TAS-MRAM based Non-volatile FPGA logic circuit

Weisheng ZHAO1, 2, Eric BELHAIRE2, 1, Bernard DIENY3, Guillaume PRENAT3 and Claude CHAPPERT2, 1
1. IEF, Univ Paris-Sud, UMR 8622, Orsay, F-91405
2. CNRS, Orsay, F-91405
3. SPINTEC, CEA, 17, av. des martyrs, 38054, Grenoble, France
E-mail: zhao@ief.u-psud.fr

Abstract

As one of the most promising Spintronics applications, MRAM combines the advantages of high writing and reading speed, limitless endurance and non-volatility. The integration of MRAM in FPGA allows the logic circuit to rapidly configure the algorithm, the routing and logic functions, easily realize the dynamical reconfiguration and multi-context configuration. However, the conventional MRAM technology based on Field Induced Magnetic Switching (FIMS) writing approach consumes very high power and large circuit surface, and produces high disturbance between memory cells. These drawbacks prevent FIMS-MRAM’s further development in memory and logic circuit. Thermally Assisted Switching (TAS) based MRAM is then evaluated to address these issues and some design techniques for FPGA logic circuits based on TAS-MRAM technology are presented. By using STMicroelectronics CMOS 90nm technology, some chip characteristic results have been calculated to demonstrate the expected performance of TAS-MRAM based FPGA logic circuits.

1. Introduction

In the last years, MRAM has been rapidly evaluated as one of the most promising Spintronics applications, which represents some advantages like non-volatility, high writing/reading speed, limitless endurance and radiation-hardness etc. The use of MgO barriers [1] in Magnetic Tunnel Junctions (MTJ) (see Fig 1.1) [2] which are the basic memory cells of MRAM, improves significantly their reading performance. Furthermore, the development of writing approaches based on Thermally Assisted Switching (TAS) [3] and Spin Transfer Torque (STT) [4] greatly reduces the power consumption and improves the writing selectivity of conventional MRAM technology, which is based on Field Induced Magnetic Switching (FIMS) [5]. The excellent writing/reading and power performance of MRAM makes it one of the best non-volatile memory candidates.

Figure 1.1. Conventional Magnetic Tunnel Junctions (MTJs) are composed of four main layers: an oxide barrier, such as MgO and AlxOy a storage magnetic layer and a reference magnetic layer (typically in CoFe alloy), an anti-ferromagnetic layer (AF1) used to pin the magnetization of the reference via the so-called “exchange bias” phenomenon. The magnetization of the storage layer can be switched by an external magnetic field either parallel or anti-parallel to that of the reference layer. When a current flows across the MTJ, a change in resistance R is observed between these two magnetic configurations (typically R/R of the order of 40% for AlxOy and 200% for MgO based MTJ). This change of resistance is named Tunnel MagnetoResistance (TMR).

SRAM based FPGA (Field Programmable Gate Arrays) logic circuits have been the object of intense development in the last twenty years. The reconfigurability property provided by this technology compatible with standard CMOS process have made this technology quite attractive for numerous applications; however SRAM is volatile, which means that all the functions have to be pre-programmed at each power-up and external non-volatile PROM memory must be integrated with the chip either in the same package or at the printed circuit board level. This increases the start-up time, the total device cost and the
printed circuit board area. Internal Flash memory [6] is now sometimes used to replace the external memory. However, it has some drawbacks such as slow reprogramming and sensing, limited number of writing cycles (up to $10^6$), which limit its lifetime and the reconfiguration speed in FPGA circuit.

MRAM based FPGA logic circuits were proposed [7]. These circuits first benefit from the non-volatility of MRAM to store the configurations, but the high writing speed of MRAM cells can have also a strong impact on the chip architecture. Intermediate data normally stored in D-flip-flops, can here be stored in MRAM cells and any block can then be safely powered off. In a SRAM based FPGA, all the SRAM cells must be initialized with the configuration information, besides most of them will not be used during some computing. In MRAM based FPGA, only the part required to process data is active, but the rest is able to be powered off. This new processing architecture could economize greatly the processing energy in the majority of applications. Moreover, the low cell area and 3D stack structure of MRAM (see Fig 1.2) allow to easily realizing the multi-context and dynamical reconfiguration architecture without much additional surface. The applications of MRAM based FPGA can be extended to aerospace and military fields thanks to its radiation hard property and its limitless switching endurance. In this field, conventional FPGA is difficult to implement, as data security is one of the most important characteristic. However the high switching power consumption of conventional MRAM writing approach (FIMS) limits the future interest of this technology. In this paper, we have investigated how TAS-MRAM can be used instead of FIMS MRAM in magnetic FPGA to circumvent this drawback. Special design techniques were used to conceive several hybrid CMOS/MTJ devices and evaluate them. In the next part, we briefly introduce how TAS writing mode works. The third section presents the design of low power non-volatile Look Up Table (LUT) structure and discusses solutions to reduce the power consumption. At last, the design and simulation of TAS-MRAM based Flip-Flop (TAS-FF) are presented in the fourth section.

2. MRAM based on Thermally Assisted Switching (TAS) writing approach

Thermally Assisted Switching (TAS) combines a local short heating of the selected bit with a single low amplitude magnetic field [8]. The Magnetic Tunnel Junction (MTJ) stack is slightly modified to be adapted to the TAS operation (see Fig 2.1). An additional anti-ferromagnetic layer (AF2) with low blocking temperature ($T_{B2}$ ~ 160°C) [9] is added in the stack above the ferromagnetic storage layer. As a result, the magnetization of the storage layer is pinned at the standby temperature and only becomes unpinned (free) when the MTJ is heated at a temperature $T_{write}$ above the blocking temperature $T_{B2}$ of this additional AF layer. In order to prevent any switching of the magnetization of the reference layer, the AF1 material which is used to pin this reference magnetization is chosen with a much higher blocking temperature $T_{B1}$ (typically 300°C) than $T_{B2}$. During the write operation, two low currents $I_h$ (~100μA to 1mA) and $I_m$ (~4mA) are required. As shown in Fig 2.2, when $I_h$ passes through the MTJ, it heats up the MTJ and in particular the storage layer by Joule heating. Within a few nanoseconds, the temperature exceeds $T_{B2}$. The ferromagnetic storage layer can then be reversed by the application of a small magnetic field, generated by a current $I_m$ passing in a current line located above the junction. $I_h$ is mono-directional whereas $I_m$ is bidirectional.

From a physical point of view, the magnetization of the storage layer can only be switched when the temperature of the storage layer exceeds $T_{B2}$. Therefore, to reduce the electrical consumption, the
current generating the field can be advantageously switched on once the MTJ has reached $T_{B2}$. However, it is important to maintain $I_m$ on during the cooling phase to insure that AF1 pins the storage layer magnetization in the desired new direction as shown in Fig 2.3. The duration of the heating phase can be significantly reduced down to a few nanoseconds by increasing $I_h$, but the duration of the cooling phase only depends on materials parameters such as the MTJ specific heat and the heat conductance from the tunnel barrier to the electrodes. This cooling duration is of the order of ten nanoseconds.

![Figure 2.2. TAS write operation needs two current: $I_h$ and $I_m$. $I_h$ passes through the Junction and $I_m$ passes above](image)

**Figure 2.2.** TAS write operation needs two current: $I_h$ and $I_m$. $I_h$ passes through the Junction and $I_m$ passes above.

**Figure 2.3.** Suitable configurations of $I_h$ and $I_m$ pulses for TAS writing. $t_{\text{heat}}$ and $t_{\text{cool}}$ respectively represents the intervals during which the heating and cooling of the MTJ cell take place.

3. **TAS-MRAM technology based non-volatile Look Up Table (TAS-LUT)**

A. **MTJ Write and read circuit in TAS-LUT**

Look Up Table (LUT) [10] is the main processing unit of FPGA logic circuits. The algorithms and logic functions it computes, are stored in its configuration memory cells. By using MRAM as memory cells, non-volatile LUT can be designed. If several MRAM cells are used by configuration bits, efficient dynamical reconfiguration function can be implemented. The nonvolatility allows to power off the unused blocks in order to reduce the standby power consumption. In our proposed design, a couple of MTJs is associated to each memory bit. The MTJs writing circuit is such that they are always in opposite magnetic configurations i.e when one is in parallel state (low resistance), the other is in anti-parallel state (high resistance). A simple 5 transistors, SRAM based Sense Amplifier (see Fig 3.1) is then employed to sense the data saved in the MTJs couple. The cell senses the magnetic information by briefly turning on the NMOS transistor MN2 switch, and then promptly turning it off (“SEN” is the reading control signal). This structure has been demonstrated to read the states of the MTJs at a very high speed of about 200 ps [7], which allows this configuration of FPGA circuits to be initialized immediately. Vdd_logic is fixed as default to 1.2v for STMicroelectronics CMOS 90nm.

![Figure 3.1. Two TAS-MTJs present 1 logic bit, SRAM based Sense Amplifier is used to read](image)

**Figure 3.1.** Two TAS-MTJs present 1 logic bit, SRAM based Sense Amplifier is used to read.

![Figure 3.2. MTJ write circuit schematic, with two current sources, generating $I_h$ and $I_m$ respectively.](image)

**Figure 3.2.** MTJ write circuit schematic, with two current sources, generating $I_h$ and $I_m$ respectively.

The write operation of TAS requires 1 bi-directional current source $I_m$ (MN3-6) to generate the writing magnetic field in the right direction and two NMOS transistors (MN7-8) to switch on-off the heat current $I_h$ (see Fig 3.2-3). The signal “EN_Heat” enables $I_h$, “EN” enables the writing field current and Input
determines its direction.

Figure 3.3. MTJ writing circuit layout: $I_m$ is generated by a bi-directional current source and passes above the couple of MTJs; $I_h$ passes through the MTJs and heats them up.

B. Advanced high programming speed and low surface architecture for TAS-LUT

As a next step, one can conceive LUT with several bits of data comprising a single magnetic field generating line ($I_m$) passing above all the MTJs. Such design reduces the total surface (a single bi-directional current source per LUT) and allows the LUT to be programmed at high speed (only one cycle of “Input”, ~20ns estimated) and with low power consumption. Fig 3.4 shows an example of such structure for 2-inputs LUT, with 4 bits of data and 8 MTJs. The signals “I0-3” are the data to be programmed; “SE” in MUX selects to output the “I0-3” or their opposite logic values. For example, if “I0-3” is “1010” and “SE” equals to ‘1’, the output of these four MUXs are “1010”. In contrast, if “SE” is ‘0’, the outputs are “0101”. This way, the programming process could be divided into two phases: data logic value ‘1’ and ‘0’. Four heat current sources are used to heat up locally the four couples of MTJs. Each couple can be switched only if the “EN_Heat” is active and the corresponding output of MUX equals logic value ‘1’ (see Fig 3.2). A single bi-directional current source is used to globally switch the whole LUT, in which “Input” determines the current direction and then the data stored in the couples of MTJs. Considering again the example mentioned above, the complete LUT programming sequence is as follows: Assuming as an example that the signals “I0-3” are “1010” and “SE” equals to ‘1’, the heat current source 0 and 2 are activated. After $t_{\text{heat}}$ (see Fig 3.5), the “Input” is set to ‘1’ and the two couples of MTJs (MTJ00-1 and MTJ20-1) are written by the magnetic field as ‘1’. Then “SE” is set to ‘0’: the heat current sources 1 and 3 are activated. After $t_{\text{cool}}$, the “Input” is set to ‘0’ and the other two couples of MTJs (MTJ10-1 and MTJ30-1) are programmed as ‘0’. Therefore, it takes just one cycle of “Input” and “SE” from ‘1’ to ‘0’ to program the whole LUT. This significantly improves the programming speed and reduces the electrical consumption.

Figure 3.4. Architecture of a TAS MRAM based 2-inputs LUT, which has 4 bits stored in 4 couples of MTJs.

Another advantage of this advanced structure is that no matter how many inputs in the LUT, the complete programming period is always in two phases. This
makes TAS-MRAM very interesting to be used in complex LUT such as 5-inputs LUT and 8-inputs LUT, which are necessary for large calculations and processing digital systems. Furthermore, whatever the number of bits in the LUT, a single bi-directional current source is needed. Therefore, the surface per bit and the total switching current per bit, including the magnetic field generating current and heat current, could be reduced significantly by increasing the number of bits in the LUT, as shown in Fig 3.6.

C. Dynamical reconfiguration and multi-context configurations for TAS-LUT

The reprogramming of TAS-MRAM cells is independent of the data processing in TAS-LUT. This allows easily performing run time configuration or dynamical reconfiguration. The switching of the couple of MTJs is controlled by “EN” and “EN_Heat” (see Fig 3.7) and the logic output of the Sense Amplifier Qm can be modified only when “SEN” is switched from ‘1’ to ‘0’. As a result, the whole LUT reconfiguration (i.e. switching the couples of MTJs in the suitable configuration) can be performed during the LUT calculation; Once the MTJs have been switched; a pulse of “SEN” (sub 1ns) reconfigures Qm with a new logic value. It takes about 20–30 ns for this dynamical reconfiguration.

Vdd_Heat could be a higher voltage than Vdd_logic such as 2V or 3.3V to increase Ih, thereby reducing t_heat and decreasing the power consumption from the heating current. The signals “EN” and “EN_Heat” are necessary to control the dephasing between the heating pulse and the field pulse so as to reduce the electrical consumption (See Fig 2.3).

Figure 3.5. By using this advanced structure, TAS-MRAM based LUT can be reprogrammed at very high speed (~20ns)

Figure 3.6. Total switching current per bit versus number of inputs in the Look Up Table (LUT)

C. Dynamical reconfiguration and multi-context configurations for TAS-LUT

Thanks to the non-volatility of TAS-MRAM, multi-context configuration architecture has been developed (Fig 3.8), which allows sub ns dynamical reconfiguration in FPGA logic circuits. In this architecture, the write operation works as in the LUT (Fig 3.4). Each couple of MTJs has its own heat current source and a global bi-directional current source switches all the MTJs in two phases. The Fig 3.8 shows an example of Four-context configuration. In order to change the configuration, one has, simply to select the address, C0 and C1, and then set “SEN” to ‘1’. Once the circuit is stable, “SEN” is reset to ‘0’ and Qm is then set to its new value. This multi-context TAS-LUT can be adapted advantageously to those systems which have some limited kinds of configurations and are not required to be reprogrammed very often.

Figure 3.7. Full schematic of a magnetic non-volatile bit in Look Up Table (LUT)
Figure 3.8. Example of Four-context configuration for a TAS-MRAM based LUT

4. TAS-MRAM technology based non-volatile Register or Flip-Flop

Configurable logic Blocks (CLBs) are the basic calculation elements in FPGA. They are composed of 4 or 8 semi slices, which include principally a LUT and a Register (see Fig 4.1) [11]. The LUT outputs the result after the processing; the register saves the result temporarily and synchronizes it with the global clock. The register is usually built with D flip-flop. The conventional structure of a D Flip-Flop includes two elements: Master and Slave, which are both clock-controlled Latch (see Fig 4.2) [12]. The master part writes the data in the Flip-Flop; the slave part keeps the precedent data during the write operation and output the data saved during the read operation, the global clock and its anti-phase clock control the process. This conventional latch based Flip-Flop is volatile, which means that it looses all the intermediate data if the electrical supply of the FPGA is switched off. This constitutes a drawback for the security of the data in the circuit and it increases the duration of the chip initialization. We have therefore developed a concept of TAS-MRAM based Flip-Flop (TAS-FF) as shown in Fig 4.3 and 4.4. Such FF is non-volatile, run at high speed and with low Power Delay Product (PDP) [12], the quality metric figure to evaluate the performance of Flip-Flop.

In conventional latch based D-Flip-Flop, two clock-controlled latches write, read and keep the data in the TAS-MRAM based non-volatile Flip-Flop (TAS-FF). At the same time, the couple of MTJs is switched with a given low frequency “NW” such as 1 MHz and 100 kHz. If the power is switched off, the data in the master latch is lost, however that saved in the MTJs remains non-volatile. In order to sense the data in the couple of MTJs, the master latch is slightly modified (Fig 4.4) to work as the SRAM based Sense Amplifier (Fig 3.1) during the data retrieve operation of TAS-FF.

Figure 4.1. Semi slice is the basic calculating cell in FPGA logic circuit. It includes a function generator LUT and a register

Figure 4.2. Conventional latch based D-Flip-Flop

Figure 4.3. TAS-MRAM based non-volatile Flip-Flop

“NW” is the frequency signal of the TAS MRAM
switching circuit. If “NW” increases, the chip data security level could be improved (E.1), \( R_{nv} \) is the data saved non-volatile ratio relative to all the data processed in the Flip-Flop, however the power dissipation rises simultaneously (E.2), so that “NW” should be defined by users according as their power provision and data security level requirement. Since the TAS switching period is of the order of 20ns, “NW” should be less than 50MHz to ensure the correct switching of MTJs.

\[
R_{nv} = \frac{f_{NW}}{f_{Clk}} \quad (E.1)
\]

\[
P_{\text{dynamics}} = f_{NW} \int_0^T V_{dd} \times I_d(t) \, dt \quad (E.2)
\]

TAS-FF combines high speed and low delay (see Fig 4.5) of conventional D-Flip-Flop with the non-volatility and low power dissipation of TAS MRAM. Thanks to the lower switching currents, smaller writing circuit surface is required than in conventional MRAM based Flip-Flop [13]. The non-volatility of MTJs and high reading speed of SRAM based SA in TAS-FF allows the FPGA logic circuit to perform a real “instant on” as shown in Fig 4.6. By using STMicroelectronics CMOS 90nm design kit, Spice simulations have been done and shown that data could be retrieved in 214ps.

Because Flip-Flops need to work at very high frequency [11], the high writing/reading speed and endurance are thus important, according to (International Technology Roadmap for Semiconductors) ITRS 2006 [14], Flash memory and Phase Change Memory (PCRAM) cannot be used in Flip-Flop due to their endurance limit and low writing/reading speed. Therefore Magnetic RAM (MRAM) appears to be the only solution to bring the non-volatility in Flip-Flops.

Figure 4.4: the full schematic of TAS-MRAM based Non-Volatile D-Flip-Flop (TASFF)

Figure 4.5: Restart simulation of TAS-FF. Data in the couple of MTJs is ‘1’ and the logic delay of this TAS-FF is about 170ps.

Figure 4.6: Restart simulation of TAS-FF. Data in the couple of MTJs is ‘0’ and can be retrieved in about 214ps.
5. Conclusions and perspectives

TAS-MRAM based FPGA logic circuits are presented in this paper, which can perform dynamical reconfiguration, multi-context configuration and “instant-on” start-up with low power dissipation and at high speed. This non-volatile FPGA logic circuit has great potential in the field of complex logic processing systems such as high performance game station and radar signal processing. They could be advantageously used in the field of aviation and space where the security of data is one of the most important considerations. The prototype of this hybrid magnetic CMOS logic circuit is under development and realization in our laboratory in collaboration with SPINTEC laboratory (Magnetic process) and STMicroelectronics (CMOS 90nm technology).

6. Acknowledgements

The work and results reported were obtained with research funding from the European Community under the sixth Framework, Contract Number 510993: MAGLOG. The views expressed are solely those of the authors, and the other Contractors and/or the European Community cannot be held liable for any use that may be made of the information contained herein.

7. References


