MOS/MTJ-Hybrid Circuit with Nonvolatile Logic-in-Memory Architecture

Masanori Natsui and Takahiro Hanyu
Research Institute of Electrical Communication, Tohoku University
2-1-1 Katahira, Aoba-ku, Sendai, Miyagi 980-8577, Japan
Phone: +81-22-217-5679, E-mail: {natsui, hanyu}@ncc.riec.tohoku.ac.jp

1. Introduction

Reduction of power consumption and interconnection delay are the two major issues for the next generation very large scale integrated circuits (VLSIs). Drastic increase of static power dissipation due to leakage current is being anticipated in beyond 45 nm complementary metal oxide semiconductor (CMOS) technology [1]. In addition, increase in the length of global-interconnection in advanced VLSIs results in further increase of both power and delay.

Logic-in-memory architecture [2], where memory elements are distributed over a logic-circuit plane, combined with nonvolatile memory is expected to realize both ultra-low-power and shorten interconnection delay [3]-[7]. However, in order to fully take advantage of the logic-in-memory architecture, it is important to implement a nonvolatile memory that has a capability of shorter access time below 10 ns, unlimited endurance, scalable write, and small dimension comparable to the employed CMOS technology. The only available candidate of a nonvolatile memory that could satisfy all the above requirements at this stage is the one using magnetic tunnel junction (MTJ) with spin-injection write [8]-[10].

In this paper, we present a basic concept of the MTJ-based nonvolatile logic-in-memory circuit structure, and a possibility of a highly-efficient power-gating scheme for fully-parallel VLSI systems using this structure.

2. Nonvolatile Logic-in-Memory Circuit Structure

Fig. 1 shows an MTJ-based logic-in-memory circuit model. It consists of three basic components; a cross-coupled keeper (CCK), a logic-circuit tree and a dynamic current source (DCS). The CCK generates complementary binary outputs, z and z', in accordance with a magnitude-comparison result between two current signals, I_z and I_{z'}, where precise current difference can be immediately detected by using the feedback circuit structure. The use of the DCS makes it possible to cut off steady current from VDD to GND, which results in low-power dissipation. Arbitrary logic circuits are realized by programming the configuration of the logic-circuit tree. For example, two-input AND and two-input OR gates are realized by using 4 NMOS transistors and two MTJ devices as shown in Fig. 1. By changing the wired-connection points of the logic-circuit tree, two different gates are simply realized.

3. MTJ-Based Nonvolatile Full Adder

The proposed MTJ-based nonvolatile logic-in-memory circuit is suitable for realizing a fully-parallel VLSI, because nonvolatile storage elements are merged into a fine-grain processing element (PE). As a typical example, we discuss about a nonvolatile full adder for an operation unit of sum of absolute differences (SAD) which is used for a motion-vector detection of an MPEG encoding [6]-[7]. Fig. 2(a) shows the circuit diagram of the full adder [11], whose logic function is represented by the table in Fig. 2(b). It consists of SUM-circuit and CARRY-circuit parts, where the symbols A (A'); the complement of A) and C_l (C'_l) are the external inputs and the symbol B (B') is a stored input. The use of a dynamic logic
style [12] controlled by clock signals, CLK and CLK’, cuts off the steady current flow from the supply voltage VDD to GND, which reduces the dynamic power dissipation of the circuit. The stored data is programmed by controlling external signals. Complementary stored inputs, B and B’, are programmed by using individual current-flow path, which is selectable by the word lines, WL1, WL2, WL3, and WL4, and the bit lines, BL and BL’. For example, in the case of storing B = 0 into the corresponding MTJ in the SUM circuit, the word line WL1 is set to the supply voltage VDD, and BL and BL’ are set to GND and VDD, respectively, which makes the current-flow path through the MTJ set up as shown in Fig. 2(a). All the external inputs and the complementary clock signals are turned off during the above write operation.

4. Impact towards Fine-Grained Power-Gating Control

Power gating scheme is one of the most useful methods to cut off the leakage current completely. In logic-in-memory architecture proposed here, data can be easily and quickly stored to NV storage elements that are distributed over the logic-circuit plane. Hence, the supply voltage can be immediately cut off without data transmission into external nonvolatile storage devices when the circuit changes to a standby mode. This property achieves very small overhead to change the operation modes in terms of time delay and power dissipation, which enables to manage the power supply of fully-parallel VLSI with fine granularity.

As one possible application suitable for using the proposed technique, we consider a motion-vector extraction system with fully-parallel architecture. In this system, reference window’s data is stored into each processing module in advance, and pixel data in the current frame is applied serially to perform SAD operation in pipelining manner.

Figure 3 shows time charts of power dissipation in the system based on a conventional CMOS process and that based on a CMOS/MTJ-hybrid process with threshold detection algorithm. Since SRAM-based volatile memory is used in the conventional CMOS-based implementation, it is impossible to cut off the power supply, which consumes large static power. In contrast, in the proposed CMOS/MTJ-hybrid implementation, there is no static power dissipation realized by fine-grained power gating. Additionally, by using the output monitoring technique, the “meaningful” activation time of PEs can be estimated more precisely, which results in the great reduction of static power dissipation.

5. Conclusion

A new circuit structure, called a MOS/MTJ -hybrid logic-in-memory circuit, has been presented and its usefulness in realizing an ultra-low-power system with a fine-grained power-gating technique has been demonstrated in a particular application.

Acknowledgements

This work was supported by Research and Development for

References