Cuda C Programming Guide Nvidia

'The CUDA Handbook' begins where 'CUDA by Example' leaves off, discussing both CUDA hardware and software in detail that will engage any CUDA developer, from the casual to the most hardcore. Newer CUDA developers will see how the hardware processes commands and the driver checks progress; hardcore CUDA developers will appreciate topics such as the driver API, context migration, and how best to structure CPU/GPU data interchange and synchronization. The book is partly a reference resource and partly a cookbook.

This book constitutes the refereed proceedings of the 14th International Conference on Intelligent Data Engineering and Automated Learning, IDEAL 2013, held in Hefei, China, in October 2013. The 76 revised full papers presented were carefully reviewed and selected from more than 130 submissions. These papers provided a valuable collection of latest research outcomes in data engineering and automated learning, from methodologies, frameworks and techniques to applications. In addition to various topics such as evolutionary algorithms, neural networks, probabilistic modelling, swarm intelligent, multi-objective optimisation, and practical applications in regression, classification, clustering, biological data processing, text processing, video analysis, including a number of special sessions on emerging topics such as adaptation and learning multi-agent systems, big data, swarm intelligence and data mining, and combining learning and optimisation in intelligent data engineering.

More useful techniques, tips, and tricks for harnessing the power of the new generation of powerful GPUs.

This book constitutes the proceedings of the 7th International Workshop on Accelerator Programming Using Directives, WACCPD 2020, which took place on November 20, 2021. The workshop was initially planned to take place in Atlanta, GA, USA, and changed to an online format due to the COVID-19 pandemic. WACCPD is one of the major forums for bringing together users, developers, and the software and tools community to share knowledge and experiences when programming emerging complex parallel computing systems. The 5 papers presented in this volume were carefully reviewed and selected from 7 submissions. They were organized in topical sections named: OpenMP; OpenACC; and Domain-specific Solvers.

Hands-On GPU-Accelerated Computer Vision with OpenCV and CUDA

GPU Gems 2

A Hands-on Approach

CUDA Fortran for Scientists and Engineers

Parallel Programming for Modern High Performance Computing Systems

Concepts and Practice

OpenCL Programming Guide

Advances in GPU Research and Practice focuses on research and practices in GPU based systems. The topics treated cover a range of issues, ranging from hardware and architectural issues, to high level issues, such as application systems, parallel programming, middleware, and power and energy issues. Divided into six parts, this edited volume provides the latest research on GPU computing. Part I: Architectural Solutions focuses on the architectural topics that improve on performance of GPUs, Part II: System Software discusses OS, compilers, libraries, programming environment, languages, and paradigms that are proposed and analyzed to help and support GPU programmers. Part III: Power and Reliability Issues covers different aspects of energy, power, and reliability concerns in GPUs. Part IV: Performance Analysis illustrates mathematical and analytical techniques to predict different performance metrics in GPUs. Part V: Algorithms presents how to design efficient algorithms and analyze their complexity for GPUs. Part VI: Applications and Related Topics provides use cases and examples of how GPUs are used across many sectors. Discusses how to maximize power and obtain peak reliability when designing, building, and using GPUs Covers system software (OS, compilers), programming environments, languages, and paradigms proposed to help and support GPU programmers Explains how to use mathematical and analytical techniques to predict different performance metrics in GPUs Illustrates the design of efficient GPU algorithms in areas such as bioinformatics, complex systems, social networks, and cryptography Provides applications and use case scenarios in several different verticals, including medicine, social sciences, image processing, and telecommunications

With breadth and depth of coverage, the Encyclopedia of Computer Science and Technology, Second Edition has a multi-disciplinary scope, drawing together comprehensive coverage of the inter-related aspects of computer science and technology. The topics covered in this encyclopedia include: General and reference Hardware Computer systems organization Networks Software and its engineering Theory of computation Mathematics of computing Information systems Security and privacy Human-centered computing Computing methodologies Applied computing Professional issues Leading figures in the history of computer science The encyclopedia is structured according to the ACM Computing Classification System (CCS), first published in 1988 but subsequently revised in 2012. This classification system is the most comprehensive and is considered the de facto ontological framework for the computing field. The encyclopedia brings together the information and historical context that students, practicing professionals, researchers, and academicians need to have a strong and solid foundation in all aspects of computer science and technology.

High-performance computing (HPC) describes the use of connected computing units to perform complex tasks. It relies on parallelization techniques and algorithms to synchronize these disparate units in order to perform faster than a single processor could, alone. Used in industries from medicine and research to military and higher education, this method of computing allows for users to complete complex data-intensive tasks. This field has undergone many changes over the past decade, and will continue to grow in popularity in the coming years. Innovative Research Applications in Next-Generation High Performance Computing aims to address the future challenges, advances, and applications of HPC and related technologies. As the need for such processors increases, so does the importance of developing new ways to optimize the performance of these supercomputers. This timely publication provides comprehensive information for researchers, students in ICT, program developers, military and government organizations, and business professionals.

Parallel Programming: Concepts and Practice provides an upper level introduction to parallel programming. In addition to covering general parallelism concepts, this text teaches practical programming skills for both shared memory and distributed memory architectures. The authors' open-source system for automated code evaluation provides easy access to parallel computing resources, making the book particularly suitable for classroom settings. Covers parallel programming approaches for single computer nodes and HPC clusters: OpenMP, multithreading, SIMD vectorization, MPI, UPC++ Contains numerous practical parallel programming exercises Includes access to an automated code evaluation tool that enables students the opportunity to program in a web browser and receive immediate feedback on the result validity of their program Features an example-based teaching of concept to enhance learning outcomes

Ray Tracing from the Ground Up

8th International Conference, MACIS 2019, Gebze, Turkey, November 13–15, 2019, Revised Selected Papers

GPU Programming in MATLAB

GPU Parallel Program Development Using CUDA

Programming Massively Parallel Processors

Parallel Programming

Taking you to the limit in Concurrency, OOP, and the most advanced capabilities of C

GPU Parallel Program Development using CUDA teaches GPU programming by showing the differences among different families of GPUs. This approach prepares the reader for the next generation and future generations of GPUs. The book emphasizes concepts that will remain relevant for a long time, rather than concepts that are platform-specific. At the same time, the book also provides platform-dependent explanations that are as valuable as generalized GPU concepts. The book consists of three separate parts; it starts by explaining parallelism using CPU multi-threading in Part I. A few simple programs are used to demonstrate the concept of dividing a large task into multiple parallel sub-tasks and mapping them to CPU threads. Multiple ways of parallelizing the same task are analyzed and their pros/cons are studied in terms of both core and memory operation. Part II of the book introduces GPU massive parallelism. The same programs are parallelized on multiple Nvidia GPU platforms and the same performance analysis is repeated. Because the core and memory structures of CPUs and GPUs are different, the results differ in interesting ways. The end goal is to make programmers aware of all the good ideas, as well as the bad ideas, so readers can apply the good ideas and avoid the bad ideas in their own programs. Part III of the book provides pointer for readers who want to expand their horizons. It provides a brief introduction to popular CUDA libraries (such as cuBLAS, cuFFT, NPP, and Thrust),the OpenCL programming language, an overview of GPU programming using other programming languages and API libraries (such as Python, OpenCV, OpenGL, and Apple's Swift and Metal,) and the deep learning library cuDNN.

Build real-world applications with Python 2.7, CUDA 9, and CUDA 10. We suggest the use of Python 2.7 over Python 3.x, since Python 2.7 has stable support across all the libraries we use in this book. Key Features Expand your background in GPU programming—PyCUDA, scikit-cuda, and Nsight Effectively use CUDA libraries such as cuBLAS, cuFFT, and cuSolver Apply GPU programming to modern data science applications Book Description Hands-On GPU Programming with Python and CUDA hits the ground running: you'll start by learning how to apply Amdahl's Law, use a code profiler to identify bottlenecks in your Python code, and set up an appropriate GPU programming environment. You'll then see how to "query" the GPU's features and copy arrays of data to and from the GPU's own memory. As you make your way through the book, you'll launch code directly onto the GPU and write full blown GPU kernels and device functions in CUDA C. You'll get to grips with profiling GPU code effectively and fully test and debug your code using Nsight IDE. Next, you'll explore some of the more well-known NVIDIA libraries, such as cuFFT and cuBLAS. With a solid background in place, you will now apply your new-found knowledge to develop your very own GPU-based deep neural network from scratch. You'll then explore advanced topics, such as warp shuffling, dynamic parallelism, and PTX assembly. In the final chapter, you'll see some topics and applications related to GPU programming that you may wish to pursue, including AI, graphics, and blockchain. By the end of this book, you will be able to apply GPU programming to problems related to data science and high-performance computing. What you will learn Launch GPU code directly from Python Write effective and efficient GPU kernels and device functions Use libraries such as cuFFT, cuBLAS, and cuSolver Debug and profile your code with Nsight and Visual Profiler Apply GPU programming to datascience problems Build a GPU-based deep neuralnetwork from scratch Explore advanced GPU hardware features, such as warp shuffling Who this book is for Hands-On GPU Programming with Python and CUDA is for developers and data scientists who want to learn the basics of effective GPU programming to improve performance using Python code. You should have an understanding of first-year college or university-level engineering mathematics and physics, and have some experience with Python as well as in any C-based programming language such as C, C++, Go, or Java.

If you need to learn CUDA but don't have experience with parallel computing, CUDA Programming: A Developer's Introduction offers a detailed guide to CUDA with a grounding in parallel fundamentals. It starts by introducing CUDA and bringing you up to speed on GPU parallelism and hardware, then delving into CUDA installation. Chapters on core concepts including threads, blocks, grids, and memory focus on both parallel and CUDA-specific issues. Later, the book demonstrates CUDA in practice for optimizing applications, adjusting to new hardware, and solving common problems. Comprehensive introduction to parallel programming with CUDA, for readers new to both Detailed instructions help readers optimize the CUDA software development kit Practical techniques illustrate working with memory, threads, algorithms, resources, and more Covers CUDA on multiple hardware platforms: Mac, Linux and Windows with several NVIDIA chipsets Each chapter includes exercises to test reader knowledge

This two volume set LNCS 9049 and LNCS 9050 constitutes the refereed proceedings of the 20th International Conference on Database Systems for Advanced Applications, DASFAA 2015, held in Hanoi, Vietnam, in April 2015. The 63 full papers presented were carefully reviewed and selected from a total of 287 submissions. The papers cover the following topics: data mining; data streams and time series; database storage and index; spatio-temporal data; modern computing platform; social networks; information integration and data quality; information retrieval and summarization; security and privacy; outlier and imbalanced data analysis; probabilistic and uncertain data; query processing.

Advances in Parallel & Distributed Processing, and Applications

Hands-On GPU Programming with Python and CUDA

Accelerator Programming Using Directives

Mathematical Aspects of Computer and Information Sciences

Programming Techniques for High-performance Graphics and General-purpose Computation

Extreme C

Multicore and GPU Programming

This book constitutes the refereed proceedings of the 8th International Conference on Mathematical Aspects of Computer and Information Sciences, MACIS 2019, held in Gebze, Turkey, in November 2019. The 22 revised papers and 14 short papers presented were carefully reviewed and selected from 66 submissions. The papers are organized in the following topical sections: algorithms and foundation; security and cryptography; combinatorics, codes, designs and graphs; data modeling and machine learning; tools and software track.

Not only does almost everyone in the civilized world use a personal computer, smartphone, and/or tablet on a daily basis to communicate with others and access information, but virtually every other modern appliance, vehicle, or other device has one or more computers embedded inside it. One cannot purchase a current-model automobile, for example, without several computers on board to do everything from monitoring exhaust emissions, to operating the anti-lock brakes, to telling the transmission when to shift, and so on. Appliances such as clothes washers and dryers, microwave ovens, refrigerators, etc. are almost all digitally controlled. Gaming consoles like Xbox, PlayStation, and Wii are powerful computer systems with enhanced capabilities for user interaction. Computers are everywhere, even when we don't see them as such, and it is more important than ever for students who will soon enter the workforce to understand how they work. This book is completely updated and revised for a one-semester upper level undergraduate course in Computer Architecture, and suitable for use in an undergraduate CS, EE, or CE curriculum at the junior or senior level. Students should have had a course(s) covering introductory topics in digital logic and computer organization. While this is not a text for a programming course, the reader should be familiar with computer programming concepts in at least one language such as C, C++, or Java. Previous courses in operating systems, assembly language, and/or systems programming would be helpful, but are not essential.

China Satellite Navigation Conference (CSNC) 2014 Proceedings presents selected research papers from CSNC2014, held on 21-23 May in Nanjing, China. The theme of CSNC2014 is 'BDS Application: Innovation, Integration and Sharing'. These papers discuss the technologies and applications of the Global Navigation Satellite System (GNSS) and the latest progress made in the China BeiDou System (BDS) especially. They are divided into 9 topics to match the corresponding sessions in CSNC2014, which broadly covered key topics in GNSS. Readers can learn about the BDS and keep abreast of the latest advances in GNSS techniques and applications. SUN Jiadong is the Chief Designer of the Compass/ BDS, and the Academician of Chinese Academy of Sciences (CAS); JIAO Wenhai is a researcher at China Satellite Navigation Office; WU Haitao is a professor at Navigation Headquarters, CAS; LU Mingquan is a professor at Department of Electronic Engineering of Tsinghua University.

Programming Massively Parallel Processors: A Hands-on Approach, Second Edition, teaches students how to program massively parallel processors. It offers a detailed discussion of various techniques for constructing parallel programs. Case studies are used to demonstrate the development process, which begins with computational thinking and ends with effective and efficient parallel programs. This guide shows both student and professional alike the basic concepts of parallel programming and GPU architecture. Topics of performance, floating-point format, parallel patterns, and dynamic parallelism are covered in depth. This revised edition contains more parallel programming examples, commonly-used libraries such as Thrust, and explanations of the latest tools. It also provides new coverage of CUDA 5.0, improved performance, enhanced development tools, increased hardware support, and more; increased coverage of related technology, OpenCL and new material on algorithm patterns, GPU clusters, host programming, and data parallelism; and two new case studies (on MRI reconstruction and molecular visualization) that explore the latest applications of CUDA and GPUs for scientific research and high-performance computing. This book should be a valuable resource for advanced students, software engineers, programmers, and hardware engineers. New coverage of CUDA 5.0, improved performance, enhanced development tools, increased hardware support, and more Increased coverage of related technology, OpenCL and new material on algorithm patterns, GPU clusters, host programming, and data parallelism Two new case studies (on MRI reconstruction and molecular visualization) explore the latest applications of CUDA and GPUs for scientific research and high-performance computing

Proceedings from PDPTA'20, CSC'20, MSV'20, and GCC'20

CUDA for Engineers

China Satellite Navigation Conference (CSNC) 2016 Proceedings: Volume III

Introduction to High Performance Scientific Computing

CUDA Programming

Parallel Processing and Applied Mathematics, Part II

GPU-based Parallel Implementation of Swarm Intelligence Algorithms

Multicore and GPU Programming offers broad coverage of the key parallel computing skillsets: multicore CPU programming and manycore "massively parallel" computing. Using threads, OpenMP, MPI, and CUDA, it teaches the design and development of software capable of taking advantage of today's computing platforms incorporating CPU and GPU hardware and explains how to transition from sequential programming to a parallel computing paradigm. Presenting material refined over more than a decade of teaching parallel computing, author Gerassimos Barlas minimizes the challenge with multiple examples, extensive case studies, and full source code. Using this book, you can develop programs that run over distributed memory machines using MPI, create multi-threaded applications with either libraries or directives, write optimized applications that balance the workload between available computing resources, and profile and debug programs targeting multicore machines. Comprehensive coverage of all major multicore programming tools, including threads, OpenMP, MPI, and CUDA Demonstrates parallel programming design patterns and examples of how different tools and paradigms can be integrated for superior performance Particular focus on the emerging area of divisible load theory and its impact on load balancing and distributed systems Download source code, examples, and instructor support materials on the book's companion website

This book constitutes the proceedings of the Kiel Declarative Programming Days, KDPD 2013, unifying the following conferences: the 20th International Conference on Applications of Declarative Programming and Knowledge Management (INAP 2013), the 22nd International Workshop on Functional and (Constraint) Logic Programming (WFLP 2013) and the 27th Workshop on Logic Programming (WLP 2013), held in Kiel, Germany, in September 2013. The 15 papers presented were carefully reviewed and selected for inclusion in this book. They cover the following topics: logic programming, constraint problem solving, programmable logic solvers, functional programming and constraint programming.

Push the limits of what C - and you - can do, with this high-intensity guide to the most advanced capabilities of C Key FeaturesMake the most of C's low-level control, flexibility, and high performanceA comprehensive guide to C's most powerful and challenging featuresA thought-provoking guide packed with hands-on exercises and examplesBook Description There's a lot more to C than knowing the language syntax. The industry looks for developers with a rigorous, scientific understanding of the principles and practices. Extreme C will teach you to use C's advanced low-level power to write effective, efficient systems. This intensive, practical guide will help you become an expert C programmer. Building on your existing C knowledge, you will master preprocessor directives, macros, conditional compilation, pointers, and much more. You will gain new insight into algorithm design, functions, and structures. You will discover how C helps you squeeze maximum performance out of critical, resource-constrained applications. C still plays a critical role in 21st-century programming, remaining the core language for precision engineering, aviations, space research, and more. This book shows how C works with Unix, how to implement OO principles in C, and fully covers multi-processing. In Extreme C, Amini encourages you to think, question, apply, and experiment for yourself. The book is essential for anybody who wants to take their C to the next level. What you will learnBuild advanced C knowledge on strong foundations, rooted in first principlesUnderstand memory

structures and compilation pipeline and how they work, and how to make most out of themApply object-oriented design principles to your procedural C codeWrite low-level code that's close to the hardware and squeezes maximum performance out of a computer systemMaster concurrency, multithreading, multi-processing, and integration with other languagesUnit Testing and debugging, build systems, and inter-process communication for C programmingWho this book is for Extreme C is for C programmers who want to dig deep into the language and its capabilities. It will help you make the most of the low-level control C gives you.

CUDA for Engineers gives you direct, hands-on engagement with personal, high-performance parallel computing, enabling you to do computations on a gaming-level PC that would have required a supercomputer just a few years ago. The authors introduce the essentials of CUDA C programming clearly and concisely, quickly guiding you from running sample programs to building your own code. Throughout, you'll learn from complete examples you can build, run, and modify, complemented by additional projects that deepen your understanding. All projects are fully developed, with detailed building instructions for all major platforms. Ideal for any scientist, engineer, or student with at least introductory programming experience, this guide assumes no specialized background in GPU-based or parallel computing. In an appendix, the authors also present a refresher on C programming for those who need it. Coverage includes Preparing your computer to run CUDA programs Understanding CUDA's parallelism model and C extensions Transferring data between CPU and GPU Managing timing, profiling, error handling, and debugging Creating 2D grids Interoperating with OpenGL to provide real-time user interactivity Performing basic simulations with differential equations Using stencils to manage related computations across threads Exploiting CUDA's shared memory capability to enhance performance Interacting with 3D data: slicing, volume rendering, and ray casting Using CUDA libraries Finding more CUDA resources and code Realistic example applications include Visualizing functions in 2D and 3D Solving differential equations while changing initial or boundary conditions Viewing/processing images or image stacks Computing inner products and centroids Solving systems of linear algebraic equations Monte-Carlo computations

Encyclopedia of Computer Science and Technology
The Definitive Guide to Programmable Real-time Graphics
20th International Conference, DASFAA 2015, Hanoi, Vietnam, April 20-23, 2015, Proceedings, Part I
Declarative Programming Days, KDPD 2013, Unifying INAP, WFLP, and WLP, Kiel, Germany, September 11-13, 2013, Revised Selected Papers
Explore high-performance parallel computing with CUDA
An Integrated Approach
11th International Conference, PCT 2017, Kazan, Russia, April 3-7, 2017, Revised Selected Papers

The book presents the proceedings of four conferences: The 26th International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'20), The 18th International Conference on Scientific Computing (CSC'20); The 17th International Conference on Modeling, Simulation and Visualization Methods (MSV'20); and The 16th International Conference on Grid, Cloud, and Cluster Computing (GCC'20). The conferences took place in Las Vegas, NV, USA, July 27-30, 2020. The conferences are part of the larger 2020 World Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE'20), which features 20 major tracks. Authors include academics, researchers, professionals, and students. Presents the proceedings of four conferences as part of the 2020 World Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE'20); Includes the research tracks Parallel and Distributed Processing, Scientific Computing, Modeling, Simulation and Visualization, and Grid, Cloud, and Cluster Computing; Features papers from PDPTA'20, CSC'20, MSV'20, and GCC'20.

These Proceedings present selected research papers from CSNC2016, held during 18th-20th May in Changsha, China. The theme of CSNC2016 is Smart Sensing, Smart Perception. These papers discuss the technologies and applications of the Global Navigation Satellite System (GNSS), and the latest progress made in the China BeiDou System (BDS) especially. They are divided into 12 topics to match the corresponding sessions in CSNC2016, which broadly covered key topics in GNSS. Readers can learn about the BDS and keep abreast of the latest advances in GNSS techniques and applications.

The book then details the thought behind CUDA and teaches how to create, analyze, and debug CUDA applications. Throughout, the focus is on software engineering issues: how to use CUDA in the context of existing application code, with existing compilers, languages, software tools, and industry-standard API libraries."--Pub. desc.

In view of the growing presence and popularity of multicore and manycore processors, accelerators, and coprocessors, as well as clusters using such computing devices, the development of efficient parallel applications has become a key challenge to be able to exploit the performance of such systems. This book covers the scope of parallel programming for modern high performance computing systems. It first discusses selected and popular state-of-the-art computing devices and systems available today, These include multicore CPUs, manycore (co)processors, such as Intel Xeon Phi, accelerators, such as GPUs, and clusters, as well as programming models supported on these platforms. It next introduces parallelization through important programming paradigms, such as master-slave, geometric Single Program Multiple Data (SPMD) and divide-and-conquer. The practical and useful elements of the most popular and important APIs for programming parallel HPC systems are discussed, including MPI, OpenMP, Pthreads, CUDA, OpenCL, and OpenACC. It also demonstrates, through selected code listings, how selected APIs can be used to implement important programming paradigms. Furthermore, it shows how the codes can be compiled and executed in a Linux environment. The book also presents hybrid codes that integrate selected APIs for potentially multi-level parallelization and utilization of heterogeneous resources, and it shows how to use modern elements of these APIs. Selected optimization techniques are also included, such as overlapping communication and computations implemented using various APIs. Features: Discusses the popular and currently available computing devices and cluster systems Includes typical paradigms used in parallel programs Explores popular APIs for programming parallel applications Provides code templates that can be used for implementation of paradigms Provides hybrid code examples allowing multi-level parallelization Covers the optimization of parallel programs

9th International Conference, PPAM 2011, Torun, Poland, September 11-14, 2011. Revised Selected Papers, Part II
Advances in GPU Research and Practice
The CUDA Handbook
Fundamentals and Principles of Computer Design, Second Edition
A Comprehensive Guide to GPU Programming
7th International Workshop, WACCPD 2020, Virtual Event, November 20, 2020, Proceedings

CUDA is a computing architecture designed to facilitate the development of parallel programs. In conjunction with a comprehensive software platform, the CUDA Architecture enables programmers to draw on the immense power of graphics processing units (GPUs) when building high-performance applications. GPUs, of course, have long been available for demanding graphics and game applications. CUDA now brings this valuable resource to programmers working on applications in other domains, including science, engineering, and finance. No knowledge of graphics programming is required—just the ability to program in a modestly extended version of C. CUDA by Example, written by two senior members of the CUDA software platform team, shows programmers how to employ this new technology. The authors introduce each area of CUDA development through working examples. After a concise introduction to the CUDA platform and architecture, as well as a quick-start guide to CUDA C, the book details the techniques and trade-offs associated with each key CUDA feature. You'll discover when to use each CUDA C extension and how to write CUDA software that delivers truly outstanding performance. Major topics covered include Parallel programming Thread cooperation Constant memory and events Texture memory Graphics interoperability Atomics Streams CUDA C on multiple GPUs Advanced atomics Additional CUDA resources All the CUDA software tools you'll need are freely available for download from NVIDIA. http://developer.nvidia.com/object/cuda-by-example.html

Break into the powerful world of parallel GPU programming with this down-to-earth, practical guide Designed for professionals across multiple industrial sectors, Professional CUDA C Programming presents CUDA -- a parallel computing platform and programming model designed to ease the development of GPU programming -- fundamentals in an easy-to-follow format, and teaches how to think in parallel and implement parallel algorithms on GPUs. Each chapter covers a specific topic, and includes workable examples that demonstrate the development process, allowing readers to explore both the "hard" and "soft" aspects of GPU programming. Computing architectures are experiencing a fundamental shift toward scalable parallel computing motivated by application requirements in industry and science. This book demonstrates the challenges of efficiently utilizing compute resources at peak performance, presents modern techniques for tackling these challenges, while increasing accessibility for professionals who are not necessarily parallel programming experts. The CUDA programming model and tools empower developers to write high-performance applications on a scalable, parallel computing platform: the GPU. However, CUDA itself can be difficult to learn without extensive programming experience. Recognized CUDA authorities John Cheng, Max Grossman, and Ty McKercher guide readers through essential GPU programming skills and best practices in Professional CUDA C Programming, including: CUDA Programming Model GPU Execution Model GPU Memory model Streams, Event and Concurrency Multi-GPU Programming CUDA Domain-Specific Libraries Profiling and Performance Tuning The book makes complex CUDA concepts easy to understand for anyone with knowledge of basic software development with exercises designed to be both readable and high-performance. For the professional seeking entrance to parallel computing and the high-performance computing community, Professional CUDA C Programming is an invaluable resource, with the most current information available on the market.

Cg is a complete programming environment for the fast creation of special effects and real-time cinematic quality experiences on multiple platforms. This text provides a guide to the Cg graphics language.

GPU programming in MATLAB is intended for scientists, engineers, or students who develop or maintain applications in MATLAB and would like to accelerate their codes using GPU programming without losing the many benefits of MATLAB. The book starts with coverage of the Parallel Computing Toolbox and other MATLAB toolboxes for GPU computing, which allow applications to be ported straightforwardly onto GPUs without extensive knowledge of GPU programming. The next part covers built-in, GPU-enabled features of MATLAB, including options to leverage GPUs across multicore or different computer systems. Finally, advanced material includes CUDA code in MATLAB and optimizing existing GPU applications. Throughout the book, examples and source codes illustrate every concept so that readers can immediately apply them to their own development. Provides in-depth, comprehensive coverage of GPUs with MATLAB, including the parallel computing toolbox and built-in features for other MATLAB toolboxes Explains how to accelerate computationally heavy applications in MATLAB without the need to re-write them in another language Presents case studies illustrating key concepts across multiple fields Includes source code, sample datasets, and lecture slides

The Cg Tutorial
Intelligent Data Engineering and Automated Learning -- IDEAL 2013
Declarative Programming and Knowledge Management
Parallel Dynamic and Transient Simulation of Large-Scale Power Systems
Professional CUDA C Programming
An Introduction to High-Performance Parallel Computing
Innovative Research and Applications in Next-Generation High Performance Computing

CUDA Fortran for Scientists and Engineers shows how high-performance application developers can leverage the power of GPUs using Fortran, the familiar language of scientific computing and supercomputer performance benchmarking. The authors presume no prior parallel computing experience, and cover the basics along with best practices for efficient GPU computing using CUDA Fortran. To help you add CUDA Fortran to existing Fortran codes, the book explains how to understand the target GPU architecture, identify computationally intensive parts of the code, and modify the code to manage the data and parallelism and optimize performance. All of this is done in Fortran, without having to rewrite in another language. Each concept is illustrated with actual examples so you can immediately evaluate the performance of your code in comparison. Leverage the power of GPU computing with PGI's CUDA Fortran compiler Gain insights from members of the CUDA Fortran language development team Includes multi-GPU programming in CUDA Fortran, covering both peer-to-peer and message passing interface (MPI) approaches Includes full source code for all the examples and several case studies Download source code and slides from the book's companion website

Based on a course developed by the author, Introduction to High Performance Scientific Computing introduces methods for adding parallelism to numerical methods for solving differential equations. It contains exercises and programming projects that facilitate learning as well as examples and discussions based on the C programming language, with additional comments for those already familiar with C++. The text provides an overview of concepts and algorithmic techniques for modern scientific computing, and is divided into six self-contained parts that can be assembled in any order to create an introductory course using available computer hardware. Part I introduces the C programming language for those not already familiar with programming in a compiled language. Part II describes parallelism on shared memory architectures using OpenMP. Part III details parallelism on computer clusters using MPI for coordinating a computation. Part IV demonstrates the use of graphical programming units (GPUs) to solve problems using the CUDA language for NVIDIA graphics cards. Part V addresses programming on GPUs for non-NVIDIA graphics cards using the OpenCL framework. Finally, Part VI contains a brief discussion of numerical methods and applications, giving the reader an opportunity to test the methods on typical computing problems.

The CUDA Handbook begins where CUDA by Example (Addison-Wesley, 2011) leaves off, discussing CUDA hardware and software in greater detail and covering both CUDA 5.0 and Kepler. Every CUDA developer, from the casual to the most sophisticated, will find something here of interest and immediate usefulness. Newer CUDA developers will see how the hardware processes commands and how the driver checks progress; more experienced CUDA developers will appreciate the expert coverage of topics such as the driver API and context migration, as well as the guidance on how best to structure CPU/GPU data interchange and synchronization. The accompanying open source code–more than 25,000 lines of it, freely available at www.cudahandbook.com–is specifically intended to be reused and repurposed by developers. Designed to be both a comprehensive reference and a practical cookbook, the text is divided into the following three parts: Part I, Overview, gives high-level descriptions of the hardware and software that make CUDA possible. Part II, Details, provides thorough descriptions of every aspect of CUDA, including Memory Streams and events Models of execution, including the dynamic parallelism feature, new with CUDA 5.0 and SM 3.5 The streaming multiprocessors, including descriptions of all features through SM 3.5 Programming multiple GPUs Texturing The source code accompanying Part II is presented as reusable microbenchmarks and microdemos, designed to expose specific hardware characteristics or highlight specific use cases. Part III, Select Applications, details specific families of CUDA applications and key parallel algorithms, including Streaming workloads Reduction Parallel prefix sum (Scan) N-body Image Processing These algorithms cover the full range of potential CUDA applications.

GPU-based Parallel Implementation of Swarm Intelligence Algorithms combines and covers two emerging areas attracting increased attention and applications: graphics processing units (GPUs) for general-purpose computing (GPGPU) and swarm intelligence. This book not only presents GPGPU in adequate detail, but also includes guidance on the appropriate implementation of swarm intelligence algorithms on the GPU platform. GPU-based implementations of several typical swarm intelligence algorithms such as PSO, FWA, GA, DE, and ACO are presented and having described the implementation details including parallel models, implementation considerations as well as performance metrics are discussed. Finally, several typical applications of GPU-based swarm intelligence algorithms are presented. This valuable reference book provides a unique perspective not possible by studying either GPGPU or swarm intelligence alone. This book gives a complete and whole picture for interested readers and new comers who will find many implementation algorithms in the book suitable for immediate use in their projects. Additionally, some algorithms can also be used as a starting point for further research. Presents a concise but sufficient introduction to general-purpose GPU computing which can help the layman become familiar with this emerging computing technique Describes implementation details, such as parallel models and performance metrics, so readers can easily utilize the techniques to accelerate their algorithmic programs Appeals to readers from the domain of high performance computing (HPC) who will find the relatively young research domain of swarm intelligence very interesting Includes many real-world applications, which can be of great help in deciding whether or not swarm intelligence algorithms or GPGPU is appropriate for the task at hand

A Developer's Guide to Parallel Computing with GPUs
Effective techniques for processing complex image data in real time using GPUs
Computer Architecture
China Satellite Navigation Conference (CSNC) 2014 Proceedings: Volume I
A High Performance Computing Solution
Database Systems for Advanced Applications
Hands-On GPU Programming with CUDA

This textbook introduces methods of accelerating transient stability (dynamic) simulation and electromagnetic transient simulation on massively parallel processors for large-scale AC-DC grids – two of the most common and computationally onerous studies done by energy control centers and research laboratories for the planning, design, and operation of such integrated grids for ensuring the security and reliability of electric power. Simulation case studies provided in the book range from small didactic test circuits to realistic-sized AC-DC grids, and special emphasis is placed on detailed device-level multi-physics models for power system equipment and decomposition techniques for simulating large-scale systems. Parallel Dynamic and Transient Simulation of Large-Scale Power Systems: A High Performance Computing Solution is a comprehensive state-of-the-art guide for upper-level undergraduate and graduate students in power systems engineering. Practicing engineers, software developers, and scientists working in the power and energy industry will find it to be a timely and valuable reference for solving potential problems in their design and development activities. Detailed device-level electro-thermal modeling for power electronic systems in DC grids; Provides comprehensive dynamic and transient simulation of integrated large-scale AC-DC grids; Offers detailed models of renewable energy system models.

This two-volume-set (LNCS 7203 and 7204) constitutes the refereed proceedings of the 9th International Conference on Parallel Processing and Applied Mathematics, PPAM 2011, held in Torun, Poland, in September 2011. The 130 revised full papers presented in both volumes were carefully reviewed and selected from numerous submissions. The papers address issues such as parallel/distributed architectures and mobile computing; numerical algorithms and parallel numerics; parallel non-numerical algorithms; tools and environments for parallel/distributed/grid computing; applications of parallel/distributed computing; applied mathematics, neural networks and evolutionary computing; history of computing.

Professional CUDA C ProgrammingJohn Wiley & Sons
With the increase in computing speed and due to the high quality of the optical effects it achieves, ray tracing is becoming a popular choice for interactive and animated rendering. This book takes readers through the whole process of building a modern ray tracer from scratch in C++. All concepts and processes are explained in detail, with the aid o
Best Practices for Efficient CUDA Fortran Programming
An Introduction to General-Purpose GPU Programming, Portable Documents
CUDA Application Design and Development
Parallel Computational Technologies
14th International Conference, IDEAL 2013, Hefei, China, October 20-23, 2013, Proceedings
CUDA by Example

This book constitutes the refereed proceedings of the 11th International Conference on Parallel Computational Technologies, PCT 2017, held in Kazan, Russia, in April 2017. The 24 revised full papers presented were carefully reviewed and selected from 167 submissions. The papers are organized in topical sections on high performance architectures, tools and technologies; parallel numerical algorithms; supercomputer simulation.

Explore different GPU programming methods using libraries and directives, such as OpenACC, with extension to languages such as C, C++, and Python Key Features Learn parallel programming principles and practices and performance analysis in GPU computing Get to grips with distributed multi GPU programming and other approaches to GPU programming Understand how GPU acceleration in deep learning models can improve their performance Book Description Compute Unified Device Architecture (CUDA) is NVIDIA's GPU computing platform and application programming interface. It's designed to work with programming languages such as C, C++, and Python. With CUDA, you can leverage a GPU's parallel computing power for a range of high-performance computing applications in the fields of science, healthcare, and deep learning. Learn CUDA Programming will help you learn GPU parallel programming and understand its modern applications. In this book, you'll discover CUDA programming approaches for modern GPU architectures. You'll not only be guided through GPU features, tools, and APIs, you'll also learn how to analyze performance with sample parallel programming algorithms. This book will help you optimize the performance of your apps by giving insights into CUDA programming platforms with various libraries, compiler directives (OpenACC), and other languages. As you progress, you'll learn how additional computing power can be generated using multiple GPUs in a box or in multiple boxes. Finally, you'll explore how CUDA accelerates deep learning algorithms, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs). By the end of this CUDA book, you'll be equipped with the skills you need to integrate the power of GPU computing in your applications. What you will learn Understand general GPU operations and programming patterns in CUDA Uncover the difference between GPU programming and CPU programming Analyze GPU application performance and implement optimization strategies Explore GPU programming, profiling, and debugging tools Grasp parallel programming algorithms and how to implement them Scale GPU-accelerated applications with multi-GPU and multi-nodes Delve into GPU programming platforms with accelerated libraries, Python, and OpenACC Gain insights into deep learning accelerators in CNNs and RNNs using GPUs Who this book is for This beginner-level book is for programmers who want to delve into parallel computing, become part of the high-performance computing community and build modern applications. Basic C and C++ programming experience is assumed. For deep learning enthusiasts, this book covers Python InterOps, DL libraries, and practical examples on performance estimation.

Using the new OpenCL (Open Computing Language) standard, you can write applications that access all available programming resources: CPUs, GPUs, and other processors such as DSPs and the Cell/B.E. processor. Already implemented by Apple, AMD, Intel, IBM, NVIDIA, and other leaders, OpenCL has outstanding potential for PCs, servers, handheld/embedded devices, high performance computing, and even cloud systems. This is the first comprehensive, authoritative, and practical guide to OpenCL 1.1 specifically for working developers and software architects. Written by five leading OpenCL authorities, OpenCL Programming Guide covers the entire specification. It reviews key use cases, shows how OpenCL can express a wide range of parallel algorithms, and offers complete reference material on both the API and OpenCL C programming language. Through complete case studies and downloadable code examples, the authors show how to write complex parallel programs that decompose workloads across many different devices. They also present all the essentials of OpenCL software performance optimization, including probing and adapting to hardware. Coverage includes Understanding OpenCL's architecture, concepts, terminology, goals, and rationale Programming with OpenCL C and the runtime API Using buffers, sub-buffers, images, samplers, and events Sharing and synchronizing data with OpenGL and Microsoft's Direct3D Simplifying development with the C++ Wrapper API Using OpenCL Embedded Profiles to support devices ranging from cellphones to supercomputer nodes Case studies dealing with physics simulation; image and signal processing, such as image histograms, edge detection filters, Fast Fourier Transforms, and optical flow; math libraries, such as matrix multiplication and high-performance sparse matrix multiplication; and more Source code for this book is available at https://code.google.com/p/opencl-book-samples/

Discover how CUDA allows OpenCV to handle complex and rapidly growing image data processing in computer and machine vision by accessing the power of GPU Key FeaturesExplore examples to leverage the GPU processing power with OpenCV and CUDAEnhance the

performance of algorithms on embedded hardware platformsDiscover C++ and Python libraries for GPU accelerationBook Description Computer vision has been revolutionizing a wide range of industries, and OpenCV is the most widely chosen tool for computer vision with its ability to work in multiple programming languages. Nowadays, in computer vision, there is a need to process large images in real time, which is difficult to handle for OpenCV on its own. This is where CUDA comes into the picture, allowing OpenCV to leverage powerful NVDIA GPUs. This book provides a detailed overview of integrating OpenCV with CUDA for practical applications. To start with, you'll understand GPU programming with CUDA, an essential aspect for computer vision developers who have never worked with GPUs. You'll then move on to exploring OpenCV acceleration with GPUs and CUDA by walking through some practical examples. Once you have got to grips with the core concepts, you'll familiarize yourself with deploying OpenCV applications on NVIDIA Jetson TX1, which is popular for computer vision and deep learning applications. The last chapters of the book explain PyCUDA, a Python library that leverages the power of CUDA and GPUs for accelerations and can be used by computer vision developers who use OpenCV with Python. By the end of this book, you'll have enhanced computer vision applications with the help of this book's hands-on approach. What you will learnUnderstand how to access GPU device properties and capabilities from CUDA programsLearn how to accelerate searching and sorting algorithmsDetect shapes such as lines and circles in imagesExplore object tracking and detection with algorithmsProcess videos using different video analysis techniques in Jetson TX1Access GPU device properties from the PyCUDA programUnderstand how kernel execution worksWho this book is for This book is a go-to guide for you if you are a developer working with OpenCV and want to learn how to process more complex image data by exploiting GPU processing. A thorough understanding of computer vision concepts and programming languages such as C++ or Python is expected.