

Reconfigurable Circuits Using Magnetic Tunneling Junction Memories

Victor Silva, Jorge Fernandes, and Horácio Neto

INESC-ID / IST / UTL, Technical University of Lisbon, Portugal
{vicant,jorge.fernandes,hcn}@inesc-id.pt

Abstract. This paper presents the first results of our work to research and develop new reconfigurable circuits and topologies based on Magnetic RAM (MRAM) memory elements. This work proposes a coarse-grained reconfigurable array using MRAM. A coarse-grained array, where each reconfigurable element computes on 4-bit or larger input words, is more suitable to execute data-oriented algorithms and is more able to exploit large amounts of operation-level parallelism than common fine-grained architectures. The architecture is organized as a one-dimensional array of programmable ALU and the configuration bits are stored in MRAM. MRAM provide non-volatility with cell areas and with access speeds comparable to those of SRAM and with lower process complexity than FLASH memory. MRAM can also be efficiently organized as multi-context memories.

Keywords: MRAM. Reconfigurable Hardware. Coarse-grained array. Programmable ALU. Multi-context.

1 Introduction

The field of reconfigurable computing has been, so far, dominated by fine-grained Field Programmable Gate Arrays (FPGA). Nevertheless, FPGA have substantial drawbacks. Since most FPGA are SRAM based, their volatility makes FPGA unappealing for applications that require reduced parts count and small footprint or rapid availability of logic or high security [1]. FPGA also have a large routing overhead due to their bit-level nature, demand large configuration memories for both processing units and routing switches and require large configuration times that make runtime reconfiguration practically unfeasible.

Reconfigurable arrays (RA) try to overcome the disadvantages of Field Programmable Gate Array (FPGA) based computing solutions by providing multiple bit-wide data-paths and complex operators instead of bit-level configurability. The wider data-paths allow for the efficient implementation of complex operators in silicon. Thus, RA offer lower overhead, exploit better operation level parallelism and are better suited for data-oriented algorithms [2].

This work proposes a MTJ-based coarse-grained reconfigurable array. The reconfigurable architecture proposed is organized as a 1-dimensional coarse-grained array. The basic processing elements (PE) are arithmetic and logic function units 4 or more

bits wide. The operation to be executed by each PE is controlled by the MRAM bits. The array is run-time reconfigurable. The MTJ structures can be rewritten while the circuit is in execution, and therefore a new reconfiguration can be loaded while the current application is running.

A number of architectural and circuit solutions are being studied, implemented and evaluated. The fundamental components of the RA have already been designed and electrically simulated. The results of the electrical simulations were on line with expectations.

A scaled down version of a coarse-grained reconfigurable array using TAS-MRAM was designed and a first prototype has been sent for fabrication and is expected to be delivered in a few weeks. This prototype will be further processed at INESC-MN facilities to lay out the layers that will form the MTJs.

2 Contribution to Technological Innovation

This work aims to answer a number of open research questions, such as:

- Is it possible to cost effectively deal with FPGA inherent volatility?
- Is there an alternative reconfigurable architectural model to the nasty FPGA bit-level architectural model?
- Is it possible to have it all, a non volatile reconfigurable device that it is more oriented for data intensive algorithms?

Magnetic RAM (MRAM) technology offers a solution to the intrinsic volatility of SRAM, with low reading and writing times, virtual limitless re-programmability and almost no area overhead compared with other technologies such as FLASH [3],[4].

At the architectural level, reconfigurable arrays (RA) try to overcome the disadvantages of FPGA based computing solutions by providing multiple bit-wide data-paths and complex operators instead of bit-level configurability. RA wider data-paths allow for the efficient implementation of complex operators in silicon. Thus, RA offer lower overhead, exploit better operation level parallelism and are better suited for data-oriented algorithms.

The objective of this research and development is, therefore, to design and develop a reconfigurable array architecture based on MRAM. This kind of device enjoys the advantage of non volatility due to be based on MRAM and at the same time offers the computational advantages that RA have over FPGA.

The use of magnetic tunneling junction (MTJ) memory cells in run-time reconfigurable hardware devices is a very promising technological solution and an emergent area of research. At the architectural level and in order to meet the increased computational requirements of data-intensive applications, reconfigurable devices are starting to evolve to coarse-grained compositions of functional units or program controlled processors, that are more able to improve performance and energy efficiency. The combination of these two aspects, as proposed in this research, can therefore provide a main contribution to technological innovation in the area of reconfigurable computing.

3 State of the Art

Since the 1990s, a number of architectures have been proposed and designed [2], [5], [6], [7], [8]. These architectures can be classified according to their basic interconnect structure, granularity and reconfiguration model. Practically all of the proposed architectures are SRAM-based and none of these is MRAM-based.

The predominant architectural arrangement has been a 2-D array or mesh, although some of the works, see for example [7], [8] have demonstrated that 1-dimensional coarse-grained linear array architectures can provide high-performance for specific application domains.

MRAM have so far been exclusively employed as elements on fine-grained FPGA look alike solutions, such as Look Up Tables (LUT) [3], [4], [9], [10], [11], [12] and no works have been published on MRAM-based coarse-grained reconfigurable arrays.

The most common MRAM cells consist of Magnetic Tunneling Junctions (MTJ) vertically integrated with silicon CMOS transistors as depicted on Fig. 1. A MTJ cell is conceptually made of two thin ferromagnetic layers separated by an ultra thin non-magnetic oxide layer [3], [4], [9]. Magnetic remanence of the ferromagnetic elements provides for non-volatility.

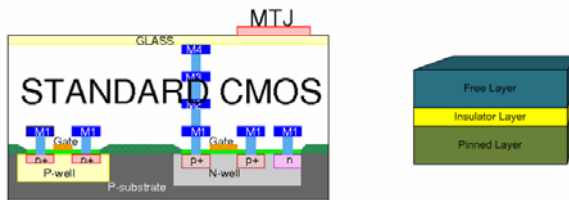


Fig. 1. Left: MTJ requires a few steps above standard CMOS. Right: Magnetic Tunnel Junction Structure.

The relative magnetic orientation of these layers defines two different values of equivalent resistance R_p (low resistance), when the free layer and the pinned layer are oriented to the same direction, and R_{ap} (high resistance), when the free layer and the pinned layer are oriented in opposite direction. These different values of equivalent resistance can be employed as state variables to represent the logic states '0' and '1'.

This resistance can be evaluated by sending current through the junction and measuring the voltages at the nodes of the MTJ. A structure is required to determine whether the junction is on a low resistivity configuration or in a high resistivity configuration.

To write information in a MTJ, a magnetic field is applied through the junction in order to force the change on the magnetic orientation of the free layer. The strength and nature of the required magnetic field depends on the writing approach. Currently, there are 3 writing approaches known as Field Induced Magnetic Switching (FIMS) [3], [4], [9], Thermally Assisted Switching (TAS) [10], [11] and Spin Transfer Torque (STT) [12].

Most MRAM produced so far are FIMS-based, but their size has been limited to 16Mb [13] because this technique has major drawbacks such as its susceptibility to

soft errors due to write selectivity, its lower scalability and its high current consumption. The TAS and STT approaches have been recently proposed to overcome the above mentioned issues. The TAS approach was preferred due to its weaker demands from the material point of view.

The TAS approach requires one bidirectional current to create the magnetic field. A local current (for each junction) is then used to enhance the effectiveness of the aforementioned magnetic field by momentarily heating the junction by Joule effect. This heating will unpin the free layer.

4 Research Contribution and Innovation

To the authors’ knowledge, this is the first R&D work where a MRAM based coarse grained reconfigurable array is proposed and a proof of concept prototype is developed and manufactured. Also, the proposed design employs a TAS writing based MRAM instead of the more conventional FIMS writing based MRAM.

The reconfigurable architecture proposed in this work [14] is organized as a 1-dimensional coarse-grained array. By providing a not too complex but very effective architecture organization 1-dimensional arrays are a good option to permit full characterization of the MRAM-based run-time reconfigurable technology. The basic processing elements (PE) are arithmetic and logic function units 4-bit wide (which can be easily scaled to larger bit-widths). The operation to be executed by each PE is controlled by the respective MRAM configuration memory bits. The array is run-time reconfigurable. The configuration memory on any PE can be rewritten while the circuit is in execution, and therefore a new reconfiguration can be loaded into any PE while the current application is running.

For computation, there are four global data buses, two input data buses and two output data buses, see Fig. 2. The A bus and the B bus feed the rALU with its operands. The S bus and the Carry out bus deliver the result of the logical-arithmetic operation executed and its eventual carry out. The exclusive access to the global data buses for each rALU is ensured by 16 selector signals, $Sel(k)$, one for each rALU.

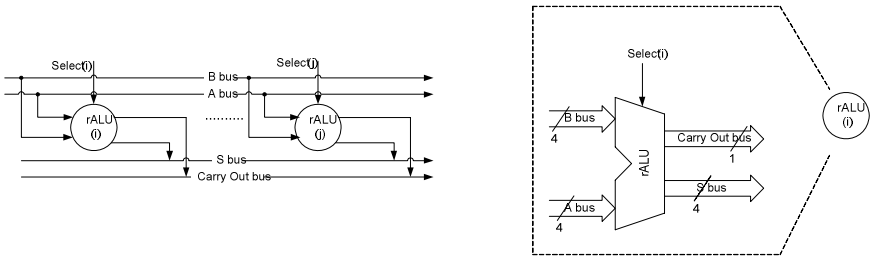


Fig. 2. General Computational Architecture

For configuration, there is one input global bus, as shown in Fig. 3. The Conf bus carries the data that will be stored on the MTJ to be used later on for configuration, thus supporting shallow dynamic re-configurability.

Sixteen selector signals determine which rALU(k) configuration memory will be written. The configuration memory contains the information that defines the ALU functionality. This configuration memory is made of an array of identical TAS-MRAM cells as the one depicted on Fig. 3.

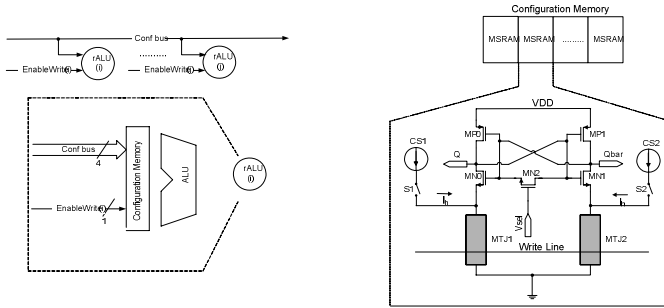


Fig. 3. Left: Configuration Architecture. Right: Configuration memory structure.

4.1 Datapath Unit

The data-path unit is implemented in a bit slice fashion. Since this module is implemented in full custom, a bit slice style makes it easier to change the available data-path width.

The 4 bit ALU is made by the concatenation of 4 identical data-path slices as the one depicted on Fig. 4. The less significant carry in bit can be set to ‘0’ or ‘1’ while the remaining carries in are connected to their direct predecessor lower order bit slice carry out. As shown in Fig. 4, the function computed by the ALU depends on the values of the five multiplexers selector values. These values depend on the data stored in the memory configuration bits and their values are easily generated from Fig. 4.

The adder is implemented as a classic Ripple Carry Adder. An adder-subtractor is implemented by adding a XOR gate to the adder input port A and by selectively setting the adder input port Carry In with a logic ‘1’.

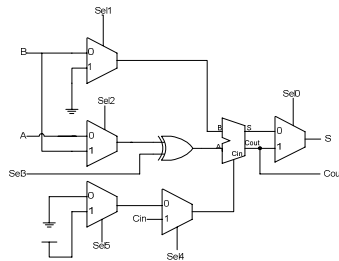


Fig. 4. Data path bit slice structure

The XOR and MUX designs are implemented using transmission gates, because of simplicity of implementation. The implementation of the arithmetic-logic functions takes advantage of the adder defining equations (1)-(2).

$$S = A \cdot B + A \cdot Cin + B \cdot Cin , \tag{1}$$

$$Cout = A \text{ xor } B \text{ xor } Cin . \tag{2}$$

For example, the operation AND is implemented by selecting the adder-subtrater output S, given by equation (1) and setting the adder-subtractor Carry-in input signal to logic ‘0’. This is accomplished by setting the controls signals, Sel0-Sel5 to logic ‘0’.

4.2 TAS-MRAM Based Storage Cell

Two circuit designs have been evaluated for the TAS-MRAM based storage cell [10], [11]. These circuits, as shown in Fig. 5 consist of:

- An Unbalanced Flip Flop (UFF) used as a sense amplifier.
- Two MTJ cells (MTJ1 and MTJ2).
- Two unidirectional current sources CS1 and CS2. These two current sources are responsible for the Joule effect on each MTJ.
- A write line that is employed to propagate the external magnetic field in either direction. This line is common to all storage cells.
- Two PMOS isolation transistors (for architecture 2).

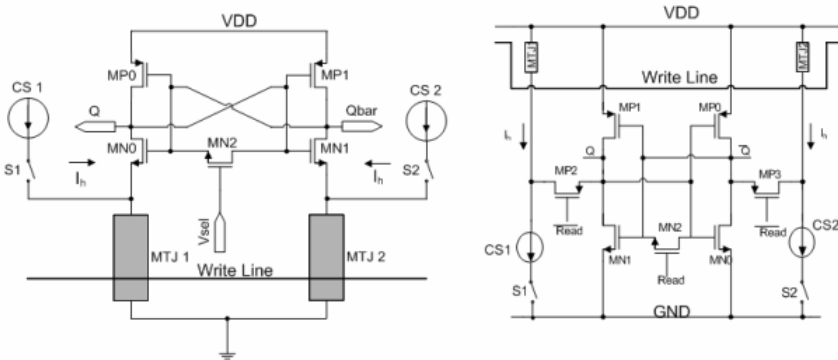


Fig. 5. TAS-MRAM based storage cell. Left Architecture 1. Right: Architecture 2.

During the read phase, the MN2 NMOS transistor acts as a short circuit, thus, the two cross-coupled inverters are pulled to a meta-stable operating state. The resistance value of each of the MTJ coupled to its respective inverter will move away the meta-stable operating point from one of the stable states and bring it closer to the other. So, when the Vsel/Read signal is released, the structure will move to the closest stable state. Afterwards, new information can be stored into the MTJ without altering the value stored in the UFF.

Due to the number of arithmetic-logic functions available on each ALU, a cluster of four memory bits is assigned to each individual ALU. Since the cluster size is small, the four control signals required are generated at the same writing cycle.

The inner working of the UFF coupled with MTJs provides the required support for run-time reconfiguration. So it is possible to write a new plane of configuration while the rALU is operating at full speed.

4.3 Bidirectional Current Generator

A bidirectional current generator capable of delivering a current higher than 20 mA has been designed to provide the current that is necessary to generate the bidirectional magnetic field responsible for switching the magnetic orientation of the MTJ's free layer and therefore write new information on the MTJ.

In order to better characterize the MTJ behavior, a digital controlled current source is used in association with a digital controlled current sink, see Fig. 6. The bidirectional current is shared among all TAS-MRAM storage cells, and therefore only one generator is required for the whole set of MTJ. The writing operation require 2 steps, in the first step the current flows in a direction that will allow writing a logical '1' on a given memory cell and in the second step, the current that flows in the write line will reverse its direction in order to allow writing a '0' in another memory cell.

Both R_{bias1} and R_{bias2} resistance are external and they are employed to change the intensity of the currents that passed across the current source and across the current sink respectively.

The switches S1 and S2 operate in complementary fashion. Their purpose is to turn on and turn off the current source as required by the logic responsible for the configuration. The switches S3 and S4 operates in complementary fashion and are responsible for turning off or on the current sink as required by the logic responsible for the configuration.

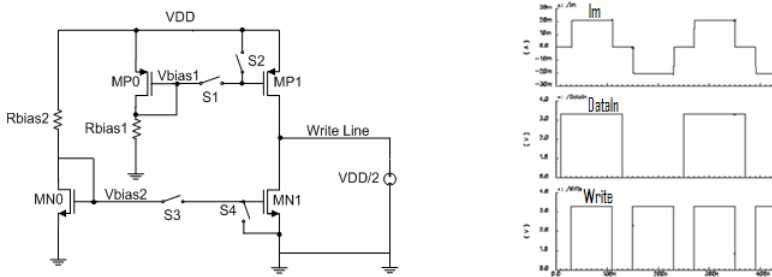


Fig. 6. Left: Bidirectional current generator. Right: Bidirectional current generator electrical behavior.

4.4 Unidirectional Current Generator

A unidirectional current generator able to deliver at least 1mA has been designed to provide the local current to momentarily heat each junction by Joule effect, as required by the TAS approach.

In order to facilitate the characterization of the MTJ behavior, a digital controlled current source has been designed, as depicted in Fig. 7. The unidirectional current generator is split into a front-end and a back-end. The Front-end is made of a resistance (R_{bias3}) and a PMOS transistor (MP0) while the back-end is made of switches (S1 and S2) and two identical PMOS transistor (MP1 and MP2). The front-end is shared among the whole set of unidirectional current generators while there is one back-end module associated with each memory cell pair of MTJ.

R_{bias3} is an external resistance that is employed to bias the back-end's driver transistors (MP1 and MP2). The switches S1 and S2 operate in complementary fashion. Their purpose is to turn on and turn off the PMOS driver transistors (MP1 and MP2). The aforementioned switches are opened and closed depending on the value of an internal generated digital control signal.

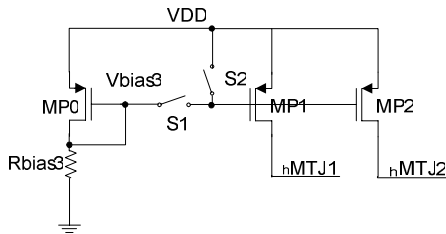


Fig. 7. Unidirectional current generator

5 Discussion of Results and Critical View

In order to fulfill our goal several architectural and circuits solutions are being studied, implemented and evaluated. The core blocks have already been designed for AMS 0.35 μm 4-Metal CMOS process technology and have been electrical simulated with a set of stimuli pre and post-layout under PVT corners. The electrical responses of the core blocks were on line with expectations.

An initial proof of concept coarse-grained reconfigurable array using TAS-MRAM was designed and a first prototype has been sent for fabrication. This prototype will be further processed at INESC-MN facilities to lay out the layers that will form the MTJs. A final test and characterization of the design will validate the viability of TAS-MRAM based memory element in the context of coarse grained reconfigurable computing and provide a study of the inner core of the basic computing unit, the reconfigurable ALU (rALU). It will also be employed to figure out how the MRAM configuration memory and the reconfigurable ALU fit together.

The circuit implements 1 of the 16 rALU that are part of the initial architecture as depicted on Fig. 2. Each ALU is capable of supporting up to 16 different operations, thus a four MRAM bit configuration memory is associated to each ALU. It is the information stored on the configuration memory that selects which operation will be carried out.

Both storage-cell circuits were simulated pre and post-layout under the same conditions, including the same set of AMS recommended PVT corners. Both performed as expected, however, architecture 1 has been preferred over architecture 2 because it

requires a lower TMR. The TMR (Tunneling Magneto Resistance) is defined by the ratio between the two MTJ equivalent resistances ($TMR = \Delta R/R$) and is a critical factor for the viability of the MRAM. According to the post-layout simulation results, architecture 2 requires a TMR of at least 40% while architecture 1 requires a TMR of 30%. On both cases the value of R_p was set to $1k\Omega$.

The whole MRAM rALU was electrical simulated post-layout under an extensive set of stimuli. In order to mitigate the ground and power supply bounce without increasing the number of both power supply pads and ground pads, external semi rings with multiple metal pads to reduce the inductance due to the bonding wires for power and ground have been connected to their respective power and ground pads. The same procedure was performed to the auxiliary voltage reference source VDD/2.

6 Conclusions and Further Work

As mentioned in section 4 and as far as the authors are aware this is the first proposal of a TAS-MRAM based coarse grained reconfigurable array. This research aims to provide a new approach to overcome the problems due to volatility that make FPGA less appealing for many applications, while maintaining the advantages inherent to coarse grain level operators. Further, it offers an open door to a very cost effective partial and full run time reconfigurability.

A scaled down prototype was sent for manufacturing at Austria Microsystems in late June 2009 and samples have been received in October 2009. A PCB board will now be designed for further evaluation and analysis of this prototype.

There are still many open questions for further work, either at circuit level or at the architectural level. For multi-context reconfiguration, further research is necessary to evaluate which of the storage cell architecture developed is more adequate for multiple planes of configuration and if there is one optimum number of configuration planes. It is also important to find if those results are fully technology scalable.

From the architectural point of view, future research should consider issues such as a fixed interconnect structure versus a fully reconfigurable one, handshake mechanisms or level of homogeneity among functional units.

Acknowledgments. We acknowledge the major contribution to this work from our partners of INESC-MN. This work is partially supported by the Portuguese Foundation for Science and Technology (FCT) through Project MRAM (PTDC/EEA-ELC/72933/2006).

References

1. Third Generation Non-Volatile FPGAs Enable System on Chip Functionality. White Paper, Lattice Semiconductor Corporation (2007)
2. Nageldinger, U.: Coarse-grained Reconfigurable Architectures Design Space Exploration. Ph.D Dissertation, C.S Department, University of Kaiserslautern (2001)
3. Bruchon, N., Cambon, G., Torres, L., Sassatelli, G.: Magnetic remanent memory structures for dynamically reconfigurable FPGA. In: International Conference on Field Programmable Logic and Applications, pp. 687–690. IEEE, Los Alamitos (2005)

4. Bruchon, N., Torres, L., Sassatelli, G., Cambon, G.: Magnetic tunnelling junction based FPGA. In: 2006 ACM/SIGDA 14th International Symposium on Field Programmable Gate Arrays, pp. 123–130. ACM, New York (2006)
5. Hartenstein, R.: A Decade of Reconfigurable Computing: a Visionary Retrospective. In: Conference on Design, Automation and Test in Europe, pp. 642–649. IEEE, Los Alamitos (2001)
6. Zain-ul-Abdin, Svensson, B.: Evolution in architectures and programming methodologies of coarse-grained reconfigurable computing. *Microprocessors and Microsystems* 33(3), 161–178 (2009)
7. Ebeling, C.: The General Rapid Architecture Description. Tech. Rep. UW-CSE-02-06-02, University of Washington (2002)
8. Goldstein, S.C., Schmit, H., Moe, M., Budiu, M., Cadambi, S., Taylor, R.R., Laufer, R.: PipeRench: A Coprocessor for Streaming Multimedia Acceleration. In: 26th International Symposium on Computer Architecture (ISCA 1999), pp. 28–39. IEEE, Los Alamitos (1999)
9. Bruchon, N., Torres, L., Sassatelli, G., Cambon, G.: New non-volatile FPGA concept using Magnetic Tunneling Junction. In: IEEE Computer Society Annual Symposium on VLSI: Emerging VLSI Technologies and Architectures (ISVLSI 2006), pp. 269–276. IEEE, Los Alamitos (2006)
10. Zhao, W., Belhaire, E., Dieny, B., Prenat, G., Chappert, C.: TAS-MRAM based Non-volatile FPGA logic circuit. In: International Conference on Field-Programmable Technology (ICFPT 2007), pp. 153–160. IEEE, Los Alamitos (2007)
11. Guillemenet, Y., Torres, L., Sassatelli, G., Bruchon, N., Hassoune, I.: A non-volatile runtime FPGA using thermally assisted switching MRAMS. In: International Conference on Field-Programmable Logic and Applications (FPL 2008), pp. 421–426. IEEE, Los Alamitos (2008)
12. Zhao, W., Belhaire, E., Mistral, Q., Nicolle, E., Devolder, T., Chappert, C.: Integration of Spin-RAM technology in FPGA circuits. In: 8th International Conference on Solid-State and Integrated Circuit Technology (ICSICT 2006), pp. 799–802. IEEE, Los Alamitos (2006)
13. DeBrosse, J., Arndt, C., Barwin, C., Bette, A., Gogl, D., Gow, E., Hoenigschmid, H., Lammers, S., Lamorey, M., Lu, Y., Maffitt, T., Maloney, K., Obermeyer, W., Sturm, A., Viehmann, H., Willmott, D., Wood, M., Gallagher, W.J., Mueller, G., Sitaram, A.R.: A 16Mb MRAM featuring bootstrapped write drivers. In: 2004 Symposium on VLSI Circuits, pp. 454–457. IEEE, Los Alamitos (2004)
14. Silva, V., Oliveira, L.B., Fernandes, J.R., Véstias, M.P., Neto, H.C.: Run-Time Reconfigurable Array using Magnetic RAM. In: 12th EUROMICRO Conference on Digital System Design, DSD 2009 (2009)