

Beyond STT-MRAM, Spin Orbit Torque RAM SOT-MRAM for High Speed and High Reliability Applications

Guillaume Prenat, Kotb Jabeur, Gregory Di Pendina, Olivier Boule,
and Gilles Gaudin

1 Introduction

For 40 years, microelectronics has been following the Moore's law, stating that the density and speed of integrated circuits would double every 18 months. However, this trend is presently getting out of breath, because of incoming insurmountable physical limits. Due to decreasing devices size, leakage current is becoming the main contributor to power dissipation of CMOS. Indeed, the increased density and reduction in die size lead to heat dissipation and reliability issues. Moreover, the dynamic power keeps on growing up with both clock frequency and global capacitance while the power supply is not scaled down accordingly. Several solutions are investigated to try to push forward these limits at technology, circuit or architecture levels. The "more than Moore" concept consists in using new devices beside or in replacement of standard CMOS transistors. For instance, the use of non-volatile devices is seen as a promising solution to reduce power consumption, improve reliability and offer new functionalities. Several technologies are intensively investigated like Phase Change Random Access Memory (PCRAM), Ferroelectric RAM (FeRAM), RedoxRAM (ReRAM) and Magnetic RAM (MRAM). In its 2010 report, ITRS identified RedoxRAM and STT-MRAM as the two most promising technologies for embedded memories at technology nodes below 16 nm. The combination of non-volatility, fast access time and endurance in MRAM technology paves the path toward a universal memory. Although an expanding attention is given to two-terminal Magnetic Tunnel Junctions (MTJ) with writing based on Spin-Transfer Torque (STT) switching as the potential candidate for future memories, it suffers

G. Prenat (✉) • K. Jabeur • G. Di Pendina • O. Boule • G. Gaudin
University of Grenoble Alpes, SPINTEC, Grenoble 38000, France
CEA, INAC-SPINTEC, Grenoble 38000, France
CNRS, SPINTEC, Grenoble 38000, France
e-mail: guillaume.prenat@cea.fr

from weaknesses. Indeed, two main shortcomings are still limiting the reliability and endurance of STT-MRAMs: i) The high current density required for writing can occasionally damage the MTJ barrier, specially for switching on the nanosecond time scale ii) It remains a challenge to fulfill a reliable reading without ever causing switching for very advanced technology nodes, since writing and reading operations share the same path, through the junction. Indeed, the smaller the MTJ the lower the writing current without having the possibility to reduce the reading current to maintain a reliable sensing. Three-terminal MTJ with writing based on Spin-Orbit Torque (SOT) approach revitalizes the hope of an ultimate RAM. It represents a pioneering way to triumph over current two-terminal MTJ's limitations by separating the reading and the writing paths, completely avoiding tunnel barrier damaging and read disturb issues.

2 MTJs Written by Spin Orbit Torque

Recently, a new effect was discovered, leading to original devices with three terminals. This effect is called Spin Orbit Torque (SOT), a generic term encompassing in reality different possible mechanisms originating from the bulk of the material and the interfaces between the different layers. These mechanisms issued from the spin orbit interaction are often named the Rashba effect and the Spin Hall Effect (SHE) [1–4] for the contribution of respectively the interfaces and the bulk. However this denomination is often a shortcut since the contribution from interfaces and bulk are intricated. SOT allows switching the magnetization of a magnetic Storage Layer (SL) by passing a current in the plane of the conductive contact line underneath the MTS, instead of through the MTJ itself (Fig. 1). This allows separating the reading and writing paths with great benefits: firstly, since no large writing current is passing through the tunnel barrier damaging it, the

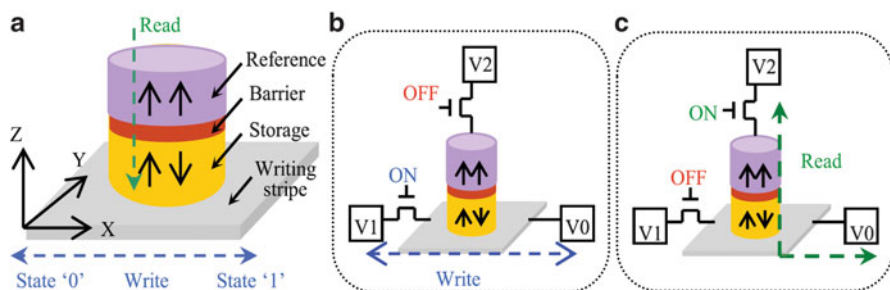


Fig. 1 (a) Schematic of the three-terminal SOT device and the two independent paths for (b) write and (c) read operations. In-plane current injection through the write line (writing stripe) induces the perpendicular switching of the storage layer

endurance can be truly infinite. Secondly, risks of accidental writing during reading are drastically reduced. Moreover, the switching duration using this scheme can be very fast, typically some hundreds of picoseconds experimentally demonstrated, providing the injection of relatively large current densities. Compared to STT [5, 6] the fast SOT-switching is deterministic and does not show any precessional behavior [7]. Today, the current density is higher than for STT (typically $5 \cdot 10^7$ A/cm²), but since it flows through a very thin line (typically 5 nm), the resulting writing current can be very small. Moreover, since the writing time can be very short, the resulting energy becomes very interesting. The main limitation of this technology compared to STT is apparently the area density because of the additional terminal. However, this picture is not so simple considering a complete memory since i) transistors can be shared ii) the area of MRAM is often limited by the size of the addressing transistors more than that of the single magnetic cell. However, since the writing current decreases proportionally to the lateral dimension of the device and not with its surface, the scalability is weaker than for STT writing, but remains much better than for magnetic field writing. An important feature of the device is the thermal stability of the device which is ensured down to 20 nm.

Another specificity of this writing scheme is that a small static longitudinal magnetic field is required to avoid stochastic switching: without it, for a given direction of the writing current, the magnetization can switch randomly in either P (Parallel) or AP (Anti Parallel) configuration. Adding this longitudinal magnetic field makes the switching deterministic, with a direction of the current writing the P state and the opposite direction the AP state. This field can be generated by the stray field of an in-plane magnetized layer, integrated to the pillar stack, acting like a permanent magnet, with no additional power consumption. This particularity can be even turned into an advantage. Indeed, if the direction of the magnetic field is switched, the operation of the device is inverted: the direction of the current previously used to write a logic “0” will write a logic “1” and vice-versa (Fig. 2). It can be useful for reconfigurable computing. In this case, the magnetization of the layer generating the magnetic field could be chosen using a current line to control it like in original Field Induced Magnetic Switching (FIMS) writing schemes. The corresponding power consumption would then depend on the reconfiguration frequency.

Although this technology is at very early stage development, these properties make it particularly interesting for high speed applications that do not require very high densities but high endurance. Typically, it can be suitable for low-cache levels replacement in processors, which is not possible with classical STT. Since this technology solves some of the major issues of STT, a lot of academic organisms and industrial companies are intensively working on it and we have good confidence that it can have a predominant place in the future of spintronics, for a lot of applications.

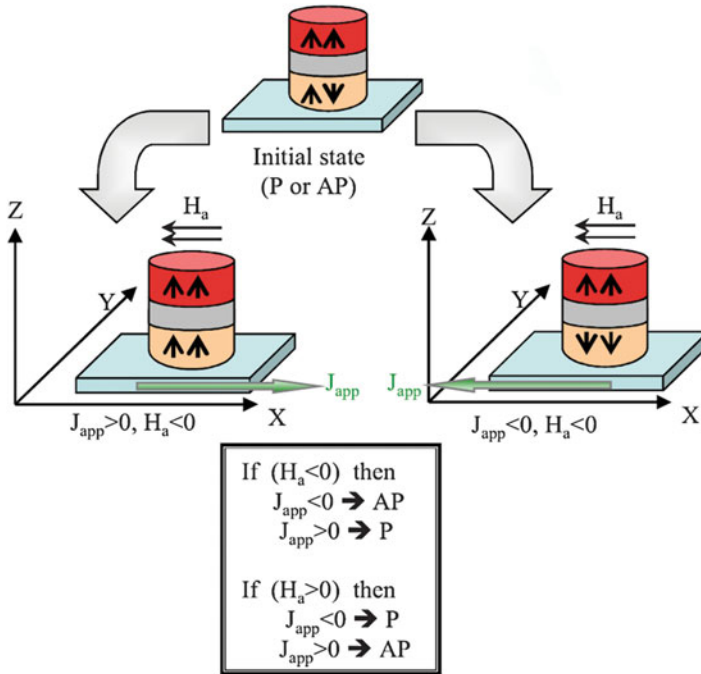


Fig. 2 Switching dependence on the external longitudinal applied field H_a and current direction

3 Applications of SOT-RAM

3.1 Introduction in the Memory Hierarchy of Processors

As seen in the previous chapters, a lot of studies have been recently carried out to introduce STT-RAM in the memory hierarchy of processors. Indeed, it is the only technology that offers an intrinsic non-volatility together with a speed suitable for use as a working memory and a quasi-infinite endurance. Adding non-volatility in the memory hierarchy has a lot of advantages, in particular the ability to ease the power gating techniques, which consist in cutting off the power supply of unused blocks to drastically reduce the leakage. Indeed, having a local non-volatility allows directly cutting off the power supply without having to save the content of the memory in distant non-volatile or very low-leakage storage devices. The first and more obvious approach consists in directly replacing the parts of the memory hierarchy whose operating speed is compatible with those of STT-RAM (DRAM main memory, or L2 cache level). Going deeper into the memory hierarchy seems difficult since STT-RAM cannot compete with fast SRAM in terms of writing speed. However, some works have shown that the better density of MRAM allows

having a larger L1 cache capacity, which can compensate the loss of speed for some applications [8]. This approach requires rethinking the architecture of processors and memory hierarchy and is adapted only to some specific applications. Other works propose the use of a hybrid SRAM/MRAM memory. Several architectures have been proposed, allowing using the memory as a standard SRAM in operation, but with the capability of backing-up the content in MTJs and restore it at any moment. This allows bringing non-volatility in low levels of cache, but with a loss in terms of density.

Since SOT-MRAM have a writing speed approaching that of SRAM, it is possible to consider directly replacing L1 cache by SOT-RAM. The result would be the introduction of non-volatility in L1 cache, without penalty in terms of speed. Moreover, thanks to its density which remains much better than the one of SRAM, larger caches could be used, resulting in probable performance improvement.

3.2 Non-volatile Flip-Flops, Normally-Off/Instant-On Computing and Memory-in-Logic

The architecture of latches and Flip-Flops (FF), distributed memory containing the active data of the logic part of the circuit, is very similar to that of SRAM. It is thus possible to introduce non-volatility in the logic itself, in the same way that SRAM can be made non-volatile [9–11]. It is particularly interesting since it allows freezing the state of the processor or any digital/mixed signal IC, and restart in the same state. It allows developing new paradigms like normally-off/instant-on computing [12], in which the circuit is off by default and power supplied only when required. It can be switched-off instantaneously and restarted in the same state. This still reduces the power consumption but also improves the reliability. It is possible for instance to implement a checkpoint for rollback operation in processors, to restore the system state with a snapshot taken in the past. Thanks to its high speed, low-energy writing and infinite endurance, SOT-MRAM could allow saving the context more often, even possibly at each clock edge, without significant energy overhead or performance loss and thus still improving the reliability.

Although the intrinsic non-volatility of MTJs naturally encourages using them as memory elements, it also offers the possibility to intrinsically mix logic and memory to shorten and multiply the communication between them, leading to a new paradigm in logic architectures. This concept, called “logic-in-memory” has been proposed in [13]. Examples of circuits based on Current Mode Logic (CML) have been proposed and implemented in silicon demonstrators giving encouraging results [14]. For this purpose, the SOT-MRAM has another advantage thanks to its third terminal offering an additional degree of freedom and the possibility to reverse its behavior by changing the direction of the longitudinal magnetic field.

4 Design Tools and Environment for the Design of Hybrid CMOS/SOT-RAM Components

4.1 Compact Electrical Model of SOT-MTJ

As the semiconductor industry is progressively going toward “hybrid CMOS” Integrated Circuits (ICs), compact model development has become a cornerstone in the circuit/system design and verification tool flow.

Recently, a compact model of the SOT-MRAM has been proposed by Spintec [15]. The most significant physical phenomena involved in the writing (SOT) and the reading (TMR) are taken into account and realistic parameters, obtained from characterization of single SOT-MRAM cells [16] were used to calibrate the model. This model is written in Verilog-A language and based on physical equations modeling the behavior of the device. The dynamics of the magnetization is described, based on the Landau-Lifshitz-Gilbert (LLG) equation [17], including additional SOT terms [1–4, 16, 18]. The dependence of the resistance upon the magnetic configuration (Tunnel Magneto-Resistance or TMR, relative resistance variation of the stack between Parallel and Anti-Parallel resistance states) and the applied voltage is described using Jullière’s model [19] and Simmons’ model [20]. Moreover, for an improved accuracy, we have integrated the dynamic conductance given by the Brinkman model [21] and we take into consideration the dependence of TMR upon bias voltage. In this model, all the physical quantities, such as the magnetization components (m_x , m_y , m_z) are represented by voltages. Figure 3 shows a physical view and a block diagram of the model, with two modules, one dedicated to the computation of the dynamics of the magnetization and the other dedicated to the TMR.

Figure 4 shows a simulation result using SPECTRE electrical simulator from Cadence. It illustrates the operation of the device, by showing the dynamics of the m_z component of the magnetization for different values of the writing current density J_{app} . In this particular case, the magnetization of the two magnetic layers constituting the MTJ is oriented perpendicularly to the layers. The value of m_z is then 1 for P state and -1 for AP state. Intermediate values correspond to transient behavior during the switching process. The duration of the writing pulse is 3 ns. We see that during the current pulse application the magnetization aligns in the plane of the layers ($m_z = 0$). Once the magnetization is stabilized, the current is removed and the magnetization switches in its final state according to the direction of the current. For $J_{app} = 10^{12}$ A/m², the SOT is not strong enough to initiate the switching and the MTJ remains in its P state. For $J_{app} = 2 \cdot 10^{12}$ A/m², the 3 ns pulse is not long enough for the magnetization to stabilize and the MTJ switches back to its initial P state. For higher values of J_{app} , switching occurs, with a duration decreasing with the increase of the current density, allowing a trade-off between speed and power consumption, depending on the targeted application. In the same way, increasing the pulse duration allows improving the reliability of the switching. We see that for high current density values ($4 \cdot 10^{12}$ A/m² for instance), the switching time can be less than 1 ns. In Fig. 5, the influence of the longitudinal magnetic field on the switching is illustrated. Without it, the switching is stochastic. When the field is applied, the

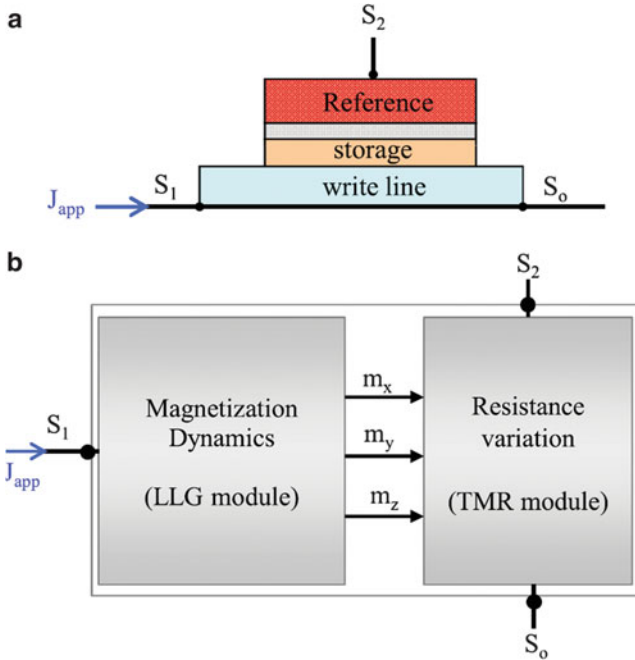


Fig. 3 Modeling strategy of the three-terminal SOT-MRAM. (a) Physical view of the SOT-MRAM (cross section). (b) Block diagram of the compact model

switching becomes deterministic, with a switching time duration decreasing with the increase of the magnitude of the field. This model was successfully validated by comparison of the results with micromagnetic simulations.

4.2 Ultra-fast SOT-MRAM Based Non-volatile Flip-Flop

In order to illustrate the potential of SOT-MRAM, we describe here a Non-Volatile Flip-Flop (NVFF) architecture based on this technology (Fig. 6), developed at Spintec [22]. It is composed of a master latch made non-volatile using a pair of SOT-MRAMs (Fig. 6b) connected to a standard slave latch (Fig. 6c). The operation is managed by a non-overlapping two-phase clock signal generator (ck1, ck2). Four transistors are used to generate a bidirectional current flowing in the writing lines of the SOT-MRAMs devices in series. The direction of the current is determined by the input data D. The current is generated during the first phase clock ck1. During the second phase clock ck2, the master latch reads the data of the SOT-MRAMs and transfers it to the output Q. This operation is illustrated in Fig. 7, showing a simulation of the NVFF using the compact model presented in the previous section. In this work, the technology node is 40 nm, for both CMOS and SOT-MRAM. In these conditions, the writing current is 60 μA , comparable to perpendicular STT-MRAM. However as

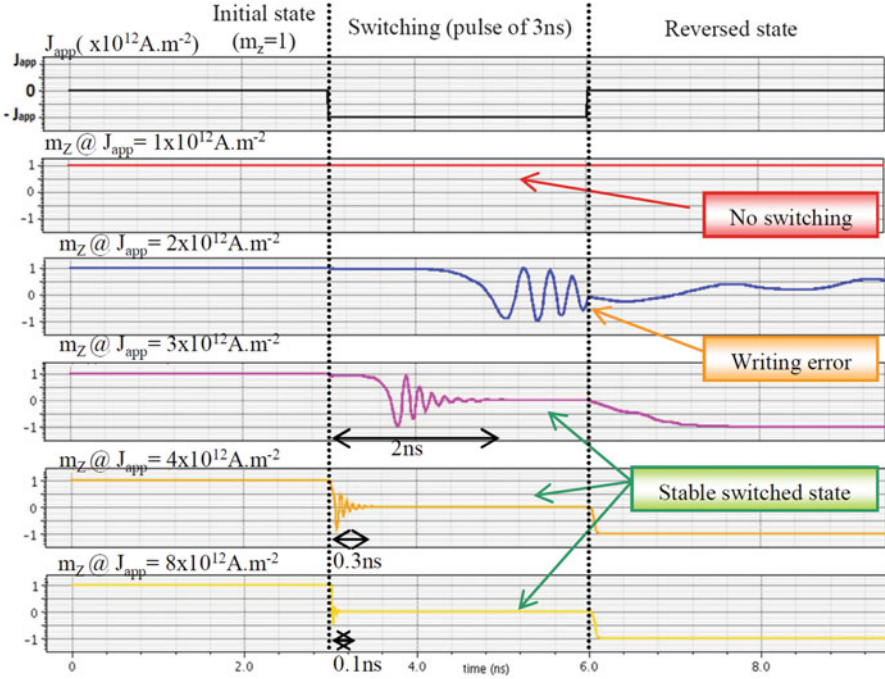


Fig. 4 Simulation of the dynamics of the perpendicular magnetization m_z of a SOT-MRAM as a function of the injected current

seen in inserts, the switching time is only 250 ps, four times faster than STT, dividing the energy by four.

The operation of this NVFF fundamentally differs from those presented so far [9–11]: the magnetic part is written at every clock cycle contrary to the previous cases, where the data was backed up in the magnetic part only when required, on demand. This is made possible because of the high writing speed and low-power consumption of the SOT technology. It means that with such a device, the power supply can be cut-off at any moment, without further action and restarted in the same state. This opens the door to the concept of “normally-off” or “instant-on” circuits [12]. This Flip-Flop can be used as a primitive cell into the ASIC design library allowing the design of non-volatile, high speed and low-power digital circuits [23, 24].

4.3 System Level Integration

To evaluate the benefits that can be expected from integrating MRAM and in particular SOT-MRAM in the different levels of processors hierarchy, it is necessary to carry-out system level simulations, using processor simulators, like Gem5 or

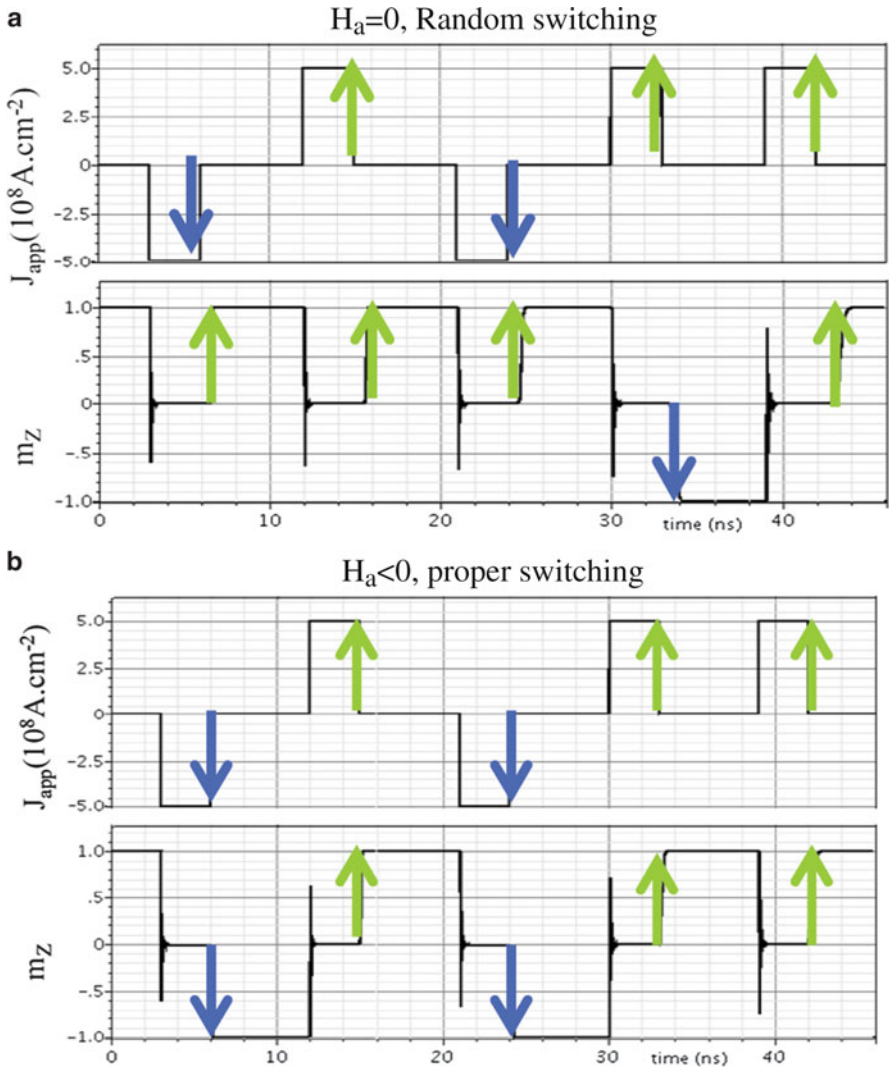


Fig. 5 Study of the dynamic behavior of the perpendicular magnetization m_z according to the value of the external applied field H_a : (a) $H_a = 0$, random switching and (b) $H_a < 0$, proper switching

simple scalar (Fig. 8). These tools take as inputs the architecture of the processor, including the memory hierarchy and the models of the memory blocks. It allows running an application and evaluating the performance in terms of CPU time, number of cycles, memory transactions (read/write, cache hit/miss. . .). The models of the memory can be provided by characterization of an existing memory (by test or simulation) or using a memory simulator, like NVSIM tool (specific for non-volatile resistive memories). This work is currently carried out for MRAM in general [8] and for SOT-MRAM in particular, giving encouraging preliminary results [25].

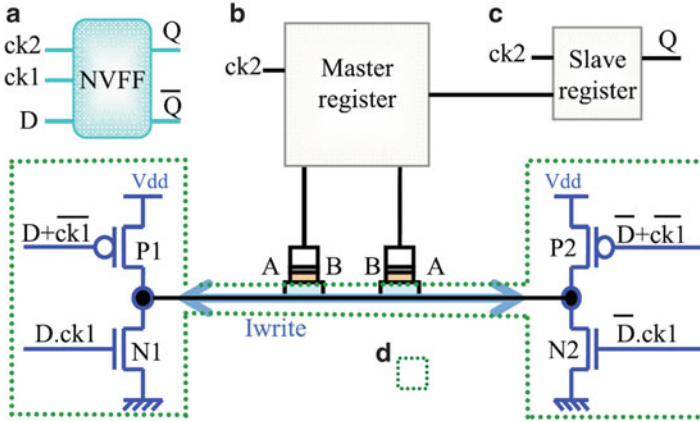


Fig. 6 SOT-MRAM based NVMMFF (a) symbol, (b) master register architecture, (c) slave register, (d) writing circuit

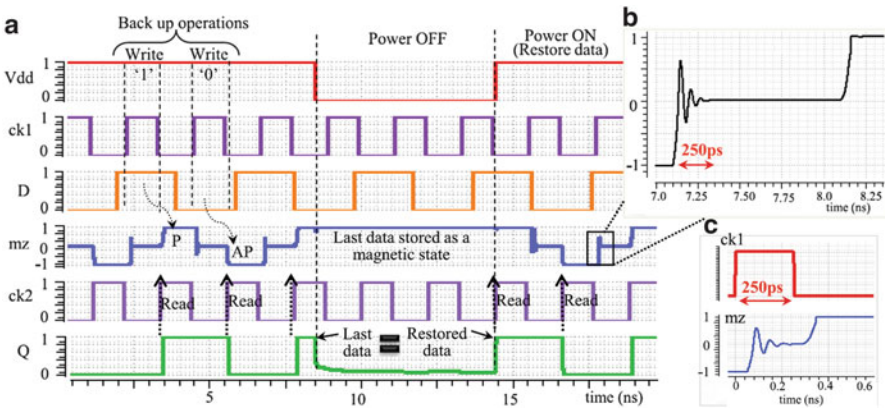


Fig. 7 SOT-MRAM based NVMMFF simulation results (a) write and read operations (b) perpendicular magnetization m_z behavior of the SOT-MRAM during a 1 ns current pulse to switch from AP ($m_z = -1$) to P ($m_z = 1$) (c) SOT-MRAM switching with only a 250 ps pulse (exactly after oscillations stabilization)

5 Conclusion

The advent of new MRAM generations has led to a keen interest in the microelectronics world for this technology as a possible solution to push forward the limits of microelectronics. Several tools are developed to allow exploring new circuits and system architectures using MRAM devices. As seen in this chapter, both STT-MRAM and SOT-MRAM technologies have their own strengths and limits and is thus suitable for a given set of applications. A wide field of investigation is open covering technology developments to improve the existing devices, discover

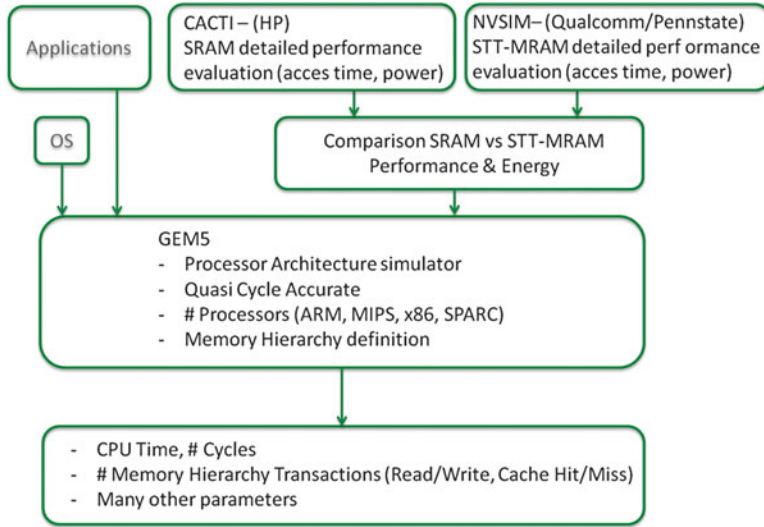


Fig. 8 System level design flow to evaluate the performance of processors embedding MRAM

new ones, and design of innovative circuits and systems. Two paths must be followed in parallel: identifying relevant target applications for each technology and adapting the system architectures to gain maximum benefit.

The performance of the most recent technologies and in particular SOT-MRAM, in terms of speed and power consumption which could compete with SRAM, paves the way towards a true revolution in the memory hierarchy. It allows dreaming of a new paradigm consisting in introducing non-volatility at all levels of this memory hierarchy and possibly writing the data in the MTJs at every clock cycle, opening the door to real normally-off/instant-on circuits.

Acknowledgement This work has been partially funded by the European Commission under the spOt project (grant agreement n318144) in the framework of the Seventh Framework Program.

References

1. M. Miron, K. Garello, G. Gaudin, P.-J. Zermatten, M.V. Costache, S. Auffret, S. Bandiera, B. Rodmacq, A. Schuhl, P. Gambardella, Perpendicular switching of a single ferromagnetic layer induced by in-plane current injection. *Nature* **476**, 89–193 (2011)
2. M. Miron, G. Gaudin, S. Auffret, B. Rodmacq, A. Schuhl, S. Pizzini, J. Vogel, P. Gambardella, Current-driven spin torque induced by the Rashba effect in a ferromagnetic metal layer. *Nat. Mater.* **9**, 230–234 (2010)
3. P. Gambardella, I.M. Miron, Current-induced spin-orbit torques. *Philos. Trans. A Math. Phys. Eng. Sci.* **369**, 3175–3197 (2011). doi:[10.1098/rsta.2010.0336](https://doi.org/10.1098/rsta.2010.0336)

4. L. Liu, C.-F. Pai, Y. Li, H.W. Tseng, D.C. Ralph, R.A. Buhrman, Spin-torque switching with the giant spin hall effect of tantalum. *Science* **336**, 555–558 (2012)
5. C. Pappas, B. Delaët, B. Rodmacq, D. Houssameddine, J.-P. Michel, U. Ebels, R.C. Sousa, L. Buda-Prejbeanu, B. Dieny, 100 ps precessional spin-transfer switching of a planar magnetic random access memory cell with perpendicular spin polarizer. *Appl. Phys. Lett.* **95**, 072506 (2009)
6. M. Marins de Castro, R.C. Sousa, S. Bandiera, C. Ducruet, A. Chavent, S. Auffret, C. Pappas, I.L. Prejbeanu, C. Portemont, L. Vila, U. Ebels, B. Rodmacq, B. Dieny, Precessional spin-transfer switching in a magnetic tunnel junction with a synthetic antiferromagnetic perpendicular polarizer. *J. Appl. Phys.* **111**, 07C912 (2012)
7. K. Garello, C.O. Avci, I.M. Miron, M. Baumgartner, A. Ghosh, S. Auffret, O. Boulle, G. Gaudin, P. Gambardella, Ultrafast magnetization switching by spin-orbit torques. *Appl. Phys. Lett.* **105**, 212402 (2014)
8. L. Torres, R. Brum, L. Cargini, G. Sassatelli, in *Trends on the Application of Emerging Nonvolatile Memory to Processors and Programmable Devices*. Circuits and Systems (ISCAS), 2013 I.E. International Symposium (2013), pp. 101–104
9. W. Zhao, E. Belhaire, V. Javerliac, C. Chappert, B. Dieny, in *A Non-volatile Flip-Flop in Magnetic Fpga Chip*. Design and Test of Integrated Systems in Nanoscale Technology, 2006 DTIS. International Conference (2006), pp. 323–326
10. N. Sakimura, T. Sugibayashi, R. Nebashi, N. Kasai, in *Nonvolatile Magnetic Flip-Flop for Standby-Power-Free Socs*. Custom Integrated Circuits Conference, CICC 2008. IEEE (2008), pp. 355–358
11. Y. Guilleminet, L. Torres, G. Sassatelli, N. Bruchon, On the use of magnetic rams in field programmable gate arrays. *Int. J. Reconfig. Comput.* **2008**, 1:1–1:9 (2008). <http://dx.doi.org/10.1155/2008/723950>
12. T. Kawahara, Scalable spin-transfer torque ram technology for normally-off computing. *IEEE Des. Test Comput.* **28**(1), 52–63 (2011)
13. A. Mochizuki, H. Kimura, M. Ibuki, T. Hanyu, Tmr based logic-in-memory circuit for low-power vlsi*. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **E88-A**(6), 1408–1415 (2005). <http://dx.doi.org/10.1093/ietfec/e88-a.6.1408>
14. S. Matsunaga, J. Hayakawa, S. Ikeda, K. Miura, H. Hasegawa, T. Endoh, H. Ohno, T. Hanyu, Fabrication of a nonvolatile full adder based on logic-in-memory architecture using magnetic tunnel junctions. *Appl. Phys. Express* **1**(9), 091301 (2008). <http://apex.jsap.jp/link?APEX/1/091301/>
15. K. Jabeur, G. Prenat, G. Di Pendina, L. Buda-Prejbeanu, I. Prejbeanu, B. Dieny, in *Compact Model of a Three Terminal Mram Device Based on Spin Orbit Torque Switching*. Semiconductor Conference Dresden-Grenoble (ISCDG) 2013 International (2013), pp. 1–4.
16. K. Garello, I.M. Miron, C.O. Avci, F. Freimuth, Y. Mokrousov, S. Blügel, S. Auffret, O. Boulle, G. Gaudin, P. Gambardella, Symmetry and magnitude of spin-orbit torques in ferromagnetic heterostructures. *Nat. Nanotechnol.* **8**, 587 (2013)
17. L.D. Landau, E. Lifshitz, On the theory of the dispersion of magnetic permeability in ferromagnetic bodies. *Phys. Z. Sowjetunion* **8**(153), 101–114 (1935)
18. S. Manipatruni, D.E. Nikonov, I.A. Young, Voltage and energy-delay performance of giant spin hall effect switching for magnetic memory and logic. **arXiv:1301.5374v1** [cond-mat.mes-hall] (Submitted on 23 Jan 2013)
19. M. Julliere, Tunneling between ferromagnetic films. *Phys. Lett. A* **54**, 225–226 (1975)
20. J.G. Simmons, Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film. *J. Appl. Phys.* **34**(6), 1793–1803 (1963)
21. W.F. Brinkman, R.C. Dynes, J.M. Rowell, Tunneling conductance of asymmetrical barriers. *J. Appl. Phys.* **41**, 1915–1921 (1970)
22. K. Jabeur, G. Di Pendina, F. Bernard-Granger, G. Prenat, Spin orbit torque non-volatile flip-flop for high speed and low energy applications. *Electron Device Lett. IEEE* **35**(3), 408–410 (2014). doi:[10.1109/LED.2013.2297397](https://doi.org/10.1109/LED.2013.2297397). <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&number=6714403&isnumber=6746036>

23. G. Di Pendina, K. Torki, G. Prenat, Y. Guillemenet, L. Torres, in *Integrated Circuit and System Design. Power and Timing Modeling, Optimization, and Simulation*. Ultra Compact Non-volatile Flip-Flop for Low Power Digital Circuits Based on Hybrid Cmos/Magnetic Technology (Springer, Berlin Heidelberg, 2011), pp. 83–91
24. G. Di Pendina, G. Prenat, B. Dieny, K. Torki, A hybrid magnetic/complementary metal oxide semiconductor process design kit for the design of low-power nonvolatile logic circuits. *J. Appl. Phys.* **111**(7), 07E350–07E350–3 (2012)
25. R. Bishnoi, M. Ebrahimi, F. Oboril, M.B. Tahoori, in *Architectural Aspects in Design and Analysis of SOT-based Memories*. Design Automation Conference (ASP-DAC), 20–23 Jan 2014 19th Asia and South Pacific, (2014), pp. 700, 707. doi:0.1109/ASPDAC.2014.6742972