

Preface

Variation in performance and power across manufactured parts and their operating conditions is an accepted reality in modern microelectronic manufacturing processes with geometries in nanometer scales. This book views such variations both as a challenge as well an opportunity to rearchitect the hardware/software interface that provides more resilient system architectures. We start with an examination of how variability manifests itself across various levels of microelectronic systems. We examine various mechanisms designers use, and can use, to combat negative effects of variability.

This book attempts a comprehensive look at the entire software/hardware stack and system architecture in order to devise effective strategies to address microelectronic variability. First, we review the key concepts on *timing errors* caused by various variability sources. We use a two-pronged strategy to mitigate such errors by jointly exposing hardware variations to the software and by exploiting flexibility made possible by parallel processing. We consider methods to predict and prevent, detect and correct, and finally conditions under which such errors can be accepted. For each of these methods, our work spans defining and measuring the notion of error tolerance at various levels, from instructions to procedures to parallel programs. These measures essentially capture the likelihood of errors and associated cost of error correction at different levels. The result is a design platform that enables us to combine these methods that enable detection and correction of erroneous results within a defined criterion for acceptable errors using a notion of memoization across the hardware/software interface. Pursuing this strategy, we develop a set of software techniques and microarchitecture optimizations for improving cost and scale of these methods in massively parallel computing units, such as general-purpose graphics processing units (GP-GPUs), clustered many-core architectures, and field-programmable gate array (FPGA) accelerators.

Our results show that parallel architectures and use of parallelism in general provides the best means to combat and exploit variability. Using such programmable parallel accelerator architectures, we show how system designers can

coordinate propagation of error information and its effects along with new techniques for memoization and memristive associative memory. This book naturally leads to use of these techniques into emerging area of approximate computing, and how these can be used in building resilient and efficient computing systems.

Berkeley, USA
Zürich, Switzerland
San Diego, USA
January 2017

Abbas Rahimi
Luca Benini
Rajesh K. Gupta

From Variability Tolerance to Approximate Computing in
Parallel Integrated Architectures and Accelerators

Rahimi, A.; Benini, L.; Gupta, R.

2017, XV, 197 p. 86 illus., 48 illus. in color., Hardcover

ISBN: 978-3-319-53767-2