Exploration of magnetic memory for ultra low-power systems-on-chip

Guillaume Patrigeon, Sophiane Senni, Pascal Benoit, Lionel Torres

Exploration of magnetic memory for ultra low-power systems-on-chip

Guillaume Patrigeon, Sophiane Senni, Lionel Torres, Pascal Benoit
LIRMM-UMR CNRS 5506, Montpellier, France
Universite de Montpellier
firstname.lastname@lirmm.fr

Abstract

Memories are currently a real bottleneck to design high speed, low area and energy-efficient systems-on-chip (SoC). An important proportion of total power is spent on memory systems. Ultra low-power (ULP) SoC often use different memory technologies to keep the advantages of each one (area, energy consumption, latency and non-volatility), however there are still penalties and this add more complexity at every development levels. MRAM (Magnetic Random Access Memory) is seen as a promising alternative solution to replace both traditional SRAM (Static Random Access Memory) and NVM (Non Volatile Memory), thanks to its high density, low read/right latency, non-volatility and negligible leakage current. The aim of this work is to explore the possibilities of using MRAM in ULP SoC at various memory levels.

1 Introduction

The permanent need of smaller and consuming less energy SoC (highly requested for Internet of Things (IoT) devices) lead to explore new technologies, architectures and techniques at every levels of the development of these systems.

Several studies explored the possibilities of MRAM and the combination of MRAM and CMOS to propose fast and low-power memories [7, 9] and new management techniques [2, 1, 6].

However, mostly of these studies focus on high-performance embedded devices, but few focus on ULP SoC where the resources are limited (small memories, small or no cache, sub-gigahertz operating frequency...) and where area and energy consumption are the main issues. The MASTA project (MRAM Based Design, Test and Reliability for Ultra-Low-Power SoC), which has received funding from the French National Research Agency under grants ANR-15-CE24-0033-01, has been launched with objective the exploration of MRAM memories in ULP SoC.

Following Section 2 presents MASTA project, Section 3 overviews microprocessors and microcontrollers architectures, design space exploration is presented in Section 4 and Section 5 concludes this paper.

2 MASTA

The objective of the MASTA project is to investigate, design, develop, and analyze hybrid CMOS-MRAM normally-off computing architectures in which MRAM are used at various levels of the memory hierarchy (memory-in-logic, register files, different levels of cache, main memory) along with traditional CMOS devices and memories to achieve ultra-low-power, high performance and low cost.

This project is divided in 3 Work Packages (WP):

- Circuit level design: designing complex circuits in advanced technologies requires adapted digital design tools. This WP aims at integrating the magnetic technology in a standard design flow (synthesis, place and route, electrical simulation, design check, layout/schematic verification...), in the form of a Magnetic Process Design Kit compatible with the standards of the industry and flexible enough to adapt to the various flavors of MRAM generations and their evolution.

- Test and reliability aspects: since MRAM memories are designed and fabricated using different principles, process, and materials, they fail differently compared to conventional CMOS memories. The aim of this WP is to understand the failure mechanisms of a hybrid CMOS/STT-MRAM technology, to develop appropriate fault models and provide efficient testing techniques to ensure the reliability of the finals systems.

- Architecture design space exploration: the objective of this WP is to study how to exploit the MRAM technology, from the software to the hardware level, to improve the reliability, resilience and the power consumption of a processor, and to create recovery strategies after fault

detect, by using the ability to store information related to non-volatile system.

3 Generic architecture for ultra-low power systems-on-chip

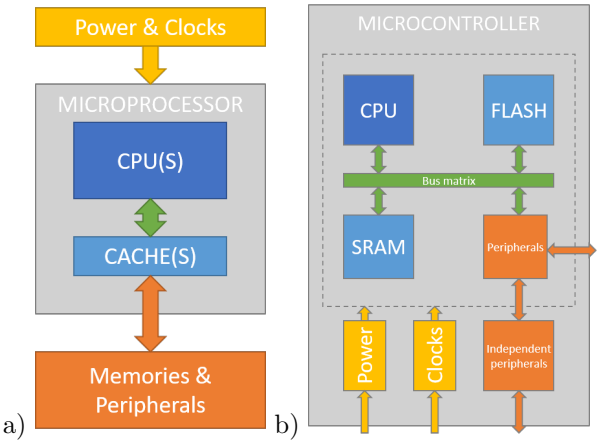

Figure 1. (a) Microprocessor architecture, (b) Microcontroller architecture (simplified)

Microcontrollers use to integrated a lot of peripherals needed by the system (power control, clocks, memories, peripherals) that are required but not integrated by common microprocessors (Figure 1), that is why microcontrollers are smaller and so easier to integrate than microprocessors. Even if microprocessors have higher performance than microcontrollers, most of application that are highly constrained by area and power (which is the case for a lot of IoT applications) use microcontrollers.

Due to the limitations of traditional memory technologies, microcontrollers have both non-volatile memories (mostly FLASH for code in today's generic microcontrollers but other technologies exist), and volatile memory (mostly SRAM without cache). Usually the code instructions are fetched directly from the NVM, and if data need to be kept after reset they are stored in the code memory or a second NVM (on-chip or external).

Also, CPU used in microcontrollers are usually lighter than microprocessor's ones, with lighter instruction set (mostly RISC), lighter arithmetical and logical unit (multiplier, divider and floating point unit are optional), and sometimes smaller data architecture (8-bit, 16-bit...).

4 Design space exploration

Dedicated tools make possible the exploration of different kind of memory technologies in a microprocessor at different levels [5] ([8] for circuit level and [3, 4] for architecture level).

The fact that microcontrollers integrates a lot of modules and peripherals, which their number, their implementation and so their impact on the final system change depending on the manufacturer and the target application, makes very hard the development of a unique tool which takes account of all the varieties of microcontrollers.

5 Conclusion

MRAM are seen as an alternative solution to solve most of current memories issues (latency, static energy consumption, area, data volatility). While several studies explore MRAM possibilities in systems-on-chip for embedded systems, only a few focus on ultra-low-power applications. The aim of this work is to explore the MRAM possibilities for ultra-low-power systems-on-chip in a context of low-power applications, taking account of the various architectures, and adapting the existing exploration methods for these architectures.

References

[1] J. W. *et al.* Oap: An obstruction-aware cache management policy for stt-ram last-level caches. *Proc. Conf. Design, Automat. Test Eur. Consort.*, pages 847–852, 2013.

[2] K. W. K. *et al.* Aware (asymmetric write architecture with redundant blocks): A high write speed stt-mram cache architecture. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions*, 22(4):712–720, 2014.

[3] N. B. *et al.* The gem5 simulator. *ACM SIGARCH Computer Architecture News*, 39(2):1–7, 2011.

[4] S. L. *et al.* Mcpat: an integrated power, area, and timing modeling framework for multicore and manycore architectures. *Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture. ACM*, pages 469–480, 2009.

[5] T. D. *et al.* Magpie: System-level evaluation of manycore systems with emerging memory technologies. *2nd International Workshop on Emerging Memory Solutions (EMS) co-located with DATE'17*, March 2012.

[6] W. K. C. *et al.* Architecture and data migration methodology for l1 cache design with hybrid sram and volatile stt-ram configuration. *Microprocessors and Microsystems*, 2015.

[7] W. X. *et al.* Design of last-level on-chip cache using spin-torque transfer ram (stt ram). *Very Large Scale Integrated Systems, IEEE Transactions*, 19(3):483–493, 2011.

[8] X. D. *et al.* Nvsim: A circuit-level performance, energy, and area model for emerging nonvolatile memory. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions*, 31(7):994–1007, 2012.

[9] X. W. *et al.* Power and performance of read-write aware hybrid caches with non-volatile memories. *IEEE Design, Automat. Test Eur. Conf. Exhibit.*, pages 737–742, 2009.