UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Posgrados

"ANALYSIS OF THE WRITING PERFORMANCE OF STT-MRAM BASED ON A SINGLE AND DOUBLE MTJ"

Diego Fernando Criollo Caisaguano

Profesor

Marco Lanuzza, Ph.D.

Trabajo de titulación de posgrado presentado como requisito

para la obtención del título de Máster en Nanoelectrónica

Quito, 01 diciembre 2019

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

COLEGIO DE POSGRADOS

HOJA DE APROBACIÓN DE TRABAJO DE TITULACIÓN

"ANALYSIS OF THE WRITING PERFORMANCE OF STT-MRAM BASED ON A SINGLE AND DOUBLE MTJ"

Diego Fernando Criollo Caisaguano

Prof. Marco Lanuzza, PhD.

Director de Trabajo de Titulación

Omar Aguirre, PhD.

Director de la Maestría en Nanoelectrónica

César Zambrano, PhD.

Decano del Colegio de Ciencias e Ingenierías

Hugo Burgos, PhD.

Decano del Colegio de Posgrados

Quito, 01 diciembre 2019

© Derechos de Autor

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Firma del estudiante: _____

Nombre: Diego Fernando Criollo Caisaguano

Código de estudiante: 203304

C. I.: 1715604482

Lugar, Fecha Quito, 01 de diciembre 2019

DEDICACIÓN

El presente documento dedico a mis padres Luis y Rocío por su incomparable esfuerzo que lo realizan día a día, por permitirme conocer un nuevo mundo y sobre todo por haber confiado en mí, ahora estoy seguro que no hay mejores amigos que mis padres.

A mis hermanos que me apoyaron para cumplir este sueño.

DIEGO.

AGRADECIMIENTO

Agradezco a Dios por permitirme tener unos padres excelentes, por darme la oportunidad de continuar mi camino en la vida profesional, y por qué siempre está presente por las sendas en las que camino.

Al país Italia que me dio la oportunidad de conocer su cultura y formar parte de su vida cotidiana, a la Universidad De La Calabria por formarme profesionalmente y a la Universidad San Francisco De Quito como una puerta al mundo.

Agradezco a Marco Lanuzza como director de tesis que me extendió su mano para desarrollar este documento, de la misma manera a Esteban Garzón por su ayuda incondicional en la culminación del proyecto.

DIEGO

RESUMEN

El consumo de energía se ha convertido en una métrica crítica en el diseño de circuitos integrados (IC). La tecnología STT-MRAM tiene características para superar las limitaciones de consumo de energía y arquitectónicas de los sistemas informáticos convirtiéndose en un candidato potencial en la aplicación de memorias y lógica de baja velocidad y alta potencia.

Esta tesis presenta un estudio del rendimiento y la escalabilidad del voltaje de una matriz STT-MRAM de 128 x 128 basada en uniones de túnel magnético simple (MTJ) y doble (DMTJ) con anisotropía magnética perpendicular (PMA). Se presenta un análisis exhaustivo para diferentes configuraciones de la celda de bits híbrida CMOS / MTJ donde se considera las variaciones tóxicas y de proceso. Además, el entorno de simulación incluye dos tipos de modelos, el modelo compacto que establece el comportamiento MTJ y el modelo MOS para transistores.

El análisis de los diferentes nodos tecnológicos, y en particular para el doble MTJ, muestra resultados significativos en los nodos tecnológicos más escalados.

Palabras Clave: STT-MRAM, MTJ, DMTJ, anisotropía magnética perpendicular, variaciones tóxicas, celda de bits híbrida.

ABSTRACT

Energy consumption has become a critical metric in integrated circuit design (IC). The STT-MRAM technology has characteristics to overcome the energy consumption and architectural limitations of computer systems, becoming a potential candidate in the application of low speed and high power memories and logic.

This thesis presents a study of the performance and voltage scalability of a 128 x 128 STT-MRAM matrix based on single (MTJ) and double (DMTJ) magnetic tunnel junctions with perpendicular magnetic anisotropy (PMA). An exhaustive analysis is presented for different configurations of the CMOS / MTJ hybrid bit cell where toxic and process variations are considered. In addition, the simulation environment includes two types of models, the compact model that establishes the MTJ behavior and the MOS model for transistors.

The analysis of the different technological nodes, and in particular for the double MTJ, shows significant results in the most scaled technological nodes.

Key Words: STT-MRAM, MTJ, DMTJ, perpendicular magnetic anisotropy, toxic variations, hybrid bit cell.

CONTENTS INDEX

TABLE INDEX

FIGURE INDEX

1 INTRODUCTION.

1.1 Volatile and non-Volatile memories

The current design of an integrated circuit considers a classic memory hierarchy built with silicon-based devices such as DRAM and SRAM. In **Figure 1**, we can see the classical memory hierarchy where DRAM memory is used at the lowest levels, such as RAM and secondary memory. On the other hand, we have SRAM memories in the cache levels. As they approach the core, the memories have a higher speed and the volume decreases. When they are far from the core, the volume increases and the speed decreases.

DRAM memory is volatile and densely dense due to the configuration of a bit cell consisting of a serially connected transistor and capacitor. To store data needs continuous power source, otherwise the data is lost. In addition, due to the high density of DRAM memory, the power dissipation is high. On the other hand, the SRAM memory has a flip-flop cell and two access transistors. This memory does not require permanent updates to store data, has a low power dissipation, high speed and is very reliable.

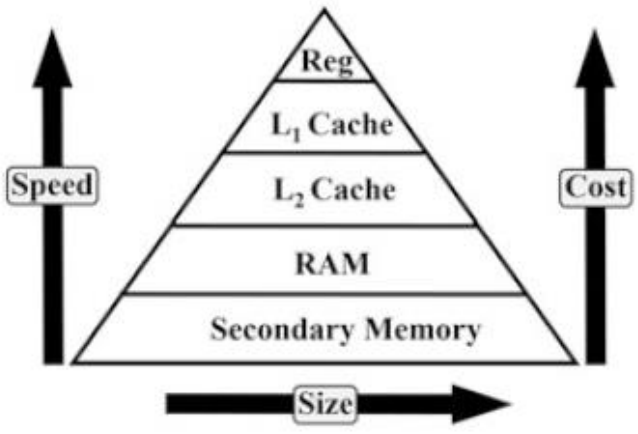


Fig 1: Conventional memory hierarchy [1]

The use of multicore processors has been on the rise in recent years, where high memory capacity is a requirement in complex applications. In addition, running multiple applications in parallel requires a high-performance memory system. Therefore, to address obstacles, you need to redesign the memory hierarchy to improve the overall performance of your systems. Flash memories like NAND and NOR show good prospects on SRAM and DRAM due to their non-volatility. On the one hand, NOR has high latencies for writing and erasing, on the other hand NAND memories have a very high latency in comparison of DRAM. This is why we need to work intensively on emerging non-volatile memories.

The new non-volatile memory technology differs in terms of materials and switching mechanism. First, for non-volatile memory materials we have ferroelectric dielectrics, ferroelectric metals, transition metal oxides and carbon materials. As for switching mechanisms, these include quantum mechanics phenomena, ion reactions, phase transition and molecular reconfiguration. On the other hand, there are new technologies such as the PCM (phase change memory) structure, which consists of a non-volatile memory element and an access device allowing 1-bit storage. In these phase-change materials, they show smaller-sized phase transition characteristics, high crystalline temperature, low thermal conductivity, and improved strength.

An example of emerging non-volatile memories following an electrical approach is ReRAM. This works according to the principle of conductivity of the dielectric material. A dielectric becomes conductive through the formation of filaments when a current is passed to a sufficiently high voltage. The broken and reformed filament represents a high and low resistance, respectively. The two-terminal storage device consists of a metal oxide inserted between two electrodes. On the other hand, we have the magnetic

approach where MRAM (magnetoresistance RAM) memories have presented favorable characteristics such as low power and high-speed operation, high density, long data storage time and easy integration with CMOS process.

MRAM, if low in the phenomenon of ST (Spin Torque), demonstrate great potential to solve problems that do not have other memory technologies. STs are classified as "spin-transfer torque" (STT) and "spin-orbit torque" (SOT). The latter, the SOT-MRAM architecture is energy efficient and has faster access to writing by optimizing the independent writing path. Another memory model is the DW-MRAM based on Domain Wall (DW) with two additional layers on the sides of the free layer, which improves write performance with respect to the STT.

In general, several emerging non-volatile memory technologies were mention. These memory technologies provide performance, maturity, and scalability. STT-MRAM-based memories and its SOT-MRAM and DW-MRAM improvements have become rivals for the next generation of non-volatile memories, thanks to its high density, strength and improved retention.[1]

1.2 Comparison Of Memories Technologies.

The non-volatile memories described at the beginning of this chapter present advantages and challenges. ReRAM has the highest integration density due to the1T memory structure. The main advantages of ReRAM are low storage consumption and high performance. The ReRAM is simple in structure, has a low cost and is of higher density, has versatile materials, structures and behaviors. Forward, the scalability of the ReRAM is better than PCM and STT-MRAM.

The most emerging non-volatile memory is STT-MRAM and is due to its best performance. Still, the problems with STT-MRAM are reliability and the thermal balance

BEOL (back-end-of-line). There is solved with the optimization of separate read and write paths as they present the SOT-MRAM and DW-MRAM; thus, improving reliability through the optimization of reading paths Separate / writing. On the other hand, PCM and ReRAM memory is better scalable than STT-MRAM. STT-MRAM has a better switching mechanism than other memories, so faster than writing and reading. In terms of power, PCM, STT-MRAM and ReRAM have a high, moderate and low write dissipation. STT-MRAM has better retention, strength and variability than other memories.

Table 1: Comparison between emerging memories.

	EMERGENCY MEMORY				**CONSOLIDATED MEMORY**	
	SOT_MRAM	STT_RAM	PCMS	RRAM	DRAM	FlashNAND
Non-Volatile	Yes	Yes	Yes	Yes	Non	Yes
Resistance	High (5x10e10)	High (10-e12)	Media (10-e8)	Low (10-e6)	High (10-e15)	Low (10-e5)
Product the last technology node (nm)	-	40 nm	20 nm	130 nm	-	15nm
Cell size (cell size F^2)	-	Media (6-12)	-	Media (6-12)	Small (6-10)	Very Small (4)
latency in reading (nm)	very fast (0.21ns)	Fast (10-20 ns)	Fast (50-100ns)	Media (250 ns)	very fast (ns)	Slow (100.00 ns)
Energy consumption	300 (pJ)	media (50pJ/bit)	Media	Media (6 nJ/bit)	Low	Very High
Price 2016 ($/Gb)	-	High ($200 - $3000/Gb)	Low (<$0.5/Gb)	High ($100/Gb)	Low (<$1/Gb)	Very low(<$0.05/Gb)
Suppliers	-	Everspin	Intel / Micron	Adesto	Micron, SK Hynix	Samsung, Toshiba, Intel

According to the memory application, there is a higher or lesser consumption market, which means that a hard disk or NAND memory is required to back up the information. On the other hand, for fast services where data is used for a short time, you need a fast memory like SRAM and stable as DRAM. In fact, the goal is to look for a possible replacement of SRAM or DRAM due to scaling. We have potential features with magnetoresistance memories (MRAM) where STT-MRAM and SOT-MRAM have low energy consumption and speed improvements according to Table 1.

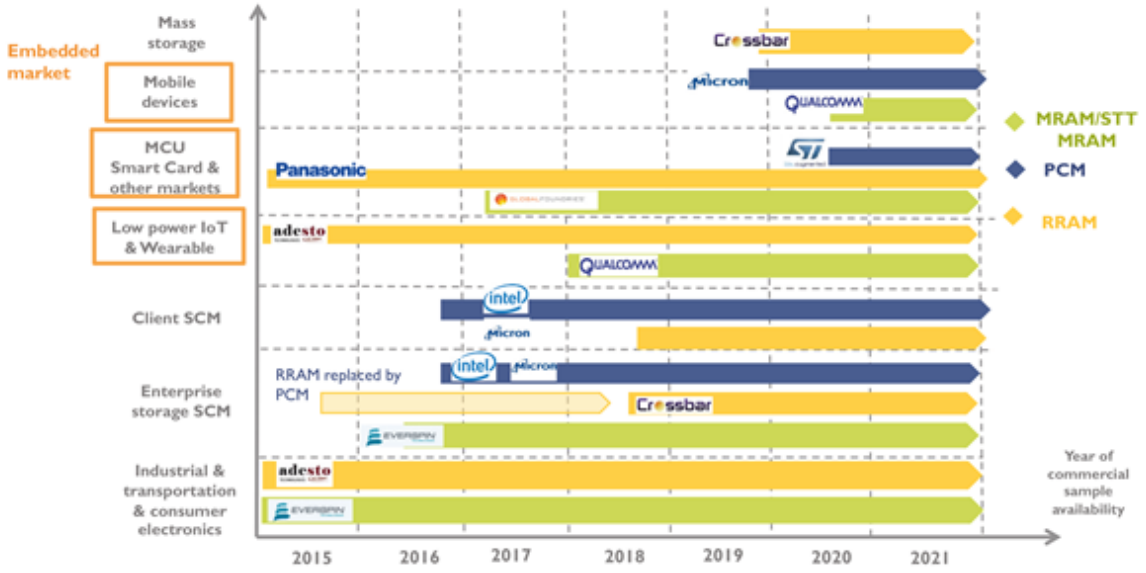


Fig 2: Market applications of NV memories *[2]*

Panasonic and SMIC have selected RRAM, while major companies such as (TSMC, GF, and Samsung) propose STT-MRAM in 2017-2018 and ST Microelectronics has selected PCM for the 28 nm node in 2020. However, many key manufacturers, including Renessa's, Infineon, Texas Instruments, Micro Chip and Cypress are not officially elected. Given the trend towards STT-MRAM among manufacturers, STT-MRAM is expected to get over the integrated market in2021. However, due to its low cost, RRAM/PCM could take a larger market share.

1.3 Spintronics-Based Memory

During the construction of transistors and other nanoscale-sized devices, processors and memory become so dense that infinitesimal currents combine producing a high-energy consumption and heat dissipation. Moreover, the quantum effects that were insignificant are now very remarkable. Spintronic devices use the third property of electrons spin. For this purpose, magnetic materials are used in spintronic devices to store information, provide non-volatility and provide resistance that cannot be compared with other memory technologies. Using the natural spin of electron spin allows the ability to control the movement of electrons, thus helping to do things such as writing and reading information.

Fig 3: 1 GB STT-MRAM (Spin-transfer Torque MRAM) [3]

It is important to note that MRAM devices have the magnetoresistive giant (GMR) effect that was discovered in the late 1980s by two European scientists from the University of Paris-South. They saw very large resistance changes of 6% and 50% on experiments at low temperatures and in the presence of very high magnetic fields. Thus, the GMR is the variation of electrical resistance in response to an applied magnetic field. In addition, the application

of a magnetic field, the multilayer implies a significant reduction in the electrical resistance of the multilayer.

1.4 Generations Overview of MRAM Technology

MRAM technology can be classified according to the switching method used to write data. Generally, the first generation of MRAM includes methods based on magnetic fields to program the array, and words memory cells. Toggle MRAM is the only first-generation mass production MRAM. The field change represents an unlimited writing resistance, so reversing the magnetization of the free layer with a magnetic field does not produce wear effects. The disadvantage of this technology to scale the size of the smaller cells, due to some factors such as the size of the switching currents and the complex geometry of the memory cell.

The second generation of MRAM uses the STT effect to program the matrix. STT switching is achieved with magnetic tunnel junction devices (MTJs) that have magnetization in the plane or perpendicular to the plane. Commercial production of the STT-MRAM device in the plane, which began in 2015 with a storage capacity of 64 MB, and in 2016 its production, grew with a storage capacity 256 MB. [4]

The third generation MRAM studies physical phenomena, including: voltage-controlled anisotropy (VCA), voltage controlled magnetism (VCM), Hall effect (SHE) and spin orbit torque switching (SOT). Reliability issues such as wear, and drift parameters should be better understood for practical VCM devices. In fact, these new devices should be implemented in a three-terminal [5] cell configuration that is not compatible with [6] high-density memory array. [4]

1.5 Working Stage Of MRAM Circuits

The latter sub-section shows some notable demonstrations of working circuits, illustrating the rapid progress in the development of technology over the past 15 years.

In 2000, Scheuerlein reported the functionality of a field-switched MRAM circuit using AlOx-based MTJ devices and Stoner-Wolfarth-type writing circuits. The basic read and write function was demonstrated in a 1 kb memory array with write pulses that were too short of 2.5 ns and with a read operation that was too short of 10 ns, using a double-cell read configuration.[7]

Durlam and col. describes the first Toggle MRAM circuit in 2003, a fully functional 4Mb circuit with 1T1MTJ architecture and AlOx-based MTO devices, indicating the effectiveness of writing to solve the problem of high writing error rates. The free-switching layer consists of a three-layer stack of synthetic antiferromagnetic (SAF) where the magnetic moments of the upper and lower layers are almost balanced. The magnetization direction of the ferromagnetic substrate 1 (FM) relative to the fixed FM layer determines the resistance state of the bit [8]. This circuit is designed to be a product, but not a demonstration, with an exclusive local bit line array architecture, a cell size of 1.55mm^2, asynchronous compatibility with 16-bit SRAM interface, and read loop times and symmetric writing less than 25 ns. This chip was introduced as a commercial product in 2006 and is manufactured today by Everspin Technologies, Inc. [9]

Significant results are obtained from several STT-MRAM demonstration circuits, using MgO-based MTJ devices. Starting with a 4KB test device in 2005, they include[10]: a 2 MB circuit and device data in 2007[11], Give stadistic array integrated da 4 KB con 70 × 210 bit

nm^2[12] and devices with perpendicular magnetization in 2008, array[13] integrated with 54 nm CMOS technology, 4 KB MRAM array with perpendicular bit switching in 2010.[14]

The first STT-MRAM product, announced in 2012, is a 64 Mb STT-MRAM circuit, the chip is manufactured with MTJ devices with floor-to-plan anisotropy in a cell with a transistor and an MTJ, integrated with 90 nm CMOS technology and introduced in a standard BGA DDR3 JEDEC. This chip is compatible with the available DDR3 memory controllers and has been market end by Ever spin Technologies, Inc. in 2015[4]. On the other hand, in 2014, we provide data on the performance of an 8 Mb STT-MRAM demonstration chip with high-energy barrier for storing information bits compatible with automotive applications and MTJ devices that tolerate high required processing temperatures for integrated memory. Although CMOS circuits have 90 nm technology, the MTJ uses devices as small in diameter as 45 nm, demonstrating reliable switching with short pulses like 2ns in a cell architecture with a transistor and an MTJ. [4]

2 Spin Transfer Torque MRAM (STT-MRAM)

One of the limitations of MRAM technology in the switching field is the difficulty of maintaining minimum error rates and high data retention when scaling MTJ devices. The effects affect the write current, error rates, and loss of data retention time. Because of these two disadvantages of scalability, the industry decides to switch from alternating to STT switching for technology nodes less than 90 nm. [4]

2.1 The Magnetic Tunnel Junction MTJ. (Storage Device)

The elementary device in an MRAM memory is the MTJ. An MTJ consists of a ferromagnetic layer that stores information (also called a "free" layer), a tunnel layer (usually AlOx or more commonly, MgO) and a ferromagnetic reference layer (also called a "fixed" layer). Within

the MTJ (Magnetic Tunnel Junction) two very important physical phenomena are generated for reading and writing in a memory cell, the first effect is the Tunneling Magneto-Resistance Ratio (TMR), and the second effect is magnetic anisotropy (PMA). To understand these two phenomena, it is important to consider the principle of the magnetoresistance effect.

MRAM is based on magnetic tunnel junction (MTJ) devices, where the simplest is called single barrier MTJ (SMTJ). The SMTJ has a non-magnetic spacer inserted between two ferromagnetic layers. This spacer is an oxide barrier. The magnetization of one layer is fixed and the other is free. The magnetic polarization in the free layer can be changed when a current is applied in the device. There are two possible states, minimum resistance (logical state "1" or anti-parallel) and maximum (logical state "0" or parallel). The TMR STT cell contains an MTJ (magnetic tunnel junction) and a MOS transistor called cell 1T-1MTJ. (WL) is connected to the transistor port to access the information stored in the MTJ. (BL) and the source line (SL) are connected to the Free Layer MTJ and the transistor source terminal, respectively. Magnetic anisotropy depends on geometry and is classified into two categories: flat magnetic anisotropy (IMA) and perpendicular magnetic anisotropy (PMA).

2.1.1 Magnetoresistance Effect

This phenomenon was discovered in the 1980s where there is a change in the electrical resistance of a material in the presence of an applied external electric field [3]. This phenomenon focuses on the behavior and properties of the magnet due to external influences. If we think of a simple magnet, it is understood that the electrons in a magnet have a spin state (with a north and south direction). We can see in **figure 4** that these electron spin states can change by rotating the poles of the magnet. You need a reference point to know the appropriate state of the electron.

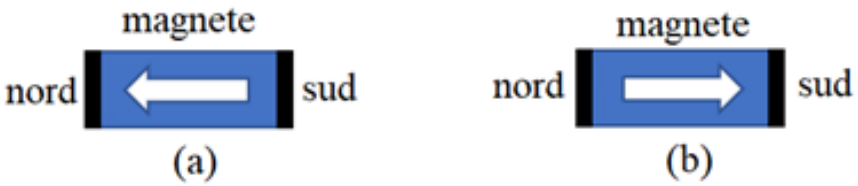


Fig 4: two configurations of the magnet: (a) south - north, (b) north - south where the black layer represents the poles.

A magnet has two zones that most strongly manifest the North Pole attraction and the south magnetic pole **figure 5**. This phenomenon is related to the high and low resistance in a ferromagnetic material. Then, you can use it in the implementation of electrical circuits **figure 6**. Basically it is said that it presents a low resistance when the 2 magnets are attracted to the opposite poles, but if we change the position of one of the magnets and encounter the same poles there is a repulsion between them, to this the high resistance is related. Thus, these effects occur in ferromagnetic materials, which are used for the development of new MRA memory technologies.

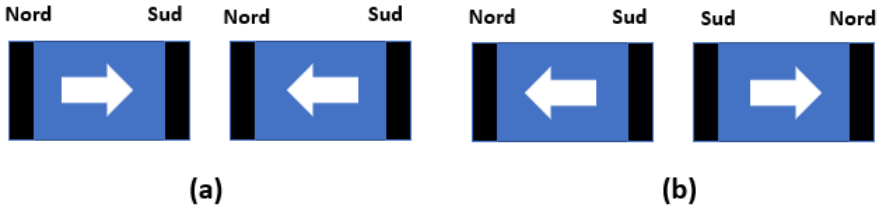


Fig 5: (a) NS-NS configuration: magnetic attraction is felt. (b) NS-SN configuration: there is a magnetic repulsion

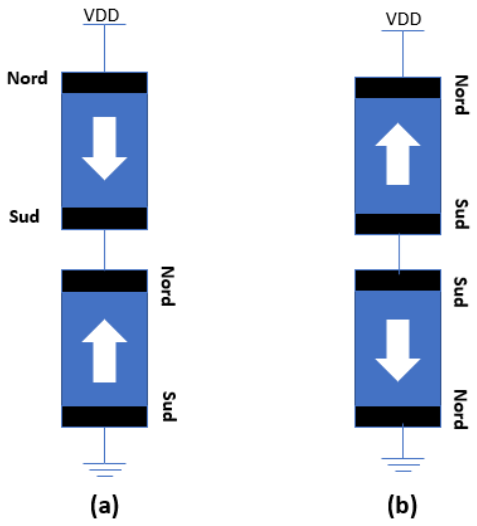


Fig 6: (a) Low resistance and (b) high resistance

2.1.2 Overview of the Conventional Structure of Magneto-Resistance Tunnel Junction (MTJ)

The MTJ structure can be classified by the type of insulation. In **figure 7,** we can see a classic MTJ structure consisting of two ferromagnetic layers and an insulating layer. The first is the fixed layer (PL): also called as the necessary reference layer when you want to change the state of the electronic rotation, a next layer is the insulator defines the type of magnetoresistance structure.

In an MTJ structure, we use an insulator (oxide), this device is a TMR. Unlike, if we use a non-magnetic metal these devices is giant magnetoresistance (GMR). In addition, both configurations are related to the effects of quantum mechanics. Another layer present in the MTJ structure is the free layer (FL) that provides the rotation configuration as Parallel (P) or Anti-Parallel (AP). In summary, PL and FL function as a polarizer and analyzer respectively as electrons cross a thin oxide tunnel barrier, which is commonly used.

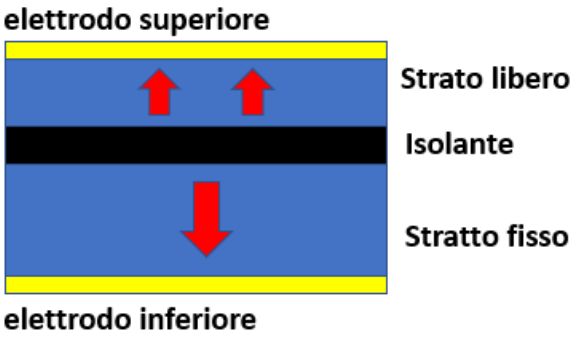


Fig 7: MTJ base configuration

MTJ can be switched between two stable states. When the free layer that the pinned layer is magnetically aligned, the configuration is called a "parallel" state (P) and when the free layer and the pinned layer are magnetically misaligned, the configuration is called a state (AP).[15]

When the MTJ is in the P state, the density of similar spin states around EF (Fermi Energy) is very high in ferromagnetic layers. Conversely, the density of spin states around EF (Fermi energy) in the ferromagnetic layers is very low when the MTJ is in AP state. Therefore, MTJ is low in state P and high in AP state. This difference is called "tunneling magnetoresistance ratio" (or TMR), which is given by:[15]

$$TMR = \frac{R_{AP} - R_P}{R_P} \tag{2.1}$$

Where $R_{AP}R_P$ and represent respectively the high and low resistance of MTJ, which corresponds to the stable states of Antiparallelo or Parallel. By increasing the MR, the better the MTJ and the better differentiation between states.

Previously we have seen two different types of structures: GMR and TMR. Between these two, the most appropriate structure is the TMR which has an MR greater than 100% [16].

However, having a high MR is not enough due to the different problems that this structure must overcome to achieve a good performance in the design of the circuit.

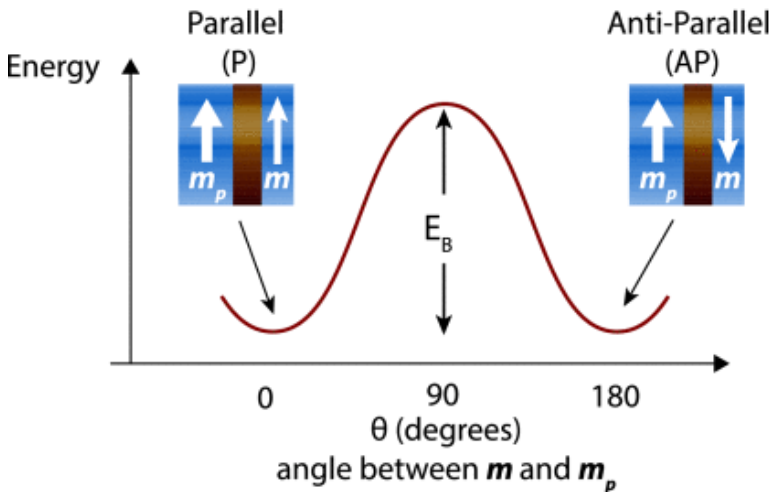


Fig 8: Energy and spin configuration [17]. The degrees represent the angle between the magnetization PL and the magnetization FL.

Figure 8 clearly shows how you can change the rotation status. This change of status is due to the crossing of an EB energy barrier. In addition, the energy is the same in the parallel state and in the anti-parallel state (EB (0°) - EB (180°)), but it is not always true due to the presence of high-order effects. When the MTJ is subject to high-order effects, the energy barrier is the minimum energy between EB (0°) and EB (180°).

2.1.3 Perpendicular Magnetic Anisotropy (PMA)

There are two technologies for applications in STT-MRAM memory, those of MTJ magnetic layers with PMA and those of the magnetic layer in MTJ with IMA, PMA have perpendicular magnetizations at the plane of the magnetic layers. MTJ technology with IMA anisotropy in which magnetic layers have magnetizations that are located on the in plane of the magnetic layers. If we talk about structure **figure 9**, we can see that the IMA has a larger area than

the PMA, so, to achieve a high integration density and a low critical switching current (improving writing efficiency), PMA is used.

The combination of the CoFeB-MgO-based magnetic tunnel with perpendicular anisotropy (p-MTJ) shows high potential for use in spintronic-based integrated circuits and magnetoresistive random access memories of spin transfer. In this paper, we examine the development of p-MTJ using the CoFeB-MgO and double CoFeB-MgO single interface structures. The performance improvement is subject to the addition of metal roofing layers in the MTJ, which present a positive improvement in the device's characteristics. The addition of layers of hedging complicates the production process, the price increases by orders of magnitude compared to the memory devices of typical consumer applications.

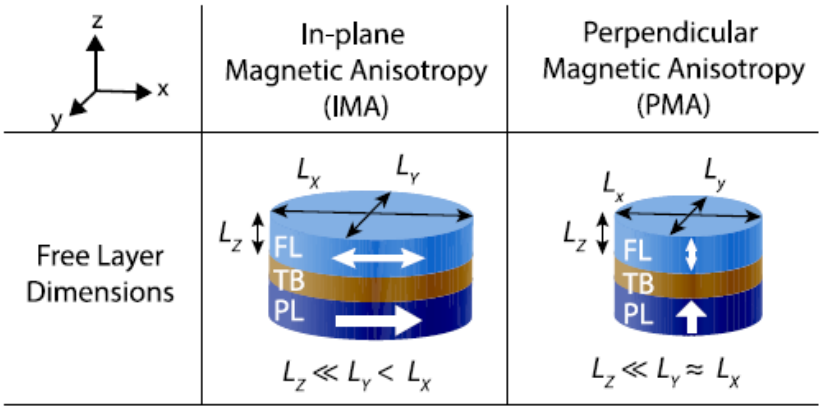


Fig 9: In-plane magnetic anisotropy "(IMA) and" perpendicular magnetic anisotropy "(PMA). It is shown that the IMA has Lx> Ly while PMA has a circular section that makes it suitable for integration. [17]

A device with P-MTJ perpendicular technology can be created using an FL and a PL, where high temperatures and a high current flowing in the direction of the PL create a large field from the PL and can modify the quantum properties of the FL. As a result, it may present an unwanted change in the storage of information. To solve this problem, two alternate PLs are built in the direction anti-parallel and positioned after the FL. This technology with

two PLs is an alternative structure that was published two years ago. However, regardless of the topology of the structure, to overcome the different stability problems in variability or reliability, the manufacturing process is the most important and is becoming very complex. Today, to build a MTJ is built with about 15-20 layers (some of these layers are of a few atoms).[18]

2.1.4 Write and Read in MJT

The most important operations of a memory are reading and writing, it is essential to know each of these operations to relate to the MTJ device and how it relates to the memory elements. **Figure 10** shows the typical resistive behavior of an MTJ device in which two logical states (high-strength states and low-strength states) are well defined. Then, in the reading operation a low polarization voltage is applied and considering the resistance value. Due to the ferromagnetic material, there is no relaxation of the material that causes a drift of resistance that can affect the storage of data [19]. The MTJ device has no such effect, so the states always remain the same and the memory lifetime is highly reliable.

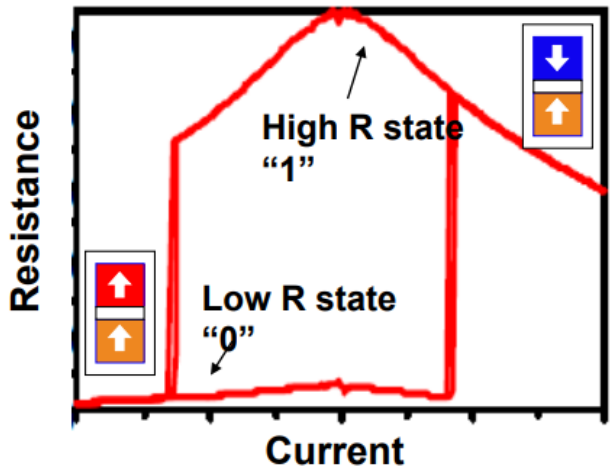


Fig 10: Read operation that checks the resistance of the low voltage device. [19]

Instead, the write operation uses the spin transfer pair (spin state change). A current flowing through the device has two possibilities. The first possibility is checked when a current circulates and passes through the device without changing the rotation state and a second possibility is checked when a current circulates and is compared to the critical current and can change the rotation state. This change is based on the spin impulse transfer, which is explained in detail below. However, hysteresis is shown in the writing operation in **figure 11**.

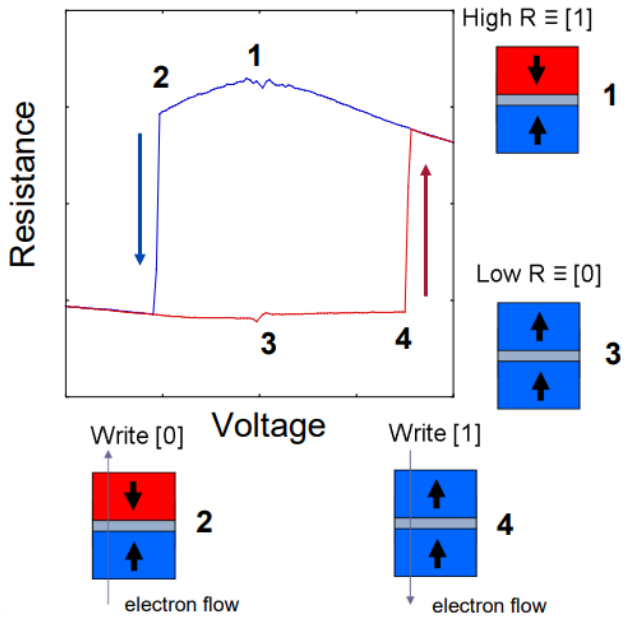


Fig 11: Writing operation that checks the resistance of the device. [19]

2.1.5 Structure Spin-Transfer Torque (STT).

Several models were made and tested during the development of MRAM devices. One of these models is based on the transition from one state to another induced by an H magnetic field; this state change is called field-induced magnetization switching (FIMS) [6]. To generate this magnetic field a current must circulate through a conductor, the change takes place with the appropriate current value. That is, due to the high number of cables and interconnections (growing due to scaling) near the MTJ, unwanted switching may occur.

Therefore, FIMS is no longer considered and to cover scalability issues the STT structure has been developed.

STT technology is focused on another type of switching. The difference between STT and FIMS is that STT is based on GMR or TMR, where the orientation of the spin rotation is changed by the magnetic field generated by a current flowing through an electrical conductor. In contrast, STT uses polarized current per rotation to change the magnetization state [20]. In this model, it is considered a stream of electrons from the PL to the FL. In addition, the ferromagnetic material corresponding to the PL has a strong polarization that can change the state of electronic rotation. After, the electrons create a tunnel through the material barrier and a pair is produced in FL magnetization causing the alignment of FL magnetization, m, with PL magnetization, MP. Conversely, if the electrons travel from FL to PL, the electrons will try to align with m. **Figure 12** shows a possible state change based on the direction of an applied current.

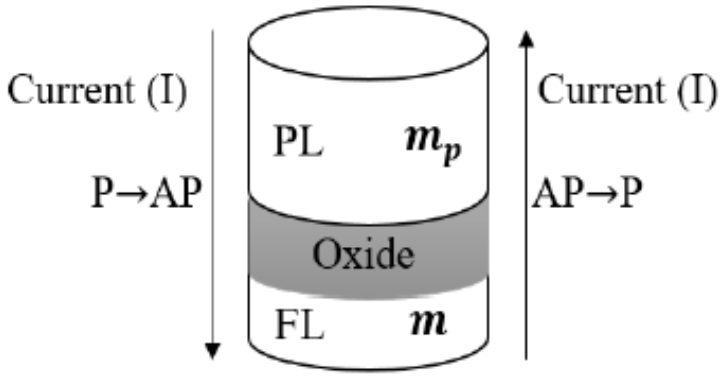


Fig 12: Physical structure MTJ considering the current direction and the corresponding switching state. [21]

After much analysis, it was concluded that magnetization occurs in the FL. To discuss the details of how STT-MRAM works, you need to introduce the way the spin-transfer pair

determines switching. The magnetization dynamics of any magnetic layer can be described by the Landau- Lifshitz-Gilbert equation (LLG): [17]

$$\frac{\partial \boldsymbol{m}}{\partial t} = -|\gamma_0|\boldsymbol{m}x\boldsymbol{h_{eff}} + \propto \left(\boldsymbol{m}x\frac{\partial \boldsymbol{m}}{\partial t}\right) + \boldsymbol{STT} \quad (2.2)$$

Where **m** is a unit vector that describes the direction of the magnetization of the mono domain, γ_0 is the electronic spin-magnetic ratio and $\propto$ is Gilbert's damping factor. An effective magnetic field, HEFF, models the forces acting on the single domain. The solution of the equation is complex for the fact of working in the micro-magnetic field, avoiding a series of differential and integral mathematical development, the expression (2) is literally exposed. In the field of simulation, MTJ modelling is based on a micro magnetic analysis or a compact analytical model, which will be explained; these patterns are described in the following equation:

$$\frac{\partial m}{\partial t} = -m \times \left[h_{eff} - \alpha\frac{dm}{d\tau} - \beta\frac{m \times m_p}{1 + c_p m.m_p}\right] h_{th} \quad (2.3)$$

Where it represents the magnetization of m_p the PL, $\gamma_0 M_S$the time is$M_S h_{th}$ the saturation magnetization, it is the density of normalized injected current, it is the thermal field in which it describes a white Gaussian noise, $\sqrt{2 \propto K_B T / \mu_0 M_S^2}$ is the intensity of thermal fluctuations, is the permeability to the vacuum μ_0T is the temperature, K_B is the constant of Boltzmann, V_{FL} it is the volume of the free layer and cp = η^2 is the description of the asymmetry of the pair of rotation, is the spin polarization factor. However, it is important to consider the influence of temperature. The temperature is always present and directly affects the MTJ switching time constant when changing the reference angle between the PL and FL magnetization. [22]

2.2 Comparation between MTJ Single and Double.

There are improvements in STT-MRAM devices, in reference to its structure, the first is known as single barrier MTJ (SB) and a second MTJ structure with double barrier (DB). Until now, there was talk of a single MTJ barrier. The single barrier device consists of an FL and a PL, from the direction that the current takes we can change it from the P state to the AP state or from the state AP to the state P. On the other hand, two fixed layers, the top fixed layer (PL_T) and the bottom fixed layer (PL_B), includes two layers of oxide (the upper oxide layer and the bottom oxide layer) and an FL that constitutes the MTJ double barrier configuration. The fixed layers are oriented in the opposite direction from each other, thus exploiting the torque force as the electrons flow through the device. In addition, an SB or DB structure is always regarded as a two-state device associated with low strength and high resistance as discussed earlier the typical SB and DB MTJ parts are shown in the **figure 13**. [22]

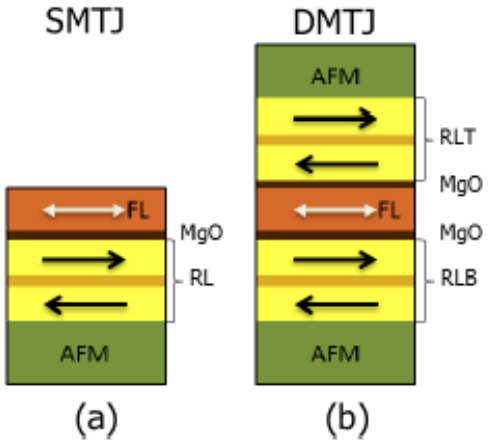


Fig 13: Various implementations of STT-MRAM cells (a) MTJ single barrier (b) dual barrier MTJ [23]

The MTJ-SB and MTJ-DB structure have the same principle of operation. Considering the structure of the DB, the incident electrons that are oriented in the same direction as the

anchored layer can pierce the oxide wall by creating a tunnel through the first oxide. Then, FL, which is in the opposite direction to the PLT, will develop a pair to change the rotation state to another state. In this rotation, change the electrons will be favored by the PLB, with this action the torque is stronger and causes a switch with greater speed, that is, less current will be needed to make the change. Therefore, the main feature of MTJ-DB is to work at low switching currents. Also, in the case of the MTJ-SB structure, the transition from P to AP is too slow as electrons enter from the FL, so we have low polarization efficiency. While in the case of DB there are fixed layers in the upper and lower terminals, where electrons will always have the presence of an PL regardless of the direction in which they take the electrons. Another advantage is the use of a low voltage source for these structures; it can be considered a low-power solution. Still, the problem is related to managing high write currents either for MTJ-DB structures or for MTJ-SB.

2.3 STT-MRAM

We have several types of configuration: lower fixed (also called standard configuration - SC) and a second higher fixed configuration (also called reverse configuration - RC) as shown in (**figure 15 a).** To access the MTJ structure, you need an access transistor by applying a VDD voltage on the transistor port. If we consider the write operation, the word line of a cell of type N (WLn) will be loaded to VDD then, a current will circulate between the bit line (BL) and the source line (SL). In the previous chapter, we mentioned that read and write operations to a TT-MRAM memory are not decoupled, these two operations occupy the same path. That is, it degrades the source (VGS < VDD) during one of the operations when the current is driven by the transistor from the source-line (SL) to the bit-line (BL).

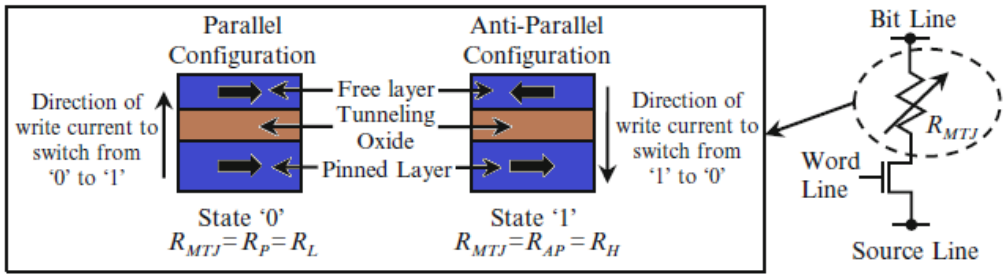
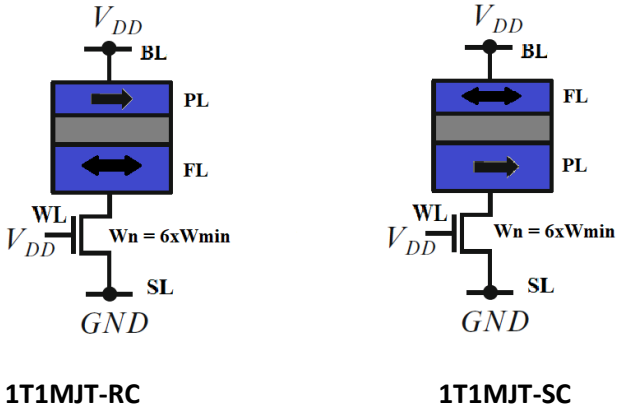
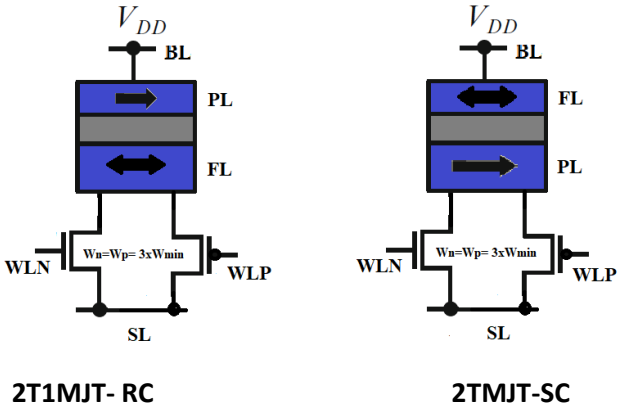


Fig 14: The storage device in the MRAM memory cell is the junction of the magnetic tunnel. The memory cell consists of an access transistor and the connected storage device. [15]



(a)



(b)

Fig 15: (a) 1T1MTJ-RC and 1T1MTJ-SC. (b) Two RC transistors (2T1MTJ-RC) and SC (2T1MTJ-SC).

Unfortunately, when the voltage source degrades causes an effect, this is the reduction of the write Iwriting current. Because the degeneration of the voltage source in the bit cell is included in **figure 16**, if we pay attention to a DB-MTJ structure, RC is considered when the

TP is attached to BL and SC when BP is attached to BL. In **figure 16**, because of the current address, we observe that the source terminal becomes the transistor terminal that is connected to the MTJ. To reduce, control, and tolerate the effect of source degeneration, a two transistors configuration is performed as shown in **(figure 15 b)**. Therefore, there are four types of configurations in which each represents an STT-MRAM bit cell. [21]

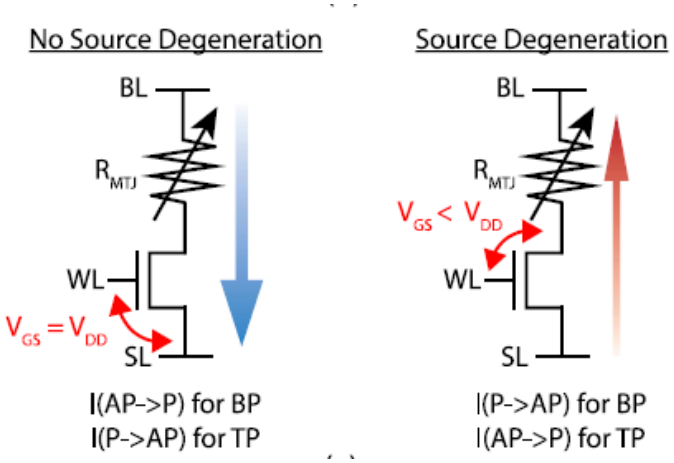


Fig 16: Cases of degeneration at source for SB and DB MTJ. The arrow represents the flow of electrons from SL to BL. [24]

The **figure 16** shows an illustration of the MRAM architecture. Arrays of multiple MRAM cells form a memory device. A typical MRAM cell has a transistor and a magnetoresistance element, very similar to a DRAM, which contains a transistor and a capacitor. While the charge stored in the capacitor of a DRAM defines its memory state, the strength of the existing element determines states 1 and 0. For each MRAM cell a transistor is required, since the absolute difference between resistances, therefore, the two-state tensions are not high enough to work without a transistor. In addition, the transistor also provides the current required for the write operation.[25]

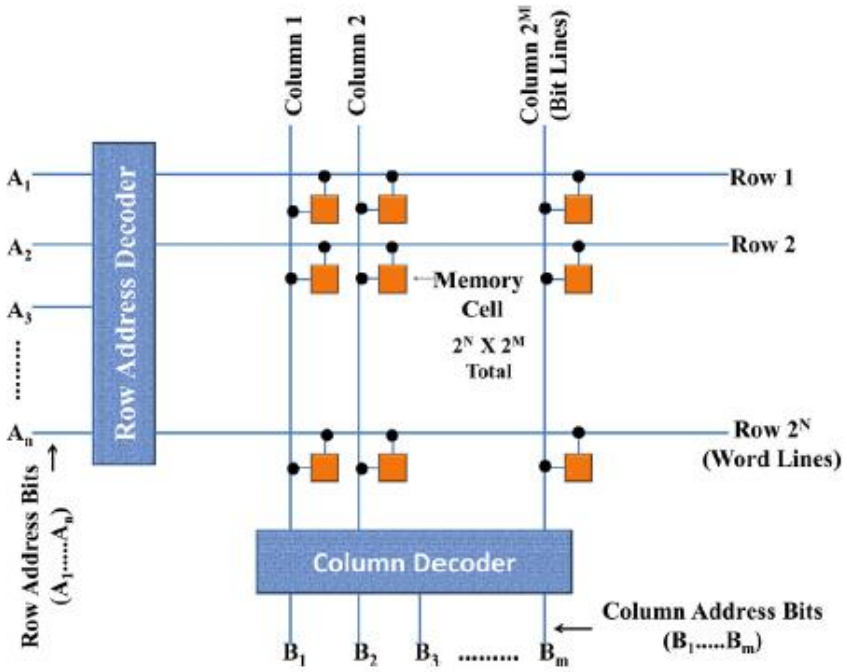


Fig 17: Scheme of a matrix of MRAM cells in a typical memory architecture. Each frame represents a cell that typically includes a transistor and a magnetic tunnel junction element. [25]

We need a column decoder (for BL and SL) and a row decoder (for WL). The WL controls N cells, so you need a write controller to have a good signal response speed. Remembering the 1T1MTJ and 2T1MTJ projects, in the case of 2T, we will have two WLs, which means that we must consider two buffers for each cell. To make a read, WL is attached to Vdd and a current read is generated to circulate in the bit cell. So, if we want to detect a voltage, a voltage drops (V-drop) is generated between BL and SL.V_{drop}. To know the state of the bit cell currently, a detection amplifier (SA) is used. SA will compare V_{drop} e V_{REF} giving a state AP ("1") When V_{drop}> V_{REF} and a state P ("0") When V_{drop}<V_{REF}. [21]

3 CMOS / MTJ Hibrid Memory Design

In this section, we will focus on the analysis and simulation methodologies for an STT-MRAM memory. We will cite an approach used to know the behavior of an MTJ structure, it is necessary to be clear about the behavior since it is a requirement to understand the

simulation of MTJ circuits. In addition, this document includes a brief description of the simulation structure. Finally, we will compare the single barrier to the double MTJ barrier.

3.1 Simulation Methodology

From this section we will talk only about simulations, the devices with their different nodes will be simulated in the Cadence software® - Virtuoso®. To get reliable simulation results, we'll look for different approaches to getting a simulation with reliable data. There are methodologies based or built with Verilog-A code. However, only one of these methodologies is used because of the lower computational load that shows the simulation.

3.1.1 MTJ Circuit Approach

The goal of the circuit approach of an MJT structure is to build a hybrid circuit design between the CMOS and MTJ technologies. In this approach, FinFET technology is considered not to glide, as the first step is to have as a basis the FinFET model or a Monte Carlo statistic available in doping; for this basic template, you can change parameters, such as finger number, length. Therefore, this basic model (FinFET model) is associated with a compact model for MTJ, which has code written in Verilog-A. Simulation results are a function of critical switching currents, and the statistical distribution of MTJ switching time in both switching transitions is used as input to a table-based MTJ model (LUT) developed in Verilog-A, which is incorporated into the Virtuous Cadence environment to perform electrical simulations of CMOS/MTJ circuits under testing. The Verilog-A code also includes the effect of changes related to the manufacturing processt_s of the MTJ structure in terms of oxide thickness variability (t_{ox}), FL thickness (t_{FL}) and TMR ratio [26]. Being a Verilog-A programming code, you can also change different parameters and size of the MTJ device such as physical size, temperature characteristics, and resistance of the magnetic device.

Still, by combining these two models, FinFET and MTJ, we can achieve hybrid circuit design simulation. This process is crucial when it is necessary to introduce a different model of an MTJ structure because to date it is not commercially available, that is, there are no business models. When we want to use a tunnel FET, it happens that there is no simulation to enter it commercially.

3.1.2 Compact Analytical Model

The use of compact macro spin models involves the use of tools in the design of advanced circuits; these compact models are used for modeling in the magnetization dynamics of the STT-MTJ structure. These models are implemented based on an equivalent electrical circuit, consisting of integrated electrical devices or formulas with a high degree of simplified complexity. Switching time probability statistics (t_S) are critical to modeling proper process behavior in switching. In general, any analytical model based on the structure of macro spin takes advantage of a simplified description of the probability distribution function (PDF) of the t_S for the rapid change regime assuming a function of normal probability distribution.

The following illustration shows in detail a complete block diagram of the compact analytical model. It is a generic block, valid for MTJ-SB and MTJ-DB configurations. The only difference in configurations is the block called "Resistance and TMR that depends on the inclination" and the "Analytic formulation". This model estimates five effects that mainly affect the switching phenomenon of an MTJ device. These effects are the transformations of the MTJ process, another effect is the asymmetry of the pair when there is a switch in the device, a subsequent effect is the temperature connection, it is considered as an effect heating or the and as the latest voltage-dependent effect due to the perpendicular

magnetic anisotropic effect. The most important parameter among all parts under repair is switching in the process, which includes a statistical switching model divided into two regimens, the first regimen called thermal activation and the second regime called rapid switching. The first regimen is for injected currents (I_{MTJ}) below the critical current (Ic) and follows the Nèel-Brown model. This current is intended as the one needed to read the data. Conversely, for the write operation, currents that are higher than the critical current are required, an extended analytical formulation is used. [21]

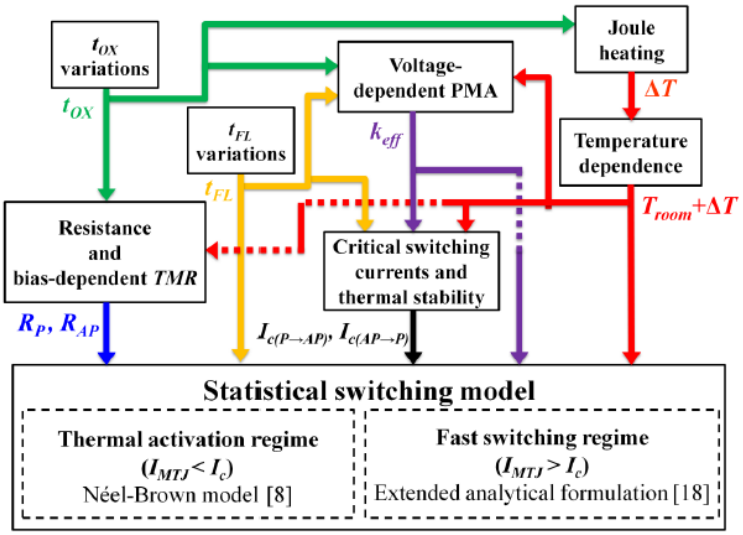


Fig 18: Analytical description of the compact model block [27]

Statistical distribution of each of the AP-P transitions or vice versa is provided by the compact model. In addition, this model can distinguish a deterministic and stochastic behavior when the initial magnetization performs a first rotation of the magnetic angle, in this way a deterministic simulation for MTJ switching is possible [28]. Until now, the analytical formulation for rapid switching is regulated for currents slightly higher than the critical current; however, there is no model that can describe the region between the thermal activation regime and the rapid switching regime. [21]

3.2 Validation of the Model

SB and DB structures make use of compact models, because it is the best option to get the appropriate response to the behavior of the STT-MRAM switching activity. To validate the model, a comparison is made between the two micro magnetic simulations and analytical predictions. Once validated, the simulation and comfort of each MTJ-SB and MTJ-DB device are indicated. For validation purposes, it is performed for certain physical parameter values of MTJ. Next, the MTJ parameters used for MTJ are indicated. [28]

Two validations are presented to validate the model. The first validation refers to the resistance model and TMR by comparing it with the experimental data presented in [43]. The second validation is based on the comparison with the STT switching analytical model that works with a micro magnetic solver. Remember that there are two MTJ-SB and MTJ-DB models, that is, it means that each device must be validated. As a demonstration example, we will validate the MTJ-DB template. **Figure 19** shows the first validation making the following parameters valid: tox, t = 0,80 nm, tox, b = 0,75 nm, a polarization voltage for TMR VH = 0,5 V, TMRT (0) = 140%, TMRB (0) = 70%, a higher RAT than the resistance area product = 100 Ω, μm2 and a lower RAB of the resistance area= 50 Ω · μm2. [28] [21]

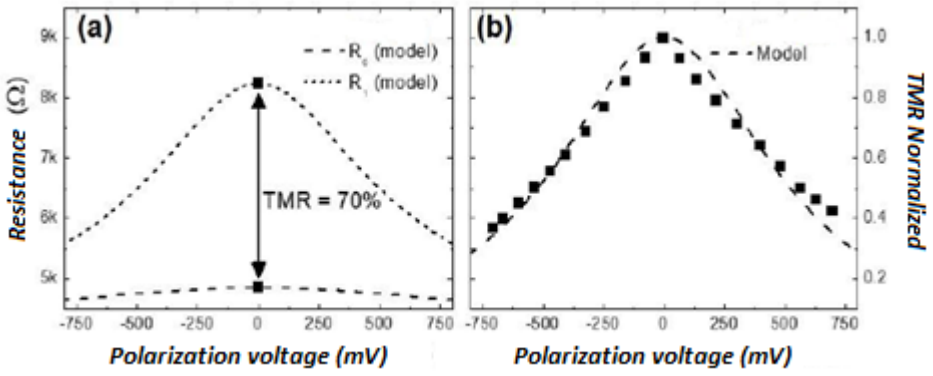


Fig 19: Validation of the resistance model and TMR with respect to experimental data [21]

On the other hand, to validate the STT-DB switching model, **Figure 20** explains it with three different Radius MTJs (r = 12 nm, r = 10 nm e r = 7 nm). In addition, the following parameters are considered: A saturation magnetization Ms = 10^6 A/m, α = 0,03, Ku = 1,1×106 J/m^3,thickness of the free layer t_{FL}= 1.2 nm, η = 0,67. We can see that the moments follow the results of the micromagnetic solver. [21] [28]

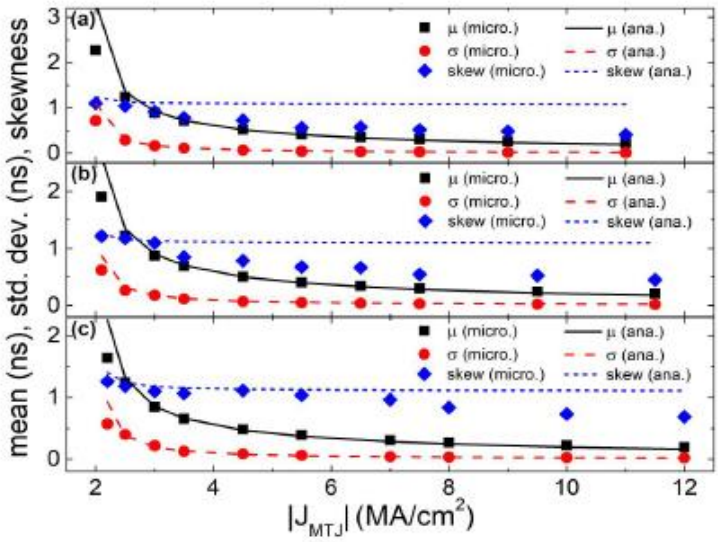


Fig 20: Validation of data for a DB-MTJ with average value (μ), standard deviation (σ) and asymmetry (inclination) of the switching time as a function of the current density MTJ [28]. Three dimensions considered: (a) r = 12 nm, (b) r = 10 nm and (c) r = 7 nm

3.3 Simulation structure and CMOS / MTJ parameters

This section established the simulation methodology and how to validate the MTJ-SB and MTJ-DB. Now the hybrid CMOS/MTJ parameters used for analysis and the simulation framework are shown, which will be used in the rest of this Document.

3.3.1 CMOS / MTJ Parameters

As you can see in the following table 3, the main parameters that characterize an MTJ-SB and MJT-DB are tabulated. These parameters must match the node we want to analyze, in this case, we will look at the CMOS technology node at 28 nm, the coincidence parameter is the MTJ radius, we choose r plus 14 nm. In addition, current and thermal stability have

been adjusted according to experimental data [29]; In addition, a percentage variability is included for different parameters; these variations are included in the compact MTJ model.

Parameters SB / DB	Description	Value 28 nm	Unit
M_S*	Saturation magnetization	1000×10^3	A/m
α*	Magnetic damping	0.05	—
R*	MTJ radius	14	nm
k_u*	Uniaxial anisotropy constant	8.8×10^5	J/m^3
$A^*(\sigma/\mu)$	MTJ surface (variability)	6.16×10^{-16} (5%)	m^2
Δ*	Thermal stability	59.14	—
$t_{OX}(\sigma/\mu)$	SMTJ oxide thickness (variability)	0.85 (1%)	nm
$t_{OX,T}(\sigma/\mu)$	DMTJ Top oxide thickness (variability)	0.85 (1%)	nm
$t_{OX,B}(\sigma/\mu)$	DMTJ Bottom oxide thickness (variability)	0.65 (1%)	nm
$t_{FL}^*(\sigma/\mu)$	SMTJ & DMTJ free layer thickness (variability)	1.2 (1%)	nm
RA	SMTJ resistance-area product	5.0	$\Omega \cdot \mu m^2$
RA_t	DMTJ resistance-area product of top barrier	5.0	$\Omega \cdot \mu m^2$
RA_b	DMTJ resistance-area product of bottom barrier	1.0	$\Omega \cdot \mu m^2$
R_p	SMTJ resistance in P state	8.12	kΩ
R_{ap}	SMTJ resistance in AP state at 0V	20.3	kΩ
R_0	DMTJ resistance in P state at 0V	8.97	kΩ
R_1	DMTJ resistance in AP state at 0V	20.6	kΩ
$TMR^*(\sigma/\mu)$	SMTJ TMR ratio (variability)	150% (3%)	—
	DMTJ TMR ratio (variability)	130% (3%)	—
$\lvert I_{c0(P\rightarrow AP)}\rvert$	SMTJ P→AP critical current	40.2	µA
$\lvert I_{c0(AP\rightarrow P)}\rvert$	SMTJ AP→P critical current	15.3	µA
$\lvert I_{c0(P\leftrightarrow AP)}\rvert$	DMTJ P↔AP critical current	11.8	µA

Table 2: These data were entered based on the experimental data reported on [29]

Now, we will look at the CMOS technology node at 24nm and 20nm. In addition, current and thermal stability have been adjusted according to experimental data [29].

Parameters SB / DB	Description	Value	Unit
		24 nm	
M_S*	Saturation magnetization	1000×10^3	A/m
α*	Magnetic damping	0.05	—
R*	MTJ radius	12	nm
k_u*	Uniaxial anisotropy constant	9.3×10^5	J/m^3
$A^*(\sigma/\mu)$	MTJ surface (variability)	4.52×10^{-16} (5%)	m^2
Δ*	Thermal stability	51.21	—
$t_{OX}(\sigma/\mu)$	SMTJ oxide thickness (variability)	0.85 (1%)	nm
$t_{OX,T}(\sigma/\mu)$	DMTJ Top oxide thickness (variability)	0.85 (1%)	nm
$t_{OX,B}(\sigma/\mu)$	DMTJ Bottom oxide thickness (variability)	0.65 (1%)	nm
$t_{FL}^*(\sigma/\mu)$	SMTJ & DMTJ free layer thickness (variability)	1.2 (1%)	nm
RA	SMTJ resistance-area product	5.0	$\Omega \cdot \mu m^2$
RA_t	DMTJ resistance-area product of top barrier	5.0	$\Omega \cdot \mu m^2$
RA_b	DMTJ resistance-area product of bottom barrier	1.0	$\Omega \cdot \mu m^2$
R_p	SMTJ resistance in P state	11.1	kΩ
R_{ap}	SMTJ resistance in AP state at 0V	27.6	kΩ
R_0	DMTJ resistance in P state at 0V	12,2	kΩ
R_1	DMTJ resistance in AP state at 0V	28,1	kΩ
$TMR^*(\sigma/\mu)$	SMTJ TMR ratio (variability)	150% (3%)	—
	DMTJ TMR ratio (variability)	130% (3%)	—
$\lvert I_{c0(P\rightarrow AP)}\rvert$	SMTJ P→AP critical current	31.34	µA
$\lvert I_{c0(AP\rightarrow P)}\rvert$	SMTJ AP→P critical current	11.92	µA
$\lvert I_{c0(P\leftrightarrow AP)}\rvert$	DMTJ P↔AP critical current	8.64	µA

Table 3: These data were enter based on experimental data reported on [30]

Parameters SB / DB	Description	Value	Unit
		20 nm	
M_S*	Saturation magnetization	1000×10^3	A/m
α*	Magnetic damping	0.05	—
R*	MTJ radius	10	nm
k_u*	Uniaxial anisotropy constant	1.01×10^5	J/m^3
$A^*(\sigma/\mu)$	MTJ surface (variability)	3.14×10^{-16} (5%)	m^2
Δ*	Thermal stability	44.41	—
$t_{OX}(\sigma/\mu)$	SMTJ oxide thickness (variability)	0.85 (1%)	nm
$t_{OX,T}(\sigma/\mu)$	DMTJ Top oxide thickness (variability)	0.85 (1%)	nm
$t_{OX,B}(\sigma/\mu)$	DMTJ Bottom oxide thickness (variability)	0.65 (1%)	nm
$t_{FL}^*(\sigma/\mu)$	SMTJ & DMTJ free layer thickness (variability)	1.2 (1%)	nm
RA	SMTJ resistance-area product	5.0	$\Omega \cdot \mu m^2$
RA_t	DMTJ resistance-area product of top barrier	5.0	$\Omega \cdot \mu m^2$
RA_b	DMTJ resistance-area product of bottom barrier	1.0	$\Omega \cdot \mu m^2$
R_p	SMTJ resistance in P state	15.9	kΩ
R_{ap}	SMTJ resistance in AP state at 0V	39.8	kΩ
R_0	DMTJ resistance in P state at 0V	17.6	kΩ
R_1	DMTJ resistance in AP state at 0V	40.5	kΩ
$TMR^*(\sigma/\mu)$	SMTJ TMR ratio (variability)	150% (3%)	—
	DMTJ TMR ratio (variability)	130% (3%)	—
$\lvert I_{c0(P\rightarrow AP)}\rvert$	SMTJ P→AP critical current	22.35	µA
$\lvert I_{c0(AP\rightarrow P)}\rvert$	SMTJ AP→P critical current	8.5	µA
$\lvert I_{c0(P\leftrightarrow AP)}\rvert$	DMTJ P↔AP critical current	6.16	µA

Table 4: These data were enter based on experimental data reported on [30]

The node to use CMOS technology is the one available in the software that is a FinFET. Table 5 shows the important FinFET parameters used regardless of the device we use, which can be NMOS or PMOS. In the case of values n_{fin} o m, are used for the entire analysis. On the other hand, n_f is the only parameter that varies in simulations. That is, the transistor area and consequently the cell area will change. [21]

Parameters	Description	Value	Unit
L	Gate length	28, 24, 20	nm
nfin**	Number of fins for Finger	2	--
nf*	Finger number	1	--
M	Multiplier - number of parallel MOS devices	1	--

**It corresponds to the width of each finger and is expressed in whole units.*
*** Corresponds to the finger number of the ports presented in the circuit*

Table 5: FinFET parameters used for the access transistor in the memory cells [21].

As we get the parameters of our hybrid model, as a next step is to establish the STT-MRAM design parameters. In general, when measuring the area of the bit cell, the size of MTJ does not matter, in fact the area is limited by the size of the transistor or the pitch of the metal [41]. Design parameters are used to calculate the minimum size of the technology function (F).

3.3.2 Simulation Structure.

All simulations are based on the compact analytical model described in the previous sections. With this model, the designed memory is transformed into a deterministic and statistical model, in which process variations are represented using Monte Carlo simulations. Process variations for MTJ are included in the analytical model written in Verilog-A, while in the case of FinFET, the foundry provides statistical models.

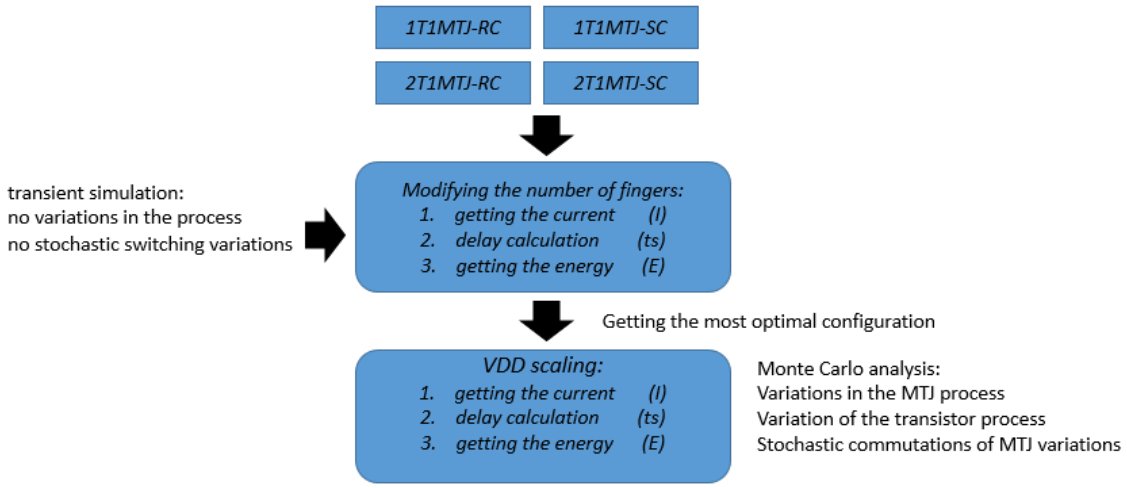


Fig 21: General workflow for the analysis of STT-MRAM writing independently if it is MJT-SB or MJT-DB.

Write and read operations can be performed on memory. For the write operation, it is described by a data stream shown in **Figure 23**. Let's start with the four-cell bit configurations studied above as part of a transient analysis, we choose the optimal setting to look for the working point where the energy is optimal. Then, by scaling Vdd you can find the optimal configuration where the minimum power point is located. In fact, the entire process is performed using an MTJ-SB or MTJ-DB.

All simulations that were performed implicitly contain data such as energy, delay, area of a device, in fact, we will use 3 different nodes, these are 20nm, 24nm, 28nm, which are available in the Cadence software, the operation is analyzed only in different nodes with different structures, resulting in energy and delay results.

4 STT-MRAM Analysis

In this last chapter, we will run the simulations of the 4 devices described in order to provide the main results of the thesis. As mentioned in the previous chapter, the technology of the nodes is taken from the foundry, in this case we will take 3 different nodes, these are 20nm, 24nm, and 28 nm, and a different true -A code for each node for

the MTJ device. For these 4 nodes, four 1T1MTJ-RC, 1T1MTJ-SC, 2T1MTJ-RC, and 2T1MTJ-SC configurations are scanned. It is important to remember that the critical current of SB and DB is considered very high, considering the STT-MRAM memory in a very pessimistic case. The result of the simulations of the 4 devices is made for the writing mode, and the optimal curve between energy and delay will be verified. In this way, it is possible to understand which structure is the best and which has the lowest energy consumption.

4.1 STT-MRAM Writing Analysis .

The write analysis is divided into two parts. The first analysis is carried out by varying the integration capacity, expressed in area units in F^2. Starting with a nominal simulation and Monte Carlo, the results are layered. It should be noted that SMTJ has low performance when using small transistors. On the contrary, we see that 2T configurations, and in general the DMTJ, perform better than we expected.

4.1.1 Initial Considerations and Preliminary Analysis

The memory under test is a topology of a memory array of 128 to 128. In our case for simulation, a single-bit cell is built with buffer lines and peripheral capabilities for each line (WL, BL and SL). In the MJT block shown in the example in **figure 25**, terminal T1 represents the PL, terminal 2 represents the FL, and the terminal called "State" allows us to know the status of the MTJ. In addition, capacity values depend on the number of fingers (nf) access transistor. Therefore, by increasing the bit cell area, capacity will also increase, and the suburbs will see greater capacity. [21]

In circuit design, the first step is capacity design. Having a 128x128-bit memory block, we know that peripherals like WL have a load of 128 transistors, in the same way that SL has a load of 128 terminals. It is true that two transistors are used in the 2T configuration, but in the end, each transistor has the same load, 128 WL and 128 SL. All capacity values are

tabulated, in the case of 1T and 2T structures obtained from test transistors (that is, 20nm, 24nm, and 28nm nodes), pre-terminate FinFET values should be considered for these capacity values. They are tabulated in Table 5 of the previous section. In the case of 1T1MTJ-RC and 1T1MTJ-SC the capabilities are the same and are obtained as $C_w = (C_{gs} + C_{gd}) \times 128$, $C_{sL} = (C_{sd} + C_{sg}) \times 128$ e $C_{BL} = C_{SL}/10$, Not knowing the capacitive effectC_{BL} MTJ structure, with simulation criteria a decade has been reduced based on the value C_{sL}. On the other hand, for design in 2T configurations are of the shape$C_{WLn} = (C_{gsn} + C_{gdn}) \times 128$, $C_{WLp} = (C_{gsp} + C_{gdp}) \times 128$, $C_{sL} = 2(C_{sd} + C_{sg}) \times 128$ e $C_{BL} = C_{SL}/10$. In the bit cell design does not include buffs, since we will use the deflectors available in the lab or in the case in the CADENCE software, that is, we do not know the capacity values. In general, according to other projects, you may notice that the SL buffer is less strong than drive (half) than the WL buffer; That is, it is because CSL capacity is almost half of the CWL. Finally, the BL buffer is generally 10 times smaller than the SL buffer. [21]

1T NMOS 28 nm				
Parameters	*description*	*NMOS*	*PMOS*	*Unit*
		Value		
C_{WL}	word line capability	35.14	23.72	fF
C_{SL}	source line capacity	17.59	11.86	fF
C_{BL}	bit line capacity	1.759	1.186	fF

Table 6: Capacitance values extracted from the 28nm transistor for 1T NMOS

2T NMOS/PMOS 28 nm				
Parameters	*description*	*NMOS*	*PMOS*	*Unit*
		Value		
C_{WLn}	word line capability NMOS	*35.14*		*fF*
C_{WLp}	word line capability PMOS	23.72		*fF*
C_{SL}	source line capacity	35.17		*fF*
C_{BL}	bit line capacity	3.517		*fF*

Table 7: Capacitance values extracted from the 28 nm transistor for 2T NMOS and PMOS

1T NMOS 24 nm				
Parameters	description	NMOS	PMOS	Unit
		Value		
C_{WL}	word line capability	33.71	23.72	fF
C_{SL}	source line capacity	16.88	11.86	fF
C_{BL}	bit line capacity	1.688	1.186	fF

Table 8: Capacitance values extracted from the 24 nm transistor for 1T NMOS

2T NMOS/PMOS 24nm				
Parameters	description	NMOS	PMOS	Unit
		Value		
C_{WLn}	word line capability NMOS	33.71		fF
C_{WLp}	word line capability PMOS	23.72		fF
C_{SL}	source line capacity	33.76		fF
C_{BL}	bit line capacity	3.37		fF

Table 9: Capacitance values extracted from the 24 nm transistor for 2T NMOS and PMOS

1T NMOS 20 nm				
Parameters	description	NMOS	PMOS	Unit
		Value		
C_{WL}	word line capability	32.24	23.75	fF
C_{SL}	source line capacity	16.16	11.88	fF
C_{BL}	bit line capacity	1.61	1.18	fF

Table 10: Capacitance values extracted from the 20 nm transistor for 1T NMOS

2T NMOS/PMOS 20nm				
Parameters	description	NMOS	PMOS	Unit
		Value		
C_{WLn}	word line capability NMOS	32.23		fF
C_{WLp}	word line capability PMOS	23.75		fF
C_{SL}	source line capacity	32.31		fF
C_{BL}	bit line capacity	3.37		fF

Table 11: Capacitance values extracted from the 20 nm transistor for 2T NMOS and PMOS

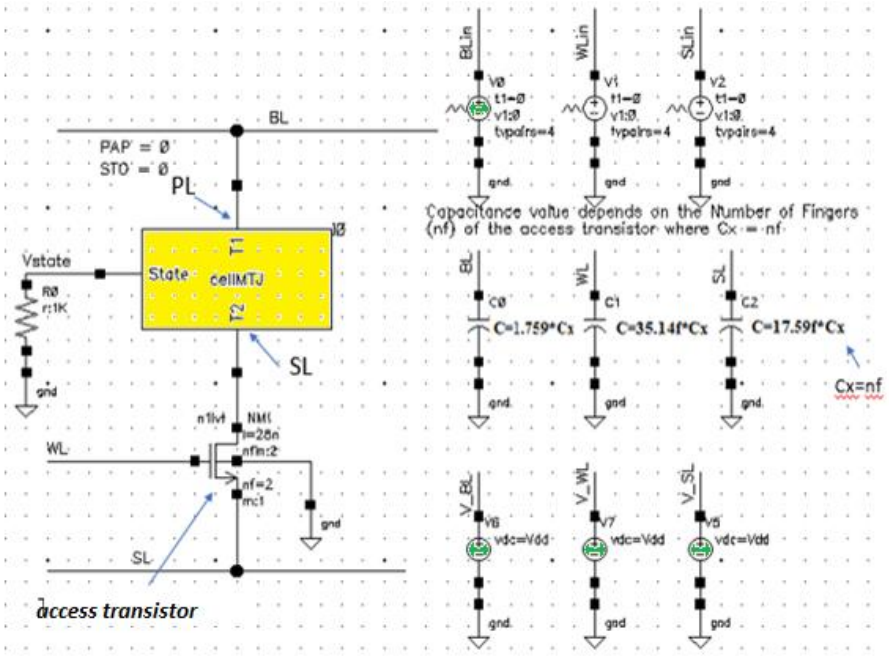


Fig 22: Simulation scheme of a STT-MRAM 128 × 128 structure that uses the 1T1MTJ-RC, 1T1MTJ-SC bit cell for 28nm, 24nm and 20nm nodes.

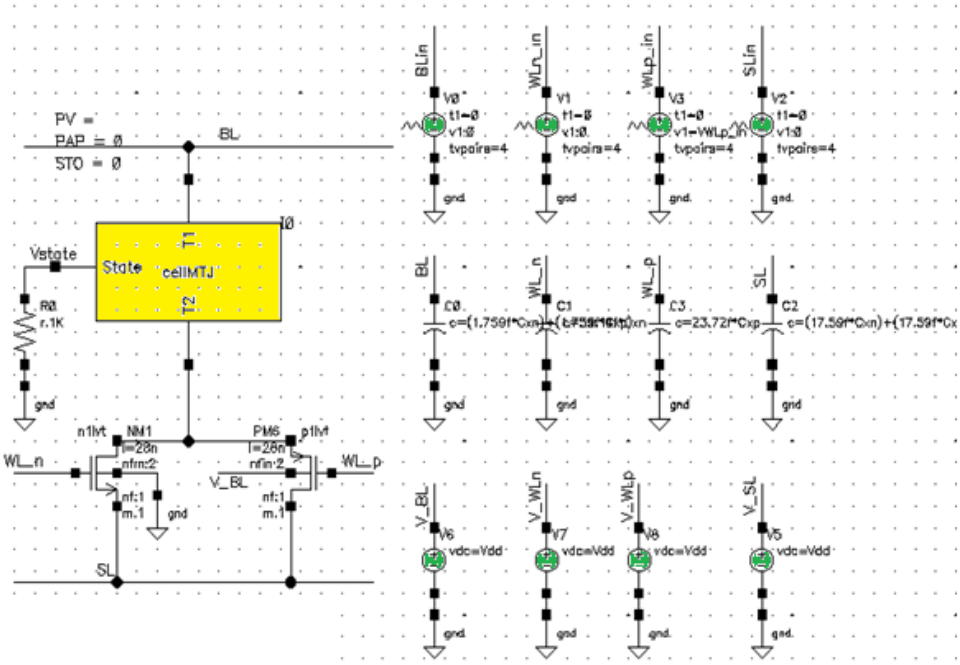


Fig 23: Simulation scheme of a STT-MRAM 128 × 128 structure that uses the 2T1MTJ-RC bit cell, 2T1MTJ-SC for 28nm, 24nm and 20nm nodes.

Now let's consider the energy calculation. For this, the peripheral circuit, that is, the buffers, is considered. For example, when we finish the write operation, we access a row within the array 128 x 128 STT-MRAM, if we consider 3 Lines like WL, SL, and BL. The energy on the WL line is the sum of all the MTJs presented on the line (considering that all 128 are active). In contrast, for SL and BL lines that contain the transient signal, which travels through the access line to the bit cell considering all the energy of the SL and BL buffers; in other words, all the energy of the lines is needed to write to the bit cell.

Prior to the writing study, a pre-analysis is carried out to understand the general behavior of the MTJ. In Table 3 we can see that the SB and DB structures have the same values as FL and PL (referring to PL in the case of DB) and that the results for SB structure will not be presented. Therefore, the parameter that varies in this pre-analysis is the PL ($t_{ox,b}$) DB, which is shown in **figure 26**. This curve gives us information about the TMR, and you can see that the TMR changes by changing the ($t_{ox,b}$) this means that one barrier is more resistive than the other and changing the$t_{ox,b}$, the total strength of the MTJ is changing, that is, while ($t_{ox,b}$) TMR is reduced, and vice versa. This allows for an increase in the current. Therefore, you can say that the DB can be adjusted by varying the smaller oxide layer. In fact, we cannot overcome the conditions, because a collapse of the MTJ can occur; however, this goes beyond the scope of our analysis.

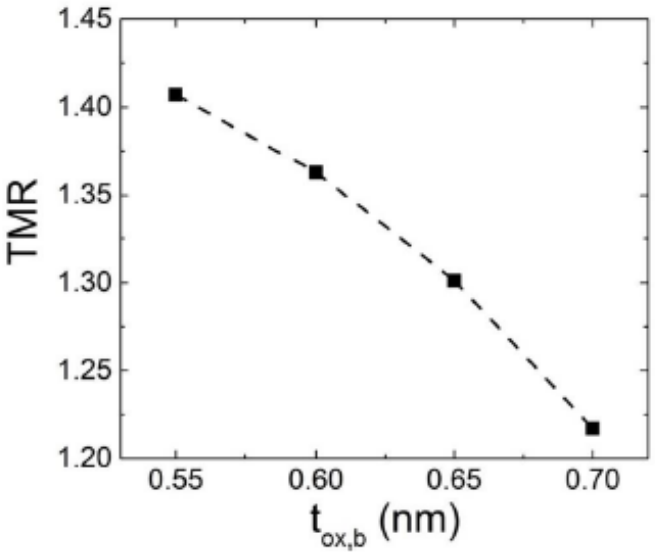


Fig 24: TMR vs. tox,b DB MTJ. [21]

4.1.2 Writing Deterministic Analysis

The analysis follows the flowchart presented in **figure 24**. The purpose is to vary the integration density with the number of fingers, which translates into cell area units (F^2). The simulation starts deterministically where the current can be obtained while the integration density is variable. In the **figure 25**, we can see the results of the different structures or models described in the previous chapters. When analyzing a chart, you can get valid information as incorrect if we do not have analysis criteria. In this simulation the best criterion we can see, refers to the SB structure, it can be concluded that it does not have a good performance for small transistors, that is, the performance decreases for a small number of fingers. In addition, it can be concluded that 2T1MTJ configurations show better performance than 1T1MTJ configurations.

The problem with the high writing currents of the STT-MRAM must be checked. In fact, one possible way to mitigate this issue is to use a DMTJ, which makes a faster change. On the other hand, the DMTJ is compared with the SMTJ in the performance and scalability

analysis for a 128 x 128 ST-MRAM matrix with the corresponding peripheral capabilities and buffer lines.

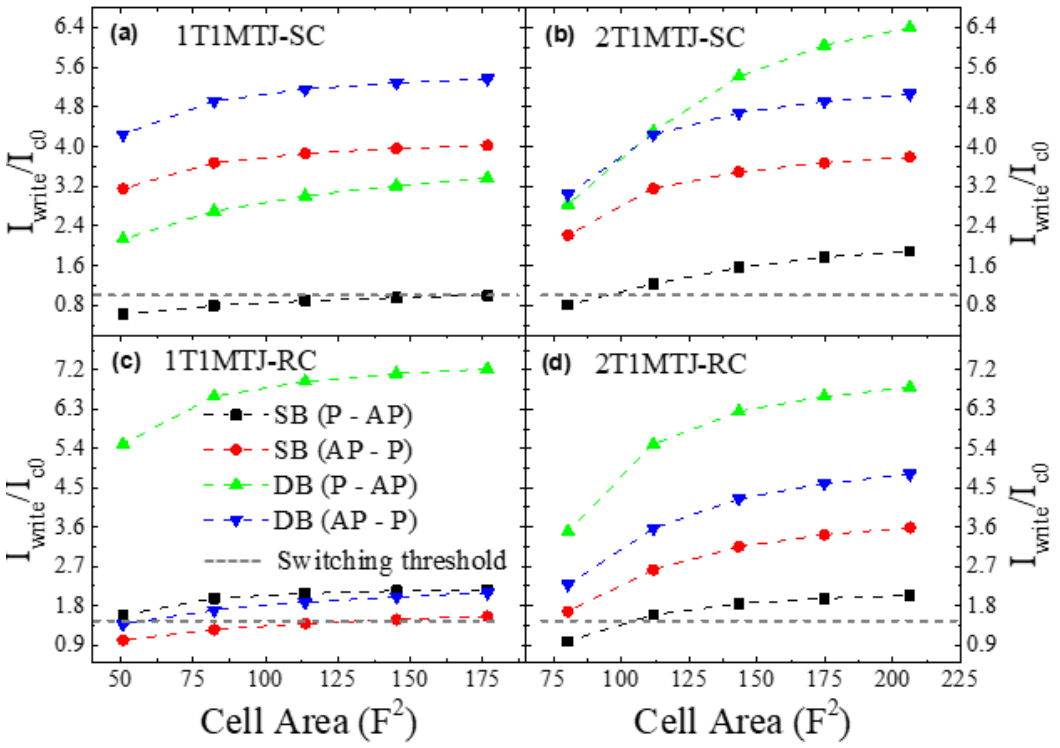


Fig 25: Current Report I_{write}/I_{c0} vs. Cellular Area (F^2) of the results of the nominal analysis. (a) 1T1MTJ-SC, (b) 2T1MTJ-SC, (c) 1T1MTJ-RC and (d) 2T1MTJ-RC for 28nm

Depending on the placement of PL or PLT by bit line (BL), bit cells can be categorized in reverse configuration (RC) or originally (SC). On the other hand, the FinFETs used have a channel length equal to the diameter of MTJ for 28 nm. $\mathrm{FinFET} = \phi\mathrm{MTJ} = 28\ \mathrm{nm}$. In this section, we analyze the performance of the STT-MRAM array for the possible bit cell configurations mentioned in the last section. The analysis is performed for all technology nodes, the first node analyzed is 28nm, considering the write performance with a nominal VDD = 0.8 V. With these results, we know that the current is proportional to the area. In addition, the device with double DB structure starts with two or three times the critical current, confirming one of the problems mentioned in the previous chapters, which is the

presence of high writing currents. Finally, for SB and DB structures, many parallel simulations were made to this deterministic or transient analysis. The current increases when the oxide decreases and the TMR shown in the preliminary analysis improves performance and, as a result, the bit cell improves. Now, we have the current bitcell for the different topologies and considering a write error rate (WER) of 1×10-7.

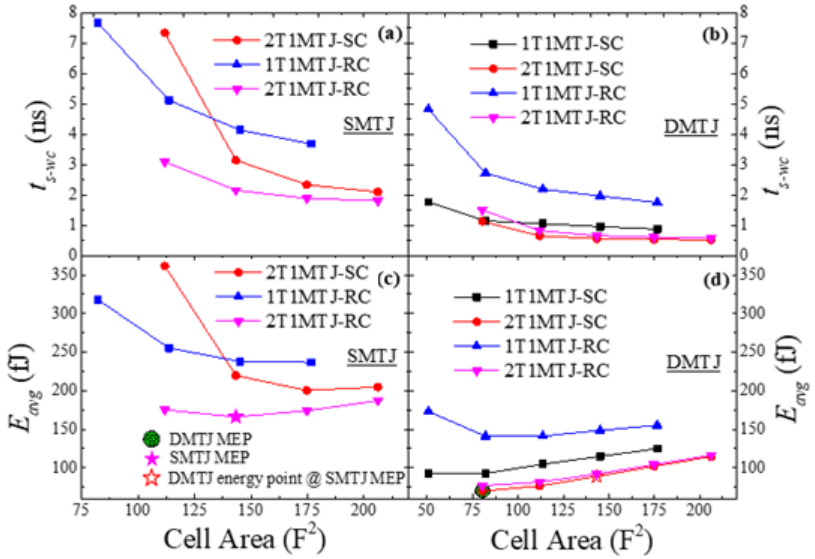


Fig 26: Result of the nominal Energy and Delay Analysis. (a)-(b) switching delay (t_{s-wc}) vs. Cellular area (F^2) for SMTJ end DMTJ respectively, (c)-(d) average energy (E_{avg}) vs. Cellular area (F^2) (MEP) for SMTJ end DMTJ for 28nm.

Now, for each current value, corresponds a certain cell area, then the switching delay or the cell switching time (ts) is calculated for different configurations and with the help of the current previously calculated with respect critical current SB or DB, we'll know which configurations will write and which won't. For the calculation of t_s an external calculation is performed, where with a MATLAB script the ts is calculated for AP - P and P - AP transitions. The script runs a CDF switching time, which indicates the error of switching probability. Keep in mind that t_s is calculated for AP - P and P - AP transitions, so at this point you need to consider the maximum delay between these two. This maximum delay is considered the worst case t_{s-wc}.

Finally, the average energy is calculated for the largest delay (the worst case) in a $\mathrm{WER} = 1 \times 10\text{-}7$ between the two transitions (AP - P e P - AP) and the results are obtained from the simulation, are in the **figure 26 c-d**. Calculated average energy considers STT-RAM peripherals. With these results, we can achieve the most optimal configuration in terms of energy. According to the charts, the best configurations are 2T1MTJ-RC and 2T1MTJ-SC for SB and DB. At this stage, the dimension defines how it fixes the ability to integrate memory. If we analyze the optimal energy point for single barrier (SB) and double barrier (DB) cases by choosing a certain area we will notice which structure is best, in our case in the simple barrier structure (SB) the optimal energy point is in a cell area $140F^2$. If we do the same analysis for a double barrier structure (DB) the optimal energy point is in a cell in area $81F^2$. In the case of SB does not fit for maximum integration capacity. During performance analysis, it was considered the technology node at 28 nm. In the next section, we present the results for 20 nm, 24 nm, and 28 nm technology, and the voltage scalability of the STT-MRAM considering Monte Carlo simulations.

4.1.3 Write Scalability

When we analyzed performance for 2T1MTJ-RC structures or configurations for SB and 2T1MTJ-SC for DB for the write operation, we achieved optimal energy and performance results, now we do a voltage and technology scaling to analyze write scalability. In **figure 27**, we can see the results of writing the scalability for MTJ-SB and DMTJ-DB as we scale from 28 nm to 20 nm, the calculated MEP is in a smaller region. If we look at the results **figure (27 a)**, the capacity of the 28 nm area of the SMTJ changes from143 F^2 a 112 F^2. In fact, the structure that offers significant energy savings is the DMTJ compared to the SMTJ. Table II provides a summary of write performance when resizing the voltage and technology node. If we only analyze the 28 nm node with reference to the region, we see

that the DMTJ has a saving in the write operation of about 70% compared to the SMTJ, in addition, we can see greater energy savings when we reduce from 28 nm to 20 nm in a of the two structures, that is, that between the technologies there is an energy saving of about 54%, we will always notice better performance.

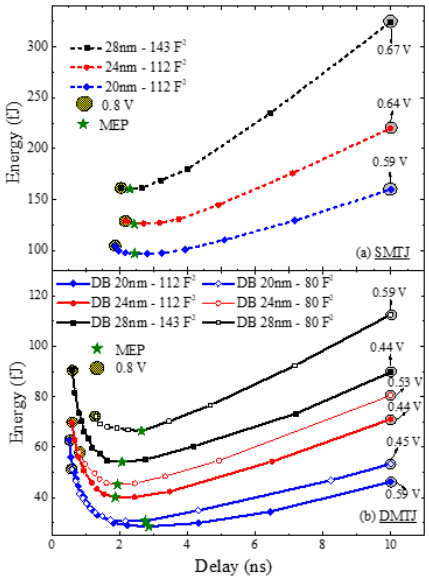


Fig 27: (a) Delay between energy and SMTJ and (b) Delay between energy and DMJ for technological nodes 20 nm, 24 nm and 28 nm and capacity of the area. The minimum energy point (MEP) is a green star and all points correspond to 1000 MC samples.

		Write Operation		
Bitcell	**Area F^2**	**Vmep (mV)**	**Write delay (ns)**	**Write energy (fJ)**
28nm				
SMTJ	143	775	2.31	160.5
DMTJ	143	525	2.07	54.1
DMTJ	80	675	2.65	66.4
24 nm				
SMTJ	112	775	2.43	126.1
DMTJ	112	525	1.87	54.1
DMTJ	80	625	1.94	66.4
20nm				
SMTJ	112	725	2.45	96.9
DMTJ	112	425	2.86	28.5
DMTJ	80	500	2.76	30.7

Table 12: Write and read performance for STT-MRAM DE 128 X 128 arrays for 28nm, 24nm, y 20nm nodes.

For 3 different nodes we performed the write operation, where we tabulated variables such as area, minimum energy point, delay and energy consumption, when we reduce the node, we have a substantial reduction in energy when we write the memory There is a energy reduction when we change technology, SMTJ to DMTJ. If we pay attention to the nodes of 24nm and 20nm, the areas are the same, but the energy reduction in the SMTJ write mode varies from 126. 1 fJ to 96. 9 fJ, even if we compare DMTJ with an area of 80 F^2 between the 28nm and 20nm nodes reduce the energy from 66.4 fJ to 30.7fJ, in fact the delay in writing does not have greater variability, it is still a quick access memory.

Conclusions

In the design of state-of-the-art devices, they have a high energy consumption due to scaling. These devices have effects of loss, reliability and variability. The high density of a chip translates to the area that uses a memory, resulting in a high dissipation of power. This energy consumption is a concern, in fact, there are metrics in new projects that focus on reducing power. In addition, research into new memory technologies such as MRAM presents potential projects to contain technological problems. By looking for new technologies, you get positive results by creating potential devices, such as STT-MRAM, which combines CMOS technology, achieving compatibility in semiconductor production.

STT-MRAM memory was characterized in terms of write performance for three technology nodes (28nm, 24nm and 20nm) in which information was obtained considering parameters such as speed (delay), energy, and occupation of the area. According to the characterization in writing and considering only the structure of the simple SB barrier, when you change the integration capability, you conclude that it cannot be written in small transistors or for certain configurations, especially for 1T1MTJ cases. In the dual barrier structure DB has better writing performance and is due to low critical currents. Therefore, by resizing the area, the MEP resizes the critical current from 3 to 4 times. It concludes that SB and DB structures maintain the best energy terms. The 2TRC adapts better in SB due to the different critical currents in the two transitions, while the dominant factor of why 2TSC is better in the DB are the different resistances. Finally, when the t_{OX} Decreases, the current Increases and the TMR it's getting better, as well as the structure MTJ.

When a node is reduced, regardless of structure, a substantial reduction in energy consumption is verified. Now, if we look at different memory structures STT-MRAM such

as the DB dual barrier structure, not if you see a reduction in delay, but energy consumption decreases considerably. If we combine the reduction of a node with area scaling, the savings are large in write operations.

References

[1] S. Verma end B.K. Kaushik, "Next Generation Spin Torque Memories", *SpringerBriefs in Applied Sciences and Technology*, 2014.

[2] Yole Développement, "Storage-class memory will be the clear go-to market for emerging non-volatile memory in 2021", July 2016. [En línea]. Available: http://www.yole.fr/Emerging_NVM_Market.aspx#.XUXr--gzbIU.

[3] ANANDTECH, "Everspin Begins Production of 1Gb STT-MRAM", June 2014. [En línea]. Available: https://www.anandtech.com.

[4] D. Apalkov, Bernard Dieny end J. Slaughter, "Magnetoresistive Random Access Memory", vol. 104, *Proceedings of the IEEE*, 2016.

[5] M. Weisheit, S. Fahler end A. Marty, "Electrical Field-Induced Modification of Magnetism in Thin-Film Ferromagnets", vol. 315, *SCIENCE*, February 2007.

[6] M. Cubukcu, O. Boulle end M. Drouard, "Spin-orbit torque magnetization switching of a three-terminal perpendicular magnetic tunnel junction", *Department of Materials, ETH Zurich, Schafmattstrasse 30, CH-8093 Zürich, Switzerland*, October 2013.

[7] R. Scheuerlein , W. Gallagher end S. Parkin, "A 10ns read and write non-volatile memory array using a magnetic tunnel junction and FET switch in each cell", *IEEE International Solid-State Circuits Conference*., February 2000.

[8] M. Durlam , D. Addie end B. Butcher, "A 0.18 μm 4Mb toggling MRAM", *IEDM Technical Digest* , May 2003.

[9] T. W. Andre , J. J. Nahas end C. K. Subramanian, "A 4-Mb 0.18-/spl mu/m 1T1MTJ toggle MRAM with balanced three input sensing scheme and locally mirrored unidirectional write drivers", vol. 40, *IEEE Journal of Solid-State Circuits*, January 2005.

[10] M. Hosomi , H. Yamagishi end T. Yamamoto, "A novel nonvolatile memory with spin torque transfer magnetization switching: spin-ram", *IEEE InternationalElectron Devices Meeting, December* 2005.

[11] T. Kawahara , R. Takemura end K. Miura, "2Mb Spin-Transfer Torque RAM (SPRAM) with Bit-by-Bit Bidirectional Current Write and Parallelizing-Direction Current Read", *IEEE International Solid-State Circuits Conference*, February 2007.

[12] R. Beach, T. Min end C. Horng, "A statistical study of magnetic tunnel junctions for high-density spin torque transfer-MRAM (STT-MRAM)", *IEEE International Electron Devices Meeting*, December 2008.

[13] T. Kishi, H. Yoda end T. Kai, "Lower-current and fast switching of a perpendicular TMR for high speed and high density spin-transfer-torque MRAM", *IEEE International Electron Devices Meeting*, December 2008.

[14] D. C. Worledge, G. Hu end P. L. Trouilloud, "Switching distributions and write reliability of perpendicular spin torque MRAM", *IEEE International Electron Devices Meeting,* December 2010.

[15] R. Topaloglu, "More than Moore Technologies for Next", *Springer, Electronics & Electrical Engineering*, February 2015.

[16] L. Loong *et al*, "Strain-enhanced tunneling magnetoresistance in MgO magnetic tunnel junctions", *Institute of Materials Research and Engineering – Scientific Reports*, March 2014.

[17] X. Fong et al, "Spin-Transfer Torque Memories: Devices, Circuits, and Systems", vol. 104, *Proceedings of the IEEE*, July 2016.

[18] T. Liu, "Thermally robust Mo/CoFeB/MgO trilayers with strong perpendicular anisotropy", *Nature Scientific Reports*, November 2014.

[19] L. Thomas, "Basic Principles, Challenges and Opportunities of STT-MRAM for Embedded Memory Applications", *TDK – Headway Technologies*, Junio 2017.

[20] I.K. Yanson, "Current-field diagram of magnetic states of a surface spin valve in a point contact with a single ferromagnetic film", vol. 39, *Kamerlingh Onnes Laboratory, Leiden University*, December 2012.

[21] E. Garzón , "Perpendicular STT-MTJs with Double Reference Layers and its Application to Downscaled Memory Cells", *Unical*, September 2018.

[22] R. De Rose, "Variability-Aware Analysis of Hybrid MTJ/CMOS Circuits by a Micromagnetic-Based Simulation Framework", *IEEE Transactions on Nanotechnology,* March 2017.

[23] A.V. Khvalkovskiy *et al*, "Basics principles of the STT-MRAM cell operation in memory arrays", *Journal of Physics D Applied Physics*, March 2015.

[24] X Fong end Yusung Kim, "Spin-Transfer Torque Memories: Devices, Circuits,and Systems", *in Proceedings of the IEEE*, July 2016.

[25] S. Bhatti *et al*, "Spintronics based random access memory", *ELSEVIER,* April 2017.

[26] R. De Rose *et al*., "Impact of voltage scaling on STT-MRAMs through a variability-aware simulation framework", *International Conference on Synthesis*, July 2017.

[27] R. De Rose, "A Compact Model with Spin-Polarization Asymmetry for Nanoscaled Perpendicular MTJs", vol. 64, *IEEE Transactions on Electron Devices*, 2017.

[28] R. De Rose, "A Compact Model for Perpendicular Spin-Transfer-Torque Magnetic Tunnel Junctions with Double Reference Layers", vol. 16, *Proceedings of the IEEE*, March 2017.

[29] Y. Zhang, "Compact Model of Subvolume MTJ and Its Design Application at Nanoscale Technology Nodes", *IEEE Transactions on Electron Devices*, June 2015.

[30] S. Ikeda *et al*., "Perpendicular-anisotropy CoFeB-MgO based magnetic tunnel junctions scaling down to 1X nm", *in Proceedings of the IEEE*, February 2015.

[31] E. Garzón, et. al. Assessment of STT-MRAM Performance at Nanoscaled Technology Nodes Using a Device-To-Memory Simulation Framework, Journal of Microelectronics Engineering, Vol. 215, 2019

[32] E. Garzón et al., "Exploiting Double-Barrier MTJs for Energy-Efficient Nanoscaled STT-MRAMs," 2019 16th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD), Lausanne, Switzerland, 2019, pp. 85-88