Open access • Journal Article • DOI:10.1109/JPROC.2016.2590142

Magnetoresistive Random Access Memory — Source link

Dmytro Apalkov, Bernard Dieny, J. M. Slaughter

Institutions: Samsung, University of Grenoble

Published on: 29 Aug 2016

Magnetoresistive Random Access Memory

Dmytro Apalkov[1], Bernard Dieny[2], J. M. Slaughter[3]

1. Samsung Electronics, Semiconductor R&D Center, San Jose, California, USA.

2. SPINTEC, Grenoble Alpes Univ, CEA, CNRS, CEA/Grenoble, INAC, Grenoble, FRANCE

3. Everspin Technologies, Inc., Chandler, Arizona, USA

Abstract

A review of the developments in MRAM technology over the past 20 years is presented. The various MRAM generations are described with a particular focus on Spin-Transfer-Torque MRAM (STT-MRAM) which is currently receiving the greatest attention. The working principles of these various MRAM generations, the status of their developments, and demonstrations of working circuits, including already commercialized MRAM products, are discussed.

Keywords: MRAM, spintronics, spin electronics, magnetic tunnel junctions, tunnel magnetoresistance, spin transfer torque, STT-MRAM, toggle, thermally assisted MRAM

OUTLINE

1. Introduction to MRAM technology

Magnetoresistive random access memory (MRAM) is a class of solid-state storage circuits that store data as stable magnetic states of magnetoresistive devices, and read data by measuring the resistance of the devices to determine their magnetic states. In practice, the magnetoresistive devices are integrated with CMOS circuitry to make chips that are compatible with mass-produced semiconductor electronics. Such circuits have been designed around a variety of magnetoresistive devices, but commercially-produced MRAM products, and the vast majority of MRAM technologies being developed for future commercial MRAM technologies, are based on magnetic tunnel junction (MTJ) devices. All of these circuits are resistive memories in terms of the read operation; it is the method of writing the magnetic state that sets apart the different types of MRAM technology. Some of the heavily-studied write methods include Stoner-Wolfarth-type field switching, Savtchenko switching (also a field-switching method), spin-torque switching, and thermally-assisted switching (heat with field or spin torque). Two methods have so far been commercialized: toggle MRAM, which uses Savtchenko switching[1], has been in mass production since 2006,[2] and spin-torque switched MRAM is in the early-stages commercial production[3].

Advances in MRAM technology have been closely linked with advances in the understanding of magnetic and magneto-transport properties of ultra-thin films, including: tunneling magnetoresistance (TMR), MgO-based MTJ materials for giant TMR, synthetic antiferromagnet (SAF) structures, interfacial perpendicular magnetic anisotropy (PMA), and spin-transfer torque (STT). The application of scientific discovery to commercial technology seen in this field is striking in its breadth and speed of adoption. In this paper we review the major developments that are driving accelerating interest and adoption of MRAM, key considerations for functionality and scaling to higher densities, and the status of the major technology types.

1.1 Magnetic Tunnel Junction Devices for MRAM

Figure 1 shows the most basic magnetic tunnel junction structure, two ferromagnetic layers separated by a dielectric spacer layer, the tunnel barrier. When the tunnel barrier is very thin, typically < 2nm, quantum mechanical tunneling of electrons through the barrier makes the MTJ behave like a resistor having a resistance that depends exponentially on the barrier thickness and is proportional to the inverse of the in-plane barrier area. The tunneling current is spin-polarized, due to the asymmetric band structure of the ferromagnetic electrodes, giving rise to the tunneling magnetoresistance as shown in Figure 1. The relative orientation of the magnetizations in these two layers determines the resistance of the MTJ device. For most materials, the resistance is low when the magnetizations of the two layers are parallel, because the majority band electrons can tunnel into the majority band on the opposite side of the barrier. When the orientation is antiparallel, the resistance is high since the majority band electrons have to tunnel into the minority band of the opposite electrode.

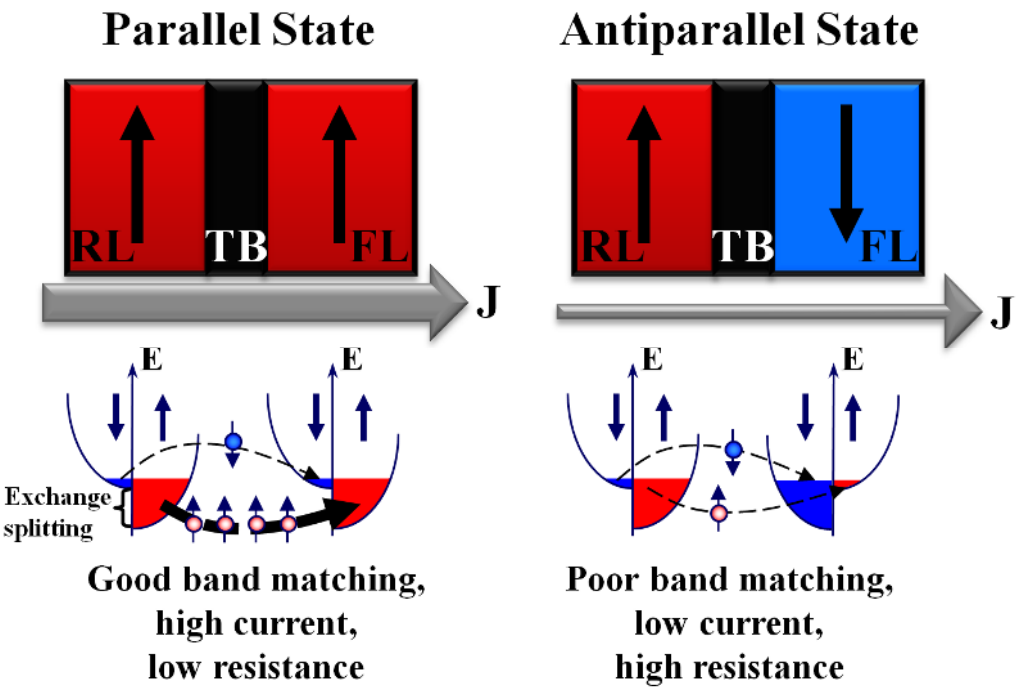

Figure 1. Sketch of a magnetic tunnel junction: two ferromagnetic layers (in red/blue) typically made of CoFe based alloys separated by a tunnel barrier (black) most often made of MgO about 1nm thick. Explanation of the tunneling magnetoresistance effect: parallel state has good band matching and low resistance, whereas antiparallel state has poor band matching: either absence of available carriers of a given spin or inadequacy of available states in the receiving electrode.

The simplest possible STT-MRAM design has the following components:

- Free layer (FL)
- Tunnel Barrier (TB)
- Reference layer (RL)

The free layer, sometimes called recording layer or storage layer, is the ferromagnetic layer retaining the stored information. This layer is often made of CoFeB material of different compositions[4]. The tunnel barrier is a thin (around 1 nm) insulating non-magnetic layer, that provides means to switch and read the state of the free layer with a spin-polarized tunneling current. In the past ten years, MgO tunneling barriers have been most heavily studied due to the giant TMR effect, though other materials, such as AlOx, and TiO have been used [5] and AlOx is used in production toggle MRAM. The other magnetic layer – the reference layer – provides a stable reference magnetization direction for the FL reading and switching. This layer is designed to have magnetic anisotropy much higher than the FL so that it never switches during memory operation.

Although multi-bit concepts exist, the MTJ devices that have been demonstrated in practical MRAM circuits have two stable magnetic states that store one bit of data as a parallel/antiparallel magnetic state with a low/high resistance as described above. To achieve this, such a magnetic device has a free layer engineered to have a uniaxial

magnetic anisotropy, so that the magnetization tends to lie along an easy axis in one or the other direction. Since the magnetization of ordinary ferromagnetic thin films is constrained to be in the film plane by the thin-film shape anisotropy, the in-plane easy axis is created mainly by patterning the free layer into a shape with a long direction (easy axis) and short direction (hard axis). The intrinsic anisotropy of the material may also contribute to the total uniaxial anisotropy. On the other hand, devices having their easy axis perpendicular to the film plane can have a number of advantages, discussed further in the sections that follow. A free layer engineered to have a perpendicular easy axis has stable states with the magnetization either up or down, with respect to a horizontal film plane, and the in-plane directions are hard. Such layers employ materials having a strong perpendicular anisotropy that can overcome the thin-film shape anisotropy, also discussed below. For any of these devices, the stability of the stored data over time and temperature is determined by the energy barrier between the two stable magnetic states, which is in turn related to the strength of the uniaxial anisotropy and the volume of magnetic material involved in the reversal process.

To create a memory array from MTJ devices, each device is typically integrated with an isolation transistor that can be turned on to pass a current selectively through the MTJ devices of interest, such as during the read operation. For spin-torque MRAM, the same transistor is used to pass the switching current through the target MTJ devices. Since each memory cell typically has one transistor and one MTJ, this particular architecture is known as the 1T-1MTJ MRAM architecture. Other architectures have been proposed and evaluated for various purposes, but the 1T-1MTJ cell is the most commonly used.

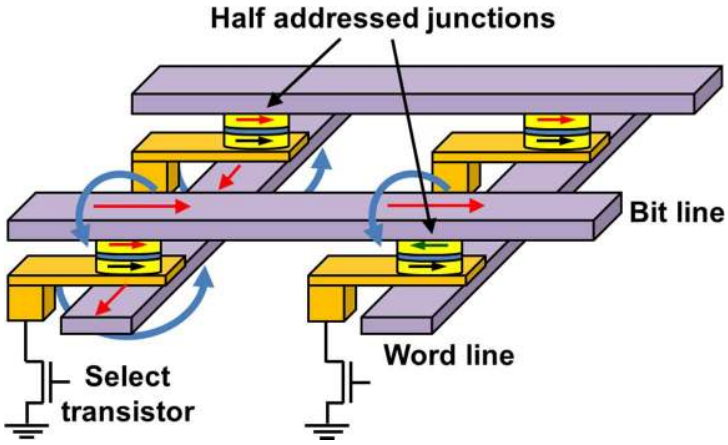

Fig.2: Sketch of the organization of field-written MRAM. Each cell comprises an MTJ connected in series with a select transistor.

1.2 Overview of MRAM technology generations

MRAM technology can be classified by the switching method employed to write data. First-generation MRAM is generally understood to include methods using magnetic fields to program the array. Toggle MRAM is the only first-generation MRAM in mass production [2], [6], [7]. A strong advantage of field switching is unlimited write endurance, since reversing the free-layer magnetization with a magnetic field does not create any wear-out effects. A disadvantage is the difficulty in scaling to smaller cell sizes, due to several factors including the magnitude of the

required switching currents and the somewhat complex memory cell geometry. A detailed discussion of this technology is presented in the section on Field-Switched MRAM.

Second-generation MRAM uses spin-transfer torque (STT) to program the array. STT switching can be accomplished with reasonable efficiency with MTJ devices having either in-plane or perpendicular-to-the-plane magnetization. STT-MRAM with in-plane devices began commercial production in 2015 with a 64Mb product [3] and is expected to ramp to volume production in 2016 with a 256Mb product[8]. STT-MRAM with perpendicular MTJ devices is under intensive development at a number of companies around the world. A discussion of technology demonstration vehicles, the fundamentals of the technology, and relative advantages/disadvantages is presented in several sections below.

A number of physical phenomena are under investigation for use in third generation MRAM, including: voltage-controlled anisotropy (VCA), voltage-controlled magnetism (VCM)[9]–[12], spin Hall effect (SHE) and spin-orbit torque switching (SOT)[13]–[17]. The potential for improved MRAM scaling and performance comes from idea that switching could be accomplished while passing little or no electrical current through the MTJ device. However, each effect has challenges to overcome for use in practical MRAM circuits. Some examples of those challenges are as follows. VCA does not directly lead to deterministic switching between two stable states, but is more likely to be applied in combination with another innovation for use in an MRAM array. Reliability issues like wear-out and parameter drift need to be better understood for practical VCM devices. SHE does not lend itself to efficient switching in devices with perpendicular magnetization, and SOT appears to require innovation in materials to increase the magnitude of the effect for improved switching efficiency. In many cases, these new devices would need to be deployed in a three-terminal cell configuration that is not compatible with high-density memory arrays. Nonetheless, the possibility of switching with little or no charge current passing through the tunnel barrier is stimulating a growing interest and investment due to the potential for use in low-power circuits and high-endurance applications.

2. Major advancements that enabled MRAM and stimulated commercial development

2.1 Tunneling Magnetoresistance

The development of magnetic tunnel junction (MTJ) material with high room-temperature magnetoresistance ratio, MR~10%, reported in 1995 [18], [19] offered a much improved read signal for MRAM technology. This major development in tunneling magnetoresistance (TMR) was followed by intensive work around the world that resulted in dramatic improvements in the techniques and materials, so that record TMR values have increased steadily, most markedly with the development of MgO tunnel barriers [4], [20], [21] to over 600%[22]. In addition to having a high magnetoresistance, MTJ materials also provide the ability to match the resistance of the storage device to that of the CMOS transistors used in the circuitry. This capability is essential to the technology since the transistors used in the MRAM read circuit result in a series resistance, typically in the kΩ range, that reduces the relative resistance change at the point of sensing. The giant magnetoresistance (GMR) and anisotropic magnetoresistance (AMR) devices, that

were available prior to practical MTJ devices, are inherently low resistance since they are metallic and therefore not well suited to integration with CMOS circuitry.

The promise of practical, high-TMR, materials motivated new MRAM R&D programs around the world, leading to the first commercial product [23], [2] approximately 11 years after the initial discovery [18],[19]. These materials also had a large positive impact on the hard disk drive (HDD) industry where MTJs were quickly commercialized for improved read sensors. Although the detailed requirements for MTJ devices in read sensors are quite different from the MRAM case, there is no doubt that the huge development effort in MTJ materials for HDD readers greatly accelerated the progress in MTJ materials for MRAM.

2.2 Synthetic antiferromagnet (SAF) structures and their application to MRAM

The discovery and intense study of oscillatory magnetic coupling (RKKY-coupling) in synthetic layered structures, in the 1980s, showed that it was possible to engineer thin film structures with separate ferromagnetic layers that are coupled either ferromagnetically or antiferromagnetically to each other through a nonmagnetic spacer layer separating them, with a coupling strength and sign determined by the thickness of the spacer [25]. This understanding led to the use of synthetic antiferromagnet (SAF) structures engineered to provide the required magnetic properties for both reference magnetic layers and free layers in MTJ devices for MRAM. Almost all MTJ devices for MRAM use a SAF for the reference layer. A SAF reference layer has strong antiferromagnetic (AF) coupling between ferromagnetic (FM) layers of nearly equal magnetic moment, resulting in a magnetically rigid system. When used as a free layer, the spacer layer thickness is set for weaker AF coupling to make a less rigid structure that is more responsive to applied magnetic fields as described further in sections that follow.

The simplest design of an MTJ is represented in Fig.3a. It exhibits tunnel magnetoresistance. However it has undesired magnetostatic interactions between the RL and the FL, which tend to bias the FL toward the antiparallel (AP) state (for in-plane magnetization).

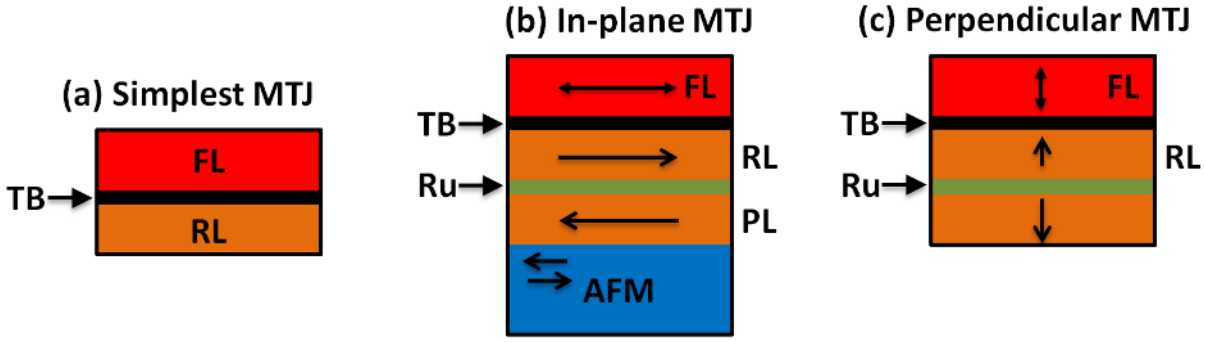

Figure 3. (a) Simplest MTJ design. (b): Typical in-plane MTJ with a synthetic antiferromagnetic reference layer pinned by an antiferromagnetic (AFM) pinning layer; (c) perpendicular MTJ with synthetic antiferromagnetic reference layer.

A typical in-plane MTJ with a SAF reference structure is illustrated in 3b. In this example, the pinned layer of the SAF (PL) is pinned in a specific direction by direct contact with an underlying antiferromagnetic (AFM) material, and the reference layer of the SAF (RL) is in contact with the tunnel barrier providing the reference direction for the MTJ. The SAF performs two major functions: it provides a stable magnetic reference direction since its low net moment is less affected by external fields than a single ferromagnetic layer, and the imbalance of magnetic moment between the two layers can be easily adjusted to provide a negative dipolar field that cancels the positive interlayer coupling between the reference layer and the free layer [26]. The interlayer coupling can include Neel coupling, due to magnetic poles created by correlated roughness at the surfaces of the layers, [27] and an electron-mediated coupling associated with the polarized electrons tunneling through the barrier, both of which favor ferromagnetic alignment of the reference and free layers. A typical MTJ with perpendicularly magnetized SAF and free layer is shown in 3c. The same principles apply in this geometry although, due to the extremely high anisotropies achievable with engineered perpendicular materials, the pinning AFM layer is not necessary.

A SAF structure for the free layer serves a different purpose and has very different design requirements compared to a SAF reference structure. The main use of a SAF free layer is to enable the Savtchenko switching technique used in toggle MRAM, which is described further in the section below.

2.3 Savtchenko Switching

Writing data into a memory array requires the ability to select the storage devices to be written within a large array of such devices, switch those target devices with extremely high reliability, and leave all other devices unaffected with extremely high reliability. For field-switched MRAM, this proved to be a significant challenge [26]. The basic idea of field switching is to use rows and columns of write lines, placed under and over the MTJ devices, that create magnetic fields large enough to reverse the free layer magnetization for the devices located at the cross point where both a row and column line are energized (See Fig.2). The fundamental problem with this approach is that all of the non-target bits along the energized lines also see a field that, although smaller in magnitude, causes a non-negligible probability of switching those half-selected bits. This is known as the “half-select problem.”

The half-select problem for field switching was solved with the invention of Savtchenko switching, [1] which cleverly employs a SAF free layer and a sequence of write-current pulses that creates a rotating field to switch it. MRAM that uses Savtchenko switching is usually called “toggle MRAM” because the switching technique toggles the SAF free layer from one stable state to the other using the same write pulse sequence, as opposed to using different fields to write the high or low state. The first commercially-produced MRAM product was a 4Mb toggle MRAM which began production in 2006 by Freescale Semiconductor. That part and a family of products based on the same technology platform, now produced by Everspin Technologies, account for nearly all of MRAM production and sales today, with the remainder being Everspin’s 64Mb STT-MRAM product.

2.4 Spin-Torque Switching

The spin-transfer torque effect was theoretically predicted independently by J. Slonczewski and L. Berger in 1996 [28], [29]. When a current is passed between two ferromagnetic layers separated by a non-magnetic spacer, it exerts a torque on the magnetization of each of the layers. If the current is high enough, this can result in switching of the free layer. The first demonstration of STT induced magnetization switching was done by Katine et al. using a spin valve device made from all-metal GMR stack in 2000 [30]. The first STT switching demonstration with MTJ devices was by Huai, et al. [5] in 2004 on AlOx-based tunneling junctions and 2005 on MgO-based MTJ [31]–[33].

2.5 Interfacial perpendicular magnetic anisotropy

Magnetic materials can exhibit magnetic anisotropy large enough to overcome the thin-film demagnetization energy and to pull the magnetization of the film out-of-plane. Many scientific studies in the late 1980s and early 1990s laid the foundations for rapid application of such materials in data storage and later in MRAM development. Perpendicular magnetization has been demonstrated in multilayered ultra-thin films and superlattices having a ferromagnetic material layered with a nonmagnetic material, such as Au, Pt, or Pd [34], [35], [36], and for similar structures made from two different ferromagnetic materials, Co and Ni [37]. This perpendicular anisotropy can have several origins associated with spin-orbit interactions: electronic hybridization effects, interfacial strain, magnetocrystalline anisotropy etc. A comprehensive survey of experimental studies on these materials can be found in reference [38].

In perpendicular MTJs, both the storage layer and reference layer are magnetized out-of plane (Fig.3c). For the reference layer, synthetic antiferromagnetic multilayers have become the preferred materials due to their strong and tunable perpendicular anisotropy energy and compatibility with fabrication techniques normally used for MTJ materials. As discussed above, practical use requires a SAF reference structure in MTJ devices for MRAM. Such SAF materials have been successfully and extensively developed and integrated into MgO-based pMTJ materials[39]–[41].

In addition to the SAF reference structure, pMTJ materials for MRAM also require a perpendicular free layer with soft magnetic properties that enable efficient spin-torque switching between the two stable states (up and down). Of particular importance for STT-MRAM was the discovery that a quite strong perpendicular anisotropy exists at the interface between magnetic metal and oxides such as (Co/AlOx or CoFe(B)/MgO) [42], [43]. This anisotropy has been shown to result from hybridization effects between the transition metal (Co,Fe) dz^2 orbitals and the oxygen orbitals of the oxide tunnel barrier [44], [45]. In STT-MRAM, this strong interfacial anisotropy provides a good thermal stability of the free layer magnetization together with a sufficiently low write current as explained in section 6.

Development of CoFeB-based free layers with first one MgO interface, [43] and later two MgO interfaces, [46], [47], combined with the multilayer-based perpendicular SAF materials, has led to practical pMTJ devices with the desired low switching current and high energy barrier needed for nonvolatile memory. STT-MRAM circuits and technology based on such pMTJ devices are in advanced development at a number of companies worldwide.

3. Demonstrations of working MRAM circuits

The major advancements described in the previous section have been successfully incorporated into a number of MRAM demonstration circuits and commercial products. Some notable demonstrations of working circuits are described below, illustrating the rapid progress in technology development over the past 15 years.

In 2000, Scheuerlein, et al. reported the functionality of a field-switched MRAM circuit using AlOx-based MTJ devices and Stoner-Wolfarth-type writing circuits [48]. Basic read and write functionality were demonstrated in a 1 kb memory array with write pulses as short as 2.5ns and with read operation as short as 10ns, using a twin-cell read configuration.

Durlam, et al. described the first Toggle MRAM circuit in 2003, a fully-functional 4Mb circuit with 1T1MTJ architecture and AlOx-based MTJ devices, demonstrating the effectiveness of Savtchenko writing to solve the problem of high write error rates due to the half-select in Stoner-Wolfarth-type field switching[49]. This circuit was designed to be a product, not simply a demonstration, with a unique local-bitline array architecture, a cell size of 1.55 μm^2, asynchronous 16-bit SRAM interface compatibility, and demonstrated symmetric read and write cycle times of less than 25ns [6]. This chip was introduced as a commercial product in 2006 and continues to be produced today by Everspin Technologies, Inc.

Significant results from several ST-MRAM demonstration circuits, all using MgO-based MTJ devices, have been published, beginning with a 4kb test vehicle in 2005 [50] followed by many others including: a 2Mb circuit and device data in 2007 [51], statistical data on 4 kb integrated arrays with 70×210 nm^2 bits [52] and devices with perpendicular magnetization [53] in 2008, arrays integrated with 54nm CMOS technology [54], 4kb MRAM arrays with perpendicular bit switching in 2010 [55].

The first ST-MRAM product, announced in 2012, is a 64Mb, DDR3, ST-MRAM circuit [56], [3]. Shown in Figure 4, the chip is fabricated using in-plane MTJ devices in a 1T1-MTJ cell, integrated with 90nm CMOS technology and packaged in a JEDEC standard DDR3 BGA. This chip is compatible with available DDR3 memory controllers and became commercially available from Everspin Technologies, Inc. in 2015.

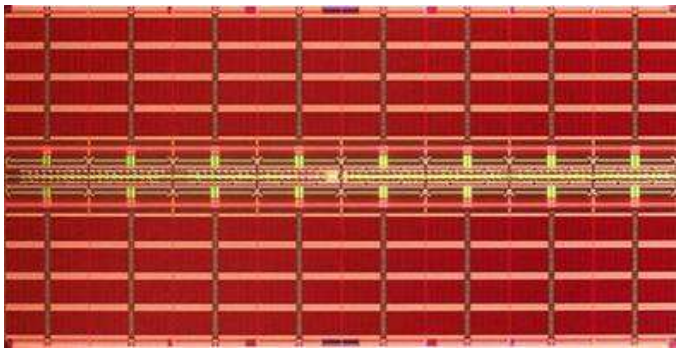

Figure 4. Everspin 64Mb DDR3 ST-MRAM Die Photo. Word line drivers run vertically through the center of each of eight 8Mb banks. Strips of column circuits run horizontally, dividing each bank into 8 sub-arrays.

In 2014 Thomas, et al. reported performance data for an 8Mb STT-MRAM demonstration chip with high energy barrier for data retention compatible with automotive applications and MTJ devices that are tolerant to high processing temperatures required for embedded memories [57]. Although CMOS circuitry is 90 nm technology, MTJ devices with diameter as small as 45nm were used, demonstrating reliable switching with pulses as short as 2ns in a 1T1MTJ cell architecture.

The continuous improvement reported in these papers, and many others, reflects significant innovation and progress toward the use of spin-torque switching in a wider range of products, including stand-alone memories and embedded memory in a multiple applications.

4. Requirements for reliable read in an array

The MTJ device should have a resistance in the kΩ range to minimize the effect of the series resistance from the isolation transistor. This resistance typically leads to a tunnel barrier thickness on the order of 1 nm. The exponential thickness dependence of RA creates a challenge for producing MTJ material that is repeatable and uniform over the area of wafers used in semiconductor production. The issues of cross-wafer uniformity and wafer-to-wafer repeatability have largely been solved by the manufacturers of MTJ productions tools that are capable of <1% uniformity and repeatability for these ultra-thin films. However, array uniformity, which is a measurement of bit-to-bit resistance variation within a memory array, is a continuing challenge as device sizes are scaled ever smaller.

During the read operation, a typical read circuit compares the resistance of the bit to a reference resistance in order to determine if the bit is in the high or low state. Figure 5 illustrates the effect of random resistance variations within an array on this type of read circuit. In the low state, bits with resistance on the high side of the distribution will fall closer to the reference than the average bit. Conversely, the bits on the low side of the distribution in the high resistance state will also be closer to the reference. To have working memories with bit counts of Mb or more, the circuit must be able to correctly read the state of these bits in the tails of the distributions. If bits that are statistically separated 5σ from the mean are unreadable, there will be approximately one bad bit in 2Mb, a marginally unacceptable result requiring error correction.

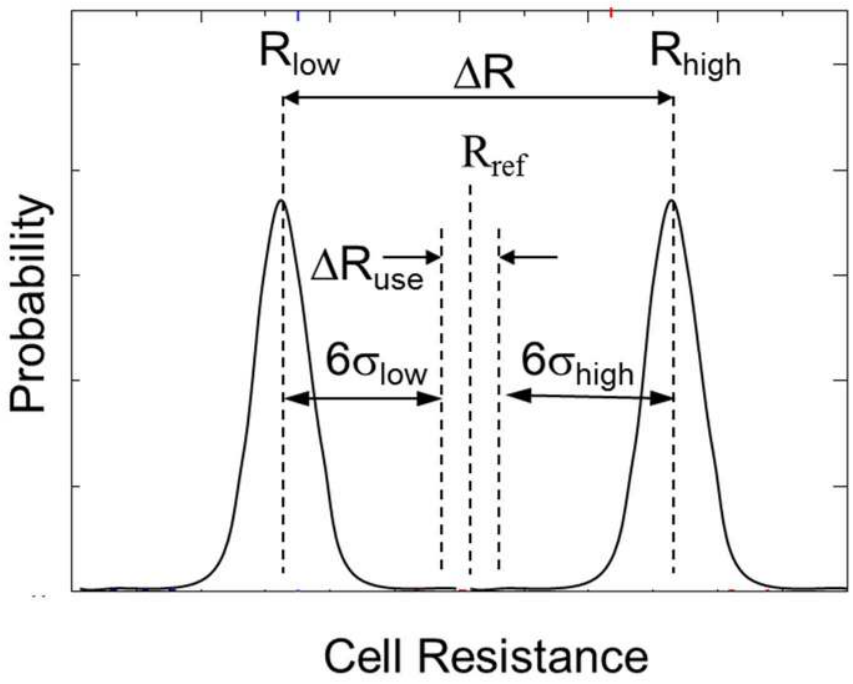

Figure 5. Schematic representation of Gaussian resistance distributions for a large array of MTJ cells in the low- and high-resistance states. A midpoint reference, represented here by R_{ref}, is generated by the read circuitry to evaluate the state of the MTJ device by comparing the measured resistance to R_{ref}. The average change in resistance from the low to high state is ΔR, but the circuit must be capable of determining the resistance change for cells in the tails of the distributions, meaning the resistance change available to the circuit is ΔR_{use}.

A reasonable criterion for feasibility is 6σ separation from the midpoint, or 12σ separation between the mean high and low resistance states. Since the circuit must be able to correctly read the low 6σ bits in the R_{high} distribution, and the high 6σ bits in the R_{low} distribution, the useable resistance change for the circuit ΔR_{use} is the difference between the midpoint reference and the 6σ tails as illustrated in Figure 5. A larger ΔR_{use} allows a more aggressive circuit design with faster access time. It will make the read-out more robust, less sensitive to noise, and can be used for higher speed operation. To account for high-speed reading with normal process variations in the circuitry, significantly more than 12σ is required. For example, Toggle MRAM in production has demonstrated over 20σ separation [49].

A number of factors can contribute to the width of the resistance distribution. Any variation in the bit area, e.g. due to lithography or etch variations, will directly cause variations in bit resistance. Process damage or veils created during the etch process also may contribute. However, it has been shown that the quality of the MTJ tunnel barrier itself can make a significant contribution to bit-to-bit resistance variation [23]. As the size (area) of MTJ devices is reduced, a given variation in the MTJ dimensions from the patterning process will create a larger relative bit-to-bit variation in the resistance. In addition, a given level of inhomogeneity in the tunnel barrier will result in a

similar increasing relative variation with decreasing device size. As a result, maintaining a given level of separation in the resistance distributions for a robust mid-point read operation requires improvements in both the patterning process and the quality of the tunnel barrier.

For cases when separation between the high and low resistance distributions is not sufficient for reading with a mid-point reference scheme in a 1T-1MTJ architecture, other methods and/or cell architectures can be employed. A twin-cell architecture, has two MTJ devices, in a 2T-2MTJ memory cell or in two adjacent 1MTJ cells, that are always written in opposite states to effectively double the resistance change. A circuit employing this cell was used to demonstrate high-speed MRAM read in one of the first MRAM demonstration circuits [48]. The trade-off for the higher read performance is array area; twice as much area is needed for a given memory density.

Another method for mitigating the effect of wider distributions is the self-referenced read [58]–[62]. There are different techniques for self-referencing, but they all involve comparing the resistance of the MTJ in its initial state (stored data) to its resistance after being switched or rotated to a known state. For example, the resistance can be measured and stored, then compared to its resistance after a known switching pulse. Since the resistance of a device is compared to the same device in a different state, the bit-to-bit resistance distribution no longer plays a primary role. In another approach, based on big differences in resistance dependence on voltage in parallel and antiparallel states, two measurements are done at different voltage values and results are compared[63]. In all the self-referenced read schemes, the trade-off for the tolerance to wide resistance distributions is longer time for the read operation, due to the extra pulse and the second read access time.

5. Field-written toggle MRAM

A memory array for field-written MRAM typically has rows and columns of copper lines above and below the MTJ devices that generate magnetic fields when current pulses are passed through the lines (See Fig.2). Three key innovations that enable the successful commercialization of field-written MRAM are: Savtchenko switching for Toggle MRAM operation, improvements in tunnel barrier quality and bit patterning fidelity for narrow and well-behaved bit-to-bit MTJ resistance distributions, and the process for cladding of the write lines. As described briefly in section 2.3 and in more detail in the section on Toggle MRAM below, Savtchenko switching solved the half-select problem through the use of a SAF free layer and timed write pulses. Another method of addressing the half-select problem, by using a combination of magnetic field and heat for thermal-assisted switching, is described in section 7. The importance of narrow read distributions is described in the preceding section and is important to all types of MRAM circuits.

Cladding involves the addition of permeable ferromagnetic material around three sides of a write line to focus the magnetic field toward the fourth side [64]. This produces a stronger field at the MTJ devices located near the fourth side, above or below the line, and minimizes the stray fields in other directions. The effect is to significantly

reduce currents needed for the write operationand to suppress possible write errors due to stray fields experienced by rows or columns of bits adjacent to the row or column of bits to be written.

5.1. Operating principle of Toggle MRAM

The use of Savtchenko switching in MRAM arrays provides a number of benefits including: low write error rates, reduced sensitivity to external fields, and a wide operating region for the write currents that reduces sensitivity to manufacturing process variations [7]. It relies on the unique behavior of a synthetic antiferromagnet (SAF) free layer, formed from two nearly identical ferromagnetic layers separated by a non-magnetic coupling spacer layer. As shown schematically in Figures 6a and b, a single free layer responds to an external field by aligning with it, while the behavior of a SAF free layer is quite different. For a SAF having some net anisotropy H_k in each layer, there exists a critical spin flop field H_{sw} at which the two antiparallel layer magnetizations will rotate (flop) to be approximately orthogonal to the applied field H, with each layer scissoring slightly in the direction of H [65]. For fields $H \geq H_{sw}$, the SAF can lower its total energy by decreasing its dipole energy with a flop and scissor, even though the antiferromagnetic exchange energy is increased by the same scissoring. As the applied field strength is increased, the layers of the SAF will scissor further until they eventually become parallel and aligned with the field at the saturation field H_{sat}.

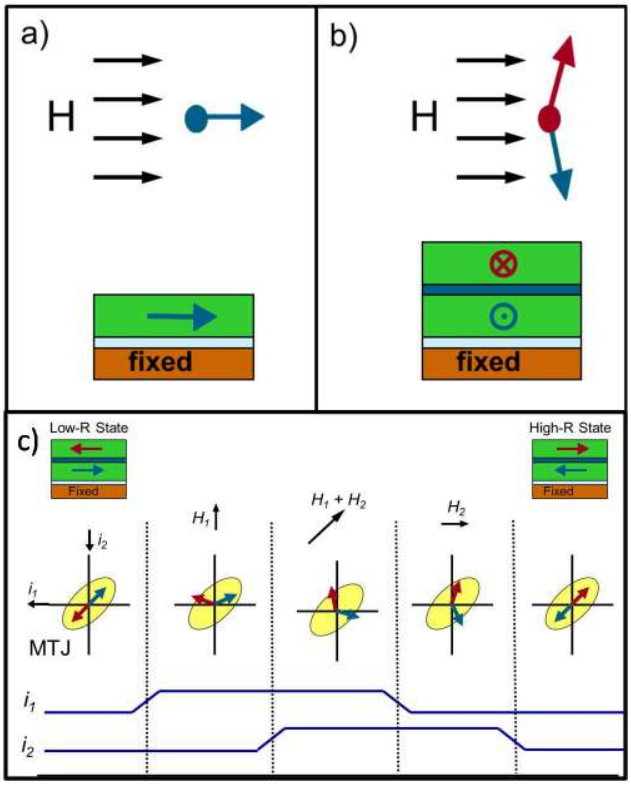

Figure 6. (a) A single layer of ferromagnetic material responds to an external field (**H**) by aligning with it, while the behavior of a SAF is quite different. A SAF free layer (b) rotates such that the moments of the layers are nearly orthogonal to **H** when the magnitude of **H** reaches the flop field of the SAF, with the moments canting to produce a net moment approximately

parallel to the field. As the field strength is increased, the moments scissor toward each other until they are parallel when the SAF is saturated. (c) Schematic of a toggle MRAM bit with the field sequence used to switch the free layer from one state to the other. The fields, H_1, $H_1 + H_2$, and H_2 are produced by passing currents, i_1 and i_2, through the write lines.

The programming pulse sequence and resulting response of the free layer are shown in Figure 6c. To toggle the bit from an initial '0' to a final '1' the currents i_1 and i_2 are pulsed with the timing indicated in the figure, such that the vector sum of the two magnetic fields effectively rotates through 90°. The SAF responds by orienting nominally orthogonal to this field and rotating with it such that, when i_2 is turned off, the moments relax to their easy-axis the magnetization of each layer now reversed. Because the TMR only depends on the direction of the layer that is in contact with the tunnel barrier, the resistance switches states each time the toggle pulse sequence is applied.

The toggle write mode requires that the memory circuit first determine the state of the target bit, and execute the toggle pulse sequence only if it is not in the desired state, i.e. only if the new data differs from the existing. If the pre-read determines that the target bit is already in the desired state for the new data, then that bit is not toggled. This approach increases the total write cycle time by adding a read operation, but has benefits in limiting the overall power consumption and improves array efficiency because the unipolar current allows the use of smaller transistors in the write drivers.

The elimination of the half-select problem can be understood by considering the energy barrier to reversal of bits experiencing a field from a single line (½-selected bits). For a SAF free layer, the single-line field raises the energy barrier of those bits, so that they are stabilized against reversal during the field pulse. This is in marked contrast to the conventional approach using a simple ferromagnetic layer for storage, where all of the ½-selected bits have their switching energy reduced and are therefore more susceptible to disturbance (thermal or external field).

5.2. Performance of Toggle MRAM

A major feature of Toggle MRAM is the essentially unlimited write endurance, since there is no wear-out mechanism related to switching the magnetization state of the free layer with magnetic fields. During the read operation, the bias applied to the MTJ is at a level far below the breakdown voltage of the device, but finite bias leads to a finite probability of eventual dielectric breakdown. For product reliability, two MTJ barrier failure modes must be controlled: time dependent dielectric breakdown (TDDB) and resistance drift. Dielectric breakdown is an abrupt drop in resistance, or shorting, of the tunnel barrier, while resistance drift is a gradual reduction of the junction resistance over time that can eventually lead to increased read error rates.

Many applications in industrial and automotive applications, require consistent performance and robust reliability over a broad operating temperature range, typically -40 °C < T < 125 °C. Rigorous reliability analysis done for the Everspin 4Mb MRAM product showed that it exceeds the reliability requirements of industrial and automotive products[66]–[68]. Extensive operational characterization and accelerated stress testing validated its full functionality and high reliability in this environments. Figure 7 demonstrates that the Toggle switching operating region is maintained at the highest temperatures with more than ample margin.

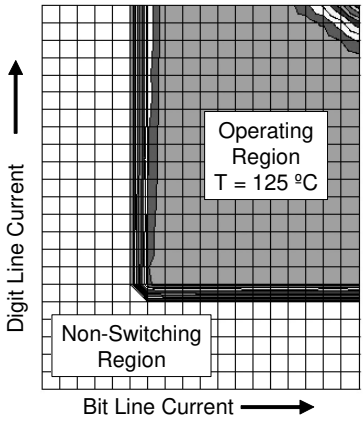

Figure 7 Toggle switching map at elevated temperature (T= 125 ºC) showing large operating region. Saturation of the free layer SAF exchange can start to be observed at over driven currents at the upper right in the figure. (From Ref. [68])

In the operating region, it is possible to reliably read data and write to the array within the 35ns read/write cycle time for this part. The large region with no switching errors is a demonstration of the ability of Savtchenko switching to eliminate half-select disturbs, even when the elevated temperature (125C) causes some reduction of the SAF coupling strength.

Everspin's commercial Toggle MRAM uses an MTJ device with AlOx tunnel barrier and NiFe SAF free layer. The SRAM-like parts are specified for full cycle time of 35ns for both read and write, but selected parts have been demonstrated to operate to <25ns. Improved read speed has been demonstrated by replacing AlOx with an MgO-based MTJ [24]. Although the relative resistance distributions with MgO are wider than for AlOx, with sigma~1.5% as compared to sigma~0.9% for AlOx, the higher MR provides more separation between the high and low resistance distributions, resulting in more usable signal for the read circuit. The MRAM cells using MgO and the NiFe SAF had double the MR compared to AlOx-NiFe, and 42% larger MR/σ, resulting in a reduction of the read cycle time from 21 ns to a circuit-limited 17 ns. Much faster read times are possible for circuits optimized to take advantage of the higher read signal.

5.3 Applications of Toggle MRAM

Toggle MRAM parts are used in a variety of different markets including: data storage, transportation, networking, and industrial automation. The combination of nonvolatility, unlimited endurance, high random access performance, and a wide operating temperature range, provides specific market advantages in many of these applications. Often the MRAM chip is replacing another type of memory as well as a battery or capacitor, providing a significant reliability benefit. The main factor limiting further market penetration is the limited memory density, with 16 Mb being the highest density in production. The next higher density MRAM chip currently in production is the 64 Mb part using STT switching [3]. Toggle MRAM is also used as an embedded memory in system-on-chip (SoC) applications. Embedded memories are often much smaller density than stand-alone memories, making Toggle

MRAM competitive with incumbent technologies. Since the MTJ layer is inserted between metal layers in the back-end part of the process, fabricating embedded MRAM is no different than fabricating standalone memory.

The application that consumes the largest number of MRAM chips today is RAID data storage systems. MRAM chips with densities usually in the 1Mb to 4Mb range are used in these systems to store metadata, such as restart vectors, controller parameters, system configuration data, and various tables. In case of power failure, this system information is inherently saved in the nonvolatile memory and can be accessed quickly upon restart. This application requires fast, unlimited reads and writes, combined with nonvolatility and high reliability. In this and similar applications, a Toggle MRAM part with a parallel interface often replaces some type of SRAM combined with a battery or capacitor to enable data retention when power is lost, providing higher reliability and a lower cost solution.

In transportation and industrial applications, the wide operating temperature range and impressive data retention time are important factors. Toggle MRAM arrays, based on 180 nm technology, have demonstrated data retention times over 20 years at 125 C, as well as reliable operation from -40 C to 125 C [68]. Since MTJ devices store data with a magnetic state, rather than charge, they have advantages over charge-based memories for space and military applications where exposure to ionizing radiation is an issue. The same MRAM arrays used for industrial grade MRAM are integrated with special radiation-hard CMOS to address those markets.

The higher densities enabled by STT switching is enabling MRAM technology to move into additional storage applications, including caching and buffering, as well as other sockets typically filled by a DRAM.

6. Spin transfer torque MRAM (STT-MRAM)

6.1. Introduction

One of the limitations of field-switched MRAM introduced at the beginning of this review is the difficulty of maintaining low error rates and high data retention while reducing the size of the MTJ devices. The first effect of shrinking the device dimension is an increase in the switching field distribution width, affecting write currents and error rates, and further shrinking results in loss of data retention time. Although toggle MRAM at the 180nm or 130nm technology nodes has a very large thermal stability factor Δ, maintaining a sufficiently large Δ for smaller free-layer volume would eventually require a larger anisotropy field H_K that is difficult to achieve with shape anisotropy. These two scaling issues have led industry to shift from toggle switching to STT switching for technology nodes of 90nm and below.

To understand the fundamental physics behind the spin transfer torque effect, we can use the free-electron model. Let us consider a structure consisting of two ferromagnetic layers separated by a non-magnetic spacer layer (Fig. 8). When a current goes through one ferromagnetic layer, the electrons become spin-polarized along the magnetization of this layer due to spin-dependent scattering, as shown in Figure. 8. As the spin-polarized current enters the other ferromagnetic layer, the spin of transmitted electrons precess incoherently around the local exchange field which is along the magnetization of this layer. This exchange field results from the exchange interactions between the spin of the conduction electrons and those responsible for the local magnetization. As a result of that, within a very short distance of the order of 1 nm, the electron current becomes repolarized along the magnetization of the second ferromagnetic layer. Due to momentum conservation, the difference between the momentum of the incoming and outgoing electron currents yields a torque acting on the magnetization of the second ferromagnet – this torque is known as *spin-transfer torque* [28], [29]. Several materials can be used to separate the two ferromagnetic layers. When this material is conducting, the corresponding trilayer structure (ferromagnet-conductor-ferromagnet) is called a spin-valve (SV)[69]. Alternatively, if the material is insulating the resultant structure becomes a magnetic tunneling junction (MTJ).

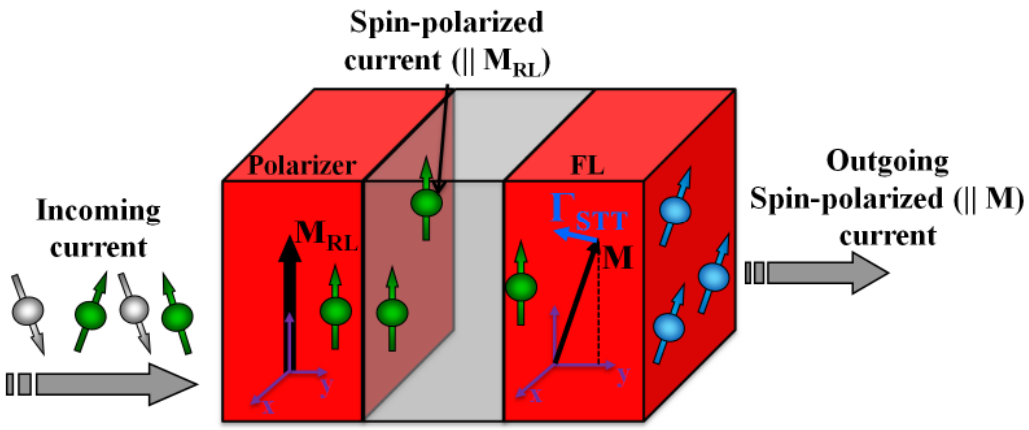

Figure 8. Free-electron model for spin transfer torque. The electrons get spin-polarized in the first ferromagnetic layer (polarizer or RL – reference layer) with spin aligned parallel to magnetization of the polarizer M_{RL}. As the electrons transverse the free layer (FL), their spin is repolarized towards the magnetization of the FL. Due to the momentum conservation, the difference of the incoming and outgoing spin-polarized current gives rise to STT torque.

Magnetoresistance and spin-transfer torque can be viewed as closely-related effects**.** In magnetoresistance, the magnetic structure (parallel or antiparallel orientation of magnetizations) affects the electrical properties of the stack, notably resistance (see Fig. 1). In STT effect, it is the electric current being passed through magnetic structure that can change its magnetic state. In both cases, the fundamental origin lies in the exchange interactions between the spin of conduction electrons (determining electrical properties) and the localized spins (responsible for the magnetic properties).

6.2. Fundamentals of STT-MRAM

The application of spin-transfer torque writing in MRAM allowed a large reduction of switching current as compared to field-switched MRAM (Fig. 9). As the bit becomes smaller, the writing current decreases proportionally to the MTJ area down to a minimum value dependent on the cell thermal stability factor and other parameters characteristic of the material used for the storage layer.

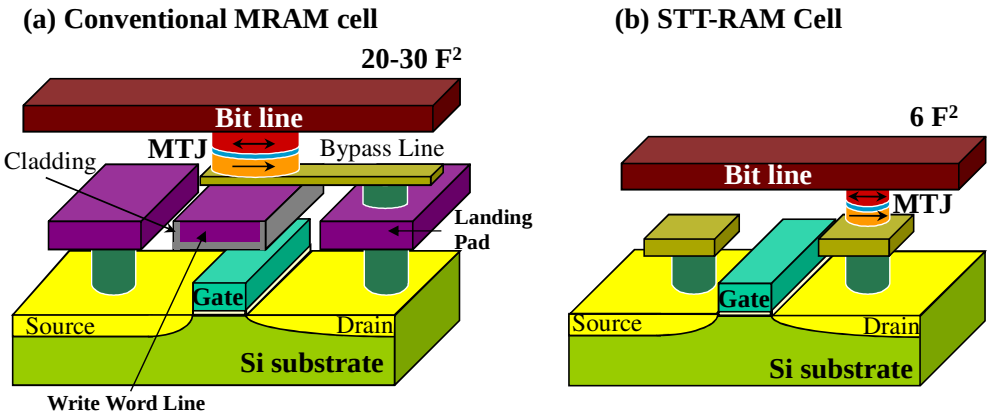

Figure 9. Structure of conventional field-switched MRAM cell as compared to STT-MRAM cell.

In addition to that, STT-MRAM has a simpler geometry (Fig. 9), eliminating landing pad, word and bypass lines that allows scaling down to cell size of 4-6F^2, where F is the semiconductor feature size. For reducing the cell size even further, it is possible to introduce multiple MTJ elements in one STT-MRAM cell [70]. Such MLC (Multiple-Level Cell) designs will pose more challenges to reduce the writing/reading margins and will likely reduce speed of operation.

6.2.1 STT-MRAM trilemma

Any memory including STT-MRAM has to provide the following basic functionalities: writing the data, retaining the written data for a predefined period of time (retention) and reading the recorded information. This section describes challenges of achieving these three properties together.

6.2.1.1 Information recording: STT writing

In order to discuss details of the STT-MRAM operation, it is necessary to introduce how the spin-transfer torque results in switching. The magnetization dynamics of any magnetic layer can be described by the Landau-Lifshitz-Gilbert (LLG) equation [71]:

$$\frac{d\mathbf{m}}{dt} = \boldsymbol{\Gamma}_{prec} + \boldsymbol{\Gamma}_{damp}$$
$$\boldsymbol{\Gamma}_{prec} = -\gamma\mu_0 \mathbf{m}\times\mathbf{H}; \qquad \boldsymbol{\Gamma}_{damp} = -\alpha\gamma\mu_0 \mathbf{m}\times(\mathbf{m}\times\mathbf{H}); \tag{1}$$

where $\mathbf{m} \equiv \mathbf{M}/M_S$ is the normalized vector along the magnetization **M** of the magnetic layer with saturation magnetization M_S, **H** is the total effective magnetic field (including anisotropy and applied fields), α is Gilbert damping, γ the gyromagnetic ratio and μ_0 the vacuum permeability. The first term $\boldsymbol{\Gamma}_{prec}$ in (1) describes the precessional motion of magnetization around the effective field **H** while the second term $\boldsymbol{\Gamma}_{damp}$ describes the gradual damping of the magnetization precessional motion towards the effective magnetic field **H**, hence reducing its total energy.

When spin-transfer torque acts on a magnetic layer such as the storage layer in STT-MRAM, two additional terms appear in the LLG equation[28]:

$$\frac{d\mathbf{m}}{dt} = \boldsymbol{\Gamma}_{prec} + \boldsymbol{\Gamma}_{damp} + \boldsymbol{\Gamma}_{IP}^{STT} + \boldsymbol{\Gamma}_{P}^{STT}$$

With

$$\boldsymbol{\Gamma}_{STT}^{IP} = \gamma\mu_0\eta\frac{\hbar}{2}\frac{J}{e}\frac{1}{M_S t}\mathbf{m}\times(\mathbf{m}\times\mathbf{m}_{RL}); \quad \boldsymbol{\Gamma}_{STT}^{P} = \gamma\mu_0\eta'\frac{\hbar}{2}\frac{J}{e}\frac{1}{M_S t}\mathbf{m}\times\mathbf{m}_{RL} \tag{2}$$

where *J* is the current density, *t* is the thickness of the free layer, *e* - the electron charge, $\hbar$ is the Planck constant, η is the spin transfer efficiency (directly related to the current spin-polarization). $\boldsymbol{\Gamma}_{STT}^{IP}$ is the torque lying in the plane defined by the two vectors $\mathbf{m}$ and $\mathbf{m}_{RL}$ and hence often called in-plane torque, or damping-like torque (since it has the same form as the Gilbert damping but can change sign depending on the current direction and therefore can behave as an antidamping torque) or Slonczewski torque. The second torque $\boldsymbol{\Gamma}_{STT}^{P}$ lies perpendicular to this plane and hence is often called perpendicular spin torque or field-like torque. Concerning the magnetization dynamics induced by STT, in the most relevant case of axial symmetry when $\mathbf{m}_{RL} \parallel \mathbf{H}$, the perpendicular STT has the same

effect as the precessional torque term i.e. its main effect is to change the frequency of the magnetization precession. In contrast, the in-plane STT can either enhance or reduce the damping torque, as depicted in Fig. 10. When the in-plane STT opposes the damping torque and overcomes it in magnitude, the amplitude of the precession increases to a point where magnetization switching can occur as illustrated in Fig. 10b. The current at which this happens is commonly referred to as *critical switching current* J_{c0}.

This action of in-plane STT is often described as a "negative damping" or an "anti-damping" [72], [73]. For switching consideration, usually the in-plane torque is most important. However the perpendicular torque can become relevant in certain specific problems [74]. An example of the switching trajectory for the case of spin torque overcoming damping torque is shown in Fig.10b. The initial orientation of the magnetization is along +Z axis. The notable change in precession chirality as m_z crosses zero (counterclockwise to clockwise looking downwards on XY plane) is a signature of spin torque switching and is due to the change of sign of the anisotropy field between upper and lower hemispheres (pointing along +Z for m_z>0 and –Z for m_z < 0).

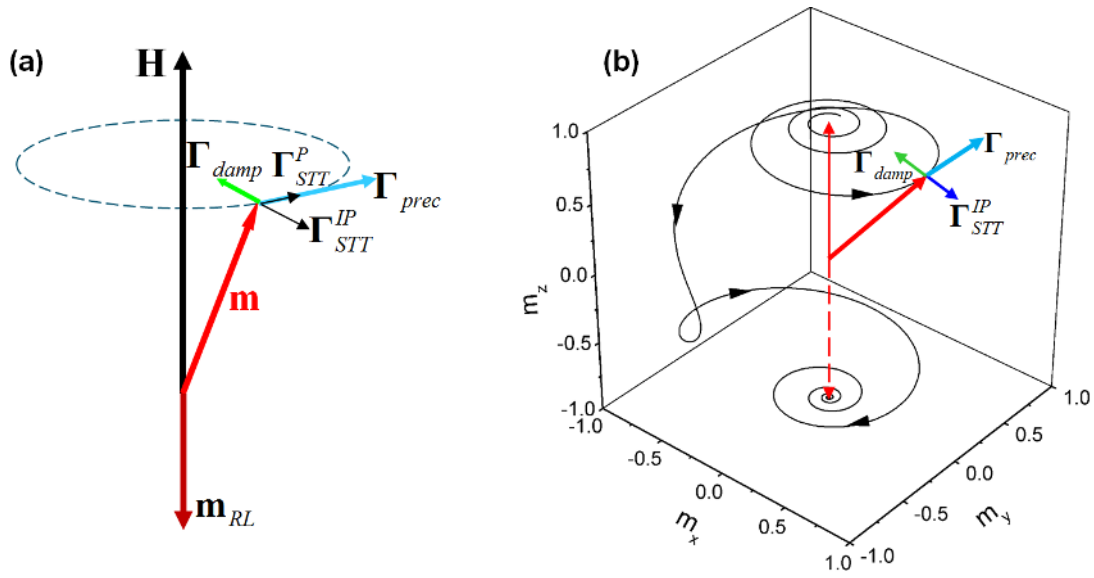

Figure 10. Spin-transfer torques acting on the magnetization (a) and resulting switching trajectory (b).

By solving the LLG equation, the following expressions for the critical switching current of in-plane (IP) and perpendicular to plane (PP) magnetized storage layer (respectively case of Figure 3b and 3c) were derived [75]:

$$J_{c0}^{IP} = \frac{1}{\eta}\frac{2\alpha e\mu_0}{\hbar}(M_s t)\left(\frac{H_{\perp eff}}{2} + H_K\right)$$
$$J_{c0}^{PP} = \frac{1}{\eta}\frac{2\alpha e\mu_0}{\hbar}(M_s t)(H_K) \quad (3)$$

where η is the spin transfer efficiency, and α, t, H_K, $H_{\perp eff}$ are respectively the magnetic damping constant, thickness, in-plane anisotropy field and out-of-plane demagnetizing field of the FL, respectively. At this value of the current, effective damping is equal to zero and any infinitesimal precession of magnetization becomes unstable. However, the switching time strongly depends on the actual value of the current density with respect to the value of J_{c0}. As a rule of thumb, if the measurements are done at T=300 K, J_{c0} is close to the average switching current density for pulse width t_p~10-20 ns. For longer pulse width, the average switching current density reduces due to thermal fluctuations, which help the magnetization to overcome the energy barrier. Two regimes in $J_c(t_p)$ are commonly observed, as shown in Figure 11:

- For very short pulses (t_p<10ns), the switching current density is larger than J_{c0} and during switching, details of individual magnetization precession are important – this is commonly referred as "precessional or ballistic regime". In this precessional regime at T = 0K, switching current for a given pulse width t_p is given by the following expression [76]:

$$J_c^{PP} = J_{c0}^{PP}\left(1+\frac{\tau}{t_p}\ln\frac{\pi}{2\theta_0}\right), \tag{4}$$

where $\tau=\left(\alpha\gamma\mu_0 H_K\right)^{-1}$ is the characteristic relaxation time, θ_0 is the initial angle between the magnetization and the easy axis when the current pulse is turned on.

From expression (3), it is clearly seen that for a given damping, spin torque efficiency and memory retention (the latter being directly related to H_k), STT-MRAM with perpendicular anisotropy should operate with lower switching current density. This and other reasons described below explain why most of the R&D in STT-MRAM is focused on cells based on out-of-plane magnetized MTJs. If the switching is performed at non-zero temperature, there will be two important effects due to random thermal fluctuations. First, the initial angle θ_0 will be distributed according to Maxwell-Boltzmann distribution [73], introducing distribution of switching current for a given pulse width. This effect is most important for short pulse width, typically for t_p<20-50 ns –see Figure.11. Second, thermal fluctuations will affect the switching process itself.

-For longer pulse width values, usually for t_p>50-100 ns – this switching regime is called thermally-activated regime (Fig.11). In this case, STT effect starts to play a secondary role by increasing effective temperature of the magnetization, thus increasing thermal fluctuations, which in turn, help the magnetization to overcome the energy barrier. The boundary between the two regimes is not very well defined and depends on the FL properties but usually this boundary corresponds to switching current near but less than J_{c0}. Based on thermal activation model, the following expression has been derived for STT-MRAM in thermally-activated regime:[76]–[78]

$$J_c = J_{c0}\left(1-\left(\frac{1}{\Delta}\ln\frac{t_{pw}}{\tau_0}\right)^{1/\xi}\right), \xi = \begin{cases} 1 \text{ for IP FL} \\ 2 \text{ for PP FL} \end{cases} \quad (5)$$

where Δ is the thermal stability factor (ratio of energy barrier for switching in stand-by $E_b=\mu_0 M_s H_k V/2$ divided by the thermal activation k_BT, V being the volume, k_B the Boltzmann constant, T the temperature), τ_0 is an intrinsic attempt time of the order of 1ns and the parameter ξ is equal to 1 or 2 for respectively in-plane and out-of-plane magnetized FL [78], [79], [77]. This dependence is often used to experimentally determine J_{c0} and Δ for IP FL by measuring median J_c at different pulse width spanning a few decades, and fitting J_c as a function of $log(t_p)$ and extrapolating to 1 ns: the slope yields the thermal stability Δ and the extrapolated J_c (1 ns) gives J_{c0}. This approach (with $\xi = 1$) has also been used for PP FL. However it was shown to give inaccurate results unless fitting with $\xi = 2$ is used [78].

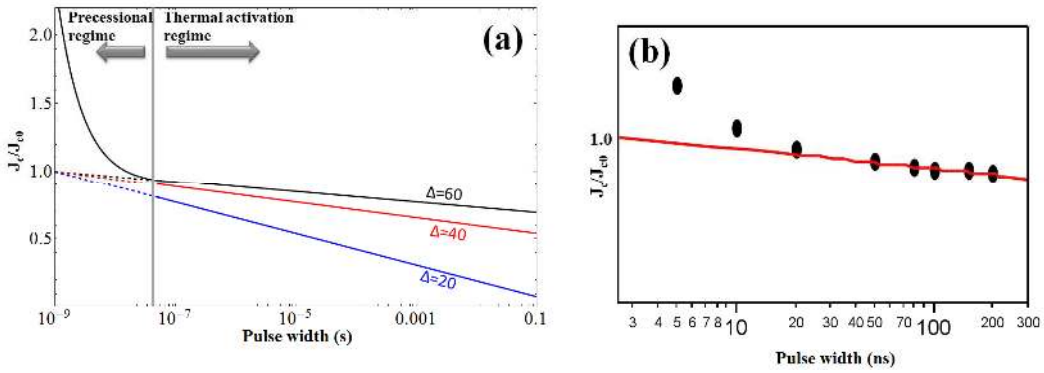

Figure 11. Dependence of switching current density on current pulse width: precessional and thermally activated regimes by modeling (a) and typical experiment (b). Behavior of the STT-MRAM cells with equal Jc0 but different thermal stability factor Δ is shown in thermally activated regime (a). The data correspond to in-plane magnetized FL. Perpendicular FL exhibit qualitatively similar behavior.

6.2.1.2. Information storing

Similar to field-switched MRAM discussed earlier in this review, the information retention capability is determined by the energy barrier E_b that the free layer magnetization has to overcome to switch to opposite state in stand-by divided by the thermal activation energy at the operation temperature [80]:

$$\Delta = \frac{E_b}{k_B T} \quad (6)$$

This expression holds for both in-plane and perpendicular STT-MRAM free layer. For small dimensions and high density, perpendicular STT-MRAM is preferred, as will be discussed later.

Typically, Δ > 60-80 is needed to guarantee reliable data retention for 10 years. For IP free layer, the energy barrier is provided by shape anisotropy (thus requiring elongated shapes) and is given by the following expression:

$$\Delta^{IP} = \frac{E_b}{k_B T} = \frac{\mu_0 H_K M_S V}{2 k_B T} \tag{7}$$

For in-plane free layer with elliptical cross section with short axis of dimension *w*, one can derive a simplified expression for the effective anisotropy field:

$$H_K^{IP} = 2\frac{M_S t(AR-1)}{wAR} \tag{8}$$

Where AR is the in-plane aspect ratio: length divided by width. Thus, the thermal stability becomes:

$$\Delta^{IP} = \frac{\mu_0 \pi (M_S t)^2 w(AR-1)}{8 k_B T} \tag{9}$$

From this expression, we can see that the stability of in-plane magnetized cells is proportional to the square of the moment per unit surface ($M_s t$) and width of the cell. $M_s t$ is a quantity that can easily be measured by VSM magnetometry. This expression, which is derived based on assumption that the free layer magnetization switches uniformly (macrospin approximation) is in good agreement with experimental observations, provided that AR is not excessively large (<2.5-3) and macrospin approximation is valid.

For FL with perpendicular magnetization, a general expression for the anisotropy energy density can be written as follows:

$$K^{PP} = K_b^{PP} + \frac{\sigma_i}{t} - \frac{1}{2}\mu_0 M_S^2 = \frac{\mu_0 H_K^{PP} M_S}{2}, \tag{10}$$

where σ_i is the perpendicular surface anisotropy per unit surface taking into account the two interfaces of the FL, K_b^{PP} is the bulk anisotropy (per unit volume). The term -1/2 $\mu_0 M_s^2$ is the demagnetizing energy which tends to bring back the magnetization in-plane.

Two classes of PMA materials are often distinguished based on the dominant origin of anisotropy [81]:

- B-PMA (bulk perpendicular magnetic anisotropy)
- I-PMA (interfacial perpendicular magnetic anisotropy)

The former usually involves materials, which have crystalline anisotropy due to reduced crystalline symmetry, making in-plane and out-of-plane directions inequivalent. Examples of such materials are FePt and CoPt alloys or multilayers. In most of these materials, binary or ternary alloys are used and at least one constituent material has high atomic (Z) number, which provides high spin-orbit coupling and correlatively large damping. While strong

anisotropy is beneficial for perpendicular STT-MRAM free layer, large damping is not desired since switching current is proportional to damping (Equation 3).

The latter, I-PMA anisotropy is related to interaction between the ferromagnetic film and the adjacent layer. An important class of I-PMA materials are based on the interface between a metallic ferromagnet and an oxide, which was demonstrated in 2002 using Pt/CoFe/AlOx structures[44], [82], [83]. Since the presence of Pt is undesirable due to an increase in damping from the spin-pumping effect, the same I-PMA at magnetic metal/oxide interface was later demonstrated in Pt-free structures suitable for STT-MRAM, namely Ta/CoFeB/MgO structures [41], [43]. This material has several important advantages as compared to B-PMA materials: very good lattice matching with MgO, low damping combined with large interfacial anisotropy and high spin-polarization yielding a large TMR amplitude. However, one challenge when using I-PMA materials is to obtain a sufficiently high thermal stability, especially at small (sub-20nm) dimensions. In B-PMA material, increasing the FL thickness while keeping everything else the same provides an easy and effective way to increase the thermal stability of the FL:

$$\Delta_{QU}^{PP-B} = \left(K_b^{PP} - \frac{\mu_0}{2} M_S^2 \right) \frac{\pi D^2 t}{4 k_B T}, \tag{11}$$

where "B" stands for B-PMA, QU stands for quasi-uniform approximation (i.e. assuming that the magnetization of the FL switches uniformly) and *D* is the diameter of the free layer.

In contrast, for I-PMA, assuming that the interfacial anisotropy is a property of the interface and does not change as the thickness of the ferromagnetic layer increases, the thermal stability decreases with thickness according to:

$$\Delta_{QU}^{PP-I} = \left(\sigma_i - \frac{\mu_0}{2} M_S^2 t \right) \frac{\pi D^2}{4 k_B T} = \kappa_i \frac{\pi D^2}{4 k_B T}, \tag{12}$$

where we have introduced $\kappa_i = \sigma_i - \frac{\mu_0}{2} M_S^2 t$ as the perpendicular magnetic anisotropy per unit area. This parameter plays a very important role in PMA free layer. It directly determines how small a given free layer can be made while still being stable, as will be discussed later in the text., The expressions of Δ in the two equations above were derived assuming coherent macrospin switching of the free layer, which works well at very small cell dimensions (typically below 30nm in diameter) For both I-PMA and B-PMA materials but has significant discrepancies with experimental data at larger dimensions (See Fig.12) [84]. Indeed, at large dimensions, the evolution of Δ versus the cell dimension observed experimentally is different from the D² quadratic behavior predicted by eq. (11) and (12). Rather, it exhibits a saturation for D>40nm, as illustrated in Figure12 below.

The two main approaches to increasing Δ for I-PMA free layers include increasing σ_i and reducing M_S. A useful approache for obtaining more interfacial anisotropy energy includes the use of two MgO interfaces, one on

each surface of the FL [46][85]. This enables the approximate doubling interfacial anisotropy acting on the FL magnetization.

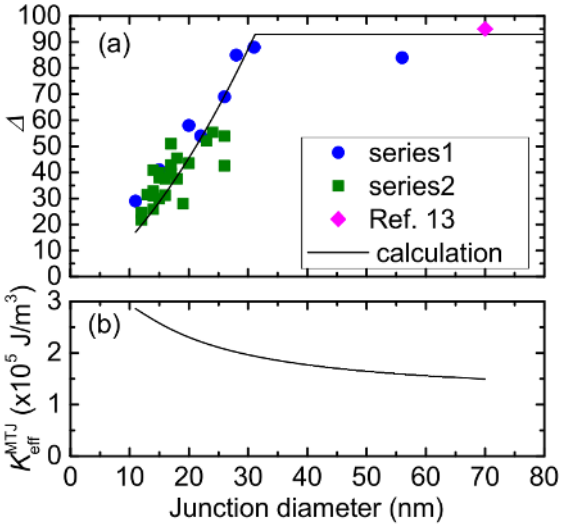

Figure 12. Experimental dependence of thermal stability Δ on the junction size (reprinted from [47] with permission). These data exhibit saturation of Δ at larger diameters indicating nonuniform magnetization reversal.

This change of behavior in the evolution of Δ versus diameter between small and large dimensions is related to a change in switching mode from coherent rotation at small dimensions to nucleation/propagation reversal of magnetization at larger dimensions [86]. As a matter of fact, at dimensions above 40nm, much less energy is required to reverse the FL magnetization by nucleating a reversed domain in the FL and propagating a domain-wall across the FL than to switch the magnetization coherently. This can be demonstrated by a Nudged Elastic Band (NEB) simulation [87]. This method consists in constructing a series of magnetic states of the free layer connecting the initial and final states (e.g. magnetization "up" and magnetization "down"). Then, each image is moved along local energy gradient while keeping the images equidistant. This procedure is repeated until convergence is found. The resultant series of images corresponds to a switching path that has minimum energy barrier. This procedure was applied to a perpendicular FL with typical parameters and it was found that even for relatively small dimensions (20-30 nm), switching by a quasi-uniform rotation results in a much larger energy barrier than switching by domain-wall formation and propagation [46], [86] (see Fig.13). Thus, equations (11) and (12) must be replaced by the following approximate expression that describes the energy needed to create a domain wall across the diameter of the free layer:

$$\Delta_{DW}^{PP} \approx \frac{4\sqrt{A_{ex}K^{PP}}Dt}{k_BT}, \tag{13}$$

where A_{ex} is the exchange stiffness in the free layer. In this expression, K^{PP} is given by expression (10).

As the diameter of the free layer changes, the energy to create a domain wall reduces linearly with diameter, whereas the energy for coherent rotation scales as the square of the diameter. Thus, at some critical dimension, the two energies cross each other yielding the observed crossover in switching behavior. For small sizes (below ~30nm diameter), quasi-uniform rotation results in smaller energy barrier and thus becomes the primary switching mechanism (see Figure 13 below 30nm). The critical dimension corresponding to the change of switching modes depends on the material parameters, e.g. increasing K^{PP} shifts it to smaller sizes.

In experiments, thermal stability often showed saturation at larger dimensions, as shown in Fig. 12 above ~30nm [47], [57]. This observation is contrary to what is expected from either single-domain model or NEB simulations (Fig. 13). A number of explanations were proposed over the years in order to improve the model to better match experimental observations, such as effective subvolume nucleation mechanism [88], edge nucleation mode [89] and others.

In another direction, researchers investigated whether experimental measurement of Δ[90] was accurate. For a typical thermal stability (above 60), direct measurement of retention (e.g. by measuring probability of thermally-activated switching of a chip at room temperature) is impossible and some acceleration techniques have to be used. Usually, they involve either modifying the energy barrier (by applying external magnetic field, spin-torque or both) or varying the temperature (chip baking at temperature significantly higher than room temperature). In all cases, a model is needed to quantify the dependence of Δ on the field, current or temperature. Until recently, the researchers have been using models which have two important assumptions:

- Uniform switching (or macrospin-like switching)
- Energy barrier dependent on current with exponent of ξ=1 (see eq. 5) [78] for both in-plane and perpendicular designs.

Both assumptions are not generally well justified for perpendicular MTJs. On the contrary, modeling shows that the thermally-activated switching occurs by domain-wall propagation (Fig. 13) and that the dependence of Δ on the field is more complicated than what was derived assuming macrospin behavior[91], see eq. (11) in [84]. Extension of Δ (H) model to domain-wall-mitigated switching was performed recently (eq. (2) in [92])) and evaluation of Δ using this model showed that indeed thermal stability is increasing linearly with increasing dimensions even well above 40 nm with excellent agreement with NEB simulations and chip data retention measurements at higher temperature without external field [92], solving the long-standing paradox of inconsistent behavior of thermal stability at larger dimensions.

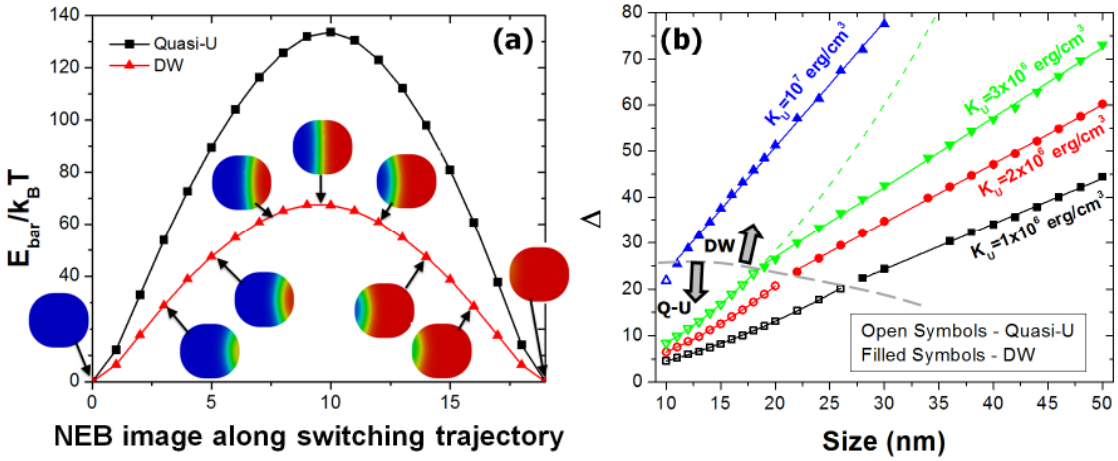

Figure 13. Nudged Elastic Band (NEB) simulation for minimum energy barrier: comparison of quasi-uniform switching (black) to domain-wall (DW) switching (red) (a). Dependence of thermal stability on size: smaller dimension – quadratic behavior, larger dimensions – linear (b). Crossover between quadratic and linear behavior moves to smaller dimensions at larger anisotropy. The short dimension of MTJ cell is 20 nm, thickness of free layer is 0.85nm[86].

6.2.1.3. Simultaneous achievement of writability, readability and retention: STT-MRAM trilemma

A key challenge for STT-MRAM is the simultaneous achievement of low switching current, high thermal stability and large TMR. For reading, some minimum TMR is needed to have a sufficient signal for determination of the resistance state by the read circuit. The actual TMR needed is dependent on the details of the read scheme and the bit-to-bit resistance distributions, but a common requirements for minimum TMR in MRAM arrays are between 150 % and 300%, sometimes more. For STT writing, the desired case is to have a switching current that is less than the saturation current of a minimum-sized transistor at the specified technology node. For storing the information, a typical requirement is to have thermal stability factor $\Delta > 80$ for 1Gbit memory array. The issue of engineering devices to meet all three of the above requirements is sometimes called the "STT-MRAM trilemma" (Figure 14), since it is often possible to improve one or two of the properties but at the expense of another. For example, for in-plane STT-MRAM, increasing the thickness of the FL provides an easy way to improve TMR and thermal stability, but the switching current increases as well. In the next section, we will discuss some approaches towards overcoming the boundaries of the trilemma using special designs.

Figure 14. An example STT-MRAM Trilemma. Simultaneously three properties have to be satisfied: reading, writing and storing. In this example is necessary to have TMR > 200 % for reliable reading, switching current less than that supplied by a small transistor for small cell size, and thermal stability factor Δ > 80 to meet a desired data retention goal.

6.2.2. Breaking STT-MRAM trilemma

6.2.2.1. Material improvements: damping and STT efficiency

One important approach to improve STT-MRAM characteristics involve development of better materials for MTJ structure. From eq. (3), reducing damping helps and increasing STT efficiency and polarization also contributes to reducing J_{c0}. Historically, for these reasons, CoFeB materials with various compositions have been used. They provide low damping constant of 0.005-0.015, good spin polarization and high TMR. At this moment, the highest TMR achieved at room temperature is 604 % for CoFeB/MgO/CoFeB structure [22]. Damping has been shown to be strongly dependent on film thickness, increasing greatly for the thinnest films due to a spin-pumping contribution that can be mitigated by using an insulating capping material [93].

In addition to CoFeB, there has been significant work on developing alternative materials to reduce damping and increase polarization. The most famous example include half-metallic materials, which at Fermi energy, have energy gap in one of the sub-bands (e.g. for spin-down) and thus not only should provide 100 % polarization, which translates to infinite TMR (at 0K) but also should give very low intrinsic damping. Very high values of TMR (>2000 %) were indeed achieved at low temperatures (4K). However, at room temperature, the half-metallic materials have failed to exceed or even reach the TMR provided by CoFeB alloys. There are a few fundamental difficulties with half-metallic materials that have to be overcome. Most of these materials are predicted or demonstrated to have half-metallicity in bulk form. However, once these materials are made as thin films (compatible with STT-MRAM FL requirements) and incorporated next to MgO, or another tunneling barrier, half-metallicity can be lost due to lattice modifications and electronic density of state changes,. In addition, the majority of half-metallic materials are cubic and do not possess perpendicular magnetic anisotropy, limiting their potential use to in-plane STT-MRAM. There have been a few approaches to address this, e.g. by using interfacial anisotropy [43] or creating a superlattice of two alternating distinct half-metallic materials [94].

6.2.2.2. Design improvements

6.2.2.2.a In-plane (IP) STT-MRAM with PPMA

One interesting approach to go beyond the boundaries of STT-MRAM trilemma consists in partially balancing the demagnetizing energy of the in-plane magnetized storage layer by a perpendicular anisotropy which has to be smaller than the demagnetizing energy. This anisotropy is often called partial perpendicular anisotropy (PPMA). If we look at Eq. (3) for J_{c0} of the in-plane FL, the term in the parenthesis involves $H_{\perp eff}$. This is the field which is needed to saturate the FL magnetization in the direction perpendicular to the plane. Without any PPMA, this field is equal to $\mu_0 M_S$. If PPMA with effective anisotropy field $H_{K\perp}$ is introduced, it becomes:

$$H_{\perp eff} = M_S - H_{K\perp} = M_S\left(1-\frac{H_{K\perp}}{M_S}\right) = M_S\left(1-h_{K\perp}\right) \tag{14}$$

where we have introduced dimensionless PPMA effective field $h_{K\perp}$ as the perpendicular anisotropy field normalized by the demagnetizing field:

$$h_{K\perp} = \frac{H_{K\perp}}{M_S} \tag{15}$$

For increasing PPMA, the switching current Jc0 can be reduced:

$$J_{c0}^{IP} = \frac{1}{\eta}\frac{2\mu_0\alpha e}{\hbar}\left(M_s t\right)M_S\left(\frac{1-h_{K\perp}}{2}+h_K\right) \tag{16}$$

where dimensionless in-plane (shape) anisotropy field $h_K = \frac{H_K}{M_S}$ was introduced. As can be seen from eq. (16) effectiveness of PPMA to reduce J_{c0} is larger for smaller in-plane H_k value, thus for larger dimensions. For smaller dimensions, we need larger H_k to maintain stability and PPMA effectiveness is reduced.

An important advantage of PPMA is that it can be introduced without change in thermal stability or TMR and can fundamentally change the boundaries of STT-MRAM trilemma. One efficient way to introduce PPMA into an existing FL is by using different capping material. This has been experimentally demonstrated [95], [96], where PPMA level was varied from 10 % to to 80 % by modifications of the cap material and thickness. PPMA is decreasing as the thickness of the FL is reducing (Figure 15a), which suggests its interfacial origin. Figure 15b demonstrates that device switching critical current J_{c0} follows the trend of PPMA.

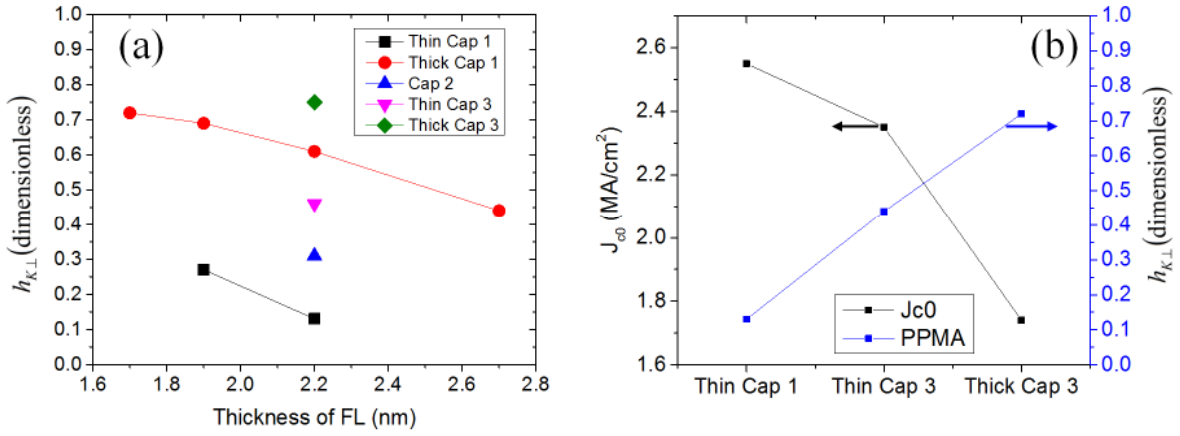

Figure 15. Experimental demonstration of PPMA due to different capping materials used (a). Reduction of J_{c0} by different capping materials (b).

Despite the fact, that PPMA was shown to be very effective to break STT-MRAM trilemma and reduce J_{c0} without changes to thermal stability or TMR, it has a few limitations. One of the limitation is that at large PPMA ($h_{K\perp} > 0.85$ for short axis of 50-60 nm), thermal stability starts to degrade. This is due to reduced demagnetizing field at edges of the FL that results in small out-of-plane tilting of the magnetization that respectively reduces in-plane component and shape anisotropy. In addition to that, even though large PPMA reduces J_{c0} and J_c at long pulses quite significantly, it becomes less effective in the precessional regime for t_p < 20 ns. This is due to a reduction of critical relaxation time τ that determines the switching dynamics in the precessional regime. Hence, effectiveness of PPMA depends heavily on target application of the STT-MRAM product, including cell dimensions, switching time and retention requirements and will likely be most applicable to dimensions above 40 nm.

6.2.2.2.b Perpendicular MTJ design (P-STT-MRAM)

In the previous section, we have been focusing on in-plane MTJ designs with equilibrium magnetization states in the plane of the free layer film. However, increasing PMA until it exceeds the thin-film demagnetizing energy results in a free layer magnetization with an easy axis perpendicular to the film plane, providing additional benefits. First, the switching current can be reduced, compared to in-plane magnetized free layers, for the same thermal stability (see eq. 3). Second, elongated bit shapes are not needed since the perpendicular anisotropy is an intrinsic property of the free layer material. This not only alleviates patterning and size control requirements but also helps to reduce the memory cell size, increasing the achievable array density in an STT-MRAM circuit. . Such devices are practical because interfacial perpendicular anisotropy as high as 2 mJ/m² has been demonstrated, enough to make a stable free layer with sub-10 nm dimensions. These benefits make perpendicular STT-MRAM most desirable for high-density (e.g. DRAM) applications.

The first observation of STT switching in perpendicular MTJ was achieved independently by Tohoku University and HGST in 2006 on GMR structures [97], [98], using Co/Ni, Co/Pt or FePt materials. Materials with high Z number, such as Pt or Pd introduce high spin-orbit interactions which increase the perpendicular anisotropy but also increase the free layer damping, which is highly undesirable. Even though the perpendicular configuration was expected to improve current density compared to in-plane devices, the high damping of these particular perpendicular materials seems to have offset the potential improvement. In addition to high damping, these PMA materials usually have 3-fold in-plane lattice symmetry, for example from FCC (111) textured film growth, which is not inherently compatible with the 4-fold symmetry of the (001) oriented MgO tunnel barrier needed for high TMR and high spin torque efficiency. The interfacial epitaxy problem can be mitigated by including amorphous CoFeB alloy in the free layer, separated from the polycrystalline layers by inserting another material, such as Ta, to break crystallographic coherence between the bulk-PMA material and CoFeB, but with a penalty in complexity and limitations on the free layer design.

The demonstration in 2010 of fully-perpendicular CoFeB-based free layers employing only the interfacial anisotropy at the magnetic metal/oxide interface [42] showed it was possible to combine large anisotropy with weak damping and good symmetry matching. Beginning with this demonstration in a simple structure with MgO on one side (Ta/CoFeB/MgO), [43], [55] this approach was further developed and widely adopted for perpendicular STT-MRAM development by researchers in industry and academia. Achieving a stable free layer in this system required the use of very thin CoFeB layers, which resulted in a significant damping increase compared to thicker layers. In addition, the achievable anisotropy strength of such free layers was found to be insufficient to provide adequate Δ for small (<30 nm) device dimensions. Both of these problems were solved by using free layers with double MgO barriers (e.g. MgO/CoFeB/MgO), which was demonstrated by Samsung [46] and Tohoku University [99]. This resulted in a doubling of the Δ and more than a twofold decrease of the damping (due to spin-pumping suppression). Higher anisotropy enables the scaling of pMTJ devices to smaller diameters while maintaining Δ needed for stable data storage. The switching efficiency figure of merit for p-STT-MRAM, defined as the ratio Δ/I_{c0} , has been shown to increase with reducing dimensions, probably due to multiple factors including reduced effective damping [47] and an increasingly coherent reversal process for smaller free layer diameter. The latter trend can be explained by different scalability of I_{c0} ($\sim D^2$) and Δ(~D), thus the figure of merit Δ/I_{c0} should increase at smaller dimensions as ~1/D with saturation below transition to quasi-uniform regime (described above). Similar behaviors of the figure of merit were reported in [57], [100]. STT switching has been demonstrated in devices with diameter as small as 11 nm [47] – the smallest dimension demonstrated to date (see Fig.16). At 11 nm dimension, the demonstrated thermal stability is $\Delta \approx 20$, which is too small for practical applications but demonstrates the potential for improved scaling. More work is needed to enhance the interfacial anisotropy to create a stable MTJ cell at sub-20 nm dimensions.

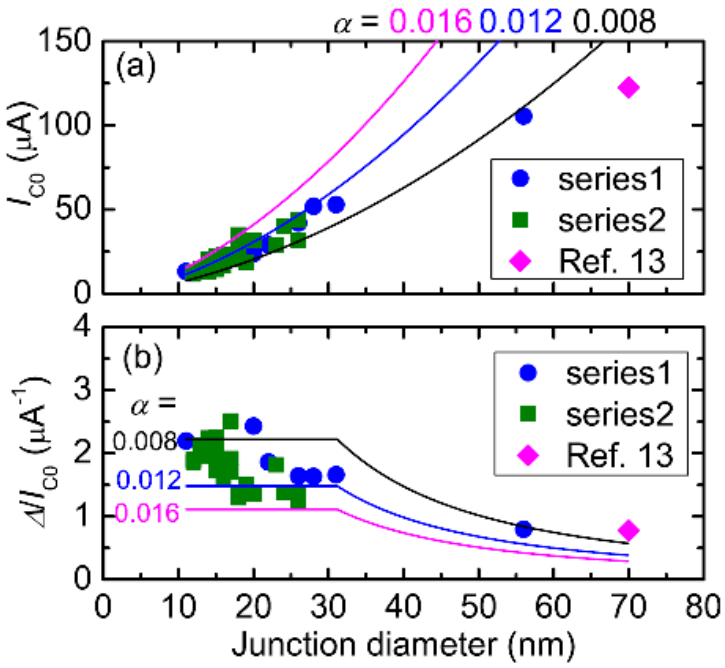

Figure 16. Critical switching current (a) and figure of merit (Δ/Ic0) (b) as a function of MTJ diameter. (Reprinted with permission from [47]).

Much work has been carried out, and much more is in progress, to increase the stability of I-PMA-type free layers by increasing the effective perpendicular anisotropy. Values of K_i as high as 1.9 mJ/m^2 have been demonstrated to date [101].

At sub-40 nm dimensions, process damage can play a dramatic role on MTJ behavior, as was shown by Samsung [102] with MTJ dimensions between 15 and 50 nm. An improved process with reduced side-wall damage allowed the improvement of Δ to 40 at 15 nm diameter.

Achieving high TMR with I-PMA-based FL is more challenging than for in-plane FL because, as previously mentioned, the effective anisotropy in I-PMA FL decreases with the FL thickness, setting an upper limit for the thickness of this layer at about 1.4nm for a single MgO interface and 2.0-2.8nm for double MgO interfaces. At this moment, TMR values above 350 % were achieved on out-of-plane magnetized MTJ, Figure 17, [101].

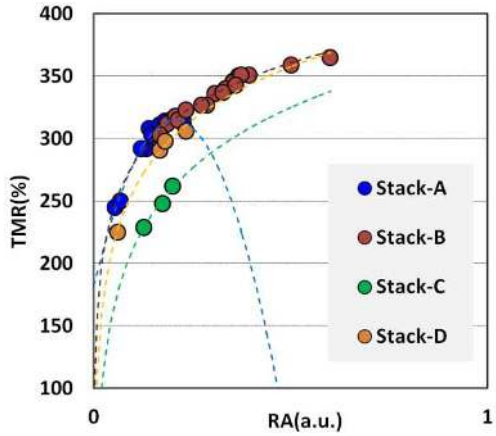

Figure 17. TMR of improved p-MTJ stack, from[101].

6.2.2.2.c MTJ structure with two MgO barriers: DMTJ

Another approach to increase STT efficiency is to introduce in the MTJ stack, on the interface of the FL opposite to the MgO barrier, a second MgO barrier together with a second reference layer magnetized antiparallel to the first reference layer, as shown in Figure 18. This creates a dual MTJ structure or DMTJ.

To explain the working principle of DMTJ, we can assume that the spin-transport across the DMTJ can be viewed as the transport across two single MTJ elements (second MTJ element has inverted FL/RL order) with shared free layer, Figure. 18. When a current goes through the first reference layer, the electrons become spin-polarized (SP) along the magnetization of this layer due to spin-dependent scattering. As this SP current enters the FL, the spin of transmitted electrons becomes repolarized along the magnetization of the FL. This results in spin-transfer torque Γ_1 acting on the magnetization of the FL. So far, the picture is the same as for a single MTJ structure. However, the spin-polarized current exiting the FL travels to the 2nd reference layer. Some of the electrons with spin direction antiparallel to magnetization of RL2 will get reflected towards the FL. As they enter the FL, these reflected electrons produce an additional spin transfer torque Γ_2 acting on the magnetization of the FL and adding to Γ_1 thereby increasing the total STT acting on the FL magnetization. If the two RL were in parallel configuration, Γ_2 would have opposite sign with respect to Γ_1 so that there would be a reduction of the net STT. Higher spin-torque is desirable for improved switching efficiency in MRAM, making the configuration with two antiparallel RL very interesting.

In similarity to PPMA effect, DMTJ does not change thermal stability of the free layer, thus addressing the write/store part of the STT-MRAM trilemma. However, with regards to reading the state of the memory, there is a difference between the two approaches since in a fully symmetric DMTJ structure there is no change in device resistance as the free layer is switched. This is due to the FL being parallel to one reference layer (low resistance) and antiparallel to the other reference layer (high resistance) regardless of the direction of the FL magnetization. This obviously makes it impossible to read the state of the device. To solve this issue, one of the junctions is made to have

higher resistance (RA) than the other, so the total resistance is dominated by the state of that single junction. This makes it possible to achieve almost the same level of TMR in DMTJ structure as compared to the single MTJ [103].

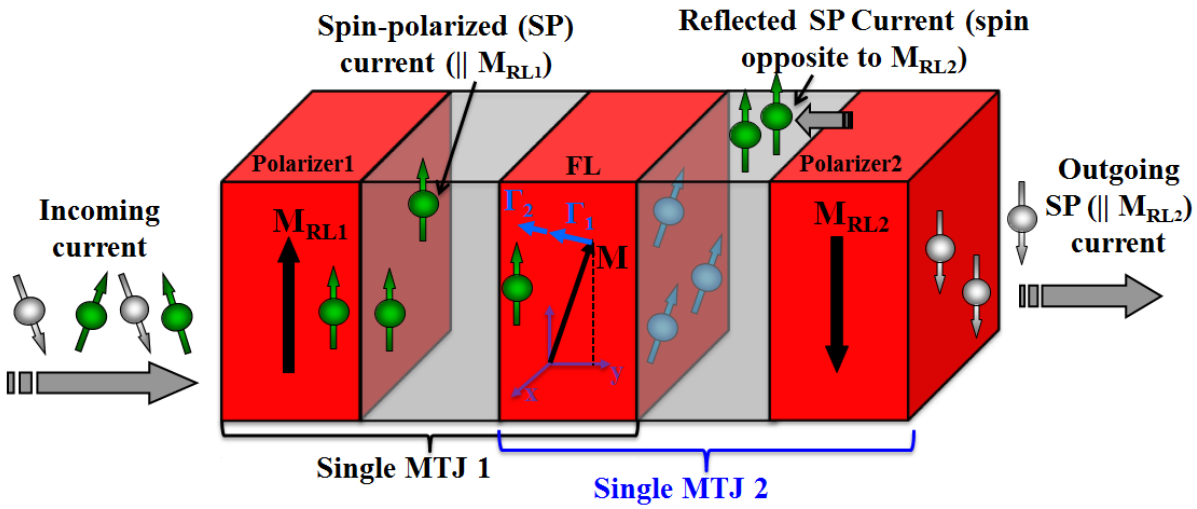

Figure 18: Schematic diagram showing how the two spin transfer torque contributions add in a dual MTJ design.

Another attractive attribute of the DMTJ design is its compatibility with I-PMA free layer designs. As described above, the presence of two MgO layers on the two FL surfaces can create a high level of PMA, providing additional energy barrier in p-MTJ FLs and a reduction of spin torque switching current in PPMA FLs. [43], [104], [105].

Experimentally, the first DMTJ structure with in-plane FL/RL was demonstrated in 2007 [106] and showed very low J_{c0} values of 1 MA/cm^2 on thermally-stable FL. It showed very symmetric switching currents between the two switching directions (AP to P and P to AP), in contrast to single MgO structures, which typically show asymmetry (ratio of P to AP current to AP to P current) of 2-3. DMTJ with perpendicular magnetization of free and reference layers was demonstrated by Samsung in 2014 [107] and IBM in 2015[108].

6.3. STT-MRAM: remaining challenges

6.3.1 Patterning process

One of the remaining challenges for STT-MRAM manufacturing is the patterning process. The fabrication of nanopillars of a certain size (down to less than 20 nm) with well-controlled dimensions, vertical sidewalls (> 80 degrees) and tight distributions is especially complicated since the MTJ involves a number of different materials: insulators, magnetic metals, non-magnetic metals, etc. A common approach to define the array of nanopillars involves deposition of a special hard-mask material and then removal of the material not covered by the hard mask to define a

two-dimensional array of MTJ cells. Two commonly used methods involve RIE (Reactive Ion Etching) and IBE (Ion Beam Etching) with both having its own advantages and challenges. For example, IBE etching method, commonly used for MTJ-reader fabrication in the hard drive industry, involves physical bombardment by ions and has the advantages of good control over the angles of the ion beam, low reactivity with the film and comparable etch rates across a wide range of materials. This method works quite well for larger dimensions and pitch between MTJ pillars but, for higher density arrays and high aspect ratio pillars, there is a limitation of the angles due to the shadowing effect (MTJ pillar being shadowed by its neighbors). The RIE methods involve both a chemical component to create volatile compounds from the etched material and a physical sputtering component. In principle, RIE methods are better suited to higher density arrays, but careful selection of the chemistry and process details are needed to equally remove all the different materials used in MTJ pillar while minimizing edge damage. Combination of the two methods also is used in a constant quest to improve the etching technique. As a recent example, a promising approach to improve the RIE patterning uniformity and edge roughness was proposed by using plasma ribbon beam etching (PRBE) [109]. In this method, a specially designed aperture is used to form a ribbon ion beam that scans through the 300 mm wafer with rotation of the wafer between the scans. The direction of the ions in the beam has an angular distribution tailored to clean the sidewall or trenches between the pillars. This process was shown to have promising results for good size control (from 100 nm to 20 nm) and small edge damage[109].

It has been demonstrated that improving the STT-MRAM fabrication process is important for reduced damage and improved cell properties, especially for very small dimensions. This is shown in Figure 19, which compares an improved MTJ process (process A) and standard process (process B). Not only the normalized TMR can be affected but also the dependence of thermal stability factor Δ or switching voltage on dimensions can be drastically changed by the process. Here, Δ was obtained by fitting the Hc distribution, as shown in the inset of 19d. The improved process results in much tighter distribution of resistances and better size control, as well, Figure 20.

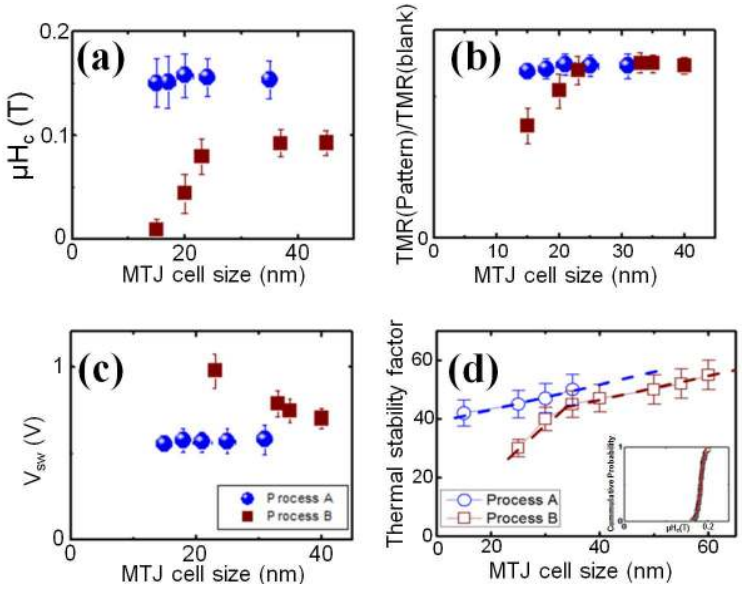

Figure 19. Comparison of various properties versus cell size between improved Process A versus conventional Process B (from [101]) (a) coercive field; (b) TMR amplitude in patterned sample normalized by TMR amplitude before patterning; (c) Switching voltage; (d) Thermal stability factor.

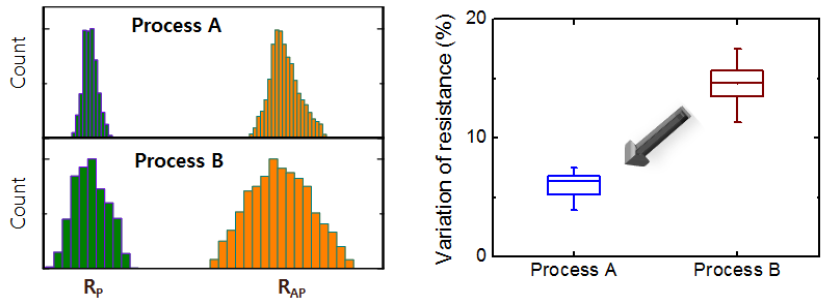

Figure 20. Improved Process A resulted in narrower MTJ cell resistance distribution (from [101]).

6.3.2. Switching probability: write and read error rates

In the discussion so far, when we were discussing about switching current density at a given pulse width, we implied average switching current density, corresponding to 50% of switching probability. For a successful memory product, this is clearly not enough: the entire memory array has to be repeatedly written with very low probability of error. For STT-MRAM, reliable writing and reading has been demonstrated a number of times. A good recent example of very low BER for writing on fully-perpendicular STT-MRAM is shown in Figure 21 on a single MTJ cell and on a set of 256 randomly selected cells (Fig. 21b)[110]. The test did not show any errors at the largest voltage applied (marked by star on Figure 21a – calculated as inverse of number of applied pulses). This result was achieved for both directions of the switching (positive and negative voltages).

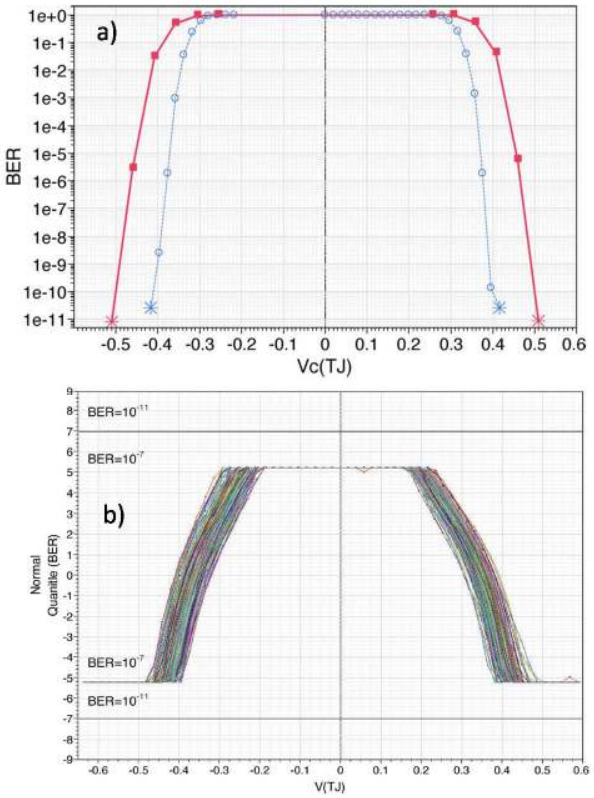

Fig. 21. (a) - Demonstration of ultra-low write error rate (BER) for 10 ns (red, square) and 50 ns writing pulses. Horizontal axis is voltage on MTJ (TJ) cell, excluding voltage drop on the transistor. (Reprinted with permission from[110]. (b) - Normal quantile plot of bit error rate for 256 randomly selected cells. Each of the electrically functioning cells (3 cells were defective and excluded) is shown to reach the requested error floor of 1e-7 (limited by test time) for both reading (NQ~5) and writing (NQ ~ -5).

6.3.3. Read disturb

In addition to being able to switch reliably, another important characteristic of STT-MRAM is being able to read without disturbing (accidentally switching) the state of the free layer. For a read current I_{read} one can derive the following expression for the probability of switching a cell in STT-MRAM chip [81]:

$$P_{RD}(I_{read}) = 1 - \exp\left(\frac{-N_{read} t_{tot}}{\tau_0 \exp\left(\Delta \left(1 - \frac{I_{read}}{I_{c0}} \right)^{\xi} \right)} \right), \quad (17)$$

where ξ =1 (in-plane magnetized MTJ) and ξ =2 (out-of-plane magnetized MTJ) [78], [84], [111], N_{read} is the number of cells which are read in parallel in one single read operation (e.g. 16kbits for DRAM-compatible specifications [112]) and t_{tot} is the total duration of all read operations during the specified chip operating time (up to 10 years for nonvolatile memory applications). Required read disturb level (eq. (17) depends on a specific application and is of the order of 10^{-9} before ECC correction and 10^{-18} for applications without ECC correction, which directly translates into minimum thermal stability for a given read pulse and read current. Characterization of read disturb and write error rates can be conveniently done on a single graph by plotting normal quantile of BER. Thus, read disturb corresponds to positive normal quantile (probability of switching < 50 %) and write error rate corresponds to negative normal quantile (probability > 50 %), Figure 21b.

6.3.4. Long-term data retention

Another important characteristic of any memory including STT-MRAM is how long the data can be retained after it is written (retention). STT-MRAM is a nonvolatile memory (i.e. it can keep the information without being electrically powered) and the retention time can be as large as 10 years. Expression 17 can be conveniently used to calculate the probability of thermally-activated data loss after time t by setting read current to zero:

$$P_{th} = 1 - \exp\left(\frac{-t}{\tau_0 \exp(\Delta)} \right) \quad (18)$$

From this expression, we can see that data retention properties are determined dominantly by the thermal stability factor. One convenient characteristic of a memory chip related to data retention is the required specification on maximum number of Failures in Time (FIT):

$$P_{th} = \frac{\mathrm{FIT}}{\mathrm{N_B}} \quad (19)$$

where N_B is the total number of bits in the chip. From the above expressions, a requirement is set on the thermal stability factor measured at room temperature as a function of retention time (t), number of bits (N_B) and maximum number of Failures in Time (FIT) (see Table 1):

$$\Delta = -Ln\left[-\frac{\tau_0}{t} Ln\left(1 - \frac{FIT}{N_B} \right) \right] \quad (20)$$

Capacity	1000 FIT @ 80C	0.1 FIT @80C	0.1 FIT @ 160C
16 Mb	70	81	99
256 Mb	73	84	103
1 Gb	75	86	105
4 Gb	76	87	107
8 Gb	77	88	108

Table 1. Required room-temperature values of thermal stability for 10 year data retention at different operating temperatures. Bit to bit distribution is not taken into account.

For practical considerations, Δ of 70-90 is usually quoted as typical minimum thermal stability for 10-year data retention. This depends on cell-to-cell distribution, operation temperature, sensitivity of FL properties to temperature, etc. Existing I-PMA-based perpendicular free layer materials provide level of anisotropy large enough for Δ=100 at 14 nm diameter of the free layer. Experimentally, values as high as 120 were demonstrated in perpendicular MTJ cell at 30 nm diameter, Figure 22. Going down below 10 nm may require using B-PMA material and special development to mitigate damping increase usually associated with B-PMA materials. Fundamentally, thermally stable bits down to 5-7 nm (size of the grain) are achievable using B-PMA as used in magnetic media in hard disk drive technology.

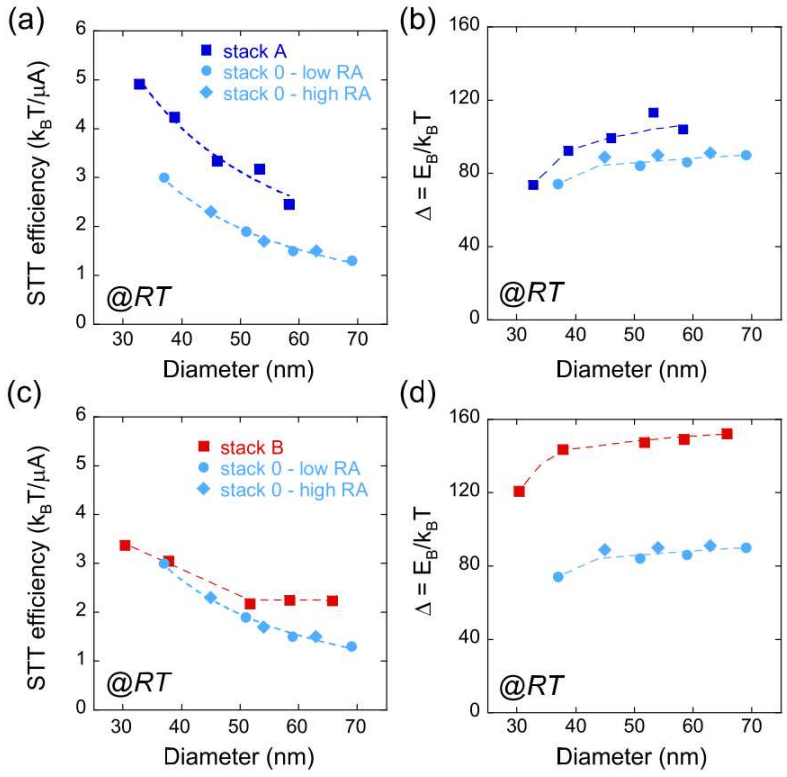

Figure 22. STT efficiency (a, c) and thermal stability factor Δ (b,d) as a function of diameter for perpendicular MTJ cell. Several designs are considered: stack A (high Figure of Merit Δ/I_c stack), stack B (high Δ design). Reprinted with permission from [57].

6.3.5. MTJ endurance and breakdown

Another important characteristic of a memory is its endurance – the number of times a particular memory cell can be read or written before irreversible degradation occurs. It can vary widely between different types of memories from virtually infinite number of cycles (for DRAM, SRAM) to 10^3-10^5 for a flash memory. The simplest measurement of the bit endurance is typically done at increased voltage and projected to operating voltage. For a given voltage, a cell is constantly written and read until it breaks down. Then, this procedure is repeated on the next cell and so on until enough statistical data are obtained. Then, the same test is repeated at increasing voltage as shown in Figure 23. The projection to operating switching voltage gives the expected single-bit median cycling endurance equivalent to 10^{20}. This means that after 10^{20} cycles, half of the bits in the memory chip are shorted.

The cycling endurance of an MRAM product depends on the allowable error rate from shorted bits, which could be in the 10^{-6} to 10^{-18} range depending on many factors. In addition, the write voltage in a circuit must be set far above the mean switching voltage to ensure a low write error rate. These practical considerations make the prediction of endurance in a product quite complicated, but it is clearly many orders of magnitude less than the mean time-to-fail of the devices.

Electron trapping/detrapping mechanism in the tunnel barrier was shown to play a key role in the early breakdown of MTJs and a correlation between endurance and 1/f electrical noise of MTJs has been demonstrated [113], [114].

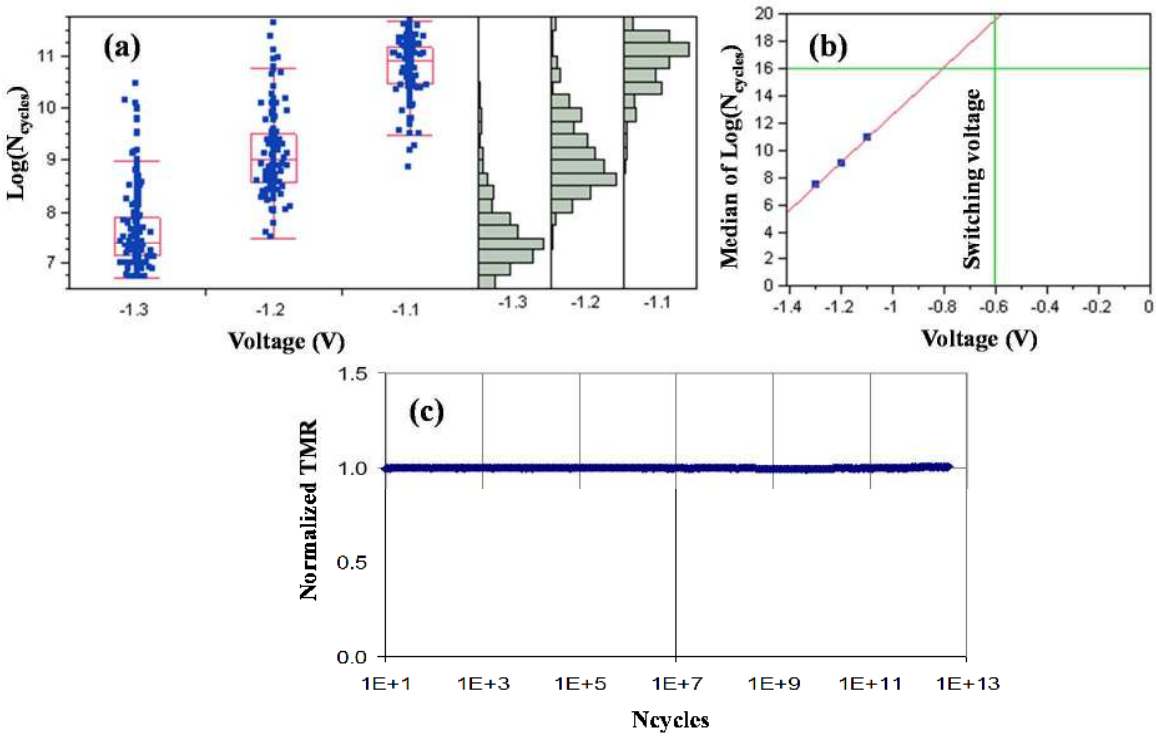

Figure 23. a) Number of write pulses to breakdown of an MTJ as a function of the pulse amplitude. Each point represents the number of pulses (cycles) to failure of a single MTJ. The within-array distribution is evident in the histogram plots shown in the three panels on the right. b) Median cycles to fail extrapolated to 1e20 cycles at V_{sw}, meaning that half of the bits fail at the switching voltage. c) Normalized TMR as a function of write cycles showing negligible degradation over a wide range prior to eventual breakdown. [81].

6.4. STT-MRAM chip demonstrations

STT-MRAM chip development and demonstrations have advanced significantly in the recent years. Two important cases are worth noting for both in-plane and perpendicular STT-MRAM. Everspin Technologies [3] demonstrated a 64Mbit chip based on 90 nm CMOS technology with in-plane magnetization. The size of the MTJ bits was 80-90 nm with aspect ratio of 2-3. MgO barrier with RA of 5-10 $\Omega*\mu m^2$ and TMR above 110 % was used. The distribution of switching voltage (1 σ) was shown to be < 10 % with separation between the breakdown and switching voltage of more than 25 σ. Sequential data rate was 1.6 GT/s, corresponding to DDR3-1600 specifications. Error-free writing of the whole chip was achieved for $5x10^5$ cycles (limited by testing time). Operation within range from 0^0C to 70^0C was shown without significant changes.

For perpendicular STT-MRAM 8Mbit chip for embedded application was demonstrated by TDK/Headway [115] using 90 nm CMOS technology and 1T-1MTJ design (single transistor per MTJ). The free layer was I-PMA-based: MgO/CoFeB-based FL/MgO to increase the stability of the FL by having interfacial anisotropy on both surfaces of the free layer. The cell size was $50F^2$ with diameter of the STT-MRAM nanopillar of 50 nm. Both writing and reading times were shown to be below 5 ns with data retention at 125^0C up to 10 years. Number of defected cells (mostly partial-shorts) was shown to be less than 10 ppm (parts per million), which is low enough to be corrected by redundancy. Reliable reading (up to 10^4 pulses) was shown in the worst-case scenario (largest read current at 125^0C). For embedded memory, STT-MRAM chip has to sustain BEOL temperature of 400°C for 30-90 mins. This is not trivial and involves optimization of the materials not only for the free layer [116] but also for reference layer[117] to prevent degradation of MTJ properties by interdiffusion. This STT-MRAM chip was shown to sustain 400°C for up to 90 min without degradation, which satisfies this requirement (Fig. 24).

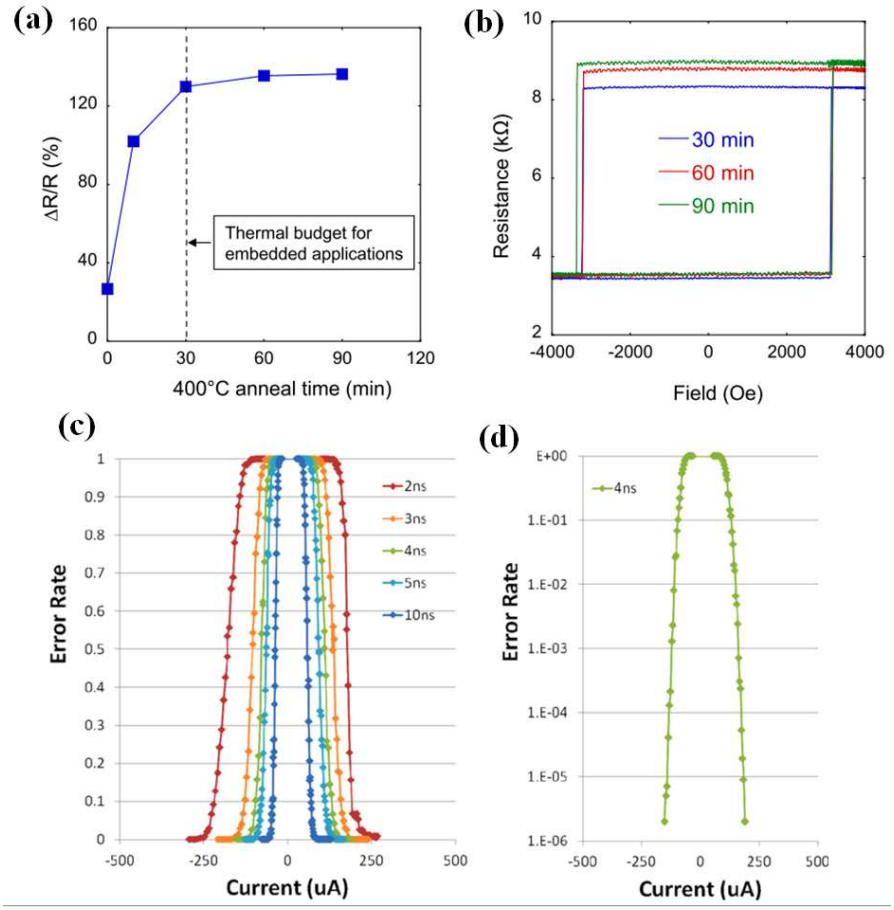

Figure 24: Data from TDK/Headway 8Mbit STT-MRAM chip (a) TMR ratio as a function of anneal time at 400^0C. Data points show the median value of more than 140 devices with a median diameter of 80 nm. (b) Hysteresis loop after various anneal times. (c, d) Bit error rate as a function of current for a 45-nm diameter device for switching pulse from 2 ns up to 10 ns. Reprinted from [57] with permission.

These two demonstrations are very important for feasibility of STT-MRAM chips but some important questions remain. One of them is the reliability of writing and reading down to very small (less than 10^{-9}) error rates. These ultra-low read and write error rates were shown in a 4kbit chip with perpendicular MTJ by IBM (see Fig.21) [110]. Very thin I-PMA (Ta/CoFeB/MgO) free layer was used, which provides better switching properties for fast (<20 ns) switching. For both directions of writing, it was shown that WER and RER are smaller than 10^{-11} for both 10 and 50 ns pulse width. The writing voltage at WER=10^{-11} was 0.5 V for 10 ns, which is in a range of voltages a typical CMOS transistor can provide, (Fig.21a,b).

One remaining challenge is achieving very fast switching speed (ideally < 1 ns) using spin transfer torque switching in perpendicular MTJ. From Fig. 24, one can see that there is a significant increase of writing current as pulse width is reduced from 10 ns to 2 ns. There has been a number of solutions proposed to alleviate this, most of them are related to increasing the initial value of the spin transfer torque and prevent long build-up of slow increase

of precession amplitude: orthogonal spin torque design[118], [119], free layer with easy-cone anisotropy [120]–[124], spin Hall effect (SHE) and spin-orbit torque switching (SOT)[13]–[17].

6.5 STT-MRAM: conclusions and outlook

In recent years, several important discoveries have been made that are enabling the creation of a new generation of magnetoresistive memory. These discoveries include prediction and demonstration of spin-transfer torque, demonstration of high TMR with MgO-based magnetic tunnel junctions, and the demonstration of high interfacial perpendicular anisotropy at the CoFeB/MgO interface. The first makes it possible to have a very efficient and fast writing scheme, the second enables fast and reliable reading, and the third opens up the possibility to scale down the size of the MTJ storage devices to very small dimensions. The STT-MRAM memory based on the foundation established by these effects has been shown to have fast read/write functionality, high density, and high reliability. The feasibility of PMA-based STT-MRAM technology at 30 nm dimension is summarized in table 2, based on available published data. Even though it was demonstrated by multiple groups that all the important parameters can be improved to achieve required values, more work is needed to demonstrate all of them on the same chip, Once that is achieved, it will allow STT-MRAM production to move from serving specialty markets to high-volume mass production serving broad markets. Thanks to its characteristics of high speed, non-volatility, and high endurance, STT-MRAM opens up unique possibilities for implementation in completely new applications not envisioned before, possibly creating new markets.

Parameter	**Typical requirements**	**Demonstrated**	**References**
MTJ diameter (nm)	< 30	<30	[57]
Delta	> 85	75-120	[57]
Jc0 (MA/cm2)	< 2.0	2.1-5.1	[57]
TMR (%)	> 200	> 300	[101]
WER	< 1e-9	< 1e-10	[110]
TDDB (cycles to median fail)	> 1e15 (chip)	> 1e16 (single-bit)	[81]
Post-annealing (BEOL)	400C, >30 min	400C, >30 min	[57]

Table 2. Summary of key parameters that indicate feasibility of PMA-based STT-MRAM at advanced technology nodes.

7) Thermally-Assisted MRAM (TA-MRAM)

7.1. Dilemma between retention and writability / benefit from thermally-assisted writing

Any memory can be viewed as a two-state system separated by an energy barrier ΔE. The retention of the memory is directly determined by the barrier height ΔE (expression 18). The higher ΔE the more stable the information i.e. the longer the memory retention. However, correlatively, the higher ΔE, the more difficult it is to switch from one state to the other i.e. to write the memory state. There is therefore a classical dilemma in memory technology between retention and writability.

Thermal assistance during write allows circumventing this dilemma. The general concept of thermally assisted writing consists in storing the information at a standby temperature at which the barrier ΔE is quite high, then temporarily increasing the temperature of the storage element during each write event to reduce the barrier height and ease the switching between the two memory states. After switching, the memory element cools down and recovers its high stability to thermal reversal.

In MRAM, this concept can be easily implemented since, in magnetic materials, the magnetic anisotropy which provides the thermal stability of the magnetization decreases with temperature so that the barrier height for switching (ΔE) decreases with temperature.

7.2. Heating in MTJ due to tunneling current

In MRAM, the heating of the storage layer can be produced in a simple way by taking advantage of the Joule dissipation around the tunnel barrier. In an MTJ, the heating is due to the inelastic relaxation of tunneling hot electrons which takes place when the tunneling hot electrons lose their excess energy while penetrating the receiving electrode. The heating power per unit area is $P = jV$ where j is the current density flowing through the tunnel barrier and V the bias voltage across the barrier. This power is released in the first nanometer of the magnetic receiving electrode in contact with the tunnel barrier. In MTJs having an RA product of the order of 30 Ω.·µm², with heating current density of the order of 10^6 A/cm², a rise in temperature ΔT of the order of 200 °C within 5 ns is typically achieved [125]

7.3. In-plane TA-MRAM

7.3.1. Write selectivity achieved by a combination of heating and field

For low density memory applications such as microcontrollers, requiring high thermal stability, in-plane magnetized MTJ with an exchange biased storage layer can be used (Fig.25.a,b). Exchange bias is a mean to pin the

magnetization of a ferromagnetic layer (F) by coupling it to an adjacent antiferromagnetic layer (AF). As long as the temperature is below the so-called blocking temperature of the antiferromagnet, the magnetization of the ferromagnet is frozen in a fixed direction. Upon heating above the AF blocking temperature, the magnetization of the F layer becomes unblocked which enables its switching with a pulse of magnetic field. The write procedure thus requires heating above the AF blocking temperature (typically in 5 to 10ns) and cooling in the presence of a magnetic field (cooling rate of the order of 7 to 20ns, depending on the heating pulse duration and exact composition of the MTJ stack). In TA-MRAM, both the reference and the storage layer are exchange biased with antiferromagnetic layers but those have quite different blocking temperatures. Typically PtMn with blocking temperature of 350°C is used in the reference layer whereas the storage layer antiferromagnet is chosen with a blocking temperature in the range 180°C-250°C depending on the requirements on the device operating temperature range (Fig.25.b). The main advantages of this write scheme combining heating and field pulses are: (1) the use of a single field selection line instead of two, as for toggle writing, (2) an important reduction in write power consumption due to low writing field and the possibility of sharing the field pulses between numerous bits as will be explained later, (3) the realization of cells with high thermal stability, i.e., much improved retention due to the exchange pinning of the storage layer.

A bit write sequence starts from a given initial orientation of the magnetization of the exchange biased storage layer, for instance representing a low resistance state "0". In this initial state, the corresponding storage layer loop is then shifted around a negative field, as seen in Fig. 25 d in the hysteresis cycle before the heating pulse is applied (red curve). The reversal of the storage layer bias is achieved by heating the AF layer above its blocking temperature with a current pulse (Fig.25c) and applying simultaneously an external magnetic field H_{sw} larger than the coercive field of the storage layer at this elevated temperature. The field is applied in the direction parallel or antiparallel to the reference layer magnetization depending whether a “0” or a “1” has to be written. In the example of Fig. 25d, a “1” is written. The heating current pulse is then stopped so that the storage layer cools in the applied magnetic field. This maintains the storage layer magnetization in the field-cooling direction during its freezing. This results in the reversal of the pinning orientation of the storage layer magnetization and correlatively a bit state change to a high resistance "1". As a result the storage layer loop is now shifted towards positive values as shown in Fig. 25d (blue curve).

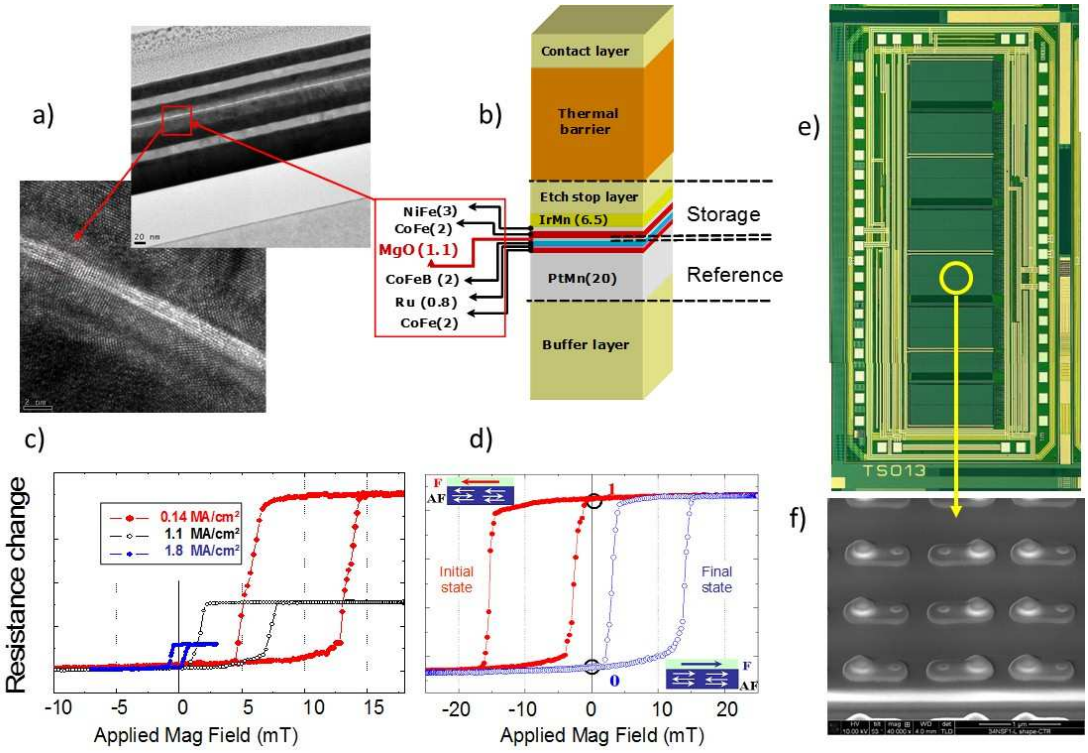

Figure 25: TA-MRAM: a) cross-sectional transmission electron microscopy images of the MTJs used in TA-MRAM. b) Typical composition of the stack. c) Evolution of the hysteresis loops of the storage layer as a function of heating current density showing the decrease in the pinning as the heating current increases. At the highest heating current density, the loop is centered, meaning that the exchange bias has vanished and low fields of a few millitesla are then sufficient to switch the storage layer magnetization. d) Example of switching from a "1" state to a "0" state: the storage layer loop shift measured at standby temperature is inverted between the two states. Note that in standby, thanks to the loop shift, only one state is stable at remanence (i.e. at zero field) which means that even if the cell is exposed to a perturbation field, it will come back to its original state once the perturbation disappears. e) Fully functional 1Mbit TA-MRAM demonstrators produced by Crocus Technology in collaboration with Tower Jazz. f) detail of the patterned MTJs (circular shape) connected by a strap to the vias emerging from the CMOS wafers (130nm technology).

Several demonstrations of TA-MRAM were realized first in micrometer-size junctions (2 µm × 2 µm) under DC currents [126] then on submicrometer cells using heating pulses down to 10 ns [127]. The company Crocus Technology in collaboration with TowerJazz has produced fully functional 1 Mbit demonstrators using this technology (Fig.25.e-f).

7.3.2. Reduced power consumption thanks to low write field and field sharing

In conventional TA-MRAM, the write selectivity is achieved by a combination of temporary heating of the storage layer which lowers its magnetic pinning energy, together with the application of a pulse of magnetic field. The excellent selectivity achieved by this write scheme is illustrated in Fig. 26. Repeated successful writing can be achieved only when heating and field pulse are combined.

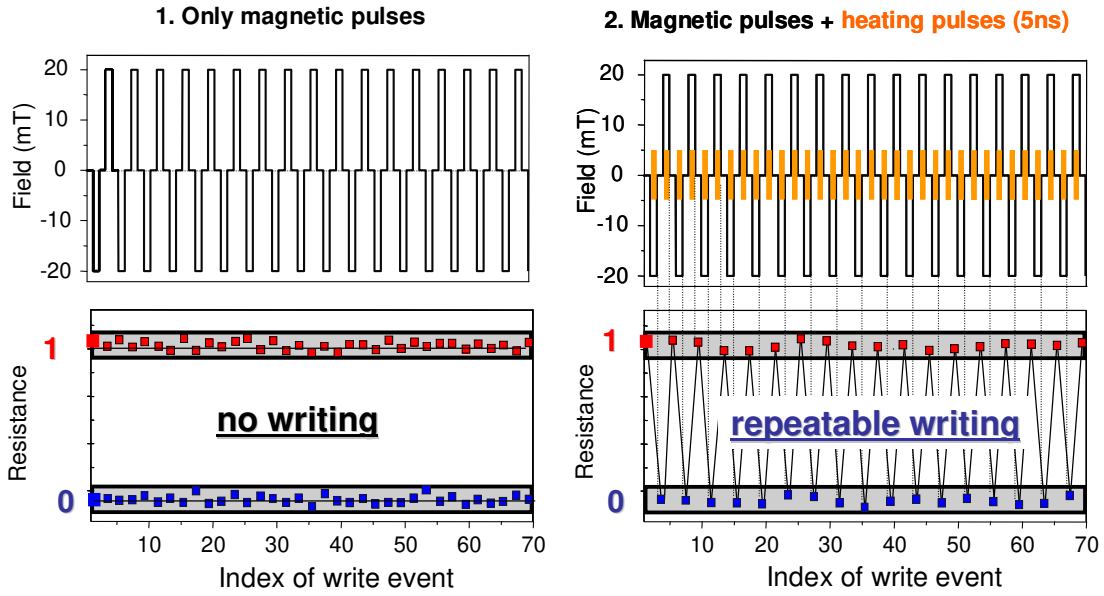

Figure 26: Illustration of the write selectivity in TA-MRAM resulting from the combination of heating pulse and application of a magnetic field. Left: Application of pulses of magnetic field alone (no writing). Right: combination of heating pulse (5ns long) and alternating pulses of magnetic field (repeatable writing of P and AP (0 and 1) magnetic states).

Another advantage of this write scheme is the protection against field erasure in standby. As seen in Fig. 25d in standby, only one state of the storage layer is stable at zero field. This means that even if the TA-MRAM chip is exposed to a perturbation field, this field may temporary switch the magnetic configuration of the memory but the latter will spontaneously return back to its original state once the perturbation disappears. This however would not be true during write. If a perturbation field representing a significant fraction of the write field is applied during write, a write error may occur. A read-after-write scheme may be used to verify that no such error has occurred during write or correct it if needed.

Field writing combined with thermal assistance is also quite advantageous in terms of power consumption in MRAM chips by offering the possibility to share the pulse of magnetic field among numerous bits. Indeed, in a RAM, the bits are usually written and read word by word and not bit per bit. Each word may contain 32 or 64 bits. To write a full word in TA-MRAM, only two pulses of magnetic field are required. The scheme for writing a word is illustrated in Fig. 27. In a first step (Fig.27a), all bits in the word which have to be written to “0” are heated simultaneously by

sending a heating current through the corresponding MTJs. The “0” field is then applied, the heating currents are switched-off, so that the heated cells cool down in the “0” magnetic field and freeze in the “0” configuration. The cells which are not heated do not switch so that they remain in their original state. In a second step, all bits which have to be written to “1” are heated simultaneously and the “1” magnetic field is applied (Fig. 27b). The heating currents are then stopped while the “1” field is still on so that the heated cells now freeze in the “1” configuration. At the end, the whole word has been written with 32 pulses of heating current (assuming a 32-bit word) --- the corresponding energy is ~0.7pJ per dot at 90nm technology --- plus two pulses of magnetic field (each pulse requiring ~30pJ which represents only 0.5pJ per bit for 64bits words).

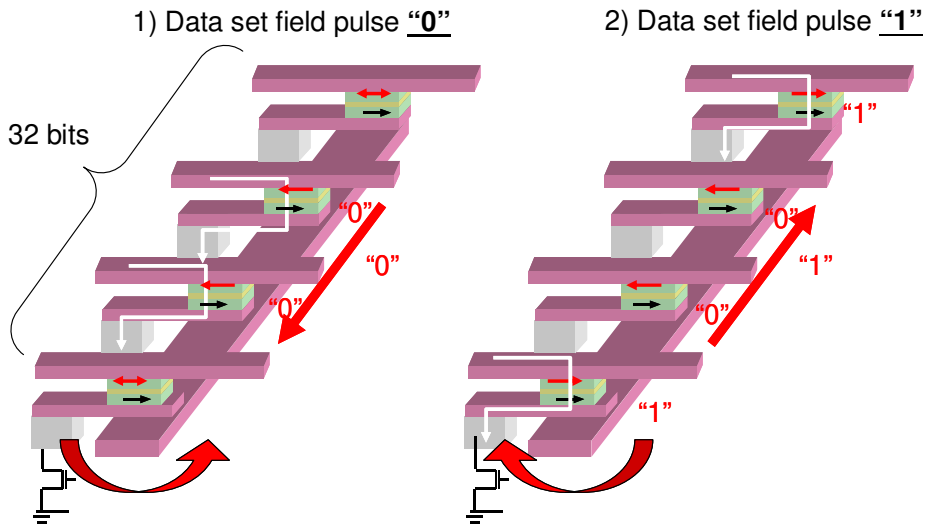

Figure 27: Illustration of write scheme with field sharing in TA-MRAM. A whole word can be written with only two pulses of magnetic fields.

Furthermore, in TA-MRAM, since the thermal stability of the storage layer magnetization in standby is provided by exchange coupling to an adjacent antiferromagnetic layer, the cell can have a circular shape. As a result, once the writing is enabled by heating the cell above the blocking temperature of the antiferromagnetic layer, only a low field of a few milliTesla (3 to 5 mT) is sufficient to switch the storage layer magnetization since the cell has no shape anisotropy.

7.4. TA-MRAM with soft reference: Magnetic logic unit (MLU)

In a second possible implementation of TA-MRAM, the reference layer is no longer pinned but is made of a soft magnetic material with easily switchable magnetization [128] The storage layer is exchange biased by an adjacent

antiferromagnetic layer as in standard TA-MRAM (see Fig.28). The writing of the storage layer is achieved similarly to the previous case by the simultaneous application of an external field and a heating pulse that allows the storage layer to be switched and then pinned as the system cools down below the antiferromagnetic blocking temperature. In contrast, the magnetization of the soft reference layer (SR) switches as soon as the field is applied independently of the fact that the cell is heated or not. This implementation allows a self-referenced reading process: The readout is performed in two steps: the SR magnetization is set in a first predetermined direction, and the MTJ resistance is measured. The SR magnetization is then switched to the opposite direction and the new resistance is measured. The resistance variation between the two measurements allows the determination of the magnetic orientation of the storage layer. In this approach, the read cycle is longer (~50 ns) but the tolerance to process variation is greatly enhanced since each bit is self-referenced. Indeed, in the standard reading scheme, the two resistance state distributions have to be well separated in order to avoid read errors. This requires narrow distributions in dots size and shape. In contrast, for the self-referenced reading scheme, the difference of resistance between the two states is used to read the junction, and is thus not sensitive to dot to dot variability. This is particularly useful for advanced technological nodes where it becomes increasingly difficult to control accurately the resistance distributions.

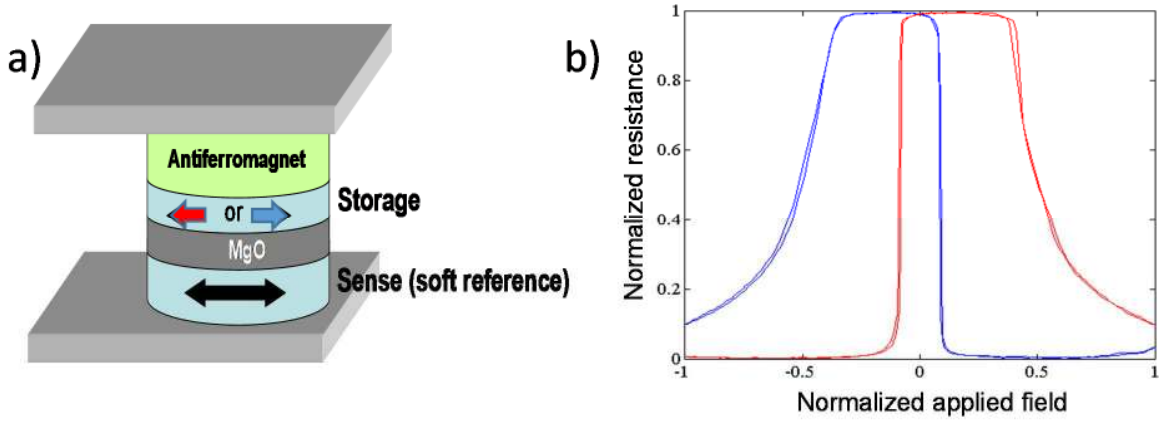

Figure 28: a) Schematic representation of a self-referenced MRAM with thermally assisted write. b) Normalized resistance response as a function of the external field for a self-referenced MRAM at standby temperature in state '0' (blue) and '1' (red). The upward or downward resistance transition at low field is associated with the switching of the soft reference layer magnetization and is used to readout the magnetic state of the storage layer (from Crocus-Technology).

Besides its MRAM application, this self-referenced cell can be used to perform logic operations with particularly interesting applications in security. Indeed, these devices combine memory and logic (XOR) functions. For this reason, they are called Magnetic Logic Units (MLU™)[128]. In addition to being storage devices, self-referenced MRAM intrinsically allows performing logic comparison functions. Considering the orientations of the storage and self-reference layer magnetization as two inputs, this magnetic stack outputs the exclusive OR function of these logic inputs through its resistance value. Both inputs are set using pulses of magnetic field, with first the storage layer being written with a combined heating pulse and then the SR layer switched at room temperature. By using a set of self-referenced MRAM cells connected in series to form a NAND chain, a Match-In-Place™ engine can be created[4], as

illustrated in Fig. 29, wherein a pattern applied to the SR layers is compared to the written pattern stored in the storage layers. Any mismatch between the two patterns systematically results in a higher resistance than for a perfect match. The expected advantages of this architecture are:

- The stored patterns are never read, and never exposed to potential hackers.
- The matching cycles are very quick and could be orders of magnitude faster than existing methods and use less power.
- Match-In-Place engines could act as a hardware accelerator, and simplify the overall chip.

The basic Match-In-Place architecture can be built at three levels: The single cell for direct matching at the bit level, NAND chains that combine multiple individual cells in series for linear Match-In-Place engines as illustrated in Fig.29, and a matrix that combine multiple NAND chains in parallel to match one input pattern to a stack of stored patterns. In this last case, Content Addressable Memories (CAM) can be realized based on this principle.

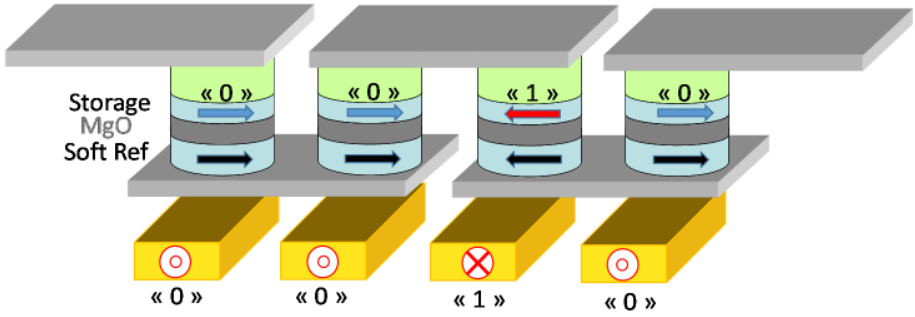

Figure 29: Example of a Match-In-Place™ engine made by the serial connection of 4 MLU in a NAND chain. The input binary pattern '0010' is compared to the stored binary pattern '0010'. Since this corresponds here to a perfect match, all resistances are at their minimum value so that the chain is at its minimum resistance value which indicates a perfect match. If one or several inputs do not correspond to the stored data, the chain resistance departs from this minimum value indicating a mismatch[4].

7.5. TA-MRAM with soft reference: Multilevel storage

A nice feature of the TA-MRAM approach with in-plane exchange biased storage layer is that it allows realizing multilevel storage using cylindrical MTJ cells. Indeed, the pinning of the storage layer magnetization can be achieved in any in-plane direction provided the write field can itself be applied in any in-plane direction during the cooling of the cell. This can be achieved with two orthogonal field lines as in toggle MRAM. The read-out is then performed by applying a rotating field with the same two current lines (Fig.30). The associated rotation of the magnetization of the self-reference layer then generates an oscillation of resistance since in MTJ, the conductance varies as the cosine of the angle between the magnetizations of the two magnetic layers sandwiching the tunnel barrier. The memory cell

output is then given by the phase of the resistance oscillation. Fig.29b illustrates the storing of 16 different states per cell representing 4 bits per cell in an MTJ of diameter 110nm [129]

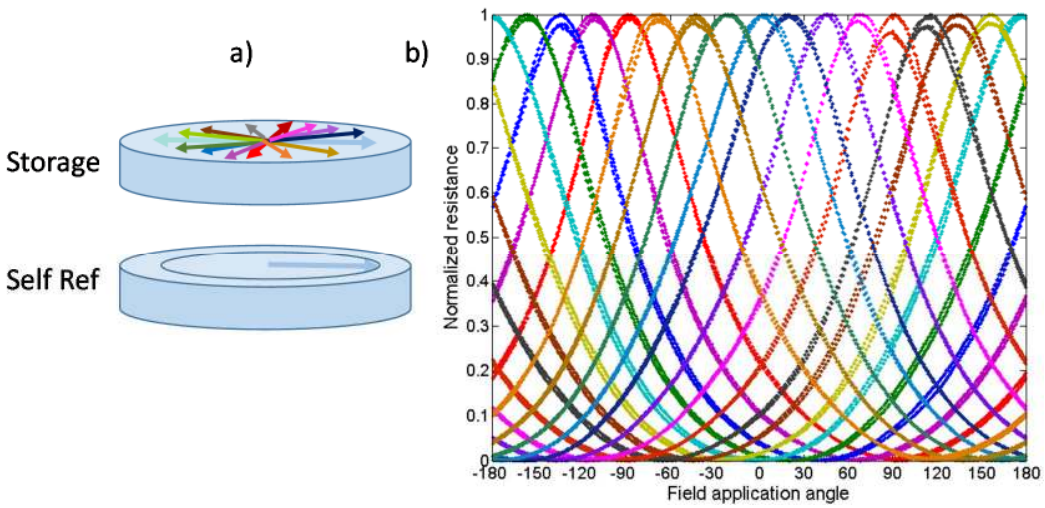

Figure 30: Illustration of the multilevel storage principle in TA-MRAM with soft reference layer. a) sketch of the storage layer and soft reference layer. The arrows of different colors represent the magnetization of the storage layer pinned in different directions (16 here) to achieve 4 bits per cell storage. The self-reference layer magnetization can easily rotate in-plane by applying a rotating field. b) Oscillatory voltage measured across the MTJ under the rotating field during readout. The different curves correspond to the different in-plane orientations of the storage layer pinned magnetization with corresponding colors between Fig.30.a and b.

This multilevel approach offers an interesting approach to increase the storage density when it becomes more difficult to reduce the memory dots size.

7.6 Thermally assisted STT-MRAM

So far in section 7, MRAM using a write scheme based on thermal assistance combined with pulses of magnetic field has been described. Alternatively, thermal assistance can also be used in combination with spin transfer torque to reduce the write current in STT-MRAM as well as to extend their downsize scalability. This is explained in the following section in the case of out-of-plane magnetized MTJs.

In STT writing, the write current which produces the spin-transfer-torque on the storage layer magnetization traverses the tunnel barrier. For RA product of the tunnel barrier around 10 to 20 Ω.µm², this STT write current also produces a significant heating of the MTJ. When the current is switched on, the temperature starts rising in the MTJ and at typical current density of 2.10^6A/cm² used in STT-MRAM, can reach ~200°C in about 5 to 10ns. This temperature rise can be advantageously used to assist the switching of the storage layer magnetization thanks to the phenomenon of thermally induced anisotropy reorientation (TIAR) [130], [131] Indeed, the magnetization of an out-of-plane magnetized layer at room temperature can reorient in the thin film plane when heated due to the different

temperature dependence of the out-of-plane anisotropy and demagnetizing energy [132]. TIAR assisted switching allows decreasing the STT critical current thanks to an optimization of the temperature dependence of the MTJ magnetic properties while keeping a satisfying thermal stability under standby conditions.

Figure 31 describes this writing scheme[131]: the MTJ exhibits a large thermal stability factor in standby over the whole operating temperature window (for instance -20°C-85°C). During write, a current pulse is sent through the junction. Heating occurs leading to a decrease in the PMA. As a result, the free layer magnetization falls into the thin film plane while the reference layer is engineered so as to keep its out-of-plane magnetization. This latter electrode polarizes the current in the out-of-plane direction. The STT, due to this spin-polarized current, pulls the free-layer magnetization in the upper or lower hemispheres depending on the current direction and induces large angle out-of-plane precessions[133]. In such a configuration, the STT effect is highest since the spin polarization is almost perpendicular to the magnetization. It results in a more reliable switching since no thermal fluctuations are required to initiate the reversal. The injected current is then gradually decreased so that the junction cools while STT is still effective. The free layer recovers its PMA, and its magnetization gets reoriented out-of-plane, in the direction defined by the hemisphere into which it was previously pulled into by STT.

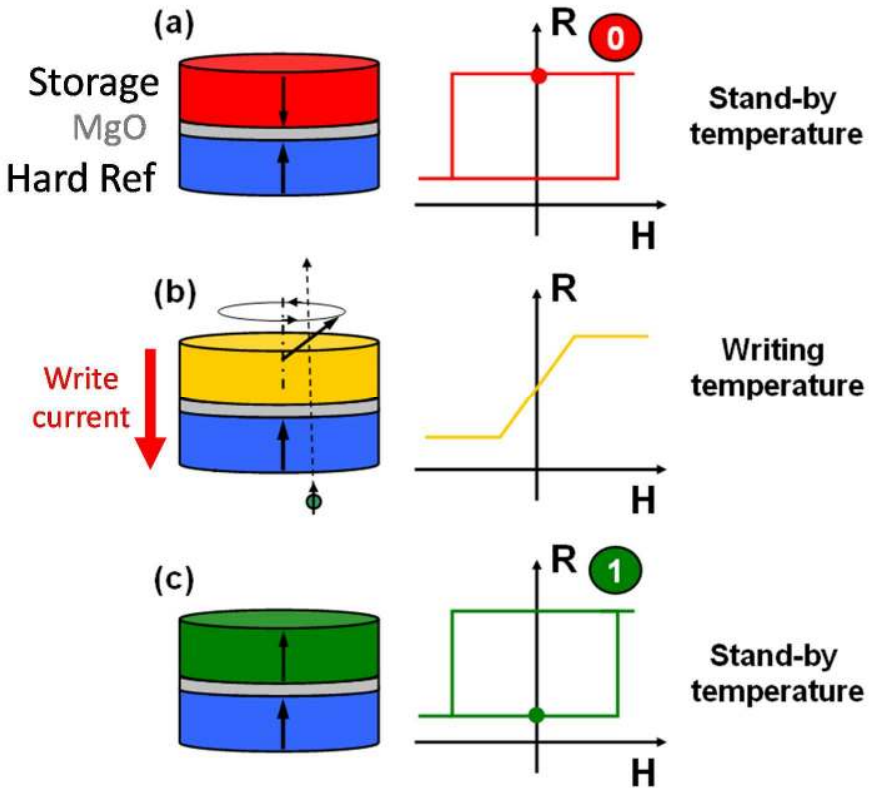

Figure 31: Principle of TIAR assisted switching: (a) The junction exhibits a large PMA at stand-by temperature. (b) A current is sent through the junction. The storage layer magnetization undergoes a TIAR due to the heating. The STT induces a large angle precessional motion of the storage layer magnetization and pulls it upwards or downwards depending on the current direction. (c) The storage layer recovers its PMA during the cooling when the current is gradually decreased to zero. Its magnetization rotates out-of-plane in the direction corresponding to the hemisphere in which it was pulled in by STT.

This switching approach reduces the write consumption as soon as the required current to heat the junction to the anisotropy reorientation temperature T_K is lower than the critical switching current in a standard MTJ (typically a few 10^6 A/cm^2). This condition is fulfilled in most common cases. In addition, the heating current density can be further reduced by adding in-stack thermal barriers on the various sides of the MTJ.

In this particular writing scheme, the voltage required to switch the junction is mainly determined by the anisotropy reorientation temperature T_K, which sets the lower limit of the writing window. T_K can be estimated theoretically by calculating K^{eff} as a function of temperature. For a cubic or uniaxial anisotropy, it can be written to first order as:

$$K(T) = K_{\perp}(0)\left(\frac{M_s(T)}{M_s(0)}\right)^m - (N_{perp} - N_{plan})\frac{\mu_0}{2}M_s(T)^2 \quad (21)$$

where $K_{\perp}(0)$ is the perpendicular magnetic anisotropy constant at 0 K, $M_S(T)$ the free layer magnetization, and $-(N_{perp} - N_{plan})\frac{\mu_0}{2}M_s(T)^2$ the demagnetizing energy, N_{perp} and N_{plan} being the demagnetizing factors of the pillar. The exponent m depends on the origin of the PMA. For a single ion anisotropy, m = i(i + 1)/2, where i denotes the anisotropy constant order, equal to 2 to first order in the most common case of a uniaxial anisotropy so that m = 3. From this equation, the K^{eff} sign changes beyond T_K, where T_K is the temperature where $K^{eff} = 0$. In that situation, the in-plane anisotropy induced by the demagnetizing energy becomes larger than the PMA, so that the free layer magnetization falls into the thin film plane. In order to tune T_K, Eq. (21) shows that the evolution of M_S as a function of T has to be controlled. This can be achieved by tuning the Curie temperature T_C of the magnetic material. (Co/NM) multilayers, where NM is a non-magnetic metal, provide a wide range of T_C, and exhibit PMA at room temperature when NM is a noble metal such as Pt, Pd or Au. In those systems, the T_C can be easily tuned by changing the Co or NM thickness, or by doping Co with Ni or non-magnetic elements [134]–[136]

However such multilayers cannot be used alone as magnetic electrodes in MTJ for MRAM applications which require relatively high TMR ratio. Indeed, (Co/NM) multilayers present an fcc (111) texture which does not match the bcc (001) MgO texture [136]. This problem can be solved by inserting a thin CoFeB layer at the MgO interface: the (Co/NM) multilayer together with the interfacial anisotropy at the CoFeB/MgO interface provides a sufficiently large PMA to pull the CoFeB magnetization out-of-plane while the CoFeB layer provides the appropriate bcc texture required to obtain large TMR ratio [137]. In such structures, TIAR can still be obtained by tuning the (Co/NM) multilayer properties.

Figure 32 illustrates the TIAR phenomenon in a multilayer of MgO / CoFeB1.5 /Ta0.2/ (Pd1.2 / Co0.3) × 3 (thickness in nanometers). In Fig.32a, hysteresis loops were measured on a full sheet sample by polar magneto-optical Kerr effect as a function of temperature, with an out-of-plane applied field. At room temperature, the loops have square-shape indicating out-of-plane anisotropy. When the temperature increases, the coercivity decreases while the signal amplitude decreases due to the decrease of magnetization. At 150 °C, perpendicular domains appear, leading to a loop with no remanence. The anisotropy reorientation occurs at around 175 (±20) °C. Above this temperature, the loops exhibit a reversible linear behavior indicating in-plane anisotropy. When the sample is cooled down, the loops

recover their squareness and coercivity, indicating that the TIAR phenomenon is reversible. Concerning the reference layer, a perpendicular SAF can be designed so that its anisotropy remains always perpendicular on the whole operating temperature range. Fig.32b illustrates the dramatic decrease in coercivity of the storage layer magnetization resulting from the temporary heating of the cell due to 30ns long voltage pulse of various amplitudes. The coercivity vanishes when the anisotropy reorientation takes place. The RA product was here of 25 Ω.µm² allowing to safely apply voltage pulses of larger amplitude than in conventional STT-MRAM cells. Since the STT switching current scales with the anisotropy, the thermally induced drop of anisotropy clearly helps the STT induced switching. During the cooling phase, the current must be gradually decreased so that the STT is still active to overcome the thermal fluctuations and force the storage layer to remain in the same hemisphere.

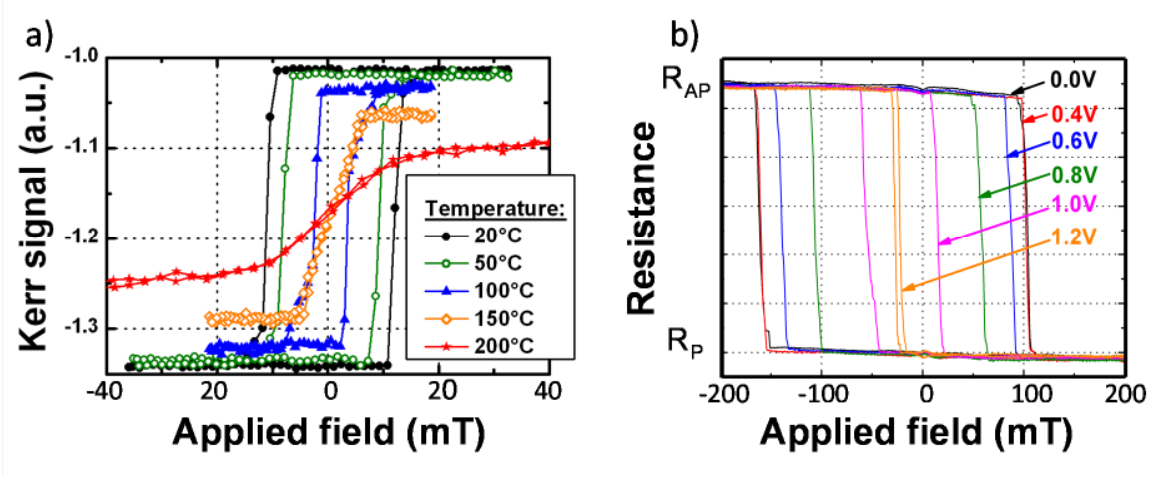

Figure 32a – a) Hysteresis loops of sheet films of CoFeB 1.5 /Ta0.2/(Pd 1.2 / Co 0.3) × 3 / Pd 2 for different measurement temperatures with out-of-plane applied field showing the temperature induced anisotropy reorientation. b) Illustration of the drop of coercivity in patterned perpendicular MTJs for different voltage pulse apmplitude (pulse 30ns long) due to the resulting heating of the cell (from Ref [101]).

The TIAR approach allows circumventing the classical dilemma in memory technology between retention (state stability) and write energy. As a result, at the time of these experiments, this approach allowed to obtain record values in terms of figure of merit Δ/Ic [131].

7.7 Remaining challenges and conclusion regarding Thermally Assisted MRAM

In conclusion, the thermal assistance in TA-MRAM helps circumventing the classical dilemma of storage technology between writability and retention. In field written TA-MRAM, the thermal assistance allows significantly reducing the power consumption as compared to toggle MRAM thanks to the use of low write field provided by the circular shape of the dots and the possibility of field sharing between numerous bits. The very good thermal stability of the

storage layer magnetization provided by the exchange bias with suitably chosen antiferromagnetic layer can allow using this type of memory at elevated operating temperature such as required for automotive applications. When using soft-reference layer, a new logic XOR functionality is added at the cell level on top of the memory function. This opens new applications such as Match-in-Place function interesting for security applications (smart cards for instance) as well as for Content Addressable Memories. In this technology, the remaining problems remain the dot to dot variability and still too large write error rates. Both phenomena seem to be associated with magnetostriction issues in the cells combined with non-uniform stress in the patterned array of MRAM cells. These issues should be solvable by careful adjustment of the stack composition to minimize magnetostriction.

When combined with STT, the thermal assistance can help reducing the write current by breaking the proportional relationship between thermal stability factor and write current. As in hard disk drive (HDD) industry where heat assisted magnetic recording is being developed to further extend the areal density on HDD, thermally assisted STT may help to extend the downsize scalability of STT-MRAM at sub-16nm nodes. This work is still in progress. In this technology, besides the variability issues also encountered in conventional p-STT-MRAM at sub-20nm nodes, additional difficulties remain. A first one is related to the reduced heating efficiency at small cells diameter due to heat leakage through the lateral edges of the MTJs pillars. To improve the heating efficiency, encapsulating the MTJ after etching within a low conductivity insulator material would be useful. Another difficulty is to avoid write errors during the cooling of the cells. Indeed as long as the cell is at elevated temperature, the anisotropy is reduced and thermal activation is enhanced so that the written state can be more easily lost. Here different schemes can be applied to reduce this effect. One is to slow down the current decrease following the switching of the magnetization so as to maintain a significant STT (proportional to current I) while the temperature decreases (proportionally to I^2) [136]. Another approach is to reduce the temperature rise during write to maintain a long enough retention at the maximum temperature to avoid back switching of the magnetization during the cooling phase. However, this is not easy to implement in chips supposed to be functional on a large range of operating temperature (0°C-80°C for instance).

8. Conclusion

MRAM is a high-speed nonvolatile memory that can provide unique solutions which improve overall system performance in a variety of areas including data storage, industrial controls, networking, and others. A series of scientific discoveries and innovations have stimulated progress in MRAM technology, and commercial investment, worldwide since the late 1990s. It is the method of writing the magnetic state of the storage device that distinguishes one technology generation from another. Each new generation offers the potential for much higher memory density and lower power operation.

Toggle MRAM is a field-switched MRAM, generally considered the first generation, that is in mass production in densities up to 16Mb. These products are generally either SRAM-like (parallel interface) or serial. They

find uses in a wide range of commercial products where the unique combination of speed, nonvolatility, and unlimited endurance provide a better system solution than other memory technologies.

The second MRAM generation employs magnetic switching by spin-transfer-torque, enabling smaller MTJ devices for higher density. A 64Mb DDR3 (DRAM-like) STT-MRAM product, based on spin-torque switching of in-plane MTJ devices, is currently in production at Everspin Technologies. Spin-transfer-torque switching of MTJ devices with magnetization perpendicular to the film plane will further extend the density of ST-MRAM, and reduce the write energy, to enable Gb-class memories in the near future. The age of commercial spin-torque MRAM (second generation) is just beginning, with intensive product development underway by large and small companies, as well as institutions, around the world.

The use of thermal assist can be complimentary with either field switching or STT switching. Multiple methods for employing heat in MTJ memory arrays have been demonstrated. In some cases heat acts as a half-select method, replacing magnetic field for example, and in other cases it can be used to enable unique device capabilities, such as an MLU. As a result, this continues to be an active area of research and development.

Challenges in developing commercially viable STT-MRAM include achieving certain key MTJ device parameters simultaneously, as well as a detailed understanding and control of bit-to-bit distributions in the memory array. Examples include: scaling the critical current for switching (Ic) while maintaining energy barrier to thermal reversal (Eb) in ever smaller devices, controlling the switching for high Eb and tight distributions, and separation of the critical voltage (Vc) distribution from the breakdown voltage (Vbd) distribution. Meanwhile, researchers continue discovering new effects that can be used to switch the magnetic state of MTJ devices in hopes of defining the third generation of MRAM.

References

[1] L. Savtchenko, B. Engel, N. Rizzo, M. Deherrera, and J. Janesky, "Method of writing to scalable magnetoresistance random access memory element," US Patent 6,545,906 B1, 2003.

[2] Electronicdesign.com, "4-Mbit Device Is First Commercially Available MRAM," *Electronic Design*, Jul-2006. [Online]. Available: http://electronicdesign.com/dsps/4-mbit-device-first-commercially-available-mram.

[3] N. D. Rizzo, D. Houssameddine, J. Janesky, R. Whig, F. B. Mancoff, M. L. Schneider, M. DeHerrera, J. J. Sun, K. Nagel, S. Deshpande, H.-J. Chia, S. M. Alam, T. Andre, S. Aggarwal, and J. M. Slaughter, "A Fully Functional 64 Mb DDR3 ST-MRAM Built on 90 nm CMOS Technology," *IEEE Trans. Magn.*, vol. 49, no. 7, pp. 4441–4446, Jul. 2013.

[4] S. S. P. Parkin, C. Kaiser, A. Panchula, P. M. Rice, B. Hughes, M. Samant, and S.-H. Yang, "Giant tunnelling magnetoresistance at room temperature with MgO (100) tunnel barriers," *Nature Mater.*, vol. 3, no. 12, pp. 862–867, Dec. 2004.

[5] Y. Huai, F. Albert, P. Nguyen, M. Pakala, and T. Valet, "Observation of spin-transfer switching in deep submicron-sized and low-resistance magnetic tunnel junctions," *Appl. Phys. Lett.*, vol. 84, no. 16, p. 3118, Apr. 2004.

[6] T. W. Andre, J. J. Nahas, C. K. Subramanian, B. J. Garni, H. S. Lin, A. Omair, and W. L. Martino, "A 4-Mb 0.18-/spl mu/m 1T1MTJ toggle MRAM with balanced three input sensing scheme and locally mirrored unidirectional write drivers," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 301–309, Jan. 2005.

[7] B. N. Engel, J. Akerman, B. Butcher, R. W. Dave, M. DeHerrera, M. Durlam, G. Grynkewich, J. Janesky, S. V. Pietambaram, N. D. Rizzo, J. M. Slaughter, K. Smith, J. J. Sun, and S. Tehrani, "A 4-Mb toggle MRAM based on a novel bit and switching method," *IEEE Trans. Magn.*, vol. 41, no. 1, pp. 132–136, Jan. 2005.

[8] S. Aggarwal, K. Nagel, S. Deshpande, M. Hossain, S. Karre, J. M. Slaughter, D. Houssameddine, R. Whig, H.-J. Chia, F. B. Mancoff, M. DeHerrera, J. Janesky, M. Lin,

and C. Mudivarthi, “Solving Manufacturing Challenges: Spin Torque MRAM on Advanced Technology Nodes,” *SEMICONWEST, Session: NTXSPT2*, Jul-2015. [Online]. Available: http://semiconwest.org/.

[9] M. Weisheit, S. Fahler, A. Marty, Y. Souche, C. Poinsignon, and D. Givord, “Electric Field-Induced Modification of Magnetism in Thin-Film Ferromagnets,” *Science*, vol. 315, no. 5810, pp. 349–351, Jan. 2007.

[10] Y. Shiota, T. Nozaki, F. Bonell, S. Murakami, T. Shinjo, and Y. Suzuki, “Induction of coherent magnetization switching in a few atomic layers of FeCo using voltage pulses,” *Nature Mater.*, vol. 11, no. 1, pp. 39–43, Nov. 2012.

[11] T. Maruyama, Y. Shiota, T. Nozaki, K. Ohta, N. Toda, M. Mizuguchi, A. A. Tulapurkar, T. Shinjo, M. Shiraishi, S. Mizukami, Y. Ando, and Y. Suzuki, “Large voltage-induced magnetic anisotropy change in a few atomic layers of iron,” *Nature Nanotech.*, vol. 4, no. 3, pp. 158–161, Mar. 2009.

[12] K. Kita, D. W. Abraham, M. J. Gajek, and D. C. Worledge, “Electric-field-control of magnetic anisotropy of Co0.6Fe0.2B0.2/oxide stacks using reduced voltage,” *J. Appl. Phys.*, vol. 112, no. 3, p. 033919, 2012.

[13] I. Mihai Miron, G. Gaudin, S. Auffret, B. Rodmacq, A. Schuhl, S. Pizzini, J. Vogel, and P. Gambardella, “Current-driven spin torque induced by the Rashba effect in a ferromagnetic metal layer,” *Nature Mater.*, vol. 9, no. 3, pp. 230–234, Jan. 2010.

[14] I. M. Miron, K. Garello, G. Gaudin, P.-J. Zermatten, M. V Costache, S. Auffret, S. Bandiera, B. Rodmacq, A. Schuhl, and P. Gambardella, “Perpendicular switching of a single ferromagnetic layer induced by in-plane current injection,” *Nature*, vol. 476, no. 7359, pp. 189–193, Jul. 2011.

[15] L. Liu, O. J. Lee, T. J. Gudmundsen, D. C. Ralph, and R. a. Buhrman, “Current-Induced Switching of Perpendicularly Magnetized Magnetic Layers Using Spin Torque from the Spin Hall Effect,” *Phys. Rev. Lett.*, vol. 109, no. 9, p. 096602, Aug. 2012.

[16] A. Brataas and K. M. D. Hals, “Spin–orbit torques in action,” *Nature Nanotech.*, vol. 9,

no. 2, pp. 86–88, Feb. 2014.

[17] M. Cubukcu, O. Boulle, M. Drouard, K. Garello, C. Onur Avci, I. Mihai Miron, J. Langer, B. Ocker, P. Gambardella, and G. Gaudin, "Spin-orbit torque magnetization switching of a three-terminal perpendicular magnetic tunnel junction," *Appl. Phys. Lett.*, vol. 104, no. 4, p. 042406, Jan. 2014.

[18] J. S. Moodera, L. R. Kinder, T. M. Wong, and R. Meservey, "Large Magnetoresistance at Room Temperature in Ferromagnetic Thin Film Tunnel Junctions," *Phys. Rev. Lett.*, vol. 74, no. 16, pp. 3273–3276, Apr. 1995.

[19] T. Miyazaki and N. Tezuka, "Giant magnetic tunneling effect in Fe/Al2O3/Fe junction," *J. Magn. Magn. Mater.*, vol. 139, no. 3, pp. L231–L234, Jan. 1995.

[20] S. Yuasa, T. Nagahama, A. Fukushima, Y. Suzuki, and K. Ando, "Giant room-temperature magnetoresistance in single-crystal Fe/MgO/Fe magnetic tunnel junctions," *Nature Mater.*, vol. 3, no. 12, pp. 868–871, Dec. 2004.

[21] Y. M. Lee, J. Hayakawa, S. Ikeda, F. Matsukura, and H. Ohno, "Effect of electrode composition on the tunnel magnetoresistance of pseudo-spin-valve magnetic tunnel junction with a MgO tunnel barrier," *Appl. Phys. Lett.*, vol. 90, no. 21, p. 212507, 2007.

[22] S. Ikeda, J. Hayakawa, Y. Ashizawa, Y. M. Lee, K. Miura, H. Hasegawa, M. Tsunoda, F. Matsukura, and H. Ohno, "Tunnel magnetoresistance of 604% at 300 K by suppression of Ta diffusion in CoFeB/MgO/CoFeB pseudo-spin-valves annealed at high temperature," *Appl. Phys. Lett.*, vol. 93, no. 8, p. 082508, 2008.

[23] J. M. Slaughter, "Materials for Magnetoresistive Random Access Memory," *Annu. Rev. Mater. Res.*, vol. 39, no. 1, pp. 277–296, Aug. 2009.

[24] R. W. Dave, G. Steiner, J. M. Slaughter, J. J. Sun, B. Craigo, S. Pietambaram, K. Smith, G. Grynkewich, M. DeHerrera, J. Akerman, and S. Tehrani, "MgO-Based Tunnel Junction Material for High-Speed Toggle Magnetic Random Access Memory," *IEEE Trans. Magn.*, vol. 42, no. 8, pp. 1935–1939, Aug. 2006.

[25] S. S. P. Parkin, "Systematic variation of the strength and oscillation period of indirect

magnetic exchange coupling through the 3 d , 4 d , and 5 d transition metals," *Phys. Rev. Lett.*, vol. 67, no. 25, pp. 3598–3601, Dec. 1991.

[26] J. M. Slaughter, R. W. Dave, M. DeHerrera, M. Durlam, B. N. Engel, J. Janesky, N. D. Rizzo, and S. Tehrani, "Fundamentals of MRAM technology," *J. Supercond.*, vol. 15, no. 1, pp. 19–25, 2002.

[27] T. Schulthess and W. Butler, "Magnetostatic coupling in spin valves: Revisiting Neel's formula," *J. Appl. Phys.*, vol. 87, no. 9, p. 5759, 2000.

[28] J. Slonczewski, "Current-driven excitation of magnetic multilayers," *J. Magn. Magn. Mater.*, vol. 159, no. 1, p. L1, 1996.

[29] L. Berger, "Emission of spin waves by a magnetic multilayer traversed by a current," *Phys. Rev. B*, vol. 54, no. 13, pp. 9353–9358, 1996.

[30] J. Katine, F. Albert, R. Buhrman, E. Myers, and D. Ralph, "Current-driven magnetization reversal and spin-wave excitations in Co/Cu/Co pillars," *Phys. Rev. Lett.*, vol. 84, no. 14, pp. 3149–3152, 2000.

[31] Z. Diao, D. Apalkov, M. Pakala, Y. Ding, A. Panchula, and Y. Huai, "Spin transfer switching and spin polarization in magnetic tunnel junctions with MgO and AlOx barriers," *Appl. Phys. Lett.*, vol. 87, no. 23, p. 232502, 2005.

[32] H. Kubota, A. Fukushima, Y. Ootani, S. Yuasa, K. Ando, H. Maehara, K. Tsunekawa, D. D. Djayaprawira, N. Watanabe, and Y. Suzuki, "Evaluation of Spin-Transfer Switching in CoFeB/MgO/CoFeB Magnetic Tunnel Junctions," *Jpn. J. Appl. Phys.*, vol. 44, no. No. 40, pp. L1237–L1240, Sep. 2005.

[33] J. Hayakawa, S. Ikeda, Y. M. Lee, R. Sasaki, T. Meguro, F. Matsukura, H. Takahashi, and H. Ohno, "Current-Driven Magnetization Switching in CoFeB/MgO/CoFeB Magnetic Tunnel Junctions," *Jpn. J. Appl. Phys.*, vol. 44, no. No. 41, pp. L1267–L1270, Sep. 2005.

[34] F. J. A. den Broeder, D. Kuiper, A. P. van de Mosselaer, and W. Hoving, "Perpendicular Magnetic Anisotropy of Co-Au Multilayers Induced by Interface Sharpening," *Phys. Rev. Lett.*, vol. 60, no. 26, pp. 2769–2772, Jun. 1988.

[35] W. B. Zeper, F. J. a M. Greidanus, P. F. Carcia, and C. R. Fincher, “Perpendicular magnetic anisotropy and magneto-optical Kerr effect of vapor-deposited Co/Pt-layered structures,” *J. Appl. Phys.*, vol. 65, no. 12, p. 4971, 1989.

[36] B. N. Engel, C. D. England, R. A. Van Leeuwen, M. H. Wiedmann, and C. M. Falco, “Interface magnetic anisotropy in epitaxial superlattices,” *Phys. Rev. Lett.*, vol. 67, no. 14, pp. 1910–1913, Sep. 1991.

[37] G. H. O. Daalderop, P. J. Kelly, and F. J. A. den Broeder, “Prediction and confirmation of perpendicular magnetic anisotropy in Co/Ni multilayers,” *Phys. Rev. Lett.*, vol. 68, no. 5, pp. 682–685, Feb. 1992.

[38] M. T. Johnson, P. J. H. Bloemen, F. J. a Den Broeder, and J. J. De Vries, “Magnetic anisotropy in metallic multilayers,” *Rep. Prog. Phys.*, vol. 59, no. 11, pp. 1409–1458, Nov. 1996.

[39] H. Sato, S. Ikeda, S. Fukami, H. Honjo, and S. Ishikawa, “Co / Pt multilayer based reference layers in magnetic tunnel junctions for nonvolatile spintronics VLSIs Co / Pt multilayer based reference layers in magnetic tunnel junctions for nonvolatile spintronics VLSIs,” *Jpn. J. Appl. Phys.*, vol. 53, p. 04EM02, 2014.

[40] S. Bandiera, R. C. Sousa, Y. Dahmane, C. Ducruet, C. Portemont, V. Baltz, S. Auffret, I. L. Prejbeanu, and B. Dieny, “Comparison of Synthetic Antiferromagnets and Hard Ferromagnets as Reference Layer in Magnetic Tunnel Junctions With Perpendicular Magnetic Anisotropy,” *IEEE Magn. Lett.*, vol. 1, pp. 3000204–3000204, 2010.

[41] D. C. Worledge, G. Hu, D. W. Abraham, J. Z. Sun, P. L. Trouilloud, J. Nowak, S. Brown, M. C. Gaidis, E. J. O’Sullivan, and R. P. Robertazzi, “Spin torque switching of perpendicular Ta/CoFeB/MgO-based magnetic tunnel junctions,” *Appl. Phys. Lett.*, vol. 98, no. 2, p. 022501, 2011.

[42] S. Monso, B. Rodmacq, S. Auffret, G. Casali, F. Fettar, B. Gilles, B. Dieny, and P. Boyer, “Crossover from in-plane to perpendicular anisotropy in Pt/CoFe/AlOx sandwiches as a function of Al oxidation: A very accurate control of the oxidation of tunnel barriers,” *Appl. Phys. Lett.*, vol. 80, no. 22, pp. 4157–4159, 2002.

[43] S. Ikeda, K. Miura, H. Yamamoto, K. Mizunuma, H. D. Gan, M. Endo, S. Kanai, J. Hayakawa, F. Matsukura, and H. Ohno, "A perpendicular-anisotropy CoFeB–MgO magnetic tunnel junction," *Nature Mater.*, vol. 9, no. 9, pp. 721–724, Sep. 2010.

[44] A. Manchon, C. Ducruet, L. Lombard, S. Auffret, B. Rodmacq, B. Dieny, S. Pizzini, J. Vogel, V. Uhlíř, M. Hochstrasser, and G. Panaccione, "Analysis of oxygen induced anisotropy crossover in Pt/Co/MOx trilayers," *J. Appl. Phys.*, vol. 104, no. 4, p. 043914, 2008.

[45] H. X. Yang, M. Chshiev, B. Dieny, J. H. Lee, A. Manchon, and K. H. Shin, "First-principles investigation of the very large perpendicular magnetic anisotropy at Fe/MgO and Co/MgO interfaces," *Phys. Rev. B*, vol. 84, no. 5, p. 054401, Aug. 2011.

[46] J. H. Park, Y. Kim, W. C. Lim, J. H. Kim, S. H. Park, J. H. Kim, W. Kim, K. W. Kim, J. H. Jeong, K. S. Kim, H. Kim, Y. J. Lee, S. C. Oh, J. E. Lee, S. O. Park, S. Watts, D. Apalkov, V. Nikitin, M. Krounbi, S. Jeong, S. Choi, H. K. Kang, and C. Chung, "Enhancement of data retention and write current scaling for sub-20nm STT-MRAM by utilizing dual interfaces for perpendicular magnetic anisotropy," *Dig. Tech. Pap. - Symp. VLSI Technol.*, vol. 721, no. 2010, pp. 57–58, 2012.

[47] H. Sato, E. C. I. Enobio, M. Yamanouchi, S. Ikeda, S. Fukami, S. Kanai, F. Matsukura, and H. Ohno, "Properties of magnetic tunnel junctions with a MgO/CoFeB/Ta/CoFeB/MgO recording structure down to junction diameter of 11 nm," *Appl. Phys. Lett.*, vol. 105, p. 062403, 2014.

[48] R. Scheuerlein, W. Gallagher, S. Parkin, A. Lee, S. Ray, R. Robertazzi, and W. Reohr, "A 10ns read and write non-volatile memory array using a magnetic tunnel junction and FET switch in each cell," *Solid-State Circuits Conf. 2000. Dig. Tech. Pap. ISSCC. 2000 IEEE Int.*, vol. 3741, no. 1997, pp. 128–129, 2000.

[49] M. Durlam, D. Addie, J. Akerman, B. Butcher, P. Brown, J. Chan, M. DeHerrera, B. N. Engel, B. Feil, G. Grynkewich, J. Janesky, M. Johnson, K. Kyler, J. Molla, J. Martin, K. Nagel, J. Ren, N. D. Rizzo, T. Rodriguez, L. Savtchenko, J. Salter, J. M. Slaughter, K. Smith, J. J. Sun, M. Lien, K. Papworth, P. Shah, W. Qin, R. Williams, L. Wise, and S.

Tehrani, “A 0.18 μm 4Mb toggling MRAM,” in *IEEE International Electron Devices Meeting*, 2002, pp. 34.6.1–34.6.3.

[50] M. Hosomi, H. Yamagishi, T. Yamamoto, K. Bessho, Y. Higo, K. Yamane, H. Yamada, M. Shoji, H. Hachino, and C. Fukumoto, “A novel nonvolatile memory with spin torque transfer magnetization switching: spin-ram,” *IEEE Int. Electron Devices Meet. Dig.*, p. 473, 2005.

[51] T. Kawahara, R. Takemura, K. Miura, J. Hayakawa, S. Ikeda, Y. Lee, R. Sasaki, Y. Goto, K. Ito, T. Meguro, F. Matsukura, H. Takahashi, H. Matsuoka, and H. Ohno, “2Mb Spin-Transfer Torque RAM (SPRAM) with Bit-by-Bit Bidirectional Current Write and Parallelizing-Direction Current Read,” in *IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, 2007, pp. 480–481.

[52] R. Beach, T. Min, C. Horng, Q. Chen, P. Sherman, S. Le, S. Young, K. Yang, H. Yu, X. Lu, W. Kula, T. Zhong, R. Xiao, A. Zhong, G. Liu, J. Kan, J. Yuan, J. Chen, R. Tong, J. Chien, T. Torng, D. Tang, P. Wang, M. Chen, S. Assefa, M. Qazi, J. DeBrosse, M. Gaidis, S. Kanakasabapathy, Y. Lu, J. Nowak, E. O’Sullivan, T. Maffitt, J. Z. Sun, and W. J. Gallagher, “A statistical study of magnetic tunnel junctions for high-density spin torque transfer-MRAM (STT-MRAM),” in *IEEE International Electron Devices Meeting*, 2008, pp. 1–4.

[53] T. Kishi, H. Yoda, T. Kai, T. Nagase, E. Kitagawa, M. Yoshikawa, K. Nishiyama, T. Daibou, M. Nagamine, M. Amano, S. Takahashi, M. Nakayama, N. Shimomura, H. Aikawa, S. Ikegawa, S. Yuasa, K. Yakushiji, H. Kubota, A. Fukushima, M. Oogane, T. Miyazaki, and K. Ando, “Lower-current and fast switching of a perpendicular TMR for high speed and high density spin-transfer-torque MRAM,” *IEEE Int. Electron Devices Meet.*, pp. 1–4, Dec. 2008.

[54] Suock Chung, K.-M. Rho, S.-D. Kim, H.-J. Suh, D.-J. Kim, H.-J. Kim, S.-H. Lee, J.-H. Park, H.-M. Hwang, S.-M. Hwang, J.-Y. Lee, Y.-B. An, J.-U. Yi, Y.-H. Seo, D.-H. Jung, M.-S. Lee, S.-H. Cho, J.-N. Kim, G.-J. Park, Gyuan Jin, A. Driskill-Smith, V. Nikitin, A. Ong, X. Tang, Yongki Kim, J.-S. Rho, S.-K. Park, S.-W. Chung, J.-G. Jeong, and S.-J. Hong, “Fully integrated 54nm STT-RAM with the smallest bit cell dimension for high

density memory application," in *2010 International Electron Devices Meeting*, 2010, pp. 12.7.1–12.7.4.

[55] D. C. Worledge, G. Hu, P. L. Trouilloud, D. W. Abraham, S. Brown, M. C. Gaidis, J. Nowak, E. J. O. Sullivan, R. P. Robertazzi, J. Z. Sun, and W. J. Gallagher, "Switching distributions and write reliability of perpendicular spin torque MRAM," in *IEEE International Electron Devices Meeting*, 2010, pp. 296–299.

[56] J. M. Slaughter, N. D. Rizzo, J. Janesky, R. Whig, F. B. Mancoff, D. Houssameddine, J. J. Sun, S. Aggarwal, K. Nagel, S. Deshpande, S. M. Alam, T. Andre, and P. LoPresti, "High density ST-MRAM technology (Invited)," in *International Electron Devices Meeting*, 2012, vol. 85224, no. 480, pp. 29.3.1–29.3.4.

[57] L. Thomas, G. Jan, J. Zhu, H. Liu, Y.-J. Lee, S. Le, R.-Y. Tong, K. Pi, Y.-J. Wang, D. Shen, R. He, J. Haq, J. Teng, V. Lam, K. Huang, T. Zhong, T. Torng, and P.-K. Wang, "Perpendicular spin transfer torque magnetic random access memories with high spin torque efficiency and thermal stability for embedded applications (invited)," *J. Appl. Phys.*, vol. 115, no. 17, p. 172615, May 2014.

[58] T. M. Maffitt, J. K. DeBrosse, J. a. Gabric, E. T. Gow, M. C. Lamorey, J. S. Parenteau, D. R. Willmott, M. A. Wood, and W. J. Gallagher, "Design considerations for MRAM," *IBM J. Res. Dev.*, vol. 50, no. 1, pp. 25–39, Jan. 2006.

[59] J. Song, J. Kim, S. H. Kang, S. Yoon, and S.-O. Jung, "Sensing margin trend with technology scaling in MRAM," *Int. J. Circuit Theory Appl.*, vol. 39, no. 3, pp. 313–325, Mar. 2011.

[60] W. Kang, L. Zhang, J. Klein, Y. Zhang, D. Ravelosona, and W. Zhao, "Reconfigurable Codesign of STT-MRAM Under Process Variations in Deeply Scaled Technology," *IEEE Trans. Electron Devices*, vol. 62, no. 6, pp. 1769–1777, 2015.

[61] Y. Zhang, Y. Li, Z. Sun, H. Li, Y. Chen, and A. K. Jones, "Read Performance: The Newest Barrier in Scaled STT-RAM," *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 23, no. 6, pp. 1170–1174, 2014.

[62] Z. Sun, H. Li, Y. Chen, and X. Wang, “Voltage Driven Nondestructive Self-Reference Sensing Scheme of Spin-Transfer Torque Memory,” *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 20, no. 11, pp. 2020–2030, Nov. 2012.

[63] Yiran Chen, Hai Li, Xiaobin Wang, Wenzhong Zhu, Wei Xu, and Tong Zhang, “A nondestructive self-reference scheme for Spin-Transfer Torque Random Access Memory (STT-RAM),” in *2010 Design, Automation & Test in Europe Conference & Exhibition (DATE 2010)*, 2010, pp. 148–153.

[64] S. Tehrani, J. Slaughter, M. Deherrera, B. Engel, N. Rizzo, J. Salter, M. Durlam, R. Dave, J. Janesky, and B. Butcher, “Magnetoresistive random access memory using magnetic tunnel junctions,” *Proc. IEEE*, vol. 91, no. 5, pp. 703–714, 2003.

[65] J. G. Zhu, “Spin valve and dual spin valve heads with syntheticantiferromagnets,” *IEEE Trans. Magn.*, vol. 35, no. 2 Part 1, pp. 655–660, 1999.

[66] M. Durlam, B. Craigo, M. DeHerrera, B. N. Engel, G. Grynkewich, B. Huang, J. Janesky, M. Martin, B. Martino, J. Salter, J. M. Slaughter, L. Wise, and S. Tehrani, “Toggle MRAM: A highly-reliable Non-Volatile Memory,” in *2007 International Symposium on VLSI Technology, Systems and Applications (VLSI-TSA)*, 2007, pp. 1–2.

[67] J. Akerman, M. DeHerrera, J. M. Slaughter, R. Dave, J. J. Sun, J. T. Martin, and S. Tehrani, “Intrinsic Reliability of AlOx-Based Magnetic Tunnel Junctions,” *IEEE Trans. Magn.*, vol. 42, no. 10, pp. 2661–2663, Oct. 2006.

[68] S. Tehrani, “Status and Outlook of MRAM Memory Technology (Invited),” in *International Electron Devices Meeting*, 2006, pp. 1–4.

[69] B. Dieny, V. S. Speriosu, S. S. P. Parkin, B. A. Gurney, D. R. Wilhoit, and D. Mauri, “Giant magnetoresistive in soft ferromagnetic multilayers,” *Phys. Rev. B*, vol. 43, no. 1, pp. 1297–1300, Jan. 1991.

[70] T. Ishigaki, T. Kawahara, R. Takemura, K. Ono, K. Ito, H. Matsuoka, and H. Ohno, “A Multi-Level-Cell Spin-Transfer Torque Memory with Series-Stacked Magnetotunnel Junctions,” *VLSI*, pp. 234–235, 2010.

[71] A. Aharoni, “Micromagnetics: past, present and future,” *Phys. B Condens. Matter*, vol. 306, no. 1–4, pp. 1–9, Dec. 2001.

[72] D. M. Apalkov and P. B. Visscher, “Slonczewski spin-torque as negative damping: Fokker–Planck computation of energy distribution,” *J. Magn. Magn. Mater.*, vol. 286, pp. 370–374, Feb. 2005.

[73] D. M. Apalkov and P. B. Visscher, “Spin-torque switching: Fokker-Planck rate calculation,” *Phys. Rev. B*, vol. 72, no. 18, pp. 1–4, 2005.

[74] Z. Li, S. Zhang, Z. Diao, Y. Ding, X. Tang, D. Apalkov, Z. Yang, K. Kawabata, and Y. Huai, “Perpendicular Spin Torques in Magnetic Tunnel Junctions,” *Phys. Rev. Lett.*, vol. 100, no. 24, pp. 1–4, Jun. 2008.

[75] J. Sun, “Spin-current interaction with a monodomain magnetic body: A model study,” *Phys. Rev. B*, vol. 62, no. 1, pp. 570–578, 2000.

[76] R. H. Koch, J. a. Katine, and J. Z. Sun, “Time-Resolved Reversal of Spin-Transfer Switching in a Nanomagnet,” *Phys. Rev. Lett.*, vol. 92, no. 8, pp. 2–5, 2004.

[77] W. H. Butler, T. Mewes, C. K. A. Mewes, P. B. Visscher, W. H. Rippard, S. E. Russek, and R. Heindl, “Advances in Magnetics Switching Distributions for Perpendicular Spin-Torque Devices Within the Macrospin Approximation,” vol. 48, no. 12, pp. 4684–4700, 2012.

[78] A. Khvalkovskiy, V. Voznyuk, S. Schäfer, D. Apalkov, R. Beach, V. Nikitin, and M. Krounbi, “Switching current in thermally-activated switching regime for perpendicular STT-MRAM structures,” in *Proceedings of the 58th MMM conference*, p. DC–08.

[79] T. Taniguchi and H. Imamura, “Current Dependence of Spin Torque Switching Barrier,” *Appl. Phys. Express*, vol. 6, no. 10, p. 103005, Oct. 2013.

[80] W. Brown, “Thermal fluctuations of a single-domain particle,” *Phys. Rev.*, vol. 130, no. 5, pp. 1677–1686, 1963.

[81] D. Apalkov, A. Khvalkovskiy, and S. Watts, “Spin-transfer torque magnetic random

access memory (stt-mram)," *ACM J. Emerg. Technol. Comput. Syst.*, vol. 9, no. 2, pp. 13:1–13:35, 2013.

[82] S. Monso, B. Rodmacq, S. Auffret, G. Casali, F. Fettar, B. Gilles, B. Dieny, and P. Boyer, "Crossover from in-plane to perpendicular anisotropy in Pt/CoFe/AlO[sub x] sandwiches as a function of Al oxidation: A very accurate control of the oxidation of tunnel barriers," *Appl. Phys. Lett.*, vol. 80, no. 22, p. 4157, 2002.

[83] B. Rodmacq, S. Auffret, B. Dieny, S. Monso, and P. Boyer, "Crossovers from in-plane to perpendicular anisotropy in magnetic tunnel junctions as a function of the barrier degree of oxidation," *J. Appl. Phys.*, vol. 93, no. 10, p. 7513, 2003.

[84] A. V Khvalkovskiy, D. Apalkov, S. Watts, R. Chepulskii, R. S. Beach, A. Ong, X. Tang, A. Driskill-Smith, W. H. Butler, P. B. Visscher, D. Lottis, E. Chen, V. Nikitin, and M. Krounbi, "Basic principles of STT-MRAM cell operation in memory arrays," *J. Phys. D. Appl. Phys.*, vol. 46, no. 7, p. 074001, Feb. 2013.

[85] H. Sato, M. Yamanouchi, S. Ikeda, S. Fukami, F. Matsukura, and H. Ohno, "Perpendicular-anisotropy CoFeB-MgO magnetic tunnel junctions with a MgO/CoFeB/Ta/CoFeB/MgO recording structure," *Appl. Phys. Lett.*, vol. 101, no. 2, p. 022414, 2012.

[86] D. Apalkov, A. Khvalkovskiy, R. Chepulskyy, S. Watts, X. Tang, V. Nikitin, and M. Krounbi, "Nudged elastic band calculation of switching modes and energy barriers in perpendicular MRAM," in *11th Joint MMM-Intermag conference*, 2013, p. FF–01.

[87] R. Dittrich, A. Thiaville, J. Miltat, and T. Schrefl, "Rigorous micromagnetic computation of configurational anisotropy energies in nanoelements," *J. Appl. Phys.*, vol. 93, no. 10, p. 7891, 2003.

[88] J. Sun, R. Robertazzi, J. Nowak, P. Trouilloud, G. Hu, D. Abraham, M. Gaidis, S. Brown, E. O'Sullivan, W. Gallagher, and D. Worledge, "Effect of subvolume excitation and spin-torque efficiency on magnetic switching," *Phys. Rev. B*, vol. 84, no. 6, Aug. 2011.

[89] Y. X. Wang, Z. Wang, H. Gan, B.K. Yen,Hui, "Perpendicular STT-MRAM Thermal

Stability Scaling and Robust Retention Design.," *MMM Conf. FE-14*, 2014.

[90] K. Tsunoda, M. Aoki, H. Noshiro, Y. Iba, S. Fukuda, C. Yoshida, Y. Yamazaki, A. Takahashi, A. Hatada, M. Nakabayashi, Y. Tsuzaki, and T. Sugii, "Area dependence of thermal stability factor in perpendicular STT-MRAM analyzed by bi-directional data flipping model," *Iedm*, vol. 2, pp. 486–489, 2014.

[91] G. D. Chaves-O'Flynn, G. Wolf, J. Z. Sun, and A. D. Kent, "Thermal Stability of Magnetic States in Circular Thin-Film Nanomagnets with Large Perpendicular Magnetic Anisotropy," *Phys. Rev. Appl.*, vol. 4, no. 2, p. 024010, 2015.

[92] L. Thomas, G. Jan, S. Le, Y. Lee, H. Liu, J. Zhu, S. Serrano-guisan, R. Tong, K. Pi, D. Shen, R. He, J. Haq, Z. Teng, R. Annapragada, V. Lam, Y. Wang, T. Zhong, T. Torng, and P. Wang, "Solving the Paradox of the Inconsistent Size Dependence of Thermal Stability at Device and Chip-level in Perpendicular STT-MRAM," vol. 20, pp. 672–675, 2015.

[93] T. Valet, "Spin barrier enhanced magnetoresistance effect element and magnetic memory using the same," US patent 7,088,609, 2006.

[94] K. Munira, J. Romero, and W. H. Butler, "Achieving perpendicular anisotropy in half-metallic Heusler alloys for spin device applications," *J. Appl. Phys.*, vol. 115, no. 17, p. 17B731, May 2014.

[95] S. Watts, X. Tang, Z. Diao, D. Apalkov, and V. Nikitin, "Low switching current in conventional in-plane STT-RAM structures with partial perpendicular anisotropy," in *Proceedings of 11th Joint MMM-Intermag conference*, 2010, p. FV–11.

[96] A. Natarajarathinam, Z. R. Tadisina, T. Mewes, S. Watts, E. Chen, and S. Gupta, "Influence of capping layers on CoFeB anisotropy and damping," *J. Appl. Phys.*, vol. 112, no. 5, p. 053909, 2012.

[97] T. Seki, S. Mitani, K. Yakushiji, and K. Takanashi, "Magnetization switching in nanopillars with FePt alloys by spin-polarized current," *J. Appl. Phys.*, vol. 99, no. 8, p. 08G521, 2006.

[98] S. Mangin, D. Ravelosona, J. A. Katine, M. J. Carey, B. D. Terris, and E. E. Fullerton, "Current-induced magnetization reversal in nanopillars with perpendicular anisotropy," *Nature Mater.*, vol. 5, no. 3, pp. 210–215, 2006.

[99] H. Sato, M. Yamanouchi, S. Ikeda, S. Fukami, F. Matsukura, and H. Ohno, "MgO/CoFeB/Ta/CoFeB/MgO Recording Structure in Magnetic Tunnel Junctions With Perpendicular Easy Axis," *IEEE Trans. Magn.*, vol. 49, no. 7, pp. 4437–4440, Jul. 2013.

[100] J. Z. Sun, S. L. Brown, W. Chen, E. a. Delenia, M. C. Gaidis, J. Harms, G. Hu, X. Jiang, R. Kilaru, W. Kula, G. Lauer, L. Q. Liu, S. Murthy, J. Nowak, E. J. O'Sullivan, S. S. P. Parkin, R. P. Robertazzi, P. M. Rice, G. Sandhu, T. Topuria, and D. C. Worledge, "Spin-torque switching efficiency in CoFeB-MgO based tunnel junctions," *Phys. Rev. B*, vol. 88, no. 10, p. 104426, Sep. 2013.

[101] M. Krounbi, V. Nikitin, D. Apalkov, J. Lee, X. Tang, R. Beach, D. Erickson, and E. Chen, "Status and Challenges in Spin-Transfer Torque MRAM Technology," *Proceedings of the 228th ECS meeting, October 11-15*, 2015. [Online]. Available: http://ecst.ecsdl.org/content/69/3/119.abstract.

[102] J. H. Kim, W. C. Lim, U. H. Pi, J. M. Lee, W. K. Kim, J. H. Kim, K. W. Kim, Y. S. Park, M. a Kang, Y. H. Kim, W. J. Kim, S. Y. Kim, J. H. Park, S. C. Lee, Y. J. Lee, J. M. Yoon, C. Oh, S. O. Park, S. Jeong, S. W. Nam, H. K. Kang, E. S. Jung, P. D. Team, R. Semiconductor, and S. E. Co, "Verification on the extreme scalability of STT-MRAM without loss of thermal stability below 15 nm MTJ cell," *2014 Symp. VLSI Technol. Dig. Tech. Pap.*, pp. 60–61, 2014.

[103] D. Apalkov, V. Nikitin, S. M. Watts, X. Tang, D. Lottis, A. Khvalkovskiy, K. Moon, R. Kawakami, X. Luo, E. Chen, A. Ong, and M. Krounbi, "Novel Ultra-Thin Dual MTJ for STT-RAM," in *Proceedings of 56th Magnetism and Magnetic Materials Conference*, 2011, p. BB–01.

[104] D. C. Worledge, G. Hu, D. W. Abraham, J. Z. Sun, P. L. Trouilloud, J. Nowak, S. Brown, M. C. Gaidis, E. J. O'Sullivan, and R. P. Robertazzi, "Spin torque switching of perpendicular Ta|CoFeB|MgO-based magnetic tunnel junctions," *Appl. Phys. Lett.*, vol.

98, no. 2, p. 022501, 2011.

[105] D. C. Worledge, G. Hu, D. W. Abraham, P. L. Trouilloud, and S. Brown, “Development of perpendicularly magnetized Ta|CoFeB|MgO-based tunnel junctions at IBM (invited),” *J. Appl. Phys.*, vol. 115, no. 17, p. 172601, May 2014.

[106] Z. Diao, A. Panchula, Y. Ding, M. Pakala, S. Wang, Z. Li, D. Apalkov, H. Nagai, A. Driskill-Smith, L.-C. Wang, E. Chen, and Y. Huai, “Spin transfer switching in dual MgO magnetic tunnel junctions,” *Appl. Phys. Lett.*, vol. 90, no. 13, p. 132508, 2007.

[107] Z. Duan, R. S. Beach, V. Nikitin, S. Schafer, V. Voznyuk, A. Khvalkovskiy, D. Apalkov, X. Tang, and J. Lee, “STT-MRAM Cell Design with Dual MgO Tunnel Barriers and Perpendicular Magnetic Anisotropy,” in *Proceedings of 59th conference on Magnetism and Magnetic Materials,* 2014, p. FE–03.

[108] G. Hu, J. H. Lee, J. J. Nowak, J. Z. Sun, J. Harms, A. Annunziata, S. Brown, W. Chen, G. Lauer, L. Liu, N. Marchack, S. Murthy, E. J. O. Sullivan, J. H. Park, M. Reuter, P. L. Trouilloud, Y. Zhu, and D. C. Worledge, “STT-MRAM with double magnetic tunnel junctions,” *IEDM Dig.*, no. 26.3.1, pp. 668–671, 2015.

[109] L. Xue, A. Kontos, C. Lazik, S. Liang, and M. Pakala, “Scalability of Magnetic Tunnel Junctions Patterned by a Novel Plasma Ribbon Beam Etching Process on 300 mm Wafers,” *IEEE Trans. Magn.*, vol. 51, no. 12, pp. 1–3, Dec. 2015.

[110] J. J. Nowak, R. P. Robertazzi, J. Z. Sun, G. Hu, D. W. Abraham, P. L. Trouilloud, S. Brown, M. C. Gaidis, E. J. O’Sullivan, W. J. Gallagher, and D. C. Worledge, “Demonstration of ultralow bit error rates for spin-torque magnetic random-access memory with perpendicular magnetic anisotropy,” *IEEE Magn. Lett.*, vol. 2, pp. 2–5, 2011.

[111] T. Taniguchi, Y. Utsumi, M. Marthaler, D. S. Golubev, and H. Imamura, “Spin torque switching of an in-plane magnetized system in a thermally activated region,” *Phys. Rev. B*, vol. 87, no. 5, p. 054406, Feb. 2013.

[112] T. Kawahara, K. Ito, R. Takemura, and H. Ohno, “Spin-transfer torque RAM technology :

Review and prospect," *Microelectron. Reliab.*, vol. 52, no. 4, pp. 613–627, 2012.

[113] S. Amara-Dababi, R. C. Sousa, M. Chshiev, H. Béa, J. Alvarez-Hérault, L. Lombard, I. L. Prejbeanu, K. Mackay, and B. Dieny, "Charge trapping-detrapping mechanism of barrier breakdown in MgO magnetic tunnel junctions," *Appl. Phys. Lett.*, vol. 99, no. 8, p. 083501, 2011.

[114] S. Amara-Dababi, H. Béa, R. C. Sousa, C. Baraduc, and B. Dieny, "Correlation between write endurance and electrical low frequency noise in MgO based magnetic tunnel junctions," *Appl. Phys. Lett.*, vol. 102, no. 5, p. 052404, 2013.

[115] G. Jan, L. Thomas, S. Le, Y. Lee, H. Liu, J. Zhu, R. Tong, K. Pi, Y. Wang, D. Shen, R. He, J. Haq, J. Teng, V. Lam, K. Huang, T. Torng, and P. Wang, "Demonstration of fully functional 8Mb perpendicular STT-MRAM chips with sub-5ns writing for non-volatile embedded memories," *Symp. VLSI Technol. Dig. Tech. Pap.*, vol. 093008, pp. 42–43, 2014.

[116] H. Honjo, H. Sato, S. Ikeda, S. Sato, T. Watanebe, S. Miura, T. Nasuno, Y. Noguchi, M. Yasuhira, T. Tanigawa, H. Koike, M. Muraguchi, M. Niwa, K. Ito, H. Ohno, and T. Endoh, "10 nm D perpendicular-anisotropy CoFeB-MgO magnetic tunnel junction with over 400C high thermal tolerance by boron diffusion control," in *2015 Symposium on VLSI Technology (VLSI Technology)*, 2015, vol. 042501, no. 2011, pp. T160–T161.

[117] K. Yakushiji, H. Kubota, A. Fukushima, and S. Yuasa, "Perpendicular magnetic tunnel junctions with strong antiferromagnetic interlayer exchange coupling at first oscillation peak," *Appl. Phys. Express*, vol. 8, no. 8, p. 083003, Aug. 2015.

[118] O. Redon, B. Dieny, and B. Rodmacq, "Magnetic spin polarization and magnetization rotation device with memory and writing process, using such a device," US 6,532,164, 2003.

[119] A. D. Kent, B. Özyilmaz, and E. del Barco, "Spin-transfer-induced precessional magnetization reversal," *Appl. Phys. Lett.*, vol. 84, no. 19, p. 3897, 2004.

[120] D. Apalkov, A. Khvalkovskiy, R. Chepulskyy, V. Nikitin, and M. Krounbi, "Anisotropy

engineering for ultrafast perpendicular STT-MRAM performance," in *The 58th annual conference on magnetism and magnetic materials (MMM)*, 2013, no. GC-09.

[121] D. Apalkov, "Method and system for providing magnetic tunneling junction elements having a biaxial anisotropy," US patent 8,374,048 B2, 2013.

[122] D. Apalkov and W. Butler, "Method and system for providing magnetic tunneling junction elements having an easy cone anisotropy," US patent 8,780,665 B2, 2014.

[123] J. M. Shaw, H. T. Nembach, M. Weiler, T. J. Silva, M. Schoen, J. Z. Sun, and D. C. Worledge, "Perpendicular Magnetic Anisotropy and Easy Cone State in Ta/Co60Fe20B20 /MgO," *IEEE Magn. Lett.*, vol. 6, pp. 1–4, 2015.

[124] R. Matsumoto, H. Arai, S. Yuasa, and H. Imamura, "Spin-transfer-torque switching in a spin-valve nanopillar with a conically magnetized free layer," *Appl. Phys. Express*, vol. 8, no. 6, p. 063007, 2015.

[125] R. C. Sousa and I. L. Prejbeanu, "Non-volatile magnetic random access memories (MRAM)," *Comptes Rendus Phys.*, vol. 6, no. 9, pp. 1013–1021, Nov. 2005.

[126] I. L. Prejbeanu, W. Kula, K. Ounadjela, R. C. Sousa, O. Redon, B. Dieny, and J.-P. Nozieres, "Thermally Assisted Switching in Exchange-Biased Storage Layer Magnetic Tunnel Junctions," *IEEE Trans. Magn.*, vol. 40, no. 4, pp. 2625–2627, Jul. 2004.

[127] J. Wang and P. P. Freitas, "Low-Current Blocking Temperature Writing of Double-Barrier MRAM Cells," *IEEE Trans. Magn.*, vol. 40, no. 4, pp. 2622–2624, Jul. 2004.

[128] B. Dieny, R. Sousa, S. Bandiera, M. Castro Souza, S. Auffret, B. Rodmacq, J. P. Nozieres, J. Hérault, E. Gapihan, I. L. Prejbeanu, C. Ducruet, C. Portemont, K. Mackay, and B. Cambou, "Extended scalability and functionalities of MRAM based on thermally assisted writing," *IEEE Int. Electron Devices Meet.*, no. 33, pp. 11–14, 2011.

[129] Q. Stainer, L. Lombard, K. Mackay, D. Lee, S. Bandiera, C. Portemont, C. Creuzet, R. C. Sousa, and B. Dieny, "Self-referenced multi-bit thermally assisted magnetic random access memories," *Appl. Phys. Lett.*, vol. 105, no. 3, p. 032405, Jul. 2014.

[130] B. Dieny, “Heat assisted magnetic write element,” US Patent 8,208,295 B2, 2008.

[131] S. Bandiera, R. C. Sousa, M. Marins de Castro, C. Ducruet, C. Portemont, S. Auffret, L. Vila, I. L. Prejbeanu, B. Rodmacq, and B. Dieny, “Spin transfer torque switching assisted by thermally induced anisotropy reorientation in perpendicular magnetic tunnel junctions,” *Appl. Phys. Lett.*, vol. 99, no. 20, p. 202507, 2011.

[132] E. R. Callen and H. B. Callen, “Anisotropic magnetization,” *J. Phys. Chem. Solids*, vol. 16, no. 3–4, pp. 310–328, Nov. 1960.

[133] D. Houssameddine, U. Ebels, B. Delaët, B. Rodmacq, I. Firastrau, F. Ponthenier, M. Brunet, C. Thirion, J.-P. Michel, L. Prejbeanu-Buda, M.-C. Cyrille, O. Redon, and B. Dieny, “Spin-torque oscillator using a perpendicular polarizer and a planar free layer,” *Nature Mater.*, vol. 6, no. 6, pp. 447–453, Jun. 2007.

[134] H. W. van Kesteren and W. B. Zeper, “Controlling the Curie temperature of Co/Pt multilayer magneto-optical recording media,” *J. Magn. Magn. Mater.*, vol. 120, no. 1–3, pp. 271–273, Mar. 1993.

[135] P. Poulopoulos and K. Baberschke, “Magnetism in thin films,” *J. Phys. Condens. Matter*, vol. 11, no. 48, pp. 9495–9515, Dec. 1999.

[136] S. Yuasa and D. D. Djayaprawira, “Giant tunnel magnetoresistance in magnetic tunnel junctions with a crystalline MgO(0 0 1) barrier,” *J. Phys. D. Appl. Phys.*, vol. 40, no. 21, pp. R337–R354, Nov. 2007.

[137] K. Mizunuma, S. Ikeda, J. H. Park, H. Yamamoto, H. Gan, K. Miura, H. Hasegawa, J. Hayakawa, F. Matsukura, and H. Ohno, “MgO barrier-perpendicular magnetic tunnel junctions with CoFe/Pd multilayers and ferromagnetic insertion layers,” *Appl. Phys. Lett.*, vol. 95, no. 23, p. 232516, 2009.