Architecture Design with STT-RAM: Opportunities and Challenges

Ping Chi[†], Shuangchen Li[†], Yuanqing Cheng[†], Yu Lu[‡], Seung H. Kang[‡], Yuan Xie[†]

[†]Department of Electrical and Computer Engineering, University of California, Santa Barbara, USA

[‡]Qualcomm Incorporated, San Diego, USA

[†]{pingchi, shuangchenli, yuanqing, yuanxie}@ece.ucsb.edu, [‡]yu.lu@qualcomm.com

Abstract— **The emerging spin-transfer torque magnetic random-access memory (STT-RAM) has attracted a lot of interest from both academia and industry in recent years. It has been considered as a promising replacement of SRAM and DRAM in the cache and memory system design thanks to many advantages, including non-volatility, low leakage power, SRAM comparable read performance and read energy consumption, higher density than SRAM, better scalability than conventional CMOS technologies, and good CMOS compatibility. However, the disadvantages of STT-RAM, such as higher write energy and longer write latency than SRAM, also bring design challenges. This paper introduces state-of-the-art architectural approaches to adopt STT-RAM in the cache and memory system design by taking advantage of the opportunities brought by STT-RAM as well as overcoming the challenges.**

I. INTRODUCTION

In the conventional memory hierarchy, SRAM and DRAM have played important roles during the evolvement of modern computer systems. However, with the coming of multi-core and many-core processor era and the continuous technology node shrinking, there are several severe design challenges. First, the power consumption of the cache and memory is a big concern in contemporary computer systems, especially for high performance and big data computing applications [1, 2]. As technology scales, SRAM and DRAM both suffer from severe leakage power dissipation. Moreover, the emerging many-core processor design demands a large capacity of cache to facilitate the data transfer among the cores. For example, Intel Iris Pro Graphics is featured with a 128MB L4 cache [3]. The large area overhead of SRAM prevents it from being adopted in large-scale caches. Furthermore, as DRAM density advances, the increasing refresh penalty results in significant performance degradation and high refresh power consumption [4].

In recent years, some emerging non-volatile memory (NVM) technologies, such as phase change memory (PCM) [5] and spin-transfer torque magnetic random access memory (STT-RAM) [6], have been developed and studied to replace SRAM and DRAM in cache and memory system design. Among them, STT-RAM has been considered as the most promising SRAM replacement to build on-chip large-scale caches and also a potential DRAM alternative to build energy-efficient main memory, thanks to its unique characteristics. The STT-RAM stores data in a magnetic tunneling junction (MTJ), which is a multi-layer structure consisting of two ferromagnetic layers and an insulating barrier sandwiched between them. This data storage mechanism has zero leakage power, and can retain data for ten years without power supply. In addition, STT-RAM has a comparable read speed with that of SRAM, and its single cell structure presents much higher density than SRAM. Therefore, as a SRAM replacement, it can enlarge the cache capacity without area overhead. Moreover, the endurance test result showed that STT-RAM cells can sustain more than 10^{12} writes, and the estimated endurance value of STT-RAM is up to 10^{15} [7]. STT-RAM is also inherently immune to high energy particle radiation from space, which is a promising property for mission critical systems such as servers for aerospace and aeronautical applications.

Although STT-RAM has many nice features, it also has some disadvantages that may be a hindrance for its adoption. First of all, the write operation of STT-RAM is generally slower and more energy-consuming than its read operation. In addition, as the technology generation advances, the write current decreases quickly, whereas the read current does not scale well, and therefore the read disturbance error increases [8]. Moreover, when the MTJ is over-stressed by the write voltage, it may not achieve the required endurance for on-chip caches [9].

In this paper, we survey the state-of-the-art research work on STT-RAM based cache and memory system design. By investigating the proposed approaches to leverage the opportunities and overcome the challenges of STT-RAM, we provide an overview of this research area from an architectural perspective.

II. BACKGROUND

STT-RAM is the second generation of Magnetoresistive RAM (MRAM). The key component of MRAM to store