Numerical Computations with GPUs

Volodymyr Kindratenko
Editor

Numerical Computations with GPUs

Springer

Editor
Volodymyr Kindratenko
National Center for Supercomputing
Applications
University of Illinois
Urbana, IL, USA

ISBN 978-3-319-06547-2 ISBN 978-3-319-06548-9 (eBook)
DOI 10.1007/978-3-319-06548-9
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014940054

© Springer International Publishing Switzerland 2014
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book is intended to serve as a practical guide for the development and implementation of numerical algorithms on Graphics Processing Units (GPUs). The book assumes that the reader is familiar with the mathematical context and has a good working knowledge of GPU architecture and its programming sufficient to translate specialized mathematical algorithms and pseudo-codes presented in the book into a fully functional CUDA or OpenCL software. In case the reader is not familiar with the GPU programming, the reader is directed to other sources, such as NVIDIA's *CUDA Parallel Computing Platform* website, for low-level programming details, tools, and techniques prior to reading this book.

Book Focus

The main focus of this book is on the efficient implementation of numerical methods on GPUs. The book chapters are written by the leaders in the field working for many years on the development and implementation of computationally intensive numerical algorithms for solving scientific computing and engineering problems.

It is widely understood and accepted that modern scientific discovery in all of the disciplines requires extensive computations. It is also the case that modern engineering heavily utilizes advanced computational models and tools. At the heart of many such computations are libraries of mathematical codes for solving systems of linear equations, computing solutions of differential equations, finding integrals and function values, transforming time series, etc. These libraries have been developed over several decades and have been constantly updated to track the ever changing architecture and capabilities of computing hardware. With the introduction of GPUs, many of the existing numerical libraries are currently undergoing another phase of transformation in order to continue serving the computational science and engineering community by providing the required level of performance. Simultaneously, new numerical methods are under development to take advantage of the revolutionary architecture of GPUs. In either case, the developers of such

numerical codes face the challenge of extracting parallelism present in numerical methods and expressing it in the form that can be successfully utilized by the massively parallel GPU architecture. This frequently requires reformulating the original algorithmic structure of the code, tuning its performance, and developing and validating entirely new algorithms that can take advantage of the new hardware. It is my hope that this book will serve as a reference implementation and will provide the guidance for the developers of such codes by presenting a collective experience from many recent successful efforts.

Audience and Organization

This book targets practitioners working on the implementation of numerical codes on GPUs, researchers and software developers attempting to extend existing numerical libraries to GPUs, and readers interested in all aspects of GPU programming. It especially targets community of computational scientists from disciplines known to make use of linear algebra, differential equations, Monte Carlo methods, and Fourier transform.

The book is organized in four parts, each covering a particular set of numerical methods. First part is dedicated to the solution of linear algebra problems, ranging from the matrix–matrix multiplication, to the solution of systems of linear equations, to the computation of eigenvalues. Several chapters in this part address the problem of computing on a very large number of small matrixes. The final chapter also addresses the sparse matrix–vector product problem.

Second part is dedicated to the solution of differential equations and problems based on the space discretization of differential equations. Methods such as finite elements, finite difference, and successive over-relaxation with the applications to problem domains such as flow and wave propagation and solution of Maxwell's equations are presented. One chapter also addresses the challenge of integrating a large number of independent ordinary differential equations.

Third part is dedicated to the use of Monte Carlo methods for numerical integration. Monte Carlo techniques are well suited for GPU implementation and their use is widening. The part also includes chapters about random number generation on GPUs as a necessary first step in Monte Carlo methods.

The final part consists of two chapters dedicated to the efficient implementation of Fourier transform and one chapter discussing N-body simulations.

Acknowledgments

This book consists of contributed chapters provided by the experts in various fields involved with numerical computations on GPUs. I would like to thank all of the contributing authors whose work appears in this edition. I am also thankful to the Directors of the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign for the support and encouragement.

Urbana, IL, USA Volodymyr Kindratenko

Contents

Part I Linear Algebra

1 Accelerating Numerical Dense Linear Algebra Calculations with GPUs 3
Jack Dongarra, Mark Gates, Azzam Haidar, Jakub Kurzak, Piotr Luszczek, Stanimire Tomov, and Ichitaro Yamazaki

2 A Guide for Implementing Tridiagonal Solvers on GPUs 29
Li-Wen Chang and Wen-mei W. Hwu

3 Batch Matrix Exponentiation 45
M. Graham Lopez and Mitchel D. Horton

4 Efficient Batch LU and QR Decomposition on GPU 69
William J. Brouwer and Pierre-Yves Taunay

5 A Flexible CUDA LU-Based Solver for Small, Batched Linear Systems 87
Antonino Tumeo, Nitin Gawande, and Oreste Villa

6 Sparse Matrix-Vector Product 103
Zbigniew Koza, Maciej Matyka, Łukasz Mirosław, and Jakub Poła

Part II Differential Equations

7 Solving Ordinary Differential Equations on GPUs 125
Karsten Ahnert, Denis Demidov, and Mario Mulansky

8 GPU-Based Parallel Integration of Large Numbers of Independent ODE Systems 159
Kyle E. Niemeyer and Chih-Jen Sung

9 Finite and Spectral Element Methods on Unstructured Grids for Flow and Wave Propagation Problems 183
Dominik Göddeke, Dimitri Komatitsch, and Matthias Möller

10 A GPU Implementation for Solving the Convection Diffusion Equation Using the Local Modified SOR Method 207
Yiannis Cotronis, Elias Konstantinidis, and Nikolaos M. Missirlis

11 Finite-Difference in Time-Domain Scalable Implementations on CUDA and OpenCL 223
Lídia Kuan, Pedro Tomás, and Leonel Sousa

Part III Random Numbers and Monte Carlo Methods

12 Pseudorandom Numbers Generation for Monte Carlo Simulations on GPUs: OpenCL Approach 245
Vadim Demchik

13 Monte Carlo Automatic Integration with Dynamic Parallelism in CUDA 273
Elise de Doncker, John Kapenga, and Rida Assaf

14 GPU: Accelerated Computation Routines for Quantum Trajectories Method 299
Joanna Wiśniewska and Marek Sawerwain

15 Monte Carlo Simulation of Dynamic Systems on GPU's 319
Jonathan Rogers

Part IV Fast Fourier Transform and Localized n-Body Problems

16 Fast Fourier Transform (FFT) on GPUs 339
Yash Ukidave, Gunar Schirner, and David Kaeli

17 A Highly Efficient FFT Using Shared-Memory Multiplexing 363
Yi Yang and Huiyang Zhou

18 Increasing Parallelism and Reducing Thread Contentions in Mapping Localized N-Body Simulations to GPUs 379
Bharat Sukhwani and Martin C. Herbordt