Non-Volatile In-Memory Computing by Spintronics

Synthesis Lectures on Emerging Engineering Technologies

Editor
Kris Iniewski, ***Redlen Technologies, Inc.***

Non-Volatile In-Memory Computing by Spintronics
Hao Yu, Leibin Ni, and Yuhao Wang
2016

Layout Techniques for MOSFETs
Salvador Pinillos Gimenez
2016

Resistive Random Access Memory (RRAM)
Shimeng Yu
2016

The Digital Revolution
Bob Merritt
2016

Compound Semiconductor Materials and Devices
Zhaojun Liu, Tongde Huang, Qiang Li, Xing Lu, and Xinbo Zou
2016

New Prospects of Integrating Low Substrate Temperatures with Scaling-Sustained Device Architectural Innovation
Nabil Shovon Ashraf, Shawon Alam, and Mohaiminul Alam
2016

Advances in Reflectometric Sensing for Industrial Applications
Andrea Cataldo, Egidio De Benedetto, and Giuseppe Cannazza
2016

Sustaining Moore's Law: Uncertainty Leading to a Certainty of IoT Revolution
Apek Mulay
2015

Non-Volatile In-Memory Computing by Spintronics

Hao Yu, Leibin Ni, and Yuhao Wang

ISBN: 978-3-031-00904-4 paperback
ISBN: 978-3-031-02032-2 ebook

DOI 10.1007/978-3-031-02032-2

A Publication in the Springer series
SYNTHESIS LECTURES ON EMERGING ENGINEERING TECHNOLOGIES

Lecture #8
Series Editor: Kris Iniewski, *Redlen Technologies, Inc.*
Series ISSN
Print 2381-1412 Electronic 2381-1439

Non-Volatile In-Memory Computing by Spintronics

Hao Yu
Nanyang Technological University, Singapore

Leibin Ni
Nanyang Technological University, Singapore

Yuhao Wang
Synopsis, California, USA

SYNTHESIS LECTURES ON EMERGING ENGINEERING TECHNOLOGIES #8

ABSTRACT

Exa-scale computing needs to re-examine the existing hardware platform that can support intensive data-oriented computing. Since the main bottleneck is from memory, we aim to develop an energy-efficient in-memory computing platform in this book. First, the models of spin-transfer torque magnetic tunnel junction and racetrack memory are presented. Next, we show that the spintronics could be a candidate for future data-oriented computing for storage, logic, and interconnect. As a result, by utilizing spintronics, in-memory-based computing has been applied for data encryption and machine learning. The implementations of in-memory AES, Simon cipher, as well as interconnect are explained in details. In addition, in-memory-based machine learning and face recognition are also illustrated in this book.

KEYWORDS

Spintronics, in-memory computing, non-volatile memory, logic-memory integration, data encryption, AES, machine learning, data analytics, hardware accelerator

Contents

Preface

The existing memory technologies have critical challenges when scaling at nanoscale due to process variation, leakage current, and I/O access limitations. Recently, there are two research trends attempting to alleviate the memory-wall issues for future big-data storage and processing system.

First, the emerging non-volatile memory technologies, such as spin-transfer torque memory, domain wall nanowire (or racetrack) memory, etc., have shown significantly reduced standby power and increased integration density, as well as close-to DRAM/SRAM access speed. Therefore, they are considered as promising next generation memory for big-data applications.

Second, due to high data-level parallelism in big-data applications, large number of application specific accelerators can be deployed for data processing. However, the I/O bandwidth limitation will still be the bottleneck in such a memory-logic integration approach. Instead, an in-memory computing platform will be highly desired with less dependence on I/Os.

In order to achieve low power and high throughput (or energy efficiency) in big-data computing, one can build an in-memory non-volatile memory (NVM) hardware platform, where both the memory and computing resources are based on NVM devices with instant-switch-on as well as to ultra-low leakage current. Such an architecture can achieve significant power reduction due to the non-volatility. Moreover, one can develop an NVM logic accelerator that can perform domain-specific computations such as data encryption and also data analytics in a logic-in-memory fashion. Compared to conventional memory-logic-integration architectures, the storage data do not need to be loaded into volatile main memory, processed by logic, and written back afterward with significant I/O communication congestion.

In this book, we discuss the following research studies on this regard. First, we introduce a non-volatile in-memory architecture that both data storage and logic computing are inside the memory block. The data storage and logic computing is located in pairs to perform a distributed in-memory computing. We illustrate the NVM-based basic memory and logic components and find significant power reduction. Second, we develop a SPICE-like simulator NVM SPICE, which implements physical models for non-volatile devices in a similar way as the BSIM model for MOSFET. We further develop the data storage and logic computing on spintronic devices. These operations can be deployed for in-memory data encryption and analytics. Moreover, we illustrate how data encryption including advanced encryption standard (AES) and Simon cipher can be implemented in this architecture. Data analytics applications such as machine learning for super-resolution and face recognition are performed on the developed NVM platform as well.

This book provides a state-of-the-art summary for the latest literature on emerging non-volatile spintronic technologies and covers the entire design flow from device, circuit to system perspectives, which is organized into five chapters. Chapter 1 introduces the basics of conven-

tional memory architecture with traditional semiconductor devices as well as the non-volatile in-memory architecture. Chapter 2 details the device characterization for spintronics by non-electrical states and the according storage and logic implementation by spintronics. Chapter 3 explores the in-memory data encryption based on STT-RAM as well as domain wall nanowire. The implementations of in-memory AES, domain wall-based Simon cipher and low power interconnect are explained in detail. Chapter 4 presents the system level architectures with data analytics applications for the emerging non-volatile memory. In-memory machine learning and face recognition are discussed in this chapter. This book assumes that readers have basic knowledge of semiconductor device physics. It will be a good reference for senior undergraduate and graduate students who are performing research on non-volatile memory technologies.

Hao Yu, Leibin Ni, and Yuhao Wang
Singapore
October 2016

Acknowledgments

The authors would like to thank their colleagues at CMOS Emerging Technology Group at Nanyang Technological University: Dr. Wei Fei, Dr. Yang Shang, Dr. Sai Manoj P.D., Mr. Hantao Huang, Mr. Zichuan Liu, Miss. Hang Xu, and Mr. Rai Suleman Khalid. The authors also acknowledge with gratitude discussions with Prof. Guangbin Huang, Prof. Anupam Chattopadhyay, Prof. Chip-hong Chang, Prof. Jianhua Yang, Prof. Kaushik Roy, Prof. Weisheng Zhao, Prof. Dennis Sylvester, Prof. Kevin Cao, Prof. Yuan Xie, Prof. Yiran Chen, Prof. Hai Li, Dr. Tanay Karnik, Dr. Jing Li, Prof. Wei Zhang, Prof. Tony Kim, Prof. Wen-Siang Lew, Prof. Kiat-Seng Yeo, Dr. Rajiv V. Joshi, and Mr. Suman Deb. Their support was invaluable to us during the writing of this book. The relevant research is funded by MOE Tier-2 (MOE2015-T2-2-013) titled "Sparse-represented Non-volatile In-memory Accelerator for Big-Data Analytics" and NRF CRP (NRF-CRP9-2011-01) titled "Magnetic Domain Wall Logic Gate Device," from Singapore.

Hao Yu, Leibin Ni, and Yuhao Wang
October 2016