Roles of Non-Volatile Devices in Future Computer Systems: Normally-Off Computers

Koji Ando
National Institute of Advanced Industrial Science and Technology (AIST), Japan

Sumio Ikegawa
Toshiba Corporation, Japan

Keiko Abe
Toshiba Corporation, Japan

Shinobu Fujita
Toshiba Corporation, Japan

Hiroaki Yoda
Toshiba Corporation, Japan

ABSTRACT

Normally-off Computer (NOC) is a computer designed with a new concept in which most parts of a computer use non-volatile functionalities. The power of NOC will be completely turned off at short intervals when computing is not required without users even being aware of it. Thus far, slow latency, limited endurance, and small capacity of non-volatile memories have been preventing the implementation of non-volatile functionalities into computer architecture. Emerging magnetic memory, Spintronic RAM (Spin-RAM), is now changing the premise of computer design. Spin-RAM will soon replace dynamic RAM, and will actualize instant-on computer. For NOC, however, inventors must develop many new technologies including non-volatile cache memories, hierarchical non-volatile memory architecture, advanced power gating, non-volatile peripheral circuits, and non-volatile displays. The authors discuss the present status of Spin-RAM technology and the challenges for achieving NOC.

DOI: 10.4018/978-1-4666-1842-8.ch005
INTRODUCTION

It takes about a few 100 ms to move one’s finger from one key to another over a keyboard. During that period, a PC only waits for input, wasting energy. When one gives a 30-minute presentation with 30 figures, a PC requires less than 1 second to prepare all the figures thanks to the powerful computing ability of today’s processors. However, a PC’s power cannot be turned off during the presentation. People have asked why powering off computers cannot be done when computing is not required. The main components of current computers are made of CMOS transistors and Dynamic RAM (DRAM). They are volatile devices and lose all information when powered off. Once powered off, one cannot resume the task.

If non-volatile devices, which can keep information without using power, replace DRAM and all or a part of CMOS transistors, computer power can be completely turned off at short intervals without a user even being aware of it. Such a computer, i.e., “Normally-Off Computer” (NOC), would drastically reduce power consumption and enable greatly prolonged use of electronic instruments. If laptop PC batteries can last a few months, new uses of PCs will emerge. Mobile PCs running on solar batteries or hand-crank dynamos would be useful for education and business in developing countries and for evacuation and rescue activities during natural disasters. Numerous tiny computers, which are embedded in walls and streets to support daily life, will also not be limited by the power supply problem.

When the concept of NOC was proposed ten years ago (Ando, 2001), the performance of non-volatile devices was low, and it seemed to be impossible to implement non-volatile functionalities into computer architecture except for storage devices. However, recent progress in non-volatile memory technologies, especially emerging magnetic memory “Spintronic RAM” (Spin-RAM), may drastically change the premise of computer design and make NOC a possibility in the near future. In this chapter, we discuss the present status of Spin-RAM technology and the challenges for implementing non-volatile functionalities into a variety of computer components.

MEMORIES IN COMPUTER ARCHITECTURE

It is interesting to point out that the precursors of the computer, such as the abacus, slide rule, and mechanical calculators, were intrinsically non-volatile. When electronic devices, such as vacuum tubes and transistors, were introduced into calculators for increasing calculation speed, the non-volatile functionality began to be phased out. Non-volatile magnetic core memory used for main memory was the last one to withstand this trend. However, it was also eventually replaced with volatile DRAM, which is produced with the Si planer integration process and has more memory capacity than core memory. This history indicates that unless fast access speed, i.e., latency, and large memory capacity are achieved simultaneously, we cannot implement non-volatile functionalities into computer architecture (Figure 1).

The slow latency of memory has been a serious problem even for DRAM. In the first PCs, the processor and DRAM were directly connected to each other because their performances were well balanced. However, in the 30 years from the first PCs, processor performance has been improved by about 20,000 times while improvement in DRAM latency was less than 10 times (Hennessy, 2006). This large gap comes from the different natures of transistor circuits and memory. All Arithmetic Logic Unit (ALU)/FlipFlop, register files, and caches are transistor-only circuits. The register files and caches are made with Static RAM (SRAM) “memory,” but SRAM is not a genuine memory. It is a transistor flip-flop circuit that emulates the memory function. The dramatic improvement in transistor