional spin-transfer torque (STT). Similarly, the main advantage of this proposal is that each cell comprises two bits that are accessed by two transistors. This results in a nearly 50% reduction of the 1-bit effective area compared to conventional SOT-MRAM. In the write operation of S-MLC, MTJ2 must be programmed before MTJ1 to avoid a final state of write disturb failures for MTJ1, which means that the programming for the two MTJs should be serial and cannot be simultaneous. The need for STT in programming MTJ2 results in low energy efficiency as STT programming requires passing current by the high resistance MTJ stack, which demands high writing voltage in addition to the large critical current of STT switching compared to SOT switching. In addition, passing a large current through the MTJ stack reduces the tunnel barrier

reliability, which jeopardizes one of the main advantages of using SOT-MRAM. The reading operation of the S-MLC is similar to P-MLC. Being an MLC requires the stack to represent four different resistance states. Thus, S-MLC uses MTJs with different cross-sectional dimensions to achieve the four different resistance states, which may result in a complex fabrication process. Furthermore, there is a need to employ low resistance MTJs in the stack to be able to supply enough current to achieve the STT switching. This results in a smaller minimum difference between the four distinct resistance states (i.e. smaller ΔR_{\min} between the four possible resistances (R11, R10, R01 and R00)). The reduced ΔR_{\min} minimizes the read margin for the S-MLC memory and thus longer reading delay.

3. Diode-based SOT-MRAM

This section presents the various diode-based or selector-based SOT-MRAM proposals in the literature. These proposals mainly rely on replacing the read transistor (Tx), highlighted in **Figure 2(a)**, by a diode or selector. As the read operation in SOT-MRAM requires mainly a relatively small and unidirectional current, a diode or a selector can successfully satisfy these requirements. The employed diodes are targeted to be nonsilicon-based, and thus, the cell silicon area would have one less transistor, as further explained below. However, employing a diode or selector may come with an energy penalty as larger read voltage may be required to overcome the diode's on-voltage.

3.1 SLC diode-based SOT-MRAM

Seo et al. [14] proposed a diode-based single-level cell (SLC) SOT-MRAM, shown in **Figure 2(b)**. In that design, the SOT-MRAM cell area is reduced by replacing the read transistor, depicted in **Figure 2(a)**, with a Schottky diode. Thus, the cell requires only one transistor to access a single bit. This design's main advantage is the reduction of the 1-bit effective area by approximately 50% compared to conventional SLC SOT-MRAM that requires two transistors to access a single bit. In this design, the 1-bit effective area is estimated to be $34.5F^2$ using the design rules in [16]. The write and read operations in this proposal are similar to that of the conventional SLC SOT-MRAM. The write operation would consume similar write energy, as each cell comprises only 1-bit (MTJ). Moreover, the read operation follows the same biasing as in the conventional SLC SOT-MRAM, where the RWL of the required row is activated and the WWL is deactivated. However, the diode usage increases the read energy as the read voltage needs to account for the additional voltage drop across the diode. In addition, there is still a margin to achieve smaller 1-bit effective area by using the diode as it is shown in the following sections.

3.2 Multi-bit per cell dedicated diode (MBC-DD) SOT-MRAM

Ali et al. [18] extended the SLC diode-based SOT-MRAM proposal to be a multi-bit per cell (MBC). Their proposed design, shown in **Figure 4**, relies on a metal-oxide-metal (MOM) diode (or also known as selector) stacked over an MTJ (forming a D-MTJ) to replace the task of the read Tx in controlling the read current flow. The cell comprises two D-MTJs with similar tunnel oxide barrier thicknesses and their free-layers are placed in contact with a common HM electrode. The two D-MTJs can be programmed with two different bits through a common electrode. The sharing of common electrode allows the use of only single transistors to access

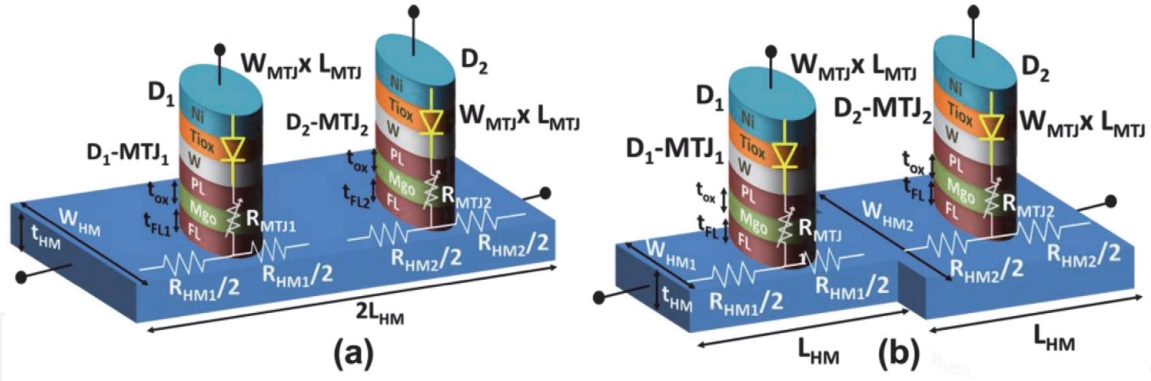


Figure 4.

3D structure of the proposed dedicated diode multi-bit per cell (MBC) DD SOT-MRAM with (a) uniform HM electrode and MTJs with different t_{F1} [18], (b) different HM width and MTJs with similar t_{F1} [18].

the two bits (2 D-MTJs). This result is the main advantage of this proposal, which is a reduced 1-bit effective area that is $4\times$ smaller than the conventional SOT-MRAM and offers at least double the density compared to any MRAM proposal in the literature. Given the design rules in [16], the 1-bit effective area is estimated to be $18F^2$. The cell also employs separate read word line (RWL) for each D-MTJ, as in **Figure 5**. On the other hand, similar to [14], employing a dedicated diode per MTJ may increase the read energy and limit the maximum acceptable diode area.

The MOM diode MTJ stack employed in this design is similar to the experimentally validated device used in the 1S1R 3D cross-point STT-MRAM developed by avalanche [19–21]. However, employing the D-MTJ device in the SOT technology would be more efficient than in the STT technology as the energy and performance degradation effects of the diode would only exist in the read operation and is avoided in the write operation. This is because the diode only exists in the read path, while in the STT technology, the write operation is dependent on the employed diode as the write current has to flow through both the MTJ and diode to achieve the STT switching.

The writing operation of this cell is similar to that in the P-MLC, where either time-dependent or current-dependent writing can be adopted as elaborated before. During the write operation, the WWL of the row comprising the required cell is asserted high. A ‘0’/‘1’ is written on the MTJ if the BL is set high/low, and SL is pulled low/high for the column including the targeted cell, allowing the charge current to flow through the HM electrode in the essential direction, as shown in **Figure 5(a)**. Furthermore, the RWLs of the row comprising the targeted cell are set low to ensure that the diodes are reverse-biased and no leakage current flows

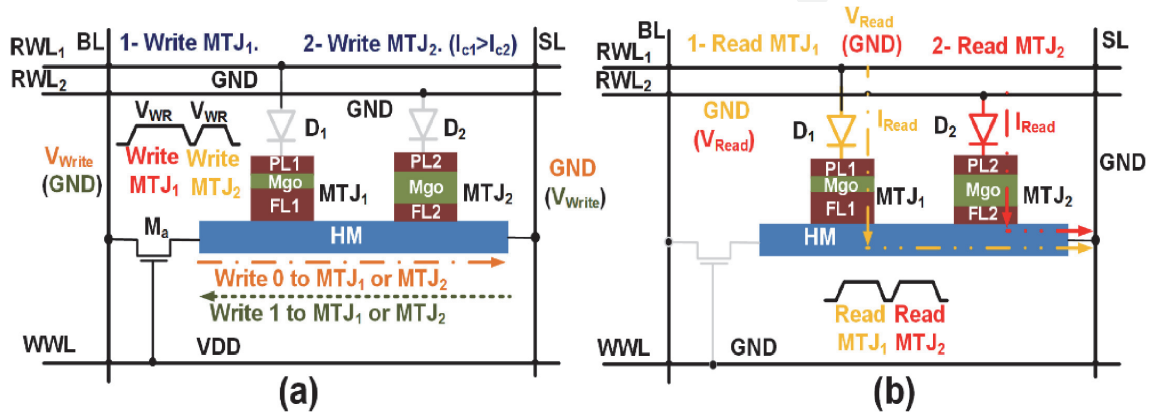


Figure 5.

Proposed MBC-DD SOT-MRAM cell in the (a) write and (b) read operations [18].

through the MTJs. It is worth mentioning that the low voltage while using the MOM diode may not be exactly zero volts, as a hold voltage ($V_{\text{hold}} \sim 0.02 \text{ V}$) might be needed.

The separate RWLs for each of the two D-MTJs permit selective reading of the two bits. During the read operation, the WWL is deactivated, and the SL of the column comprising the targeted D-MTJs is pulled to GND. The RWL of the targeted D-MTJ is then connected to the sense amplifier to forward bias the diode and read out the data stored in the MTJ, as shown in **Figure 5(b)**. Although switches are still needed to select different RWL, it has a negligible impact on the area per unit cell as these switches are shared by all the cells in the row. Furthermore, as both MTJs are sensed independently and the RMTJ is an order of magnitude larger than RHM, hence, the RHM impact on the effective TMR is minimized (average sensed resistance = RHM + RMTJ).

The realization of the MBC-DD SOT-MRAM has two main essential requirements. Firstly, employing two SOT-MTJs with different I_c to be able to write the two bits independently as discussed before. This is achieved by using either different widths of the HM below each MTJ (W_{HM}), as illustrated in **Figure 4(b)**, or using different free layer (FL) thicknesses (t_{fl}) within each MTJ, as shown in **Figure 4(a)**. Uniform HM would be preferred in high-density memories, as W_{HM} for both SOT-MTJs are limited to the technology minimum feature size (F), and any increase in the W_{HM} than F increases the overall cell area. However, it may require two MTJ stack deposition with additional lithography steps, which may increase the manufacturing cost, whereas SOT-MTJs with uniform t_{fl} is preferred from reading perspective as both MTJs would have similar TMR [22], which simplifies the read operation. It also offers lower fabrication cost due to simple processing. Nevertheless, due to the higher density, at least double other designs, it is expected that both structures would lead to lower overall cost per bit cell. Secondly, a 3D diode MTJ stack is required to have the diode replacing the read transistor without consuming silicon area. Incorporating the diode in the reading process requires applying a relatively higher read voltage (V_{Read}) that is enough to overcome the diode on-voltage (i.e. forward bias the diode) and supply the required read current. The required magnitude of the V_{Read} at a targeted read current (I_{Read}) depends mainly on two factors, which are the diode's cross-sectional area and the load resistance (R_{load}). Firstly, the smaller the diode's area, the smaller the diode supplied current at given bias voltage, as shown in **Figure 6(a)**. Thus, the needed V_{Read} increases with diode scaling down to supply the targeted I_{Read} , as depicted in **Figure 6(b)**. Secondly, the load resistance (R_{load}) affecting the diode, which is the in-parallel SOT-MTJs equivalent resistances and the existing transistors in the I_{Read} path. The smaller the R_{load} , the smaller the required V_{Read} to supply certain I_{Read} for a given diode area, as illustrated in **Figure 6(c)** and modeled in Eq. (1). Thus, smaller R_{load} aids to achieve lower read energy at a given I_{Read} .

$$V_{\text{Read}} = V_{\text{On_noload}} + I_{\text{Read}} * R_{\text{load}}, \quad (1)$$

where $V_{\text{on_noload}}$ is the diode's required bias voltage to supply certain I_{Read} with no load condition.

3.3 Multi-level cell shared diode (MLC-SD) SOT-MRAM

Ali et al. [17] also proposed another diode-based SOT-MRAM cell in which a shared diode (selector) between the two MTJs in the cell is employed, as illustrated in **Figure 7**, instead of a dedicated diode (selector) per each MTJ. Similar to MBD-DD SOT-MRAM, each cell comprises 2-bits (MTJs) accessed with a single transistor.

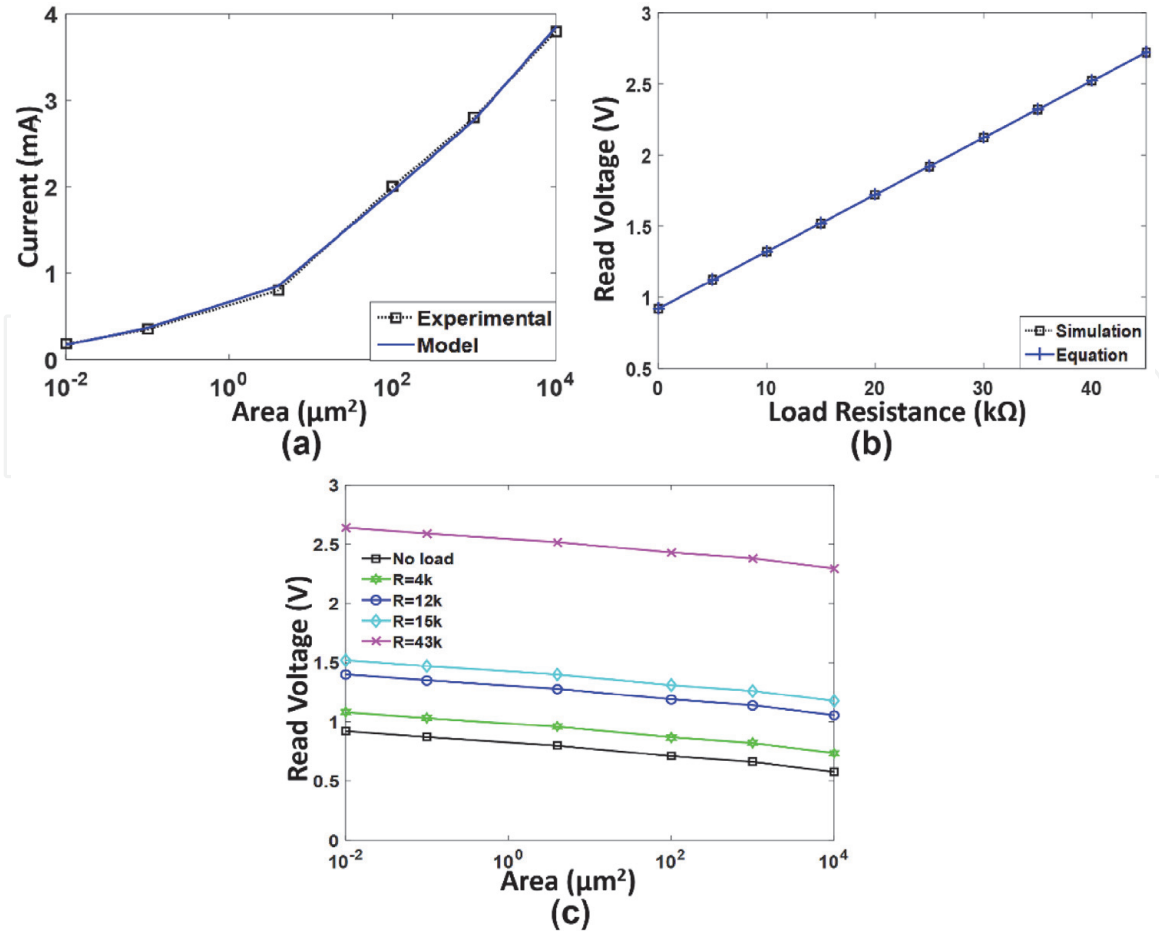


Figure 6.

(a) The diode current versus diode area @ bias voltage = 1 V and no-load resistance. The data are from the implemented diode Verilog-A model verified against experimental data in [23]. (b) The required read voltage to supply $I_{\text{Read}} = 40 \mu\text{A}$ with diode area scaling for various R_{load} . (c) The required read voltage for the diode to supply $I_{\text{Read}} = 40 \mu\text{A}$ versus different R_{load} using both Verilog-A model simulation and Eq. (1) [18].

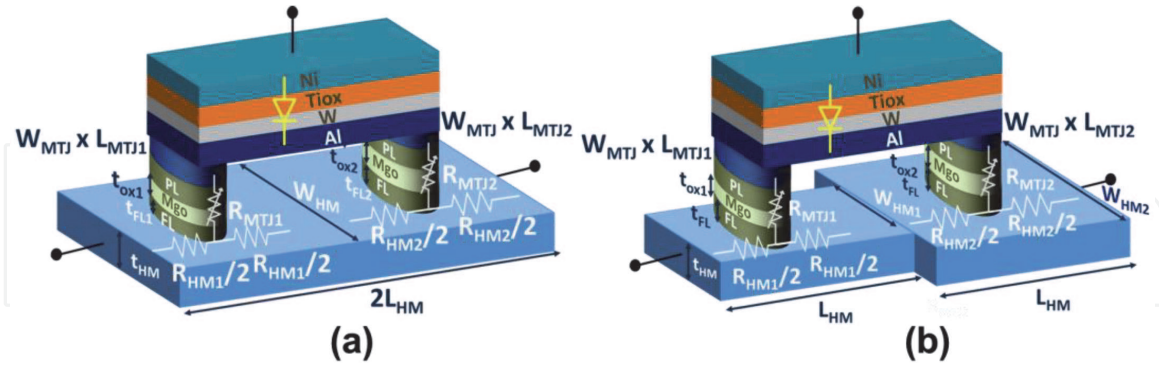


Figure 7.

3D structure of the proposed shared diode multi-level cell (MLC) SOT-MRAM with (a) uniform HM electrode and MTJs with different t_{FL} [17], (b) different HM width and MTJs with similar t_{FL} [17].

This result is a similar 1-bit effective area that is $4\times$ smaller than conventional SOT-MRAM and at least double the density compared to any of the MRAM design in the literature. The 1-bit effective area is estimated to be $17.5F^2$ using the design rules in [16]. On the other hand, employing a shared diode permits increasing the diode area as it is no longer limited by one MTJ area. However, the needed memory sensing will be MLC (i.e. the cell has four different resistance states) instead of a SLC (i.e. the cell has only two resistance states), which complicates the sensing operation and reduces the read margin.

The two MTJs in the MLC-SD cell also share a common HM electrode to be able to program the two MTJs with the energy-efficient SOT technology. This enforces using two SOT-MTJs with different I_c as well, which can be written following the same approach of P-MLC using either time or current-dependent writing. Consequently, similar to MBC-DD in the write operation, the WWL of the row comprising the required cell is asserted high. A '0'/'1' is written to the MTJ if the BL is set to high/low and SL is pulled to low/high for the column including the targeted cell, allowing the charge current to flow through the HM electrode in the essential direction, as depicted in **Figure 8(a)**. In addition, the RWL of the row comprising the targeted cell are set low to ensure that the diodes are reverse biased and no leakage current flows through the non-selected cells.

Employing a shared diode requires connecting the two MTJs in-parallel, which enforces the cell to be MLC. Consequently, the two MTJs in the cell should have different R_P and R_{AP} such that their equivalent resistance has four distinct values, which represents two-bit logic values of '00', '01', '10' and '11'. To read the bits in the cell, the WWL and the SL signals are pulled low to deactivate all the write transistors and allow the read current to flow through the targeted MTJs. The row RWL is connected to the sense amplifier and consequently is pulled up to the required read voltage (V_{Read}), as shown in **Figure 8(b)**. Thereafter, the equivalent 2 bits for the targeted MLC cell are identified sequentially by comparing the currents flowing through the sensed combined MTJs and the appropriate reference MTJs using a binary search algorithm [7]. In this algorithm, the sensed MTJs are first compared with the first reference resistance to determine the first-bit state. Based on the first-bit state, one of the two other reference resistances are then chosen to compare with the sensed MTJs to determine the second-bit state. The same technique can be used to sense the state of any of the MLC proposals illustrated above.

As can be inferred from above, the realization of MLC-SD SOT-MRAM has three main essential requirements. Similar to the MBC-DD SOT-MRAM design, it requires employing two SOT-MTJs with different I_c and a 3D diode MTJ stack. In the MLC-SD cell, as the diode is shared, the employed diode can be large, and its size can be approximately up to the whole cell area instead of just one MTJ area. A larger diode requires a smaller read voltage to supply the required I_{Read} , as depicted in **Figure 6(a)**. Moreover, the in-parallel combination of the two MTJs results in smaller overall resistance compared to a single MTJ resistance. This reduces the

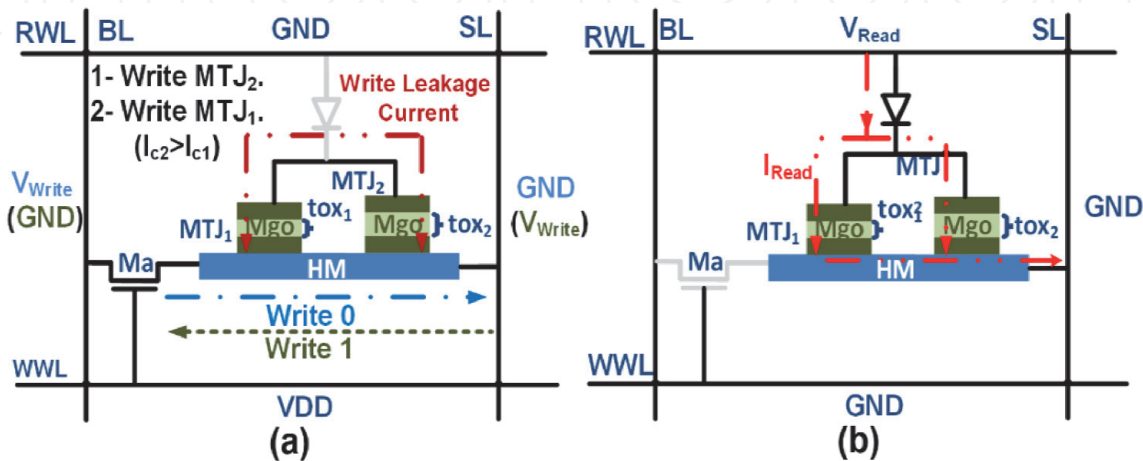


Figure 8.
Proposed 1D1T2R MLC-SD SOT-MRAM configuration and current flow direction in (a) write mode and (b) read mode [17].

diode’s R_{load} , which further decreases the required V_{Read} , as shown in **Figure 6(b)**. Hence, the two factors that affect the required V_{Read} (i.e. diode area and R_{load}) are improved in this design compared to the MBC-DD and SLC diode-based SOT-MRAM designs. Hence, the smaller required V_{Read} may permit the MLC-SD design to achieve smaller read energy consumption. Furthermore, MLC-SD cell requires employing two MTJs with different R_P and R_{AP} values as it is an MLC. Assuming that both the MTJs use the same materials, different MTJ resistances are achieved by varying either the MTJ dimensions (i.e. W_{MTJ} and L_{MTJ}), the dielectric thickness (t_{ox}), or combination of both [24]. It is essential to have a large minimum resistance difference (ΔR_{min}) between the distinct in-parallel equivalent resistance states to increase the distinguishability and read speed. However, the MLC would mainly have smaller ΔR_{min} compared to the SLC, which in-return may result in reduced reading speed and would be a competing factor with the reduced V_{Read} to decide the read energy efficiency of this cell compared to the SLC proposals.

4. Evaluation

In this section, the various proposals in the literature are evaluated using the same SOT-MTJ technology in [13] on both cell and system level perspectives. The cell level analyses are done based on a 2×2 memory array. The simulations run over Cadence Virtuoso and using the SOT-MTJ Verilog-A model demonstrated in [25] and the parameters in **Table 1**. The system-level analyses are done using the non-volatile memory simulator, known as NVSim [26]. NVSim estimates the overall memory performance, power consumption, energy, and area based on the given memory cell parameters.

Symbol	Parameter	Value
$W_{MTJ} \times L_{MTJ}$	MTJ dimensions ($W \times L$) (nm ²)	50×100
t_{FL}	Free layer thickness (nm)	1.5
$K_u V/KT$	Thermal stability	46
t_{ox}	Tunnel barrier thickness (nm)	1.8
M_s	Magnetization saturation (emu/cm ³)	1114
R_{AP}	MTJ high resistance value (k Ω)	15
TMR	Tunnel magneto-resistance ratio (%)	114
α	Gilbert damping	0.012
ρ_{HM}	Heavy metal (W) resistivity ($\mu\Omega$ -cm)	200
$W_{HM} \times L_{HM}$	HM dimensions ($W \times L$) (nm ²)	115×150
t_{HM}	Heavy metal thickness (nm)	3
J_{co}	Critical current density ($\times 10^{10}$ A/m ²)	7
θ_{SHE}	Spin hall angle	0.3
P	Spin polarization	0.6
γ	Gyromagnetic ratio (rad s ⁻¹ T ⁻¹)	1.76×10^{11}
\hbar	Reduced Planck constant (J s)	1.054×10^{34}
The parameters follow the experimental data of the SOT-MTJ in [13] and are explained in [11].		

Table 1.
SOT-MTJ device parameters.

4.1 Cell-level evaluation

Table 2 presents a comparison between the different MRAM designs. In terms of area, MBC-DD and MLC-SD SOT-MRAM do offer the smallest 2-bit cell area among the various designs, which are estimated to be $36F^2$ and $34.5F^2$, respectively, based on the rules in [16]. This is at least double the density compared to other MRAMs and achieves 75% smaller 1-bit effective area compared to conventional SOT-MRAM. From energy perspective, these designs consume at least 36% less energy compared with designs utilizing STT writing (S-MLC), due to the high energy efficiency of SOT writing. Unlike P-MLC and MLC-SD SOT-MRAM, MBC-DD has no write current leakage through the MTJ from the HM during write mode as the diodes are reverse-biased, which also leads to better energy efficiency. However, the significant area reduction for both MBC-DD and MLC-SD SOT-MRAMs comes with additional energy penalty in both worst-case write operation (i.e. writing non-identical bits) and read operation. The additional energy consumption in the worst-case write operation of non-identical bits is because of the enforced rule of using two MTJs per cell with different I_c . For instance, writing two different bits ('10' or '01') in MBC-DD and MLC-SD consume higher energy (0.88 pJ with 10.5 ns delay) compared to writing two SLC SOT-MRAM (0.76 pJ). However, to write identical bits ('00' or '11') on the MBC-DD and MLC-SD cells require only a single write pulse, which leads to better energy efficiency than SLC SOT-MRAM. This is because programming two identical bits in the SLC SOT-MRAM always requires two write pulses. Thus, if equal probability of programming '00', '01', '10', and '11' is assumed, MBC-DD and MLC-SD designs may result in similar average energy efficiency to two bits of SLC SOT-MRAM, as shown in **Table 2**. Moreover, the

		SLC SOT (2-bit)	1D1T SLC- SOT (2-bit) [14]	S- MLC [15]	P- MLC [15]	MLC-SD SOT-MRAM [17]	MBC-DD SOT-MRAM [18]
Energy per 2-bit (pJ)	Write (EW) ^a (w. case)	0.76	0.76	1.13	0.72 (0.88)	0.72 (0.88)	0.72 (0.88)
	Write leakage	0.0002	0.0001	0.0001	0.19	0.2	0.0001
	Read (Er)	0.024	0.036	0.034	0.039	0.041	0.036
Total energy/2-bit		0.78	0.8	1.17	1.11	1.11	0.9
Delay per 2-bit (ns)	Write a (w. case)	9	9	10.5	8.1 (10.5)	8.1 (10.5)	8.1 (10.5)
	Read	1	1	1.4	1.6	1.6	1
Read voltage (V)		1.2	1.8	1.2	1.2	1.3	1.8
ΔR_{min} (k Ω) ^b		5	5	3	1.4	1.4	5
Diode area (F^2) ^d		—	24	—	—	24	10
Area (A)/1bit (F^2) ^c		69	34.5	50	34.5	17.25	18
FOM (Er*Ew*A)		2.19×	1.65×	5.7×	2.5×	1.34×	1

^aEnergy and delay are the average of writing the data '00', '01', '10', '11'.
^b ΔR_{min} is $R_{AP} - R_P$ for SLC, while it is the minimum resistance difference among the four different states in MLC.
^cArea estimated based on standard design rules reported in [16].
^dF is the minimum feature size of the employed technology (i.e. In 32 nm technology, F = 32 nm).

Table 2.
Comparison of various MRAM technologies.

energy penalty for non-identical bits writing can be also minimized as discussed before. In terms of reading, MBC-DD maintains similar distinguishability to conventional SLC, as indicated by the ΔR_{\min} values in **Table 2**, because each of the two MTJs is sensed separately. This is unlike the P-MLC, S-MLC, and MLC-SD structures that have a reduced ΔR_{\min} as a result of relying on an MLC approach. However, to ensure sufficient diode drive current within the D-MTJs, a larger read voltage is needed compared to cells with read transistors, which does increase the read energy consumption compared to conventional SLC SOT-MRAM.

The diode-based SLC SOT-MRAM design [14] does offer the advantage of maintaining a similar write energy efficiency compared to the conventional SOT-MRAM (i.e. baseline from write energy perspective). Moreover, it offers 50% 1-bit effective area savings compared to conventional SOT-MRAM, which is on level with P-MLC and S-MLC designs with the advantage of maintaining an SLC sensing approach. However, diode-based SLC SOT-MRAM still consumes double the area compared to the MBC-DD and MLC-SD designs, while it still also suffers from the energy and fabrication complexity penalty of employing a diode.

Similar to diode-based SLC SOT-MRAM, P-MLC design offers 50% 1-bit effective area savings compared to conventional SOT-MRAM, whereas S-MLC does offer only 28% savings as the transistor size needs to increase to supply the required STT current through the high resistance MTJ stack. On the other hand, both P-MLC and S-MLC do not employ a diode, which may aid in reducing the read energy and avoiding the fabrication process issues related to incorporating a diode. However, they suffer from other drawbacks that increase both energy consumption and fabrication complexity. In particular from energy perspective, S-MLC consumes high energy due to using STT technology in writing one of the two MTJs per cell, whereas P-MLC shares the additional write energy penalty issue in writing non-identical bits per cell with MBC-DD and MLC-SD designs. In addition, both of the designs rely on the MLC sensing approach, which harms the sensing speed and distinguishability as reflected by the reduced ΔR_{\min} values.

Overall, if a figure-of-merit (FOM) is defined as the product of the area, read and write energy products [27], MBC-DD SOT-MRAM may outperform other designs by at least 34%, thanks to its significant area reduction and maintaining the SLC sensing approach.

4.2 System-level evaluation

As aforementioned, NVSim [26] is used to evaluate the various SOT-MRAM cells from a system-level perspective. NVSim does consider the different write/read peripherals, array organization, and routing network required in the overall memory architecture. NVSim supports various nonvolatile memories such as STT-MRAM, PCRAM, and ReRam in addition to the well-known volatile memories such as SRAM and DRAM. NVSim can be tuned to support SOT-MRAM as well. In this study, the comparison is based on the utilization of the various nonvolatile MRAM cells as cache memory and they are mainly compared to the current widely used technology as cache memory, which is the SRAM. SRAM does offer high performance; however, currently, it suffers from a significant increase in the 1-bit area and the leakage power consumption [28]. Thus, the utilization of an area-efficient, high-speed and nonvolatile SOT-MRAM would be a promising solution to replace the existing SRAM technology, especially in higher-level caches. The considered cache has 4-way set associativity and 64 Byte line size and is optimized to achieve the smallest overall silicon area. The 32 nm technology node is assumed for the various designs, in which the SRAM cell area is $170F^2$ [28], and the same cell parameters for the various MRAM technologies are maintained as stated above.

To clarify the pros and cons of the various memory technologies, three different memory capacities are considered, which are 256 KB, 1 MB, and 8 MB. **Figure 9** depicts the total leakage power consumption for the various designs with three different capacities. The volatile SRAM technology does consume significant leakage power, which increases by order of magnitudes for larger memory capacities. On the other hand, the nonvolatility of the MRAM designs results in relatively negligible leakage among the various capacities. This demonstrates the advantage of significant power consumption reduction by employing the nonvolatile MRAM designs as a replacement of the volatile SRAM, especially in battery-powered mobile devices that demand long idle durations.

From area perspective, **Figure 10** shows the comparison between the various memory technologies, estimated by NVSim, while being used as cache memory. **Figure 10(a)** reports the different MRAM designs relative to the overall SRAM memory area for the three different capacities. The figure indicates the significant reduction in the overall area by employing the various MRAM technologies in comparison to SRAM, which consumes at least 50% smaller silicon area. Furthermore, the area saving percentage increases noticeably for larger memory capacities. This is because for larger memory capacity, the impact of the cell area overtakes the impact of the periphery. The periphery area for MRAM consumes larger area than the periphery area of the SRAM due to the need for relatively larger write currents, whereas the cell area of the MRAM is much smaller than the SRAM cell area [29]. Hence, at larger capacities, the MRAMs that have significantly smaller cell area compared to SRAM would consume much lower overall silicon area. For instance, the area saving for the MBC-DD and MLC-SD cells, which are the cells with smallest footprints, can reach up to 90% smaller area compared to SRAM for large cache memory size, such as 8 MB size.

The impact of the smaller cell area is also clear while comparing the various MRAM technologies, as presented in **Figure 10(b)**. The smaller 1-bit effective area of S-MLC, SLC diode-based, P-MLC, MLC-SD, and MBC-DD SOT-MRAMs by 28, 50, 50, 75, and 74% compared to the conventional SOT-MRAM results in similar overall memory silicon area reduction. In particular, for larger memory capacity (e.g. 8 MB), where the impact of the cell area is more significant, the overall memory silicon area of the various designs in the literature relative to conventional SOT-MRAM design shows approximately equal area reduction as the 1-bit effective area reduction. Moreover, the smaller silicon area consumption of the various proposed designs in the literature compared to SRAM and conventional SOT-MRAM with equivalent capacity permits realizing the cache memory using these cells with higher capacity under the iso-area assumption. For instance, 2 MB

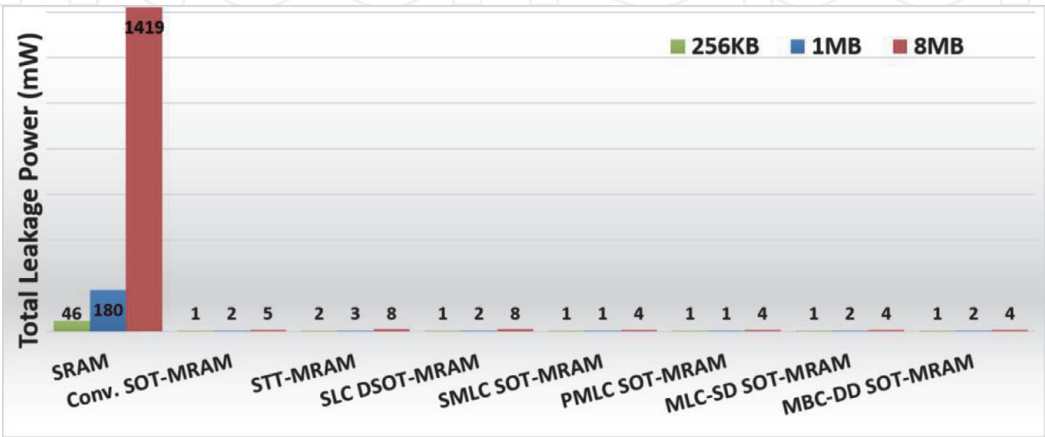


Figure 9. Total leakage power consumption of SRAM, STT-MRAM, conventional, SLC DSOT-MRAM, S-MLC, P-MLC, MLC-SD, and MBC-DD SOT-MRAMs for various memory capacities.

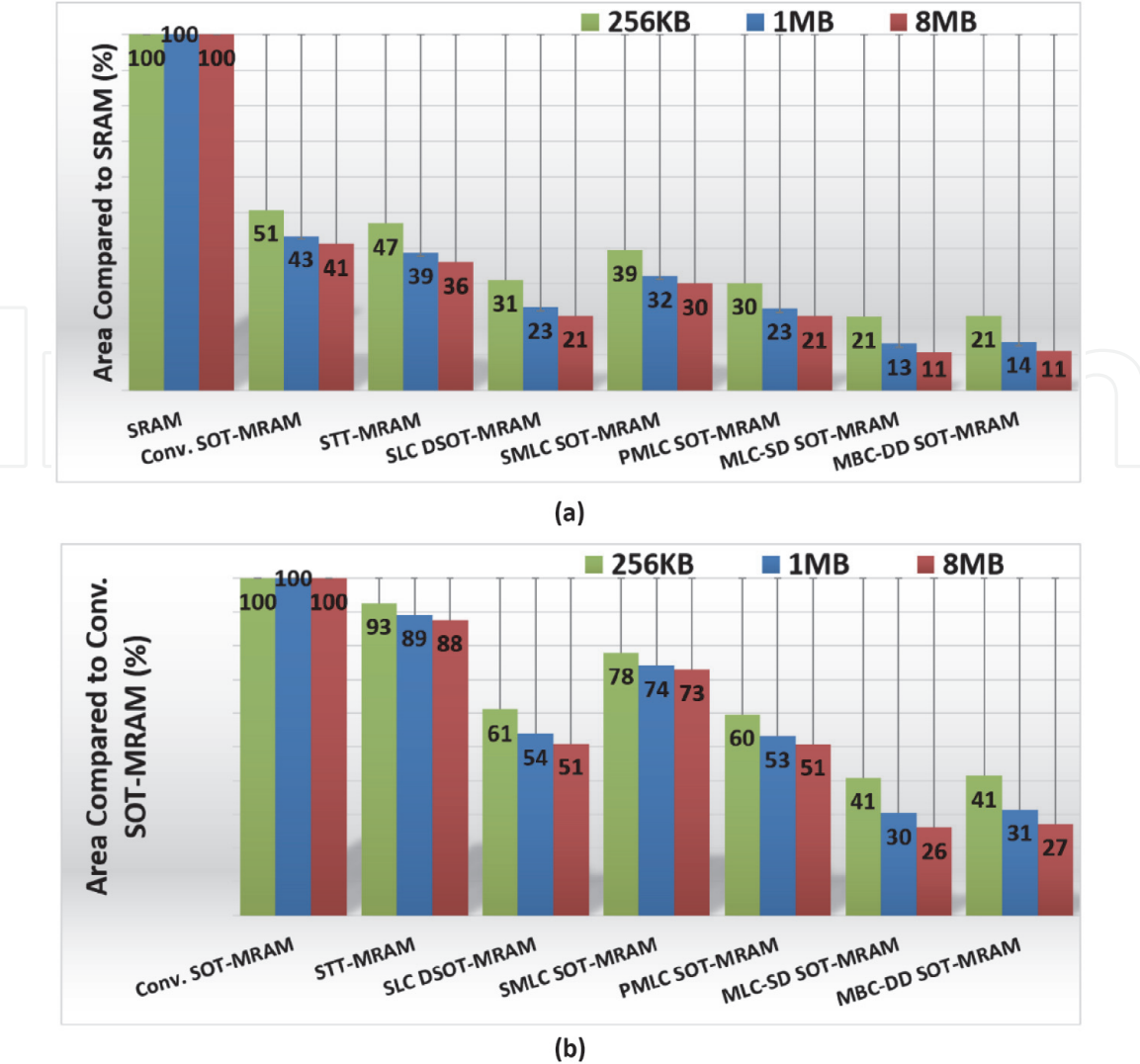


Figure 10. Area comparison of the STT-MRAM, conventional, SLC diode-based, S-MLC, P-MLC, MLC-SD, and MBC-DD SOT-MRAMs for different capacity relative to (a) SRAM and (b) conventional SOT-MRAM.

capacity of conventional SOT-MRAM consumes similar silicon area to an 8 MB capacity of MBC-DD SOT-MRAM. Higher memory capacity results in higher performance metrics, such as instruction per cycle (IPC) and energy efficiency due to reduced access counts for the off-chip memory [30].

From energy and performance perspective, SRAM can have higher hit/miss performance and energy efficiency compared to the various SOT-MRAM proposals at small memory capacity (e.g. 256 KB), as depicted by **Figure 11(a)**. This is because the impact of the large load capacitance of the SRAM cell would be minimal at these capacities [29]. However, for large memory capacity, the much higher load capacitance, parasitic, and routing complexity of the six transistors SRAM cell (as SRAM consume significantly larger silicon area compared to SOT-MRAM) causes the MRAM proposals to achieve better hit/miss performance and energy efficiency. On the contrary, the write energy for the various MRAM proposals is larger than that of the SRAM for the various memories capacity, as shown in **Figure 11(b)**. This is attributed to the larger write current requirement for the MRAM-based technologies relative to the SRAM technology. However, with improved SOT-MRAM technology such as the type-x SOT-MTJ [13] (achieves sub ns switching with write current of 100 μ A), or the presented high-performance SOT-MRAM by IMEC [12] that achieves successful switching in 100's of ps range, the write energy can be reduced significantly, as illustrated in **Figure 12**. This makes the various SOT-

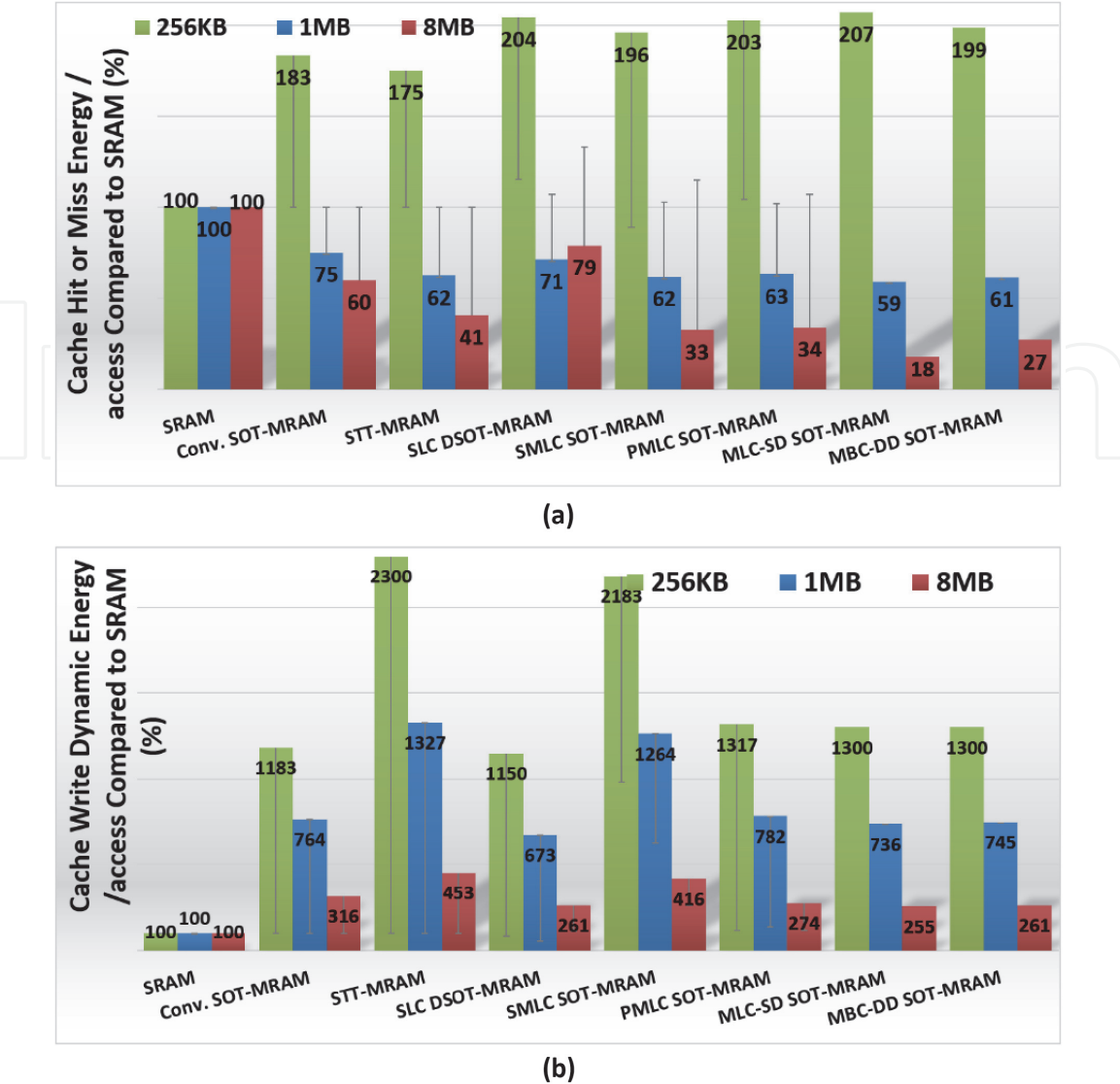


Figure 11. Comparison of the SRAM, STT-MRAM, conventional, SLC diode-based, S-MLC, P-MLC, MLC-SD, and MBC-DD SOT-MRAMs for different capacities relative to SRAM from (a) cache hit/miss energy per access perspective and (b) cache write dynamic energy per access perspective.

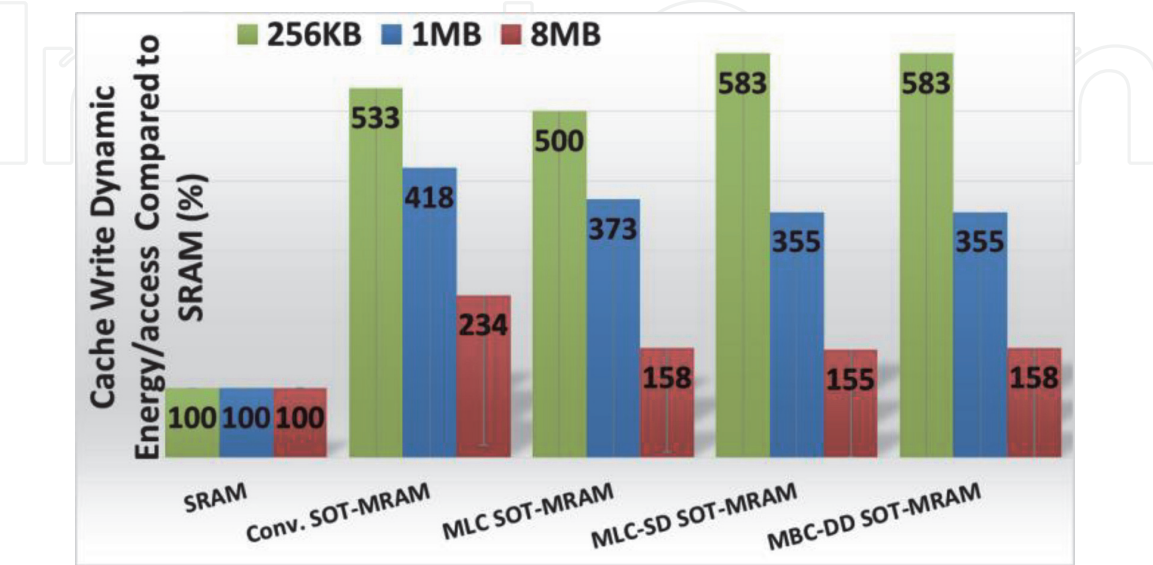


Figure 12. Cache write dynamic energy per access comparison of the SRAM, conventional, P-MLC, MLC-SD, and MBC-DD SOT-MRAMs for different capacity relative to SRAM assuming improved high-speed SOT-MTJ technology (e.g. type-x reported in [13]).

MRAM proposals a viable and realistic solution to replace the SRAM technology in certain applications.

In conclusion, the previous discussion shows that the various proposed SOT-MRAM cells do offer nonvolatility (i.e. nearly zero leakage), smaller silicon area, and high performance. It also indicates that these designs can compete with the current CMOS volatile technologies such as SRAM and DRAM. However, further development to reduce the write energy for the existing SOT-MTJ technology may be required to widen the application window for such SOT-MRAM technologies.

5. Conclusion

This chapter presents the various SOT-MRAM proposals in the literature highlighting the pros, cons, and operation of each design. SOT-MRAM relies on SOT technology, which offers various advantages such as high energy efficiency, fast switching speed, and high device reliability. However, conventional SLC SOT-MRAM requires two transistors to access a single bit. This in return results in a relatively large 1-bit effective area, which limits its application for large memory capacities. Hence, the various proposals in the literature targets reducing the 1-bit effective area compared to both conventional SOT-MRAM while maintaining the main advantages of SOT technology.

The various SOT-MRAM proposals have been divided into two main categories, which are diode-based and nondiode-based SOT-MRAM cells. These various SOT-MRAM cells have been evaluated from both cell and system level perspectives. The system-level evaluation is performed based on the utilization of the various cells as cache memory, and they are mainly compared to the current widely used technology as cache, which is the SRAM. In particular, five different proposals have been investigated. These proposals are S-MLC, P-MLC, diode-based SLC, MBC-DD, and MLC-SD SOT-MRAMs that are shown to offer 70, 79, 79, 89, and 89% reduced 1-bit effective area compared to SRAM and 28, 50, 50, 74, and 75% compared to conventional SOT-MRAM, respectively.

From energy perspective, S-MLC, P-MLC, MBC-DD, and MLC-SD consume higher write energy compared to conventional SOT-MRAM. P-MLC, MBC-DD, and MLC-SD consume higher worst-case write energy while writing nonidentical bits on the cell due to the enforced rule of employing two SOT-MTJs with different I_c . However, if an equal probability of programming the various bits options is assumed, average energy efficiency similar to conventional SOT-MRAM may be achieved, whereas S-MLC involves writing one of the two MTJs in the cell using the energy-inefficient STT technology, which also degrades the device reliability as a result of supplying large current through the MTJ stack. On the other hand, diode-based SLC SOT-MRAM may achieve similar write energy to the conventional SLC SOT-MRAM. However, similar to other diode-based designs (MBC-DD and MLC-SD), it still incorporates a diode in the read operation, which may add additional energy penalty as the read voltage needs to be large enough to overcome the diode's on-voltage.

It is noteworthy that the SLC proposals such as diode-based SLC and MBC-DD would be preferred solutions, thanks to their offered significant area reduction in addition to maintaining the advantages of the SLC sensing such as improved BER. However, that requires further improvement in the diode-MTJ stack technology such that the diode transient response would match the required read performance and the diode's on voltage would be small, and thus, the read energy will be reasonable. On the other hand, the MLC proposals such as P-MLC, S-MLC, and MLC-SD require improvements in the MLC sensing techniques to enhance its sensing distinguishability and BER such that it will meet the industry standards.

IntechOpen

IntechOpen

Author details

Karim Ali
National University of Singapore, Singapore

*Address all correspondence to: karim.ali@u.nus.edu

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Alioto M. Enabling the Internet of Things: From Integrated Circuits to Integrated Systems. Cham: Springer International Publishing; 2017
- [2] Raha A, Jaiswal A, Sarwar SS, Jayakumar H, Raghunathan V, Roy K. Designing energy-efficient intermittently powered systems using spin-hall-effect-based nonvolatile SRAM. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*. 2018;**26**(2):294-307
- [3] Sakimura N, Sugibayashi T, Nebashi R, Kasai N. Nonvolatile magnetic flip-flop for standby-power-free SoCs. *IEEE Journal of Solid-State Circuits*. 2009;**44**(8):2244-2250
- [4] Ali K, Li F, Lua SYH, Heng CH. Compact spin transfer torque non-volatile flip flop design for power-gating architecture. In: 2016 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS). 2016. pp. 119-122
- [5] Chabi D et al. Ultra low power magnetic flip-flop based on checkpointing/power gating and self-enable mechanisms. *IEEE Transactions on Circuits and Systems I: Regular Papers*. 2014;**61**(6):1755-1765
- [6] Wang W-G, Li M, Hageman S, Chien CL. Electric-field-assisted switching in magnetic tunnel junctions. *Nature Materials*. 2011;**11**:64
- [7] Jabeur K, Di Pendina G, Bernard-Granger F, Prenat G. Spin orbit torque non-volatile flip-flop for high speed and low energy applications. *IEEE Electron Device Letters*. 2014;**35**(3):408-410
- [8] Jabeur K, Di Pendina G, Prenat G. Ultra-energy-efficient CMOS/magnetic non-volatile flip-flop based on spin-orbit torque device. *Electronics Letters*. 2014;**50**(8):585-586
- [9] Kwon KW, Choday SH, Kim Y, Fong XY, Park SP, Roy K. SHE-NVFF: Spin hall effect-based nonvolatile flip-flop for power gating architecture. *IEEE Electron Device Letters*. 2014;**35**(4):488-490
- [10] Bishnoi R, Oboril F, Tahoori MB. Non-volatile non-shadow flip-flop using spin orbit torque for efficient normally-off computing. In: 2016 21st Asia and South Pacific Design Automation Conference (ASP-DAC). 2016. pp. 769-774
- [11] Wang ZH, Zhao WS, Deng EY, Zhang Y, Klein JO. Magnetic non-volatile flip-flop with spin-hall assistance. *Physica Status Solidi Rapid Research Letters*. 2015;**9**(6):375-378
- [12] Imec demonstrates manufacturability of state-of-the-art spin-orbit torque MRAM devices on 300mm Si wafers, 2018
- [13] Fukami S, Anekawa T, Zhang C, Ohno H. A spin-orbit torque switching scheme with collinear magnetic easy axis and current configuration. *Nature Nanotechnology*. 2016;**11**(7):621-625
- [14] Seo Y, Kwon KW, Roy K. Area-efficient SOT-MRAM with a Schottky diode. *IEEE Electron Device Letters*. 2016;**37**(8):982-985
- [15] Kim Y, Fong XY, Kwon KW, Chen MC, Roy K. Multilevel spin-orbit torque MRAMs. *IEEE Transactions on Electron Devices*. 2015;**62**(2):561-568
- [16] Gupta SK, Park SP, Mojumder NN, Roy K. Layout-aware optimization of STT MRAMs. In: 2012 Design, Automation & Test in Europe Conference & Exhibition (DATE). 2012. pp. 1455-1458
- [17] Ali K, Li F, Lua SYH, Heng C. Area efficient shared diode multi-level cell

SOT-MRAM. *IEEE Transactions on Magnetics*. 2018;**54**:1-5

[18] Ahmed KA, Li F, Lua SYH, Heng C. Area-efficient multibit-per-cell architecture for spin-orbit-torque magnetic random-access memory with dedicated diodes. *IEEE Magnetics Letters*. 2018;**9**:1-5

[19] Huai Y et al. High density 3D cross-point STT-MRAM. In: 2018 IEEE International Memory Workshop (IMW). 2018. pp. 1-4

[20] Yang H et al. Threshold switching selector and 1S1R integration development for 3D cross-point STT-MRAM. In: 2017 IEEE International Electron Devices Meeting (IEDM). 2017. pp. 38.1.1-38.1.4

[21] Yang H et al. 3D cross-point spin transfer torque magnetic random access memory. *SPIN*. 2017;**7**(3):1740011

[22] Lim WC, Park BG, Bae JY, Lee TD. Effects of free layer materials and thickness on TMR behaviour in magnetic tunnel junctions. *Physica Status Solidi*. 2004;**201**(8):1696-1699

[23] Li YT et al. Novel self-compliance bipolar 1D1R memory device for high-density RRAM application. In: 2013 5th IEEE International Memory Workshop. 2013. pp. 184-187

[24] Wang ZH, Zhao WS, Deng EY, Klein JO, Chappert C. Perpendicular-anisotropy magnetic tunnel junction switched by spin-hall-assisted spin-transfer torque. *Journal of Physics D: Applied Physics*. 2015;**48**(6):1-7

[25] Ali K, Li F, Lua SYH, Heng C. Energy- and area-efficient spin-orbit torque nonvolatile flip-flop for power gating architecture. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*. 2018;**26**(4):630-638

[26] Dong X, Xu C, Xie Y, Jouppi NP. NVSim: A circuit-level performance, energy, and area model for emerging nonvolatile memory. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*. 2012; **31**(7):994-1007

[27] Lee A et al. RRAM-based 7T1R nonvolatile SRAM with 2x reduction in store energy and 94x reduction in restore energy for frequent-off instant-on applications. In: 2015 Symposium on VLSI Circuits (VLSI Circuits). 2015. pp. C76-C77

[28] Daly DC, Fujino LC, Smith KC. Through the looking glass—the 2018 edition: Trends in solid-state circuits from the 65th ISSCC. *IEEE Solid-State Circuits Magazine*. 2018;**10**(1):30-46

[29] Oboril F, Bishnoi R, Ebrahimi M, Tahoori MB. Evaluation of hybrid memory technologies using SOT-MRAM for on-chip cache hierarchy. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*. 2015;**34**(3):367-380

[30] Seo Y, Kwon K, Fong X, Roy K. High performance and energy-efficient on-chip cache using dual port (1R/1W) spin-orbit torque MRAM. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*. 2016;**6**(3): 293-304