

Review

In-memory computing : synthesis and optimization

Shirinzadeh S., Drechsler R., Springer International Publishing, New York, NY, 2019. 115 pp. Type: Book

Date Reviewed: 10/28/19

For many years, von Neumann's computer architecture model was one of the most important paradigms in computer engineering. The central processing unit (CPU) with its main memory (primary storage) is connected to external storage (random access memory (RAM), disks, and so on). It would be nice to have enough RAM to store all the data processed by computers, but this is impossible because of the very high cost. Hard disks have a much higher capacity than RAM, but the disk access time is many times longer. As a result, the communication between CPU and memory causes long latency, limiting processor performance; the result is a memory bottleneck. Since it is expected that artificial intelligence (AI) will be able to process big datasets and the Internet of Things will be a new omnipresent challenge, this bottleneck is the main obstacle. Is there any solution? How can we avoid the cost of communication between CPU and memory? The authors of this book aim to convince readers that in-memory computing is the answer. How is it possible to process data inside memory? Is there a need to have new electronic technologies beyond the complementary metal-oxide-semiconductor (CMOS)?

Resistive RAM (RRAM) is the answer. It stores data in the form of electric resistance that can represent binary values. How does RRAM work? What is its basic element? The book, especially chapter 2, opens this enigmatic box and gives readers some necessary technological details. Known since the 1960s, an interesting property of an oxide insulator switched by two metal electrodes is the abrupt switching of its resistance under the presence of voltage changes. An RRAM cell is like an electrical resistor in which resistance can be switched; this is exactly the same as a memristor except the RRAM device doesn't have magnetic flow. Leon Chua proposed the concept of a memristor in 1971, as a fourth basic passive circuit element. Until 2008, there were no fabricated elements remaining in this property. Now we have at least three different memristor physical models. RRAM devices allow for in-memory logic implementation, leading to memristor-aided logic (MAGIC).

The book presents a comprehensive approach to in-memory logic synthesis and optimization for computing hardware and architecture. It consists of four main chapters. Chapter 2, except for the previously mentioned historical background, gives details about logic representations like binary decision diagrams (BDDs), and-inverter graphs (AIGs), and majority-inverter graphs (MIGs). Chapter 3 focuses on BDD optimization and approximation--BDD is used in VLSI computer-aided design (CAD)--as the main optimization technique for optimal variable ordering and low-cost circuits; here, the most important algorithm is MOB. Chapter 4 focuses on RRAM synthesis for logic-in-memory computing in relation to NAND gates. It uses BDD, AIG, and MIG (discussed in chapter 2). There are several examples of BDD-based synthesis ("the realization of an IMP-based MUX"), AIG-based synthesis (AND/NAND gates), and MIG-based synthesis (majority gate). The last chapter explains, and evaluates, how programmable logic-in-memory (PLiM) is used in practice for processing.

This short but very interesting book gives readers plenty of food for thought. Readers convinced of the omnipresence of the von Neumann architecture and the Harvard architecture in computer engineering will be presented with a possible paradigm shift. For many years in-memory computation was hard to imagine, and most of us were accustomed to the existing approaches in computer engineering and processing. This book reveals some new possibilities and uncovered areas. We will see the future of in-memory computing development, but the book is also a worthwhile read on what can be done to perform faster computations.

Reviewer: Dominik Strzalka

Review #: CR146746