

## MTJ-BASED HYBRID STORAGE CELLS FOR “NORMALLY-OFF AND INSTANT-ON” COMPUTING

**Bojan Jovanović<sup>1</sup>, Raphael M. Brum<sup>2</sup>, Lionel Torres<sup>2</sup>**

<sup>1</sup>University of Niš, Faculty of Electronic Engineering, Niš, Serbia

<sup>2</sup>LIRMM Laboratory, University of Montpellier 2, Montpellier, France

**Abstract.** *Besides increasing a computing throughput, multi-core processor architectures bring increased capacity of SRAM-based cache memory. As a result, cache memory now occupies large proportion of recent processor chips, becoming a major source of the leakage power consumption. The power gating technique applied on a SRAM cache is not efficient since it is paid by data loss. In this paper, we present two hybrid memory cells that combine a conventional volatile CMOS part with Magnetic Tunnel Junctions (MTJs) able to store a data bit in a non-volatile way. Being inherently non-volatile, these hybrid cells enable instantaneous power off and thus complete reduction of the leakage power. Moreover, given that the data bit can be stored in local MTJs and not in distant storage memories, these cells also offer instantaneous and efficient data retrieval. To demonstrate their functionality, the cells are designed using 28 nm FD-SOI technology for the CMOS part and 45 nm round spin transfer torque MTJs (STT-MTJs) with perpendicular magnetization anisotropy. We report the measured performances of the cells in terms of required silicon area, robustness, read/write speed and energy consumption.*

**Key words:** *Hybrid MTJ/CMOS cells, magnetic tunnel junction (MTJ), spin transfer torque (STT), normally-off instant-on computing*

### 1. INTRODUCTION

Conventional Von-Neumann computing architectures consist of a pure computational part (central processor unit - CPU) and a memory part in which the computing recipes (programs) and the input/output data of the calculations are stored [1]. Such complex systems have a memory hierarchy comprising different semiconductor memory types, as illustrated in Fig. 1. Dense, slow and non-volatile storage memory with limited endurance is combined with fast, volatile, power and area consuming SRAM/DRAM working memory (located close to the CPU) in order to ensure both rapid accessibility and data non-volatility. However, this sort of design hierarchy requires complex control. Start-up (booting) and shut-down procedures usually take a long time and waste a significant amount of power since

---

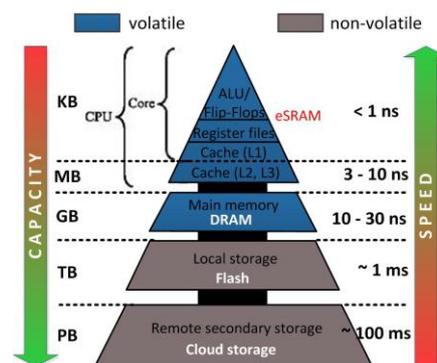
Received January 20, 2015; received in revised form March 16, 2015

**Corresponding author:** Bojan Jovanović

University of Niš, Faculty of Electronic Engineering, Niš, Serbia

(e-mail: bojan@elfak.ni.ac.rs)

they imply extensive data traffic (from storage memories to working memories and vice-versa). In recent years, both the limited clock frequency of the processor and the emergence of multi-core architectures led to a significant increase in working memory capacity. As a result, the performance and the power of the computing system became determined by working, SRAM-based, memory. It occupies most of the chip area, consumes most of the static power and is prone to soft errors caused by radiation [2]. Replacing conventional six-transistor (6T) SRAM cells with four-transistors (4T) counterparts did not solve all these issues. Although they occupy slightly less silicon area, 4T-SRAM memory cells consume more leakage power and exhibit poor data stability. Furthermore, 4T-SRAM cells still limit system performance as they require complex control and communication with the non-volatile storage elements [3].



**Fig. 1** Typical structure of a computer memory hierarchy.

To circumvent these limitations, non-volatility needs to be brought directly to the working memory cell. This would pave the way for new green computing paradigm based on “normally-off and instant-on” operation. Computing equipment could be quickly turned-off when not in use, keeping the off state with zero stand-by power as long as possible. On the computing request, the equipment could be turned on instantly, with the full performance capabilities. Such computing approach may be far more energy efficient compared with the current “normally-on” computing systems [4].

Among the non-volatile devices that are prospective candidates for co-integration with CMOS, spin-based magnetic tunnel junctions (MTJs) are the most promising [5]. Unlike the other candidates in which the position of atoms (e.g. ferroelectric RAM - FeRAM [6]) or the whole structure (e.g. phase change memory - PCM [7]) have to be changed to define a non-volatile state, spin-based MTJs are controlled only by electron spin [8]. In addition to energy efficiency (little energy is needed to change the electron spin), MTJs provide radiation immunity, high speed data switching, higher density, infinite endurance as well as the ability to continue shrinking in size [9]. Moreover, they can be very easily co-integrated with the CMOS without imposing the area overhead, as illustrated in Fig. 2a).

In this paper, we present two hybrid cells that combine CMOS transistors with perpendicular spin-transfer torque MTJs (STT-MTJs) as non-volatile storage elements. The cells can be considered as hybrid alternatives for the mainstream 4T- and 6T-SRAM cells. They can store a data bit in both volatile and non-volatile contexts. Furthermore, the cells are

able to quickly and efficiently transfer a data bit from one context to another, thus supporting the "normally-off and instant-on" computing concept.

The remainder of the paper is organized as follows: in Section 2, we analyze the evolution of the MTJ writing mechanisms. In Section 3, we introduce our hybrid cells that contain four-transistors and two-MTJs (4T-2M) and six-transistors and four-MTJs (6T-4M), explaining their structure and functionality. In Section 4, we report the measured performance of the cells in terms of required silicon area, robustness, leakage, read/write speeds and energy consumption. Finally, Section 5 is reserved for our conclusions.

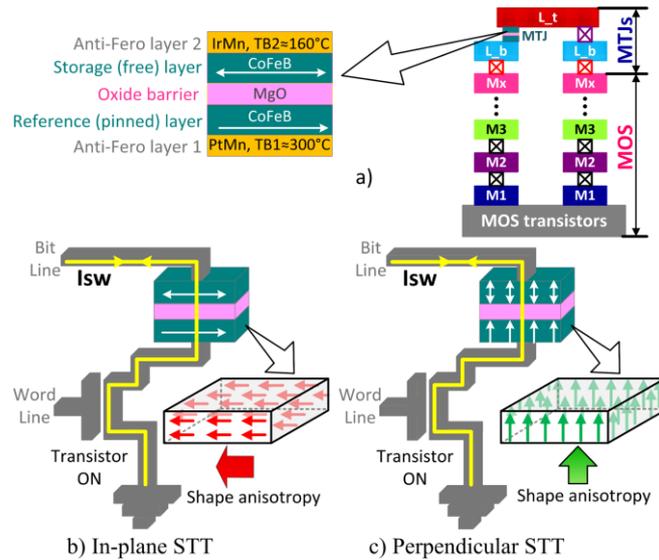
## 2. EVOLUTION OF MTJ WRITING MECHANISMS

An MTJ is a nanopillar composed of an ultra thin layer of insulator (oxide barrier) sandwiched between two ferromagnetic (FM) metals (Fig. 2a). The insulating layer is so thin that electrons can tunnel through the barrier if a bias voltage is applied between two FM electrodes. The resistance of MTJ depends on the relative orientation of the magnetization in the two FM layers. In standard applications, the magnetization of one FM layer (the reference layer) is commonly pinned, whereas the other (storage) layer is free to take a parallel (P) or an anti-parallel (AP) orientation, thus determining parallel ( $R_p$ ) or anti-parallel ( $R_{ap}$ ) MTJ resistance and storing a binary state. The relative difference between these two resistances defines the tunnel magneto-resistance (TMR) ratio,  $\Delta R/R=(R_{ap}-R_p)/R_p$ . In recent decades, much research effort has been invested in improving the TMR ratio of MTJs to make them more attractive for integration with CMOS. Today, commercial MTJs that use MgO oxide barriers have a TMR of about 200% [10], whereas some laboratory prototypes can have a TMR of up to 1000% [11].

The mechanism for switching between two MTJ states (i.e. writing non-volatile data) is also an important research field that influences the area, speed and power performances of hybrid MTJ/CMOS circuits. Early field-induced magnetic switching (FIMS) required writing currents in the order of a few milliamperes and thus very large driving transistors and write lines that penalized the die area of hybrid circuits [12].

Thermally assisted switching (TAS) has undergone improvement in terms of bit selectivity and writing efficiency. Prior to switching, MTJ stack is heated above the blocking temperature of the free layer. Afterward, the state of the MTJ is completely controlled by the external magnetic field [13]. However, due to the required heating and cooling latencies, TAS-MTJs exhibit low switching speeds (about 20ns [14]), meaning they are not efficient enough for use in "normally-off and instant-on" computing systems.

Recent current induced magnetic switching (CIMS) methods use the spin-transfer torque (STT) effect proposed by Berger [15] and Slonczewski [16]. This enables magnetization of the free layer to be switched with only one, low, spin-polarized bi-directional current passing through the MTJ stack, as illustrated in Fig. 2b) and 2c). If the density of the spin-polarized writing current is greater than the critical current density ( $J_{co}$ ), MTJ resistance is determined only by the direction of the current.



**Fig. 2** a) CMOS-MTJ co-integration; b) In-plane STT MTJ writing; c) Perpendicular STT MTJ writing.

Mature and commercialized STT-MTJs with in-plane magnetization have very fast MTJ switching speeds (up to 100 ps, according to [17]). However, with the writing currents of hundreds of micro amperes, this switching approach is still not efficient since it consumes a lot of energy and requires large driving transistors. Furthermore, it suffers from reliability issues including data thermal stability, erroneous write by read current and short retention times [18]. High error rate of reading circuits is an additional obstacle.

Emerging perpendicular STT-MTJ structures in which the magnetization direction is perpendicular to the film plane have proved to be the breakthrough technology that enables a significant reduction in the switching current required (several tens of microamperes) as well as improvements in data thermal stability. Perpendicular STT-MTJs are slightly slower than their in-plane counterparts. However, both their energy efficiency and their reported switching speeds of few ns [19], which are comparable with the write speeds of advanced SRAM cells, make them appropriate for the use in "normally-off and instant-on" computing systems [17, 19].

In the following section, we present two hybrid cells that combine perpendicular STT-MTJs as non-volatile storage elements with CMOS transistors used to store a volatile data bit.

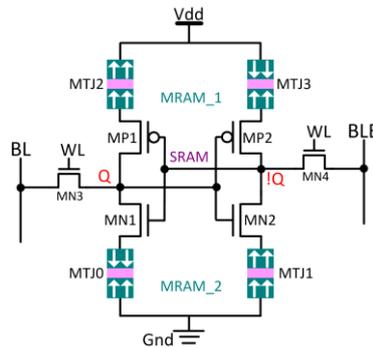
### 3. HYBRID (MTJ/CMOS) MEMORY CELLS

Here described memory cells are based on hybrid (volatile/non-volatile) cross-coupled inverters. They have perpendicular STT-MTJs "embedded" within a CMOS part which makes them suitable to replace SRAM-based volatile memory cells or flip-flops located near the processor's arithmetic logic unit (ALU). The unique feature of these cells is that while CPU is in active state, they behave as a conventional CMOS-based flip-flop or SRAM memory cells with the very high speed of operation (> 2 GHz). While CPU is in

stand-by state, data are stored in MTJs and zero stand-by power is achieved by the power gating. After power supply returns, the cell itself operates as a sense amplifier automatically restoring the data saved in MTJs into the SRAM or flip-flop. This enables the processor core to quickly become ready to start arithmetic operation. Furthermore, such cells allow run-time saving of the processors' context (non-volatile check-pointing), thus significantly improving the reliability of data processing.

### 3.1. 6T-4M hybrid cell with double non-volatile context

The first hybrid cell we propose is shown in Fig. 3. It has a structure similar to that of a conventional 6T-SRAM cell. A volatile (SRAM) data context consists of the cross-coupled inverters (CMOS latch) used to store one data bit in its electrical, complementary form ( $Q$ ,  $!Q$ ). In addition to the CMOS latch, the cell has two non-volatile (MRAM) contexts located in both pull-up and pull-down networks of the latch structure. Each MRAM context contains two perpendicular STT-MTJs that, for the correct operation of the cell, must be in mutually complementary states (Rp/Rap or vice versa).



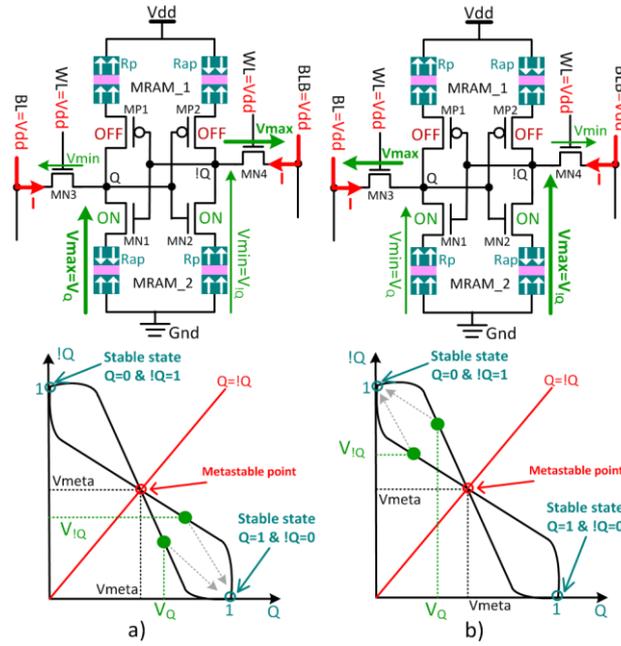
**Fig. 3** 6T-4M hybrid memory cell.

The procedure of writing a volatile data bit is exactly the same as in the conventional SRAM memory cell. The volatile data bit to be written and its complementary value are connected to the BL and BLB lines, respectively. After activation of the access transistors (MN3 and MN4) with the WL signal pulse, the volatile data bit is stored in the CMOS latch.

Reading the non-volatile data bit (i.e. restoring the MRAM context to SRAM) consists of converting the physical value (resistance) stored in MTJs into its electrical equivalent which will be stored in the CMOS latch. Fig. 4 illustrates the reading phase of MRAM\_2 context (MTJs in the pull-down network). To read this MRAM context, BL and BLB lines need to be pre-charged to Vdd. The reading phase begins with activation of WL signal ( $WL=Vdd$ ).

Consequently, pull-down transistors (MN1/MN2) of the CMOS latch are switched on, whereas the pull-up ones (MP1/MP2) are blocked (off). In both pull-down branches of the hybrid cell, there is a current flowing from the BL/BLB lines through the access transistors and NMOS pull-down transistors to the ground (Gnd). Provided that the cell is fully symmetrical (the transistors in both branches have equal on resistances since they have the same dimensions), the voltage drops on the  $Q$  and  $!Q$  nodes entirely depends on the MTJ resistances in the MRAM\_2 context that are in the path of the current. Furthermore, if both

the transistors and the MTJs are carefully sized, the voltages on the latch nodes  $Q$  and  $!Q$  can be adjusted to be one below and another above the meta-stable voltage ( $V_{meta}$ ), depending on the non-volatile data bit stored in MRAM\_2 context. As illustrated on the transfer curve in Fig. 4a), non-volatile data bit ‘1’ stored in MRAM\_2 context ( $Rp/Rp$  configuration) will cause the voltage on the  $Q$  node to be greater than the meta-stable voltage ( $V_Q > V_{meta}$ ). The opposite will occur if MRAM\_2 context stores non-volatile data bit ‘0’ ( $Rp/Rap$  configuration, Fig. 4b)):  $Q$  and  $!Q$  voltages will be below and above meta-stable voltage, respectively ( $V_Q < V_{meta}$ ;  $V_{!Q} > V_{meta}$ ).

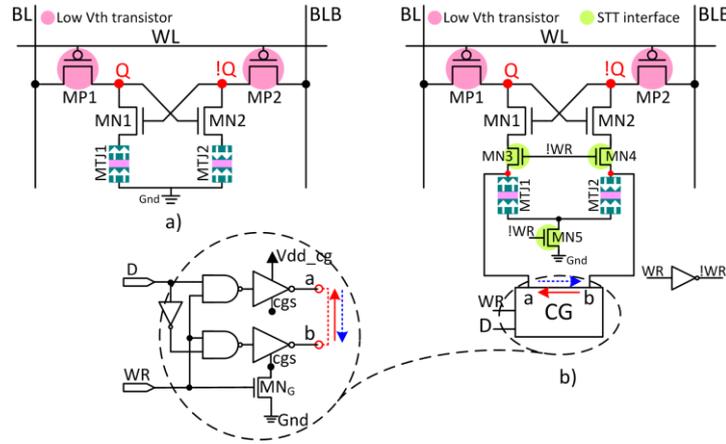


**Fig. 4** The phase of reading MRAM\_2 context that stores: a) non-volatile data bit ‘1’ ( $Rp/Rp$ ); b) non-volatile data bit ‘0’ ( $Rp/Rap$ ).

In both scenarios, at the end of MRAM reading phase when the WL signal is deactivated and the access transistors are turned off, the CMOS latch converges from an unbalanced state to one of its stable states, which is strictly determined by the state (resistance) of MTJs in MRAM\_2 context. The procedure of reading MRAM\_1 context is the same. The only difference is that, in this case, BL and BLB lines need to be pre-charged to Gnd. Consequently, the pull-up network is activated, the current flows in both branches from the power supply (Vdd) to the BL/BLB nodes (which are now on the ground potential) putting the latch in a meta-stable state. Finally, when the access transistors are deactivated, the latch converges from an unbalanced state to a stable one determined by the non-volatile data bit stored in MRAM\_1 context (MTJ2 and MTJ3).  $Rp/Rp$  configuration for MTJ2/MTJ3 stores non-volatile data bit ‘1’ whereas the  $Rap/Rp$  combination is used to store non-volatile ‘0’ bit.

### 3.2. 4T-2M hybrid cell with single non-volatile context

In order to additionally decrease required implementation area, we propose another hybrid cell with a structure similar to that of a 4T-SRAM loadless volatile memory cell. As shown in Fig. 5, it contains two PMOS access transistors (MP1 and MP2) with low threshold voltage ( $V_{th}$ ) and two cross-coupled NMOS transistors (MN1 and MN2) used to store one volatile data bit. In addition, the cell has one non-volatile (MRAM) context located in the pull-down network. It contains two perpendicular STT-MTJs that, for the correct operation of the cell, must be in mutually complementary states ( $Rp/Rap$  or vice versa).



**Fig. 5** a) The 4T-2M hybrid memory cell; b) The same cell with the STT writing interface and current generator (CG) design.

The low threshold voltage of the PMOS access transistors implies increased sub-threshold leakage current compared to the leakage of the pull-down NMOS transistors ( $I_{offp} > I_{offn}$ ). This, in turn, ensures volatile data retention when the cell is on stand-by ( $BL, BLB, WL = V_{dd}$ ).

The procedure of writing a volatile data bit is exactly the same as in conventional 4T-SRAM loadless memory cells whereas the restoring phase is similar to that of a previously described 6T-4M hybrid cell. Fig. 5b) shows STT writing interface. In addition to the current generator (CG) that supplies the bi-directional, spin-polarized current needed to write a non-volatile data bit (D), it contains the footer transistor MN5 as well as the pass transistors MN3 and MN4. In normal cell operation, these three transistors are always switched on ( $WR = '0'$ ). Conversely, during the phase of writing a non-volatile data bit ( $WR = '1'$ ), they cut the MTJs off from the ground rails and cross-coupled NMOS transistors, ensuring that spin-polarized CG current passes through both MTJs in mutually opposite directions. The direction of the CG current is strictly determined by the non-volatile data bit to be written (D). Given that in the idle state CG inverters are with the active pull-down networks (logic zero at the inverters' outputs), the volatile data bit (electrical charge) stored in NMOS cross-coupled transistors could discharge through the CG. To prevent this happening, a power-gating transistor MNG is used to cut-off the CG from the ground rails during its idle state.

## 4. EVALUATION OF HYBRID CELLS

Before measuring the performance of the cells, we implemented them in Cadence Spectre using STMicroelectronics 28 nm fully depleted silicon on insulator (FD-SOI) technology for the CMOS part [20] and 45 nm wide, round, perpendicular STT-MTJs for the non-volatile part. However, it should be said that using SOI is not essential for the proper operation of here presented hybrid cells. They could be implemented in any standard CMOS technology node.

Thanks to the presence of buried oxide in the transistor structure, FD-SOI technology has proved to be very reliable in providing high speed at low voltage [21]. For our measurements, we used a power supply of  $V_{dd}=1.1V$ . Furthermore, the buried oxide significantly reduces standby power consumption by reducing both gate induced drain leakage and junction leakage currents. In addition, the wide range back gate controllability of FD-SOI structure enables optimization of both performance and power after fabrication.

Perpendicular STT-MTJs were co-integrated with CMOS using the open source Spinlib physical model [22]. The model gives the resistances of MTJs depending on its magnetic configuration (P or AP) and its bias voltage. It also defines the current thresholds required to switch between the two configurations. Finally, the model takes the switching delays, including stochastic fluctuations, into account. To achieve high simulation accuracy, the model was calibrated with respect to the experimental data provided by Toshiba and IBM. Table 1 summarizes some of the MTJ parameters that are important for co-integration with the CMOS. As can be seen, required switching currents are few dozen microamperes, whereas switching current pulses are in the order of few nanoseconds. However, it is worth mentioning that the STT writing mechanism has the ability to adjust the amount of switching current and the duration of the switching pulse. Increasing the former entails decreasing the latter.

Thus, it would be possible to speed up non-volatile writing by increasing the amount of writing current, or to make it more energy efficient by increasing the duration of the writing current pulse. With a breakdown voltage of nearly 1 V and supply voltage of 1.1 V, STT-MTJs are in a safe area of operation (we measured 484 mV of voltage across the MTJ during the switching phase).

**Table 1** Main parameters of perpendicular STT-MTJs

Parameter	Description	Value
$R_p/R_{ap}$ [k $\Omega$ ]	P/AP MTJ resistance	3.14/9.4
$I_{sw}$ [ $\mu A$ ]	Switching currents	p $\rightarrow$ ap ~60 ap $\rightarrow$ p ~50
$t_{sw}$ [ns]	Switching speed	p $\rightarrow$ ap 4.27 ap $\rightarrow$ p 4.71
$V_{bd}$ [V]	Breakdown voltage	~1
Area	MTJ area	45nm x 45nm
$RA$ [ $\Omega \cdot \mu m^2$ ]	Resistance-area product	5
$TMR$ [%]	TMR ratio	200

To ensure the area efficiency of any target application of our hybrid memory cells, they have to be as small as possible, since they may be instanced many times. That is why our first evaluation step was to find the smallest possible hybrid cell design, i.e., the smallest

possible transistor sizes with which the cell was still operational. To this end, we used Monte Carlo (MC) analysis. The length of all the transistors in both cells was the smallest possible allowed by the technology ( $L=L_{min}=30\text{ nm}$ ). We continued to vary the width (W) of the transistors as long as we obtained 0% of conversion (non-volatile reading) errors in 5000 MC runs with  $std = 10\%$  variations in the length and width of all the transistors.

Using minimally sized hybrid cells, we continued to measure its other performances: static power consumption, the robustness of volatile data, the speed of writing the volatile data bit, the speed of restoring the non-volatile data bit as well as the dynamic energy required to restore it. The results are summarized in Table 2. Some measured parameters are also compared with the performances of conventional 4T- and 6T-SRAM cells implemented in pure 28 nm FD-SOI CMOS technology.

The total transistor area ( $W \times L$ ) of the hybrid cells is 2-3 times bigger than conventional, pure CMOS memory cells. This increase in area is mostly due to the presence of the STT writing interface. However, given that hybrid cells can store 2-3 data bits, this difference in required silicon area can be considered as expected and acceptable.

Regarding leakage power, it is calculated by the help of Cadence measurement description language (MDL) using the following formula:

$$P_s = \frac{V_{dd} \cdot \int_{t_0}^{t_1} i_{vdd}(t) \cdot dt}{\Delta t}, \quad (1)$$

where  $i_{vdd}$  is the power supply current during the idle time interval  $\Delta t = t_1 - t_0$ .

Given that 4T-SRAM cells preserve the volatile data bit with increased leakage currents coming from LVT PMOS transistors, their leakage power is significantly higher compared with 6T-SRAM leakage. Low leakage power consumed by 4T-2M hybrid cell is due to resistive MTJs that are positioned in the path of the leakage currents (pull-down network of the cross-coupled transistors). 6T-4M hybrid cell consumes more static power simply because it contains more transistors. However, unlike conventional SRAM cells, our hybrid cells can store a volatile data bit into a non-volatile context, meaning the power supply can be turned off. This, in turn, completely eliminates leakage power.

Table 2 Evaluated performance of hybrid cells

	6T-SRAM	4T-SRAM	6T-4M	4T-2M	
$W \times L$ [ $\mu\text{m}^2$ ]	0.024	0.0144	0.0624	0.0516	
Leakage [nW]	0.93	5.67	3.15	1.9	
MTJ reading [ps] <sup>§</sup>	-	-	ctx1 ctx2	33.4 60.8	92.7
$E_{rd}$ [fJ= $\mu\text{W}/\text{GHz}$ ] <sup>¶</sup>	-	-	3.94	3.17	
Vol. writing [ps] <sup>°</sup>	6.8	4.9	10	9.8	
SNM [mV] <sup>*</sup>	395	154	318	98	

<sup>§</sup> The speed of reading (restoring) a non-volatile data bit stored in MTJs

<sup>¶</sup> Dynamic energy consumed during the phase of reading non-volatile data bit

<sup>°</sup> The speed of writing a volatile data bit

<sup>\*</sup> The higher the SNM, the better the robustness

To determine the speed of restoring a non-volatile data bit to volatile context, we continued to increase the width of the reading WL pulse (by using the binary search method) as long as the first correct reading operation was detected. The measured minimum reading pulse determines the maximum possible reading speed. As can be seen from Table 2, non-volatile data bit can be read in a gigahertz regime. In 6T-4M hybrid cell, non-volatile MRAM\_1 context is faster than its MRAM\_2 counterpart due to the fact that pull-up network in our cell is less resistive the pull-down one. 4T-2M hybrid cell is slightly slower because of the position of MTJs as well as sub-threshold working regime. This influences slow reaching of the unbalanced state.

The minimum dynamic energy consumed by the hybrid cell during the phase of non-volatile reading is listed in the middle of Table 2. It was calculated by:

$$E_{rd} = V_{dd} \cdot \int_{t_0}^{t_1} i_{vdd}(t) \cdot dt - P_s \cdot \Delta t, \quad (2)$$

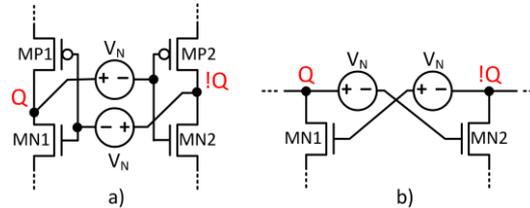
where  $i_{vdd}$  is the power supply current during the restoration phase,  $\Delta t = t_1 - t_0$  is previously determined minimum duration of the reading pulse, and  $P_s$  is leakage power consumption of the cell. Both 4T-2M and 6T-4M hybrid cells exhibit similar performance in terms of required restoration energy.

To measure the speed of writing the volatile data bit, we used binary search method to determine the minimum width of the WL pulse needed to write the volatile data bit set on the BL/BLB lines. The measured values are listed at the bottom of Table 2. It can be seen that volatile writing speeds of all the cells are below 10 ps. The presence of the STT writing interface and resistive MTJs slightly reduce the volatile writing speeds of our hybrid cells compared to both the 4T- and 6T-SRAM cells.

Hybrid cells we present here use cross-coupled inverters (6T-4M) or cross-coupled NMOS transistors (4T-2M) to store the volatile data bit. The stability of this kind of structure is typically expressed in terms of its static noise margin (SNM). Informally, the static noise margin can be understood as the minimum voltage disturbance that could flip the volatile data stored in the memory cell. Fig. 6 shows the conceptual measurement setup we used to measure SNM. DC noise sources with the value  $V_N$  were introduced between the gates of the NMOS transistors and output  $Q$ ,  $!Q$  nodes. Using Spectre MDL, we increased the noise voltage  $V_N$  as long as we detect volatile data flipping. We repeated the same procedure for both possible values of the non-volatile data bit ( $Q=1$  and  $Q=0$ ) stored in cross-coupled NMOS transistors. The measured SNMs (worst case) of hybrid cells are listed at the bottom of the Table 2. It can be seen that 4T-2M hybrid cell is the most sensitive to voltage noise.

To benefit from the dual storage facility, a significant property of the hybrid cell would be its ability to write non-volatile data bit without disturbing volatile data ( $Q$ ,  $!Q$ ). In this way, the device based on hybrid cells may profit from the run-time (on-the-fly) reconfiguration ability. During the processing of volatile data bits, some background operation may write non-volatile ones in parallel. To investigate this ability for the 6T-4M hybrid cell, we monitored the disturbance of volatile data (logic level degradation) during the non-volatile writing phase. We report logic level degradations of 92 mV and 116 mV for MRAM\_1 and MRAM\_2 contexts, respectively. Given that logic level degradations

are less than the SNM value, we conclude that both non-volatile MRAM contexts of our cell can be dynamically reconfigured.



**Fig. 6** a) Static noise margin (SNM) measurement setup for 6T-4M hybrid cell;  
b) SNM measurement setup for 4T-2M hybrid cell.

Finally, it is worth mentioning that the above presented performance analysis is not completely exhaustive. We did not take into account the influence of the MTJ process variations that become more and more critical, particularly in terms of resistance variations. Moreover, we did not consider the sensitive aspects of integrating MTJ electric signals to CMOS electronics (reliability of nearly zero run-time error is required by the logic applications) [23, 24]. This, together with the influence of voltage and temperature variations will be included in our future work.

## 5. CONCLUSION

This paper presents two hybrid cells that are able to store and process one data bit both electrically and magnetically. The cells are based on 4T- and 6T-SRAM architectures and use recently emerging perpendicular STT-MTJ nanopillars as non-volatile storage elements. Measured performance of both hybrid cells implemented in 28 nm FD-SOI technology combined with 45 nm round STT-MTJs showed that the cells are ready to be used in "normally-off and instant-on" computing systems. The cells need less than 100 ps to restore a non-volatile data bit, spending not more than 4 fJ for the operation. The volatile data bit can be written for a time below 10 ps. Moreover, 6T-4M hybrid cell presented here has a few clear advantages compared to existing hybrid cells: two non-volatile data contexts and the ability to write a volatile data bit.

The cells presented here have the potential to completely eliminate idle power consumption of a battery powered systems-on-chip. They are also suitable for non-volatile reconfigurable logic applications (non-volatile registers, processor cache, magnetic FPGAs, etc).

**Acknowledgement:** This research was sponsored in part by the French National Agency for Scientific Research (ANR), through the projects DIPMEM and MARS, as well as, by the Serbian Ministry of Science and Technological Development, through the project III-44004.

## REFERENCES

- [1] J. Rabaey, *Low Power Design Essentials*. New York: Springer-Verlag, 2009.
- [2] P. Rech, J.-M. Galliere, P. Girard, F. Wrobel, F. Saigne, and L. Dilillo, "Impact of Resistive-Open Defects on SRAM Error Rate Induced by Alpha Particles and Neutrons", *IEEE Transactions on Nuclear Science*, vol. 58, pp. 855-861, 2011.

- [3] R. Sandeep, N.T. Deshpande, and A.R. Aswatha, "Design and Analysis of a New Loadless 4T SRAM Cell in Deep Submicron CMOS Technologies", In Proceedings of the 2nd International Conference on Emerging Trends in Engineering and Technology, Nagpur, 16-18 Dec. 2009, pp. 155-161.
- [4] K. Abe, S. Fujita, and H. Lee, "Novel Nonvolatile Logic Circuits with Three-Dimensionally Stacked Nanoscale Memory Device", In Proceedings of Nanotechnology Conference, Anaheim, California, 8-12 May 2005, pp. 203-206.
- [5] Semiconductor Industry Association (SIA). (2011) *International technology roadmap for semiconductors*. San Jose, CA: Semiconductor Industry Association (SIA), <http://www.itrs.net/>. Accessed March 13 20015.
- [6] S. James, P. Arujo, and A. Carlos, "Ferroelectric Memories", *Science*, vol. 246, pp. 1400-1405, 1989.
- [7] H. Wong, S. Raoux, S. Kim et al. "Phase Change Memory", invited paper, In Proceedings of the IEEE, 2010, vol. 98, pp. 2201-2227.
- [8] C. Chappert, A. Fert, and V. Dau, "The Emergence of Spin Electronics in Data Storage", *Nature Materials*, vol. 6, pp. 813-823, 2007.
- [9] W. Zhao, E. Belhaire, C. Chappert, and P. Mazoyer, "Spintronic Device Based Non-volatile Low Standby Power SRAM", In Proceedings of IEEE Annual Symposium on VLSI Montpellier, 7-9 Apr. 2008, pp. 40-45.
- [10] S. Ikeda, H. Sato, M. Yamanouchi, et al., "Recent progress of perpendicular anisotropy Magnetic Tunnel Junctions for non-volatile VLSI", *Journal of SPIN*, vol. 2, pp. 1240003-1 - 124003-12, 2012.
- [11] T. Kawahara, K. Ito, R. Takemara, and H. Ohno. "Spin-Transfer Torque RAM Technology: Review and Prospect", *Microelectronics Reliability*, vol. 52, pp. 613-627, 2012.
- [12] W. Zhao, E. Belhaire, C. Chappert, and P. Mazoyer, "Power and area optimization for run-time reconfiguration SOPC based on MRAM", *IEEE Transactions on Magnetics*, vol. 45, pp. 776-780, 2009.
- [13] L. Torres, Y. Guillemenet, and S. Ahmed, "A Dynamic Reconfigurable MRAM-based FPGA", In Proceedings of International Conference on Engineering of Reconfigurable Systems and Algorithms, Las Vegas, Nevada, 12-15 Jul. 2010, pp. 31-40.
- [14] D. Suzuki, M. Natsui, S. Ikeda, et al., "Fabrication of a Nonvolatile Lookup-table Circuit Chip Using Magneto/Semiconductor Hybrid Structure for an Immediate-power-up Field Programmable Gate Array", In Proceedings of IEEE Symposium on VLSI Circuits, Kyoto, 16-14 Jun. 2009, pp. 80-81.
- [15] L. Berger, "Emission of spin waves by a magnetic multilayer traversed by a current", *Physical Review B*, vol. 54, pp. 9353-9358, 1996.
- [16] J. C. Slonczewski, "Current-driven excitation of magnetic multilayers", *Journal of Magnetism and Magnetic Materials*, vol. 1859, pp. L1-L7, 1996.
- [17] H. Yoda, S. Fujita, N. Shimomura, et al., "Progress of STT-MRAM Technology and the Effect on Normally-off Computing Systems", In Proceedings of IEEE International Electron Devices Meeting, San Francisco, California, 10-13 Dec. 2012, pp. 11.3.1 - 11.3.4.
- [18] R. Takemura, T. Kawahara, K. Ono, K. Miura, H. Matsuoka, and H. Ohno, "Highly-scalable Disruptive Reading Scheme for Gb-scale SRAM and Beyond", In Proceedings of IEEE International Memory Workshop, Seoul, 16-19 May 2010, pp. 1-2.
- [19] E. Kiagawa, S. Fujita, K. Nomura, et al. "Impact of Ultra Low Power and Fast Write Operation of Advanced Perpendicular MTJ on Power Reduction for High-performance Mobile CPU", In Proceedings of IEEE International Electron Devices Meeting, San Francisco, California, 10-13 Dec. 2012, pp. 29.4.1 - 29.4.4.
- [20] N. Planes, O. Weber, V. Barral, et al. "28nm FDSOI Technology Platform for High-speed Low-voltage Digital Applications", In Proceedings of the Symposium on VLSI Technology, Honolulu, Hawaii, 12-14 Jun. 2012, pp. 133-134.
- [21] T. Ishikagi, R. Tsuchiya, Y. Morita, et al. "Silicon on Thin BOX (SOTB) CMOS for Ultralow Standby Power with Forward-biasing Performance Booster", In Proceedings of the European Solid-State Device Research Conference, Edinburgh, 15-19 Sep. 2008, pp. 198-201.
- [22] Y. Zhang, W. Zhao, Y. Lakys, "Compact Modeling of Perpendicular-Anisotropy CoFeB/MgO Magnetic Tunnel Junctions", *IEEE Transactions on Electron Devices*, vol. 59, pp. 819-826, 2012.
- [23] W. Kang, W. Zhao, E. Deng et al., "A Radiation Hardened Hybrid Spintronic/CMOS Non-volatile Unit using Magnetic Tunnel Junctions", *Journal of Physics D: Applied Physics*, vol. 47, p. 405003, 2014.
- [24] W. Kang, E. Deng, J. O. Klein et al., "Separated Pre-Charge Sensing Amplifier for Deep Submicron MTJ/CMOS Hybrid Logic Circuits", *IEEE Transactions on Magnetics*, vol. 50, pp. 3400305-5, 2014.