# Spin-Orbit Torque MRAM

## Introduction

With the advent of mobile and handheld electronic devices, the demand for much smaller, faster and ultra-low power systems keeps growing.
Yet to meet such needs, the microelectronics industry cannot rely anymore on following the Moore's law like it has for the last decades. Decreasing the device size allowed to double the density and speed of integrated circuits every 18 months but also led to an exponential growth of the CMOS static and dynamic power consumption over the years.
As a result, embedded memories, that are used to further improve integrated circuit performances and which represent a major part of the circuits silicon area, have now become a major contributor to power dissipation in integrated systems.
To solve these issues, several technologies are intensively investigated to replace embedded CMOS memories. Among them, Magnetic Random Access Memories (MRAM) are considered as the most promising technology because they potentially combine a fast access time, a low leakage and a good endurance. Although Spin-Transfer Torque (STT) MRAM is currently drawing the most attention, Antaios' Spin-Orbit Torque MRAM (SOT-MRAM[1]), sometimes also referred to as "Spin-Hall", offers unique system-level value thanks to its highest working speed and a truly infinite endurance.

***In the first 15 years of my career I thought the processor and instruction set architecture were the most important things, now I think moving data in and out of the computer is the most important thing.***

*Steve Pawlowski, VP Advanced Computing, Micron Technology*

***Recently, some novel mechanism have been discovered to switch MRAM. Among them, voltage-induced magnetization switching and spin Hall effects are the most promising candidates.***

*ITRS, Beyond CMOS, 2015 Edition*

## MRAM

Antaios SOT-MRAM technology belongs to the larger family of Magnetic Random Access Memories in which the bitcell is based on a Magnetic Tunnel Junction (MTJ).
A magnetic tunnel junction is composed of two ferromagnetic electrodes separated by a tunnel barrier (thin insulating layer). The MTJ stores the binary value as the **magnetization direction** of one of the ferromagnetic thin films, which is called the "Free Layer". The other ferromagnetic layer is called the "Fixed Layer" (or "Reference Layer") for its magnetization never moves.
The electric resistance of the MTJ is larger when the magnetizations of the two ferromagnetic layers are antiparallel (coding a "0"), while it is lower when the two magnetizations are parallel (coding a "1"). This effect is called the tunnel magnetoresistance. **Writing therein consists in switching the magnetization of the free layer from one direction to the other, while reading consists in measuring the resistance state of the MTJ.**
Due to their magnetic nature, MRAM are **non-volatile memories** which means that they do not lose the information when the power is turned off.
Besides, the passive nature of the MRAM bitcell provides a much lower leakage than mainstream CMOS bitcells (SRAM, DRAM) which is interesting for **ultra-low power applications**.

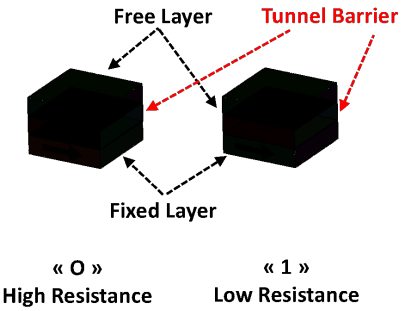


*Figure 1. Magnetic Tunnel Junction (MTJ)*

[1] The denomination « Spin-Orbit Torques » regroups different physical phenomenon among which the Spin-Hall effect (SHE) is the most well-known. Therefore Spin-Orbit Torques RAM and Spin Hall Effect RAM refer to the exact same technology.

## The shortcomings of STT-MRAM

All MRAM generations use the same read mechanism, i.e. the magnetoresistance of the MTJ. What differs between each technology generation is the write mechanism.

In early MRAM marketed by Everspin Technologies (a spin-off from Freescale Semiconductors), switching is performed by magnetic fields. Whilst efficient, this is neither power-savvy, nor scalable to advanced technology nodes.

In STT-MRAM, the write mechanism is the Spin Transfer Torque (STT) effect. The torque is generated by a current pulse flowing through the MTJ and which transfer magnetic momentum from a magnetic reservoir (the fixed layer) to the free layer, where it creates a (magnetic) torque on the magnetization.

In this technology, the current follows the same path for both read and write operations, which makes it very dense. However, it also puts the memory in a 3-way tug of war, whereupon endurance (the maximum number of R/W cycles), speed and data retention cannot be fulfilled simultaneously. As such, a fast magnetization reversal requires a high current which **limits endurance** due to the voltage-induced damages on tunnel barrier. Inversely, a low write voltage can be used only at the cost of a reduced speed and/or "weaker" magnetic layers, hence a **poor data retention, as well as potential read disturb**, e.g. unwanted write during the read cycle.

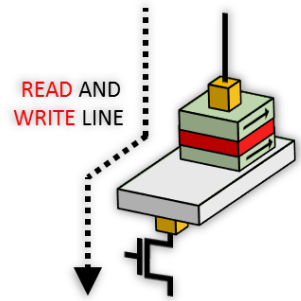


*Figure 2. STT-MRAM bitcell*

## SOT-MRAM

Antaios SOT-MRAM rely on the newly discovered Spin-Orbit Torque (SOT) phenomenon, where the magnetization of a bit cell is switched by an underlying, in-plane current that does not go through the MTJ. Hence SOT-MRAM bitcell uses two separated injection lines for the read and for the write operations, which makes it a three terminal device.

Differentiating the current read and write paths leads to great advantages with respect to STT-MRAM. Since both can be optimized independently, it *de-facto* solves the read disturb issue. By offering **intrinsic unlimited endurance** (no voltage stress on the MTJ during write), it breaks the 3-way tug of war and allows to achieve simultaneously high speed, large endurance and full data retention.

Although innovative in its underlying principles, SOT-MRAM rely on similar core technologies (materials and process) as the currently developed STT MRAM, which makes it fully **CMOS compatible and easy to implement** in semiconductor fabs already involved in STT MRAM.

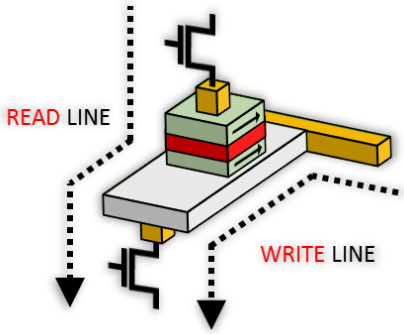


*Figure 3. SOT-MRAM bitcell*

## State of the art writing of SOT-MRAM single cells

Ultra-fast (down to 180 ps) bipolar and deterministic writing of perpendicular three-terminal spin-orbit torque SOT-MRAM single cells have been demonstrated[1][2][3]. The switching current density rises significantly as the pulse shortens below 10 ns which translates into a write energy minimum in the ns range.

In the studied tantalum based MTJ structure the switching current can be estimated to be around **180 µA at 1.5 ns for a 50 nm diameter** MTJ element. Contrary to conventional STT-MRAM that is subject to a large thermal activation time, in SOT-MRAM this incubation time is **negligibly small**. This is a reason why extremely short switching time in the **~100 ps range** can be obtained.

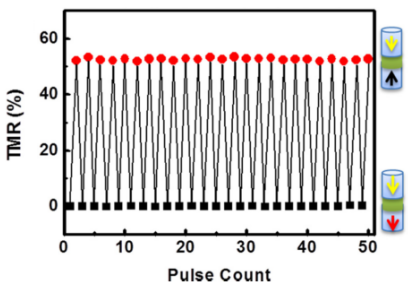


*Figure 4. SOT-MRAM single cell resistance level (TMR) after the injection of positive (black squares) and negative (red circles) current pulses.*

## Comparison with other memory technologies

To effectively compare SOT-MRAM to various memory technologies, the simulated performances of a 512kB memory case study are summarized in Figure 5.[4][5]

As showed, SOT-MRAM is comparable to SRAM in terms of performance and is **largely superior when it comes to energy consumption and area**. Overall, the SOT-MRAM performances are:

- High speed (GHz)
- Lower area (<< SRAM)
- Non-volatility
- Unlimited endurance
- Low power
- Ultra-low leakage
- Very high reliability

Which such a set of performance Antaios embedded SOT-MRAM is **the only memory technology that can replace SRAM in microprocessor caches** at a lower cost and with a largely improved power consumption.

| Attributes | SRAM | STT MRAM | SOT MRAM |
|---|---|---|---|
| Area [mm²] | 2.8 | 1.6 | 1.8 |
| Read Latency [ns] | 2.2 | 1.2 | 1.1 |
| Write Latency [ns] | 2.1 | 11.2 | 1.4 |
| Read Energy [pJ] | 587 | 260 | 247 |
| Write Energy [pJ] | 355 | 2337 | 334 |
| Leakage [mW] | 932 | 387 | 254 |

*Figure 5. Performance estimations for a 512 kB memory in TSMC 65nm GP Process using NVSim. Memories are optimized for latency.*

## SOT-MRAM for Cache Memories in Microprocessors

The impact of Antaios SOT- MRAM technology at the cache memory level has been simulated for both a single-core and a multicore microprocessor, both depicted in Fig.6 and 7[4][5]. In the single-core architecture without L3-cache both the L1 and L2 SRAM cache memories have been tentatively replaced by Antaios SOT-MRAM blocks. At constant (imposed) performances, replacing L2 SRAM with SOT-MRAM provides **significant area savings (up to 40%)** and **the average energy consumption is importantly reduced by 60%** compared to an SRAM-only solution. Using SOT-MRAM for both cache-levels does not offer additional gain.

In the multicore configuration, the shared L3 cache is implemented either with SRAM or with SOT-MRAM. Replacing the 16MB SRAM memory by a 16MB SOT-MRAM memory offers **a considerable area (~45%) and energy advantage (~70%) over SRAM.** When looking for performances, some of the important savings offered by SOT-MRAM can be used to increase the size of the L3-cache. **In this case doubling the SOT-MRAM cache capacity improves the performance by more than 4% compared to a 16MB SRAM cache,** while the area is still considerably **smaller (~15%) and the energy savings are still very impressive (~70 %).**

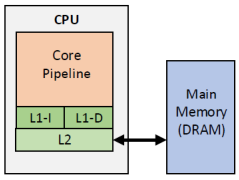


*Figure 6. Single core configuration: Processor: single core, 3 GHz L1: 32 KB L2: 512 KB*

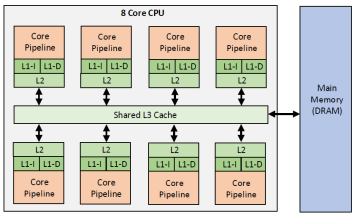


*Figure 7. Multi core configuration: 8 times the single core Shared L3 cache of 16/32 MB*

## References


[1] Cubukcu, M. *et al.* Spin-orbit torque magnetization switching of a three-terminal perpendicular magnetic tunnel junction. *Appl. Phys. Lett.* **104,** (2014).

[2] Cubukcu, M. *et al.* Ultra-fast magnetization reversal of a three-terminal perpendicular magnetic tunnel junction by spin-orbit torque. arXiv:1509.02375 (2015).

[3] Garello, K. *et al.* Ultrafast magnetization switching by spin-orbit torques. *Appl. Phys. Lett.* **105,** (2014).

[4] Oboril, F. *et al.* Evaluation of hybrid memory technologies using SOT-MRAM for on-chip cache hierarchy. *IEEE Trans. Comput. Des. Integr. Circuits Syst.* **34,** 367–380 (2015).

[5] Prenat, G. *et al.* Ultra-Fast and High-Reliability SOT-MRAM: From Cache Replacement to Normally-Off Computing. *IEEE Trans. Multi-Scale Comput. Syst.* **2,** 49–60 (2016).