SpringerBriefs in Computer Science

SpringerBriefs present concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 pages, the series covers a range of content from professional to academic.
Typical topics might include:

- A timely report of state-of-the art analytical techniques
- A bridge between new research results, as published in journal articles, and a contextual literature review
- A snapshot of a hot or emerging topic
- An in-depth case study or clinical example
- A presentation of core concepts that students must understand in order to make independent contributions

Briefs allow authors to present their ideas and readers to absorb them with minimal time investment. Briefs will be published as part of Springer's eBook collection, with millions of users worldwide. In addition, Briefs will be available for individual print and electronic purchase. Briefs are characterized by fast, global electronic dissemination, standard publishing contracts, easy-to-use manuscript preparation and formatting guidelines, and expedited production schedules. We aim for publication 8–12 weeks after acceptance. Both solicited and unsolicited manuscripts are considered for publication in this series.

More information about this series at http://www.springer.com/series/10028

Arthur Francisco Lorenzon
Antonio Carlos Schneider Beck Filho

Parallel Computing Hits the Power Wall

Principles, Challenges, and a Survey of Solutions

Springer

Arthur Francisco Lorenzon
Department of Computer Science
Federal University of Pampa (UNIPAMPA)
Alegrete, Rio Grande do Sul, Brazil

Antonio Carlos Schneider Beck Filho
Institute of Informatics, Campus do Vale
Federal University of Rio Grande
do Sul (UFRGS)
Porto Alegre, Rio Grande do Sul, Brazil

ISSN 2191-5768 ISSN 2191-5776 (electronic)
SpringerBriefs in Computer Science
ISBN 978-3-030-28718-4 ISBN 978-3-030-28719-1 (eBook)
https://doi.org/10.1007/978-3-030-28719-1

This book is dedicated to the memory of Márcia Cristina and Aurora Cera.

Preface

Efficiently exploiting thread-level parallelism from modern multicore systems has been challenging for software developers. While blindly increasing the number of threads may lead to performance gains, it can also result in a disproportionate increase in energy consumption. In the same way, optimization techniques for reducing energy consumption, such as DVFS and power gating, can lead to huge performance loss if used incorrectly. In this book, we present and discuss several techniques that address these challenges. We start by providing a brief theoretical background on parallel computing in software and the sources of power consumption. Then, we show how different parallel programming interfaces and communication models may affect energy consumption in different ways. Next, we discuss tuning techniques to adapt the number of threads/operating frequency to achieve the best compromise between performance and energy. We finish this book with a detailed analysis of a representative example of an adaptive approach.

Alegrete, Brazil — Arthur Francisco Lorenzon
Porto Alegre, Brazil — Antonio Carlos Schneider Beck Filho

Acknowledgments

The authors would like to thank the friends and colleagues at Informatics Institute of the Federal University of Rio Grande do Sul and give a special thanks to all the people in the Embedded Systems Laboratory, who have contributed to this research since 2013.

The authors would also like to thank the Brazilian research support agencies, FAPERGS, CAPES, and CNPq.

Contents

Acronyms

CMOS	Complementary metal oxide semiconductor
DCT	Dynamic concurrency throttling
DSE	Design space exploration
DVFS	Dynamic voltage and frequency scaling
EDP	Energy-delay product
FFT	Fast fourier transform
FIFO	First-in first-out
FU	Function unit
GPP	General-purpose processors
HC	High communication
HPC	High-performance computing
ILP	Instruction-level parallelism
ISA	Instruction set architecture
LC	Low communication
MPI	Message passing interface
OpenMP	Open multi-programming
PAPI	Performance application programming interface
PPI	Parallel programming interface
PThreads	POSIX threads
RAM	Random access memory
SMT	Simultaneous multithreading
SoC	System-on-chip
TBB	Threading building blocks
TDP	Thermal design power
TLP	Thread-level parallelism