Document Status

유형	Draft (초안)	보안	Confidential

OpenCL is a registered trademark of Apple Inc. used by permission by Khronos Group. All references to OpenCL components in this document are referenced from the publicly available OpenCL specification on the Khronos web-site at: http://www.khronos.org/opencl

NVIDIA, the NVIDIA logo, CUDA, and GeForce are trademarks or registered trademarks of NVIDIA Corporation.

Information furnished is believed to be accurate and reliable. However, PSC Group assumes no responsibility for the consequences of use of such information nor for any infringement of patents or other rights of third parties which may result from its use. No license is granted by implication or otherwise under any patent or patent rights of PSC Group. Specifications mentioned in this publication are subject to change without notice. This publication supersedes and replaces all information previously supplied. PSC Group products are not authorized for use as critical components in life support devices or systems without express written approval of PSC Group.

The PSC Group logo is a registered trademark of PSC Group.

All other names are the property of their respective owners

© 2011 PSC Group - All rights reserved

목차

Intro	duc	cing	4
분석	자로	₹	4
	1.	GPGPU의 기반이 되는 Architecture	4
	2.	GPGPU 적용을 위한 Application 분석	.11
:	3.	기반이 되는 Framework로서 적합한 Hadoop	.12
개발	계호	틱(가칭 PSC-framework : Personal Super Computing framework)	.14
:	1.	Hadoop System 적용	.14
2	2.	PSC-framework	.14
:	3.	PSC-framework Block Diagram	.16
4	4.	PSC-framework Data Flow	.16
관련:	자로	<u></u>	.17
	1.	GPGPU를 이용한 어플리케이션	.17
:	2.	GPGPU 관련 서적	.20
:	3.	GPGPU 관련 IEEE 자료	.24

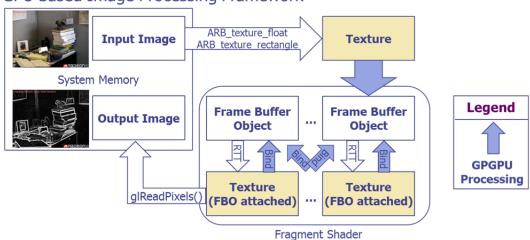
Introducing

본 문서는 '공개소프트웨어 기반의 개인용 슈퍼 컴퓨팅 플랫폼 구축 및 커뮤니티 운영' 과제에 대한 개발 분석과 계획에 대하여 서술한 자료임.

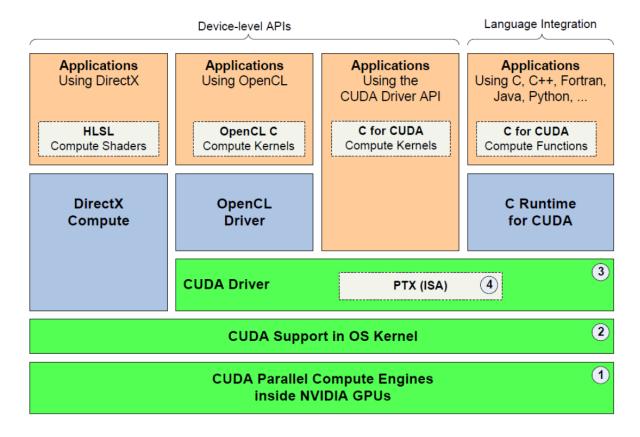
분석자료

- 1. GPGPU의 기반이 되는 Architecture
 - 1.1 초기 GPGPU Framework

: GPU based Image Processing Framework*



- ◈ Render-to-texture를 활용한 off-screen 렌더링 방법
 - 입력 영상의 해상도와 동일한 크기를 갖는 frame buffer 생성
 - 입력 영상은 texture data로 지정되어 비디오 메모리에 저장
 - Vertex shader에서는 기본적인 행렬변환, fragment shader에서 영상 처리
 - 결과값은 FBO 형식의 비디오 메모리에 저장되고 최종적으로 시스템 메모리로 복사
 - 영상처리가 아닌 데이터의 처리 역시 같은 Framework로 mapping 됨
- * Ref. : 이만희, 박인규, 원석진, 조성대, "GPU를 이용한 DWT 및 JPEG2000의 고속 연산", 전자 공학회 논문지, Vol.44-SP, No.6, pp.9-15, 2007년 11월
- 1.2 엔비디아 CUDA 구조
 - 1.2.1 Block Diagram



1.2.2 CUDA Software 개발환경

Libraries	Advanced libraries that include BLAS, FFT, and other functions optimized for the CUDA architecture
C Runtime	The C Runtime for CUDA provides support for executing standard C functions on the GPU and allows native bindings for other high-level languages such as Fortran, Java, and Python
Tools	NVIDIA C Compiler (nvcc), CUDA Debugger (cudagdb), CUDA Visual Profiler (cudaprof), and other helpful tools
Documentation	Includes the CUDA Programming Guide, API specifications, and other helpful documentation
Samples	SDK code samples and documentation that demonstrate best practices for a wide variety GPU Computing algorithms and applications

1.2.3 CUDA 적용 Language

Fortran:

- o Fortran wrapper for CUDA http://www.nvidia.com/object/cuda_programming_tools.html
- o FLAGON Fortran 95 library for GPU Numerics http://flagon.wiki.sourceforge.net/
- o PGI Fortran to CUDA compiler http://www.pgroup.com/resources/accel.htm

Java:

o JaCuda – http://jacuda.wiki.sourceforge.net

o Bindings for CUDA BLAS and FFT libs - http://javagl.de/index.html

Python:

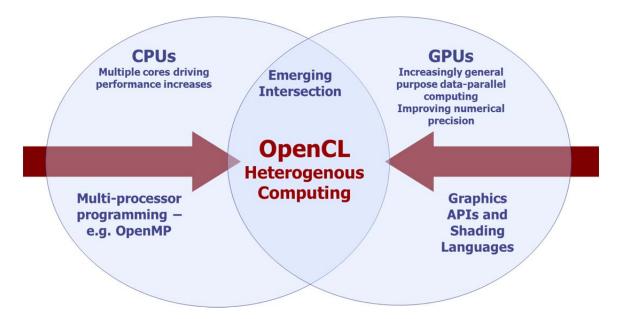
o PyCUDA Python wrapper – http://mathema.tician.de/software/pycuda

.NET languages:

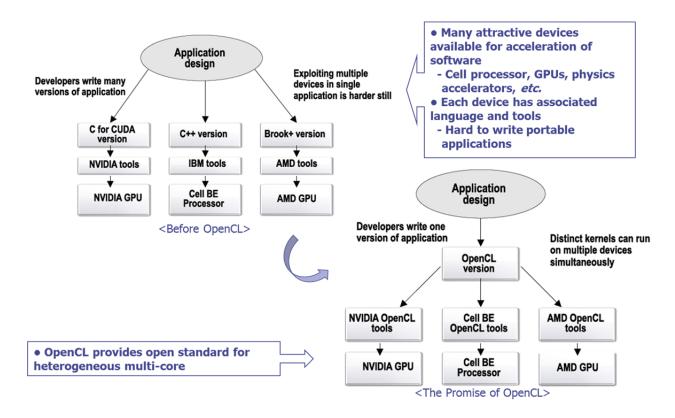
o CUDA.NET - http://www.gass-ltd.co.il/en/products/cuda.net

Resources for other languages:

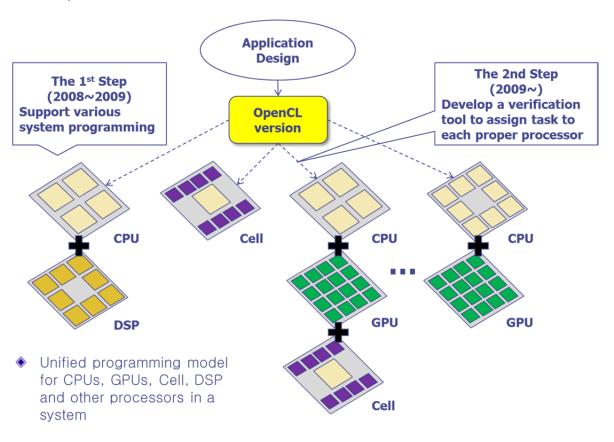
- SWIG http://www.swig.org (generates interfaces to C/C++ for dozens of languages)
- 1.3 OpenCL (OpenCL은 CUDA에 비해 검증이 부족하므로 본 과제에서는 기술 참조용임)
 - 1.3.1 OpenCL Open Computing Language



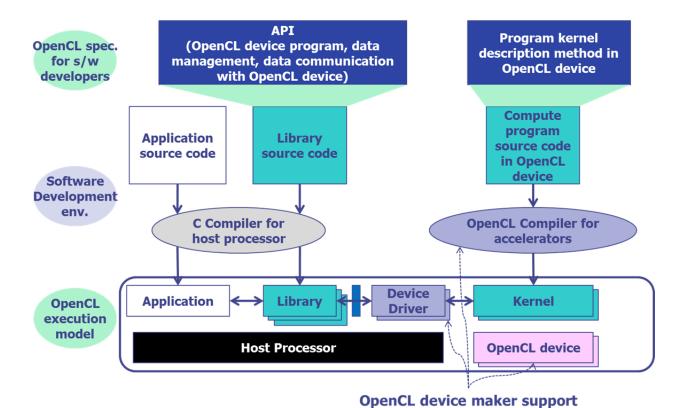
1.3.2 OpenCL 적용 이전과 이후



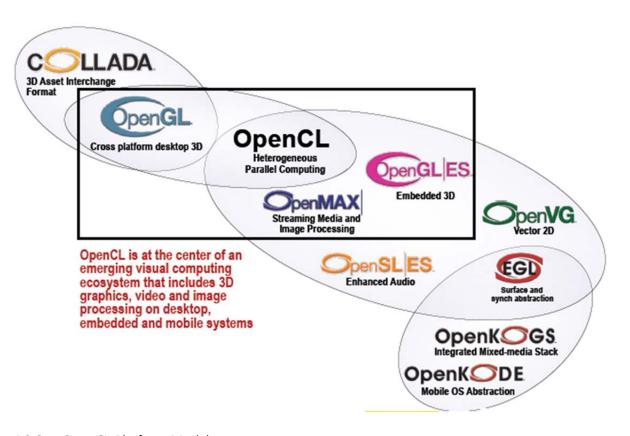
1.3.3 OpenCL 조합



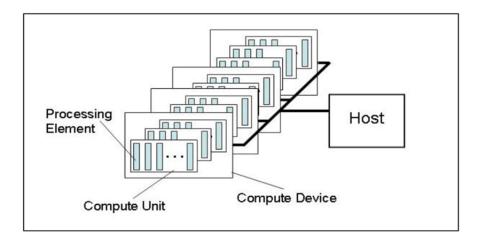
1.3.4 OpenCL Software



1.3.5 OpenCL과 기타 표준과의 관계

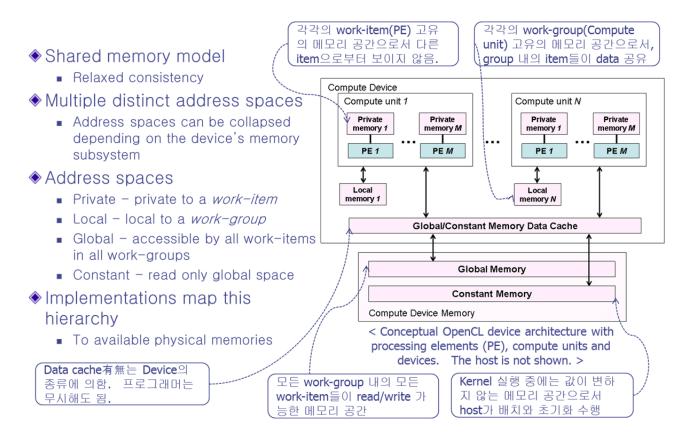


1.3.6 OpenCL Platform Model

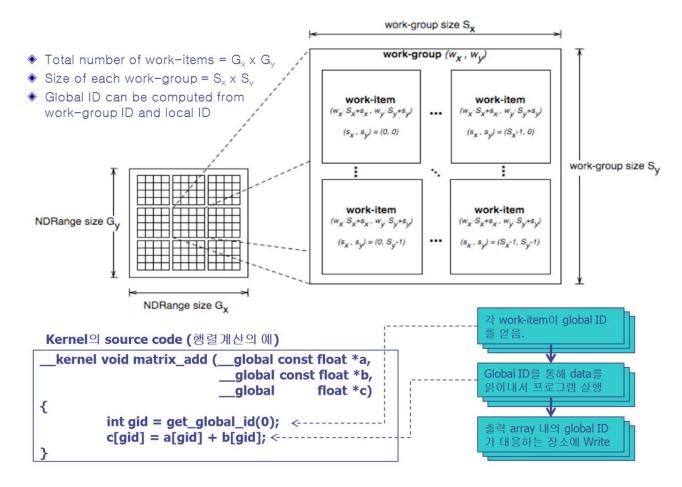


- ◆One <u>Host</u> + one or more <u>Compute Devices</u>
 - Each Compute Device is composed of one or more Compute Units
 - Each Compute Unit is further divided into one or more <u>Processing</u> Elements

1.3.7 OpenCL Memory Model



1.3.8 OpenCL Kernel 실행



1.3.9 OpenCL SDK 제공

Intel OpenCL SDK: http://software.intel.com/en-us/articles/opencl-sdk/#resource

NVIDIA OpenCL SDK: http://developer.nvidia.com/opencl

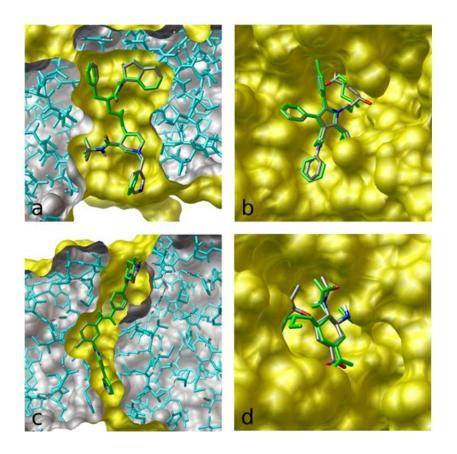
AMD OpenCL SDK: http://developer.amd.com/sdks/AMDAPPSDK/Pages/default.aspx

1.3.10 OpenCL Reference.

- * OpenCL The Open Standard for Heterogeneous Parallel Programming, Neil Trevett(Khronos President) , Multicore Expo, March 2009
- * OpenCL Technical Overview, John Roberts(NVidia), Multicore Expo, March 2009
- * OpenCL as an intermediate language for heterogeneous multi-core programming, Alastair F. Donaldson(Codeplay Software Ltd.), Multicore Expo, March 2009
- * "Apple主導 の OpenCL プロセ サに 自由 をもたらす, Nikkei Electronics, Dec. 08, 2008
- * Khronos Group: http://www.khronos.org

- * The OpenCL Specification, version : 1.0, Document Revisions : 43, Khronos OpenCL Working Group
- * Open Standard Industry Update from handhelds to supercomputers, Jon Peddie Research
- 2. GPGPU 적용을 위한 Application 분석.
 - 2.1 AutoDock Vina
 - 2.1.1 기본 설명
 - O AutoDock Vina는 신약개발을 위한 Open Source Project
 - O Molecular docking, virtual screening 기능
 - O multi-core capability, high performance, 정확도 향상, 사용 용이성 제공
 - 2.1.2 개발자
 - O Dr. Oleg Trott in the Molecular Graphics Lab at The Scripps Research Institute
 - 2.1.3 결과 형상

flexible docking (green)의 결과가 crystal structures (a) indinavir, (b) atorvastatin, (c) imatinib, and (d) oseltamivir 각각 에 올려있는 모습



2.1.4 참고자료

http://vina.scripps.edu/

O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, *Journal of Computational Chemistry 31 (2010) 455-461*

3. 기반이 되는 Framework로서 적합한 Hadoop¹

하둡(Hadoop)은 대량의 자료를 처리할 수 있는 큰 컴퓨터 클러스터에서 동작하는 분산응용 프로그램을 지원하는 자유 자바 소프트웨어 프레임워크이다. 원래 너치의 분산처리를 지원하기 위해 개발된 것으로, 아파치 루씬의 하부 프로젝트이다. 분산처리 시스템인 구글 파일 시스템을 대체할 수 있는 하둡 분산 파일 시스템(HDFS: Hadoop Distributed File System)

_

¹ http://ko.wikipedia.org/wiki/Hadoop

과 맵리듀스를 구현한 것이다

3.1 Hodoop 분산 파일 시스템(HDPS: Hadoop Distributed File Sytem)

하둡 분산 파일 시스템은 다음과 같은 시스템에서 잘 동작하는 것을 목표로 하고 있다

3.1.1 하드웨어 오동작

하드웨어 수가 많아지면 그 중에 일부 하드웨어가 오동작하는 것은 예외 상황이 아니라 항상 발생하는 일이다. 따라서 이런 상황에서 빨리 자동으로 복구하는 것은 HDFS의 중요한 목표다.

3.1.2 스트리밍 자료 접근

범용 파일 시스템과 달리 반응 속도보다는 시간당 처리량에 최적화되어 있다.

3.1.3 큰 자료 집합

한 파일이 기가바이트나 테라바이트 정도의 크기를 갖는 것을 목적으로 설계되었다. 자료 대역폭 총량이 높고, 하나의 클러스터에 수 백개의 노드를 둘 수 있다. 하나의 인스턴스에서 수천만여 파일을 지원한다.

3.1.4 간단한 결합 모델

한번 쓰고 여러번 읽는 모델에 적합한 구조이다. 파일이 한번 작성되고 닫히면 바뀔 필요가 없는 경우를 위한 것이다. 이렇게 함으로써 처리량을 극대화할 수 있다.

3.1.5 자료를 옮기는 것보다 계산을 옮기는 것이 비용이 적게 든다.

자료를 많이 옮기면 대역폭이 많이 들기 때문에 네트워크 혼잡으로 인하여 전체 처리량이 감소한다. 가까운 곳에 있는 자료를 처리하게 계산 작업을 옮기면 전체적인 처리량이 더 높아진다.

3.1.6 다른 종류의 하드웨어와 소프트웨어 플랫폼과의 호환성

서로 다른 하드웨어와 소프트웨어 플랫폼들을 묶어 놓아도 잘 동작한다.

3.1.7 네임노드와 데이터노드

HDFS는 마스터/슬레이브(master/slave) 구조를 가진다. HDFS 클러스터는 하나의 네임노드와, 파일 시스템을 관리하고 클라이언트의 접근을 통제하는 마스터 서버로 구성된다. 게다가 클러스터의 각 노드에는 데이터노드가 하나씩 존재하고, 이 데이터 노드는 실행

될 때마다 노드에 추가되는 스토리지를 관리한다. HDFS는 네임스페이스를 공개하여서 유저 데이터가 파일에 저장되는것을 허락한다. 내부적으로 하나의 파일은 하나 이상의 블록으로 나누어져 있고, 이 블록들은 데이터노드들에 저장되어 있다. 네임노드는 파일과디렉터리의 읽기(open), 닫기(close), 이름 바꾸기(rename) 등, 파일시스템의 네임스페이스의 여러 기능을 수행한다. 또한, 데이터 노드와 블록들의 맵핑을 결정한다. 데이터 노드는 파일시스템의 클라이언트가 요구하는 읽기(read), 쓰기(write) 기능들을 담당한다. 또한데이터 노드는 네임노드에서의 생성, 삭제, 복제 등와 같은 기능도 수행한다.

네임노드와 데이터노드는 GNU/Linux OS를 기반으로 하는 상용머신에서 실행하기 위해 디자인된 소프트웨어의 일부이다. HDFS는 Java 언어를 사용하므로 Java가 동작하는 어떠 한 컴퓨터에서나 네임노드나 데이터노드 소프트웨어를 실행할 수 있다.

3.2 MapReduce

MapReduce는 구글에서 개발한 소프트웨어 프레임워크(방법론)이다. 이 프레임웍은 페타바이트 이상의 대용량 데이터를 신뢰할 수 없는 컴퓨터로 구성된 클러스터 환경에서 병렬 처리를 지원하기 위해서 개발되었다. 이 프레임웍은 함수형 프로그래밍에서 일반적으로 사용되는 Map과 Reduce라는 함수 기반으로 주로 구성된다.

개발계획(가칭 PSC-framework : Personal Super Computing framework)

- 1. Hadoop System 적용
 - 1.1 필요성
 - 1.1.1 개인용 슈퍼 컴퓨팅 프레임워크를 구축해서 많은 개발자들이 쉽게 사용하기 위해서 는 검증된 기술이 필요
 - 1.1.2 특히 2대 이상의 Work
- 2. PSC-framework
 - 2.1 필요성
 - 2.1.1 엔지니어, 과학자, 분석가, 기술 전문가들이 GPU를 이용한 데이터 분석 작업에 연산 성능을 최대한으로 이용할 수 있도록 지원. 최근 증대되는 이미지 데이터, 복잡한

분석 알고리즘, 수치해석 모델 등이 기존 CPU의 성능으로는 부족함을 느끼고 있기 때문에 GPU를 추가로 이용하는 것이 필요함.

2.1.2 복잡한 GPGPU를 쉽게 용이하게 응용할 수 있는 framework의 제공이 필요함.

2.2 필요 기능

2.2.1 기존 CPU 프로그래밍을 하던 개발자들도 GPU 사용을 쉽게 할 수 있도록 지원해야 하며 이를 위한 해결 방안은 대용량 병렬 알고리즘이나 GPU 메모리 관리의 용이한 개발 환경이 필요함.

가속된 Mathematical, Image Filtering, Simulation Functions

Structural mechanics, fuid dynamics, earth, science, biosciences, medial/diagnostic imaging, financial engineering.

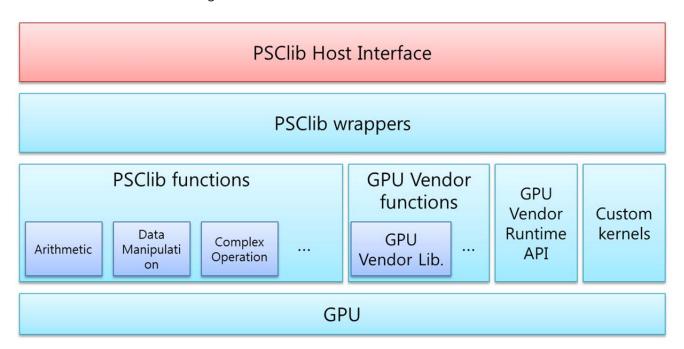
2.2.2 오작동 복구, 클라우드 컴퓨팅, 분산처리, 큰 자료집합, 대용량 처리, 스트리밍 자료 접근, 다양한 하드웨어/소프트웨어 플랫폼 지원, 간단한 작동 모델, 등에 적용이 가 능한 프레임워크가 바람직함.

2.3 참고자료

CUDA Web Site: http://www.nvidia.com/cuda

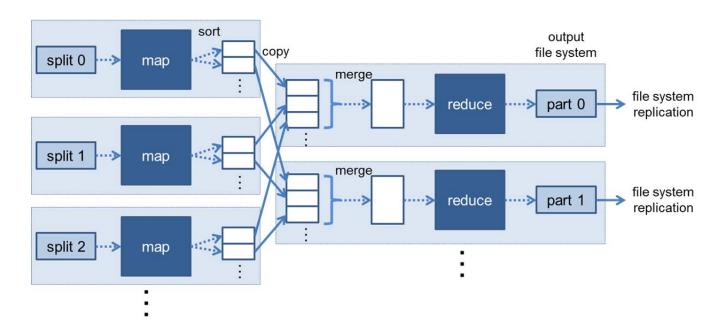
기타 본 문서 '관련자료' 참조

3. PSC-framework Block Diagram



<Fig. PSC-framework Block Diagram>

- 4. PSC-framework Data Flow
 - 4.1 설명
 - 4.1.1 Hadoop의 MapReduce를 일부 변형한 기반에서 동작.
 - 4.2 Data Flow



<Fig. PSC-framework Data Flow>

4.2.1 Hodoop의 MapReduce 방식을 기반으로 위의 그림과 같이 Reduce가 여럿일 때 Map task는 각 Reduce task만큼 파티션을 생성하고 Map의 결과를 그 파티션으로 분배시킴. 각 파티션에는 여러 키(그리고 상응하는 값)들이 존재하지만 각 키에 대한 레코드는 하나의 파티션에 모두 존재. Partitioning algorithm은 일반적으로 해시 함수를 이용하여 키들을 bucket에 분배해주는 기본적인 partitioner를 사용하거나 사용자 정의 파티셔닝 함수를 사용할 수 있음. 그러나 GPGPU에서는 Hardware의 물리적인 구조와 메모리의 한계로 인해 더 단순하게 처리될 수 있음.²

관련자료

1. GPGPU를 이용한 어플리케이션

category	title
Government & Defense	RealityServer
Government & Defense	Ikena: Imagery Analysis and Video Forensics
Government & Defense	Signal Processing Library: GPU VSIPL
Government & Defense	IDL and MATLAB Acceleration: GPULib

² Hadoop 완벽 가이드, O'Reilly Yahoo Press, 한빛미디어, 톰 화이트 지음, 심탁길, 김우현 옮김.

Government & Defense	GIS: Manifold
Government & Defense	MATLAB GPU Computing: MathWorks
Government & Defense	MATLAB Plugin: Accelereyes
Molecular Dynamics, Computational Chemistry	OpenMM library for accelerating molecular dynamics on GPUs
Molecular Dynamics, Computational Chemistry	GROMACS using OpenMM
Molecular Dynamics, Computational Chemistry	NAMD molecular dynamics
Molecular Dynamics, Computational Chemistry	VMD visualization of molecular dynamics
Molecular Dynamics, Computational Chemistry	HOOMD molecular dynamics
Molecular Dynamics, Computational Chemistry	Acellera: ACEMD bio-molecular dynamics package
Molecular Dynamics, Computational Chemistry	BigDFT: DFT (Density functional theory) electronic structure code
Molecular Dynamics, Computational Chemistry	MDGPU
Molecular Dynamics, Computational Chemistry	GPUGrid.net
Molecular Dynamics, Computational Chemistry	MATLAB GPU Computing: MathWorks
Molecular Dynamics, Computational Chemistry	MATLAB plugin: Accelereyes
Life Sciences, Bio-informatics	GPU HMMER
Life Sciences, Bio-informatics	DNA Sequence alignment: MUMmerGPU
Life Sciences, Bio-informatics	LISSOM: model of human neocortex using CUDA
Life Sciences, Bio-informatics	Silicon Informatics: AutoDock
Life Sciences, Bio-informatics	MATLAB plugin: Accelereyes
Electrodynamics and electromagnetic	Acceleware: FDTD Solver
Electrodynamics and electromagnetic	Acceleware: EM Solutions
Electrodynamics and electromagnetic	Remcom XStream FDTD
Electrodynamics and electromagnetic	SPEAG Semcad X
Electrodynamics and electromagnetic	CST Microwave Studio
Electrodynamics and electromagnetic	Quantum electrodynamics library
Electrodynamics and electromagnetic	GPMAD : Particle beam dynamics simulator
Medical Imaging, CT, MRI	<u>RealityServer</u>
Medical Imaging, CT, MRI	GPULib:IDL acceleration
Medical Imaging, CT, MRI	Acceleware: Imaging Solutions
Medical Imaging, CT, MRI	Digisens: SnapCT tomographic reconstruction software
Medical Imaging, CT, MRI	Techniscan: Whole Breast Ultrasound Imaging System
Medical Imaging, CT, MRI	NVPP: NVIDIA Performance Primitives (early access)
Medical Imaging, CT, MRI	MATLAB GPU Computing: MathWorks
Medical Imaging, CT, MRI	MATLAB plugin: Accelereyes

Oil & Gas	RealityServer
Oil & Gas	Acceleware: Kirchoff and Reverse Time Migration
Oil & Gas	SeismicCity: 3D seismic imaging for prestack depth migration
Oil & Gas	OpenGeoSolutions: Spectral decomposition and inversion
Oil & Gas	Mercury Computer systems: 3D data visualization
Oil & Gas	ffA: 3D Seismic processing software
Oil & Gas	Headwave: Prestack data processing
Oil & Gas	GIS: Manifold
Oil & Gas	MATLAB GPU Computing: MathWorks
Oil & Gas	MATLAB plugin: Accelereyes
Financial computing and options pricing	SciComp: derivatives pricing
Financial computing and options pricing	Hanweck: options pricing
Financial computing and options pricing	Exegy: Risk Analysis
Financial computing and options pricing	Aqumin: 3D Visualization of market data
Financial computing and options pricing	Level 3 Finance
Financial computing and options pricing	OnEye (Australia): Accelerated Trading Solutions
Financial computing and options pricing	Arbitragis Trading
Financial computing and options pricing	Enabling GPU Computing in the R Statistical Environment
Financial computing and options pricing	MATLAB GPU Computing: MathWorks
Financial computing and options pricing	MATLAB plugin: Accelereyes
Math	CUDA Acceleration for MATLAB
Math	Accelereyes: Jacket engine for MATLAB
Math	GPULib: mathematical functions for IDL and MATLAB
Math	Integrating Simulink with CUDA using S-functions
Math	Enabling GPU Computing in the R Statistical Environment
Math	Mathematica plug-in for CUDA
Math	Using NVIDIA GPUs with National Instruments LabView
Electronic Design Automation	Agilent EESof: ADS SPICE simulator_
Electronic Design Automation	Synopsys: Sentaraus TCAD
Electronic Design Automation	Gauda: Optical proximity correction (OPC)
Weather and Ocean Modeling	CUDA-accelerated WRF code
Video, Imaging, and Vision Applications	Axxon Intellect Enterprise Video Surveillance Software
Video, Imaging, and Vision Applications	Pflow CUDA Plugin for Autodesk 3ds Max
Video, Imaging, and Vision Applications	RUINS Shatter CUDA Plug-in for Maya

Video, Imaging, and Vision Applications Bullet 3D Multi-Physics Library with CUDA Support	
Video, Imaging, and Vision Applications <u>CUDA Voxel Rendering Engine</u>	
Video, Imaging, and Vision Applications	NVPP: NVIDIA Performance Primitives (early access) Volume Rendering with CUDA for VTK / Slicer3
Video, Imaging, and Vision Applications	Furryball: Direct3D GPU Rendering Plugin for Maya
Video, Imaging, and Vision Applications	For consumer CUDA applications, visit NZone

출처 : CUDA-Accelerated Applications (http://www.nvidia.com/object/cuda app tesla.html)

2. GPGPU 관련 서적

Front Cover	Title	Autor / Editor	Date	Publisher
Programming Guide Adda Made: Number Sister Fraciley C. Statton: State	Opencl Programming Guide	Aaftab Munshi; Benedict Gaster; Timothy G. Mattson; James Fung	2011-7-25	Addison- Wesley Professional
GPU PRO2	GPU Pro 2	Wolfgang Engel	2011-02-14	A K Peters/CRC Press
COMPUTING GEMS Emerada Entrori WE-MEI W. HWU white from trace MK 200	Gpu Computing Gems Emerald Edition	Wen-mei W. Hwu	2011-2-7	Elsevier Science & Technology

Scientific Computing with Multicore and Accelerators Item to provide Kingah Daniel Kingah Balate Jank Dengana	Scientific Computing with Multicore and Accelerators (Chapman & Hall/CRC Computational Science)	Jakub Kurzak, David A. Bader, Jack Dongarra	2010-12-07	CRC Press
CUDA BY EXAMPLE A relationaries in the relative and the	Cuda By Example: An Introduction to General- Purpose GPU Programming (OpenCL 입문 - GPU 와 멀티코어 CPU 병렬 프로그래밍)	Jason Sanders; Edward Kandrot	2010-7-29	Addison- Wesley
GPU PRO	GPU Pro: Advanced Rendering Techniques	Wolfgang Engel	2010-07-01	A K Peters
OpenCL APT	Opencl 入門 - GPU&マルチコア CPU 並列 プログラミング for MacOS Windows Linux (OpenCL 입문 - GPU 와 멀티코어 CPU 병렬 프로그래밍)	奥薗隆司	2010-5-1	インプレスジャパン

The OpenCL Programming Book Radial Registering for Halling and GPU Radial Registering Facility Statement Fa	The OpenCL Programming Book	Ryoji Tsuchiyama, Takashi Nakamura, Takuro Iizuka, Akihiro Asahara, Satoshi Miki	2010-03-31	Fixstars Corporation
CUDA 高速GPU プログラミング入門	Cuda 高速 gpu プログラミ ング入門 (CUDA 고속 GPU 프로그래밍 입문)	岡田賢治	2010-3-1	秀和システム
Programming Massively Parallel Processors A Ministran Approach	Programming Massively Parallel Processors	David Kirk; Wen-mei W. Hwu	2010-2-5	Elsevier Science & Technology
OpenCLAFF **********************************	Opencl 入門 - マルチコア CPU・GPU のた めの並列プログラミング (OpenCL 입문 - 멀티코어 CPU, GPU 를 위한 병렬 프로그래밍)	株式会社フィックスターズ, 土山 了士, 中村 孝史, 飯塚 拓郎, 浅原 明広, 三木 聡	2010-1-22	インプレスジャパン

OpenCL 並列プログラミング TASESTALUTION SERVED TO THE TERM	Opencl 並列プログラミング ニャルチコア CPU/GPU のための標準フレームワーク(OpenCL 병렬 프로그래밍 - 멀티코어 CPU/GPU를 위한표준 프레임워크)	池田成樹	2010-1-1	カットシステム
CUDA プログラミング	はじめての Cuda プログラミング - 驚異の開発環境[GPU+CUDA]]を使いこなす! (처음 배우는 CUDA 프로그래밍 - 경이적인 개발환경[GPU+CUDA]을 자유자재로!)	青木尊之; 額田彰	2009-11-1	工学社
RP (日本 日本 日	GPU 高性能运算之 CUDA (GPU 고성능 연산 CUDA)	张舒, 褚艳利	2009-10-1	DynoMedia Inc.
基于GPU的多尺度 函数模拟并行计算	基于 GPU 的多尺度离散模拟 并行计算 (GPU 기반의 다척도 이산시뮬레이션 병렬연산)	多相复杂系列国家重点实验 室,多尺度离散模拟项目组	2009-1-1	科学出版社

PROCESS ALGEBRA FOR PARALLEL AND DISTRIBUTED PROCESSING MODIFICATION WILLIAM GARRIES WILLIAM GARRIES	Process Algebra For Parallel And Distributed Processing (Chapman & Hall/Crc Computational Science Series)	Michael Alexander; William Gardner	2008-12-22	Chapman & Hall/CRC
THE ART MULTIPROCESSOR PROGRAMMING	The Art of Multiprocessor Programming	Maurice Herlihy, Nir Shavit	2008-03-14	Morgan Kaufmann

- * 출처 : 1) NVIDIA CUDA & GPU 컴퓨팅 서적 (http://www.nvidia.co.kr/object/cuda_books_kr.html)
 - 2) 아마존 (http://www.amazon.com, http://www.amazon.cn, http://www.amazon.co.jp)
 - 3) FIXSTARS (http://www.fixstars.com/en/company/books/opencl/)

3. GPGPU 관련 IEEE 자료

1	Parallel Exact Inference on a CPU-GPGPU Heterogenous System
2	Optimal loop unrolling for GPGPU programs
3	Exploring GPGPU workloads: Characterization methodology, analysis and microarchitecture evaluation implications
4	Parallel connected-component labeling algorithm for GPGPU applications
5	The optimization of parallel Smith-Waterman sequence alignment using on-chip memory of GPGPU
6	GPGPU-based Latency Insertion Method: Application to PDN simulations
7	Many-Thread Aware Prefetching Mechanisms for GPGPU Applications
8	GPGPU implementation of a synaptically optimized, anatomically accurate spiking network simulator
9	Migrating real-time depth image-based rendering from traditional to next-gen GPGPU
10	GPGPU supported cooperative acceleration in molecular dynamics
11	A GPGPU-Based Collision Detection Algorithm
12	A Case Study of SWIM: Optimization of Memory Intensive Application on GPGPU
13	Neuromorphic models on a GPGPU cluster

14	Parallelizing Simulated Annealing-Based Placement Using GPGPU
15	Fast implementation of Wyner-Ziv Video codec using GPGPU
16	GPGPU-FDTD method for 2-dimensional electromagnetic field simulation and its estimation
17	6.8: Presentation session: Neuroanatomy, neuroregeneration, and modeling: "GPGPU implementation of a synaptically optimized, anatomically accurate spiking network simulator"
18	Accelerating Particle Swarm Algorithm with GPGPU
19	GpuWars: Design and Implementation of a GPGPU Game
20	Enabling Energy-Efficient Analysis of Massive Neural Signals Using GPGPU
21	Efficient scan-window based object detection using GPGPU
22	Barra: A Parallel Functional Simulator for GPGPU
23	High-speed electromagnetic field simulation by HIE-FDTD method with GPGPU
24	An implementation and its evaluation of password cracking tool parallelized on GPGPU
25	GPGPU-Aided Ensemble Empirical-Mode Decomposition for EEG Analysis During Anesthesia
26	FSimGP^2: An Efficient Fault Simulator with GPGPU
27	Parallel implementation of a Quantization algorithm for pricing American style options on GPGPU
28	Development of nonlinear filter bank system for real-time beautification of facial video using GPGPU
29	Emerging technology about GPGPU
30	Parallel implementation of Quantization methods for the valuation of swing options on GPGPU
31	Acceleration of Streamed Tensor Contraction Expressions on GPGPU-Based Clusters
32	Acceleration of Functional Validation Using GPGPU
33	Optimizing vehicle routing problems using evolutionary computation on gpgpu
34	Fast Disk Encryption through GPGPU Acceleration
35	Preliminary implementation of VQ image coding using GPGPU
36	Optimum real-time reconstruction of Gamma events for high resolution Anger camera with the use of GPGPU
37	Implementation of Sequential Importance Sampling in GPGPU
38	Effectiveness of a strip-mining approach for VQ image coding using GPGPU implementation
39	hiCUDA: High-Level GPGPU Programming
40	A performance prediction model for the CUDA GPGPU platform
41	A Program Behavior Study of Block Cryptography Algorithms on GPGPU
42	Linear genetic programming GPGPU on Microsoft's Xbox 360
43	Parallel implementation of pedestrian tracking using multiple cues on GPGPU
44	Fast parallel analysis of dynamic contrast-enhanced magnetic resonance imaging on GPGPU
45	A design case study: CPU vs. GPGPU vs. FPGA
46	Synthetic Aperture Radar Processing with GPGPU
47	Auto-tuning Dense Matrix Multiplication for GPGPU with Cache
48	Parallelization of spectral clustering algorithm on multi-core processors and GPGPU
49	Size Matters: Space/Time Tradeoffs to Improve GPGPU Applications Performance
50	Hard Data on Soft Errors: A Large-Scale Assessment of Real-World Error Rates in GPGPU

51	SIFT-Cloud-Model for object detection and pose estimation with GPGPU acceleration
52	Performance Debugging of GPGPU Applications with the Divergence Map
53	GPGPU-based Gaussian Filtering for Surface Metrological Data Processing
54	Processing of synthetic Aperture Radar data with GPGPU
55	Message passing for GPGPU clusters: CudaMPI
56	Recent trends in software and hardware for GPGPU computing: A comprehensive survey
57	Accelerating PCG power/ground network solver on GPGPU
58	Nonnegative Tensor Factorization Accelerated Using GPGPU
59 60	An Interior Point Optimization Solver for Real Time Inter-frame Collision Detection: Exploring Resource-Accuracy-Platform Tradeoffs Design and Implementation of a Uniform Platform to Support Multigenerational GPU Architectures for Lligh Performance Stream Recod Computing
61	Architectures for High Performance Stream-Based Computing Fast Two Dimensional Convex Hull on the GPU
62	Profiling General Purpose GPU Applications
63	Planetary-Scale Terrain Composition
64	CUDA implementation of McCann99 retinex algorithm
65	GridCuda: A Grid-Enabled CUDA Programming Toolkit
66	GPU Accelerated Lanczos Algorithm with Applications
67	String Matching on a Multicore GPU Using CUDA
68	Object oriented framework for real-time image processing on GPU
69	In Situ Power Analysis of General Purpose Graphical Processing Units
70	A fast GPU algorithm for graph connectivity
71	Statistical Testing of Random Number Sequences Using Graphics Processing Units
72	Implementation of Ant Colony Algorithm Based on GPU
73	Fast Deformable Registration on the GPU: A CUDA Implementation of Demons
74	Theoretical and Empirical Analysis of a GPU Based Parallel Bayesian Optimization Algorithm
75	A Translation Framework for Virtual Execution Environment on CPU/GPU Architecture
76	RankBoost Acceleration on both NVIDIA CUDA and ATI Stream Platforms
77	High Performance Hybrid Functional Petri Net Simulations of Biological Pathway Models on CUDA
78	Heuristic Optimization Methods for Improving Performance of Recursive General Purpose Applications on GPUs
79	Shape Manipulation on GPU
80	Parallel and distributed seismic wave field modeling with combined Linux clusters and graphics processing units
81	Scalable, High Performance Fourier Domain Optical Coherence Tomography: Why FPGAs and Not GPGPUs
82	Barnes-hut treecode on GPU
83	Parallel processing between GPU and CPU: Concepts in a game architecture
84	Implementation of TFT inspection system using the common unified device architecture (CUDA) on modern graphics hardware
85	OpenCL: Make Ubiquitous Supercomputing Possible
86	SSE Vectorized and GPU Implementations of Arakawa's Formula for Numerical Integration of Equations of Fluid Motion

87	A Hybrid Computational Grid Architecture for Comparative Genomics
88	Task Scheduling of Parallel Processing in CPU-GPU Collaborative Environment
89	Toward Harnessing DOACROSS Parallelism for Multi-GPGPUs
90	Massively Parallel Neural Signal Processing: A Case for Analysis of EEG with Absence Seizure
91	An Architecture for Improving the Efficiency of Specialized Vertical Search Engine Based on GPGPUs
92	Reducing IO bandwidth for GPU based moment invariant classifier systems
93	An algorithmic incremental and iterative development method to parallelize dusty-deck FORTRAN HPC codes in GPGPUs using CUDA
94	Throughput-Effective On-Chip Networks for Manycore Accelerators
95	Fast seismic modeling and Reverse Time Migration on a GPU cluster
96	Fast Motion Estimation on Graphics Hardware for H.264 Video Encoding
97	Community Structure Discovery algorithm on GPU with CUDA
98	Multi-agent traffic simulation with CUDA
99	Data structure design for GPU based heterogeneous systems
100	GPU acceleration of method of moments matrix assembly using Rao-Wilton-Glisson basis functions
101	A High-Performance Multi-user Service System for Financial Analytics Based on Web Service and GPU Computation
102	Improving Hybrid OpenCL Performance by High Speed Networks
103	Accelerating spatial clustering detection of epidemic disease with graphics processing unit
104	Automated development of applications for graphical processing units using rewriting rules
105	Statistical testing of random number sequences using CUDA
106	A Dynamic Resource Management and Scheduling Environment for Embedded Multimedia and Communications Platforms
107	Fast acoustic computations using graphics processors
108	GPU acceleration of the dynamics routine in the HIRLAM weather forecast model
109	An approach of tool paths generation for CNC machining based on CUDA
110	Implementation and optimization of image processing algorithms on handheld GPU
111	GPU-based high-speed and high-precision visual tracking
112	CUDA Memory Optimizations for Large Data-Structures in the Gravit Simulator
113	Mapping High-Fidelity Volume Rendering for Medical Imaging to CPU, GPU and Many-Core Architectures
114	Hybrid Map Task Scheduling for GPU-Based Heterogeneous Clusters
115	Power-Efficient Work Distribution Method for CPU-GPU Heterogeneous System
116	Fast Variable Center-Biased Windowing for High-Speed Stereo on Programmable Graphics Hardware
117	Accelerating Phase Correlation Functions Using GPU and FPGA
118	GP-GPU: Bridging the Gap between Modelling & Experimentation
119	How GPUs Work
120	XMalloc: A Scalable Lock-free Dynamic Memory Allocator for Many-core Machines
121	Speeding up K-Means Algorithm by GPUs
122	Accelerator-Oriented Algorithm Transformation for Temporal Data Mining

123	A tile-based parallel Viterbi algorithm for biological sequence alignment on GPU with CUDA
124	Massively parallel implementation of cyclic LDPC codes on a general purpose graphics processing unit
125	Accelerating Simulations of Light Scattering Based on Finite-Difference Time-Domain Method with General Purpose GPUs
126	CaravelaMPI: Message Passing Interface for Parallel GPU-Based Applications
127	Non-intrusive Performance Analysis of Parallel Hardware Accelerated Applications on Hybrid Architectures
128	Accelerating System-Level Design Tasks Using Commodity Graphics Hardware: A Case Study
129	A package for OpenCL based heterogeneous computing on clusters with many GPU devices
130	Computation of Voronoi diagrams using a graphics processing unit
131	High throughput multiple-precision GCD on the CUDA architecture
132	CANSCID-CUDA
133	CUDA-BLASTP: Accelerating BLASTP on CUDA-Enabled Graphics Hardware
134	GPU Accelerated Path-Planning for Multi-agents in Virtual Environments
135	An efficient implementation of Smith Waterman algorithm on GPU using CUDA, for massively parallel scanning of sequence databases
136	A CUDA-Based Implementation of Stable Fluids in 3D with Internal and Moving Boundaries
137	GPU Acceleration of Runge-Kutta Integrators
138	Program Optimization of Array-Intensive SPEC2k Benchmarks on Multithreaded GPU Using CUDA and Brook+
139	Optimize or Wait? Using Ilc Fast-Prototyping Tool to Evaluate CUDA Optimizations
140	Hierarchical Agglomerative Clustering Using Graphics Processor with Compute Unified Device Architecture
141	A simple and efficient way to compute depth maps for multi-view videos
142	Real-time parallel remote rendering for mobile devices using graphics processing units
143	Compute Unified Device Architecture Application Suitability
144	Accelerated multi-view stereo using parallel processing capababilities of the GPUS
145	Parallel Zigzag Scanning and Huffman Coding for a GPU-based MPEG-2 Encoder
146	Real-time stereo matching: A cross-based local approach
147	Parallel Lexicographic Names Construction with CUDA
148	Scalability of Higher-Order Discontinuous Galerkin FEM Computations for Solving Electromagnetic Wave Propagation Problems on GPU Clusters
149	A Parallel Algorithm for Dot Product over Word-Size Finite Field Using Floating-Point Arithmetic
150	A Parallel Gibbs Sampling Algorithm for Motif Finding on GPU
151	Support Vector Machines on GPU with Sparse Matrix Format
152	Hybrid OpenCL over high speed networks
153	A Real-Time Soft Shadow Rendering Algorithm by Occluder-Discretization
154	GPU-based acceleration of MPIE/MoM matrix calculation for the analysis of microstrip circuits
155	3-SAT on CUDA: Towards a massively parallel SAT solver
156	Data-parallel algorithms for large-scale real-time simulation of the cellular potts model on graphics processing units
157	Efficient design and implementation of visual computing algorithms on the GPU

158	Performance and Scalability of GPU-Based Convolutional Neural Networks
158	Accurate Measurements and Precise Modeling of Power Dissipation of CUDA Kernels toward
159	Power Optimized High Performance CPU-GPU Computing
160	Program Optimization of Stencil Based Application on the GPU-Accelerated System
161	Hybrid OpenCL: Connecting Different OpenCL Implementations over Network
162	A Comparative Study on ASIC, FPGAs, GPUs and General Purpose Processors in the O(N^2) Gravitational N-body Simulation
163	Simultaneous and fast 3D tracking of multiple faces in video by GPU-based stream processing
164	Stream-Centric Stereo Matching and View Synthesis: A High-Speed Approach on GPUs
165	On accelerating iterative algorithms with CUDA: A case study on Conditional Random Fields training algorithm for biological sequence alignment
166	Streaming Algorithms for Biological Sequence Alignment on GPUs
167	Fast computation of general Fourier Transforms on GPUS
168	Efficient Collision Detection and Physics-Based Deformation for Haptic Simulation with Local Spherical Hash
169	NQueens on CUDA: Optimization Issues
170	Depth-of-Field Blur Effects for First-Person Navigation in Virtual Environments
171	Novel Computing Architectures
172	Memory Saving Discrete Fourier Transform on GPUs
173	The fast evaluation of hidden Markov models on GPU
174	A Multi-GPU Spectrometer System for Real-Time Wide Bandwidth Radio Signal Analysis
175	Exploiting Computational Resources in Distributed Heterogeneous Platforms
176	Accelerate Cache Simulation with Generic GPU
177	High-Speed Implementations of Block Cipher ARIA Using Graphics Processing Units
178	Parallel Dense Gauss-Seidel Algorithm on Many-Core Processors
179	Kernel Fusion: An Effective Method for Better Power Efficiency on Multithreaded GPU
180	The method of improving performace of the GPU-accelerated 2D FDTD simulator
181	GPU Accelerated Adams?Bashforth Multirate Discontinuous Galerkin FEM Simulation of High-Frequency Electromagnetic Fields
182	GPU Computing
183	An Analytical Approach to the Design of Parallel Block Cipher Encryption/Decryption: A CPU/GPU Case Study
184	Distributed computer emulation: Using OpenCL framework
185	Using GPU to Accelerate Cache Simulation
186	Design and Performance Evaluation of Image Processing Algorithms on GPUs
187	FFT Implementation on a Streaming Architecture
188	GPU-Accelerated KLT Tracking with Monte-Carlo-Based Feature Reselection
189	Scalable instruction set simulator for thousand-core architectures running on GPGPUs
190	Accelerating global sequence alignment using CUDA compatible multi-core GPU
191	Power analysis and optimizations for GPU architecture using a power simulator
192	A high-performance fault-tolerant software framework for memory on commodity GPUs
193	Pretty Good Accuracy in Matrix Multiplication with GPUs
i i	

105	
195	Solving k-Nearest Neighbor Problem on Multiple Graphics Processors
196	A Neighborhood Grid Data Structure for Massive 3D Crowd Simulation on GPU
197	Empowering Visual Categorization With the GPU
198	Practical examples of GPU computing optimization principles
199	Larrabee: A Many-Core x86 Architecture for Visual Computing
200	Record Setting Software Implementation of DES Using CUDA
201	Count Sort for GPU Computing
202	Automatic Dynamic Task Distribution between CPU and GPU for Real-Time Systems
203	Scaleable Sparse Matrix-Vector Multiplication with Functional Memory and GPUs
204	Efficient parallelized particle filter design on CUDA
205	Implicit Feature-Based Alignment System for Radiotherapy
206	Real-time Minute Change Detection on GPU for Cellular and Remote Sensor Imaging
207	Data handling inefficiencies between CUDA, 3D rendering, and system memory
208	Analyzing throughput of GPGPUs exploiting within-die core-to-core frequency variation
209	Event-driven gate-level simulation with GP-GPUs
210	Performance of Optical Flow Techniques on Graphics Hardware
211	An Efficient Acceleration of Symmetric Key Cryptography Using General Purpose Graphics Processing Unit
212	A comprehensive analysis and parallelization of an image retrieval algorithm
213	Efficiently Using a CUDA-enabled GPU as Shared Resource
214	Building a Personal High Performance Computer with Heterogeneous Processors
215	Title Page i
216	EASEA parallelization of tree-based Genetic Programming
217	GPU detectors for interference cancellation in chaos-based CDMA communications
218	Preliminary implementation of two parallel programs for fractal image coding on GPUs
219	Practical Pre-stack Kirchhoff Time Migration of Seismic Processing on General Purpose GPU
220	vCUDA: GPU accelerated high performance computing in virtual machines
221	Evolving a CUDA kernel from an nVidia template
222	Many-Core vs. Many-Thread Machines: Stay Away From the Valley
223	Real-time 3D reconstruction and pose estimation for human motion analysis
224	Enhancing Ubiquitous Systems through System Call Mining
225	Linear optimization on modern GPUs
226	Scaling Hierarchical N-body Simulations on GPU Clusters
227	Fast Dynamic Voronoi Treemaps
228	CuPP - A framework for easy CUDA integration
229	TransCAIP: A Live 3D TV System Using a Camera Array and an Integral Photography Display with Interactive Control of Viewing Parameters
230	Physically-based interactive schlieren flow visualization
231	Aspects of GPU for general purpose high performance computing
232	Visualizing complex dynamics in many-core accelerator architectures

233	Efficient characterizations of composite materials electrical properties based on GPU accelerated finite difference method
234	Fast generating of a digital hologram using general-purpose computation on graphics processing units
235	Parallel Approaches for SWAMP Sequence Alignment
236	HPP-Controller: An intra-node controller designed for connecting heterogeneous CPUs
237	Single-Chip Heterogeneous Computing: Does the Future Include Custom Logic, FPGAs, and GPGPUs?
238	Design and implementation of software-managed caches for multicores with local memory
239	A 57mW embedded mixed-mode neuro-fuzzy accelerator for intelligent multi-core processor
240	Compressing Floating-Point Number Stream for Numerical Applications
241	Image processing applications on a low power highly parallel SIMD architecture
242	A code motion technique for accelerating general-purpose computation on the GPU
243	Multi-dimensional characterization of temporal data mining on graphics processors
244	Tutorial 3: Methodologies and Performance Impacts of General Purpose Computing on GPUs
245	Calculation of weight vectors for wideband beamforming using Graphics Processing Units
246	Discrete-event Execution Alternatives on General Purpose Graphical Processing Units (GPGPUs)
247	An efficient GPU implementation of the revised simplex method
248	Potential of General Purpose Graphic Processing Unit for Energy Management System
249	Stream processing of moment invariants for real-time classifiers
250	A GPU-based calculation using the three-dimensional FDTD method for electromagnetic field
251	analysis Exploring scalability of FIR filter realizations on Graphics Processing Units
252	Accelerating Linpack Performance with Mixed Precision Algorithm on CPU+GPGPU
232	Heterogeneous Cluster
253	Improving the performance of PIR Protocol in Outsourced Databases
254	VolQD: direct volume rendering of multi-million atom quantum dot simulations
255	Diagnosis, Tuning, and Redesign for Multicore Performance: A Case Study of the Fast Multipole Method
256	A GPU implementation for two MIMO-OFDM detectors
257	Architecting graphics processors for non-graphics compute acceleration
258	Scalable and Parallel Implementation of a Financial Application on a GPU: With Focus on Out-of-Core Case
259	Inter-block GPU communication via fast barrier synchronization
260	Exploiting SPMD Horizontal Locality to Improve Memory Efficiency
261	Exploring New Architectures in Accelerating CFD for Air Force Applications
262	Evolving GeneChip correlation predictors on parallel graphics hardware
263	An efficient, model-based CPU-GPU heterogeneous FFT library
264	MultiGPU computing using MPI or OpenMP
265	MITHRA: Multiple data independent tasks on a heterogeneous resource architecture
266	Parallel 3D Finite Difference Time Domain Simulations on Graphics Processors with Cuda
267	Design optimization of automotive electronic control unit using the analysis of common-mode current by fast electromagnetic field solver
268	Formal Description and Optimization Based High - Performance Computing on CUDA

269	Model-T: Rethinking the OS for terabit speeds
270	Financial Derivatives Modeling Using GPU's
271	Parallelizing Motion JPEG 2000 with CUDA
272	Illustrative Volume Visualization Using GPU-Based Particle Systems
273	Multimodal collaboration and human-computer interaction
274	A characterization and analysis of PTX kernels
275	Towards smart-pixel-based implementation of wideband active sonar echolocation system for multi-target detection
276	Flexible Pixel Compositor for Plug-and-Play Multi-Projector Displays
277	High Performance Computing via a GPU
278	Software-Based Algorithm for Modeling and Correction of Gradient Nonlinearity Distortions in Magnetic Resonance Imaging
279	Front matter(Image and Vision Computing New Zealand, 2009. IVCNZ '09. 24th International Conference)
280	Canny edge detection on NVIDIA CUDA
281	Where is the data? Why you cannot debate CPU vs. GPU performance without the answer
282	Fast and Efficient Dense Variational Stereo on GPU
283	Accelerating H.264 inter prediction in a GPU by using CUDA
284	To GPU synchronize or not GPU synchronize?
285	Using parallel GPU architecture for simulation of planar I/F networks
286	Scalable Software Defined FM-radio receiver running on desktop computers
287	Improving numerical reproducibility and stability in large-scale numerical simulations on GPUs
288	Physically-Based Interactive Flow Visualization Based on Schlieren and Interferometry Experimental Techniques
289	GPU accelerated fast FEM deformation simulation
290	Introduction to GPU Computing and CUDA Programming: A Case Study on FDTD [EM Programmer's Notebook]
291	In the News(Intelligent Systems, IEEE, 24 Jul 2009, Ingebretsen, M)
292	Application of the OpenCL API for Implementation of the NIPALS Algorithm for Principal Component Analysis of Large Data Sets
293	Challenges of mapping financial analytics to many-core architecture
294	Graphic processors to speed-up simulations for the design of high performance solar receptors
295	GPU-Based Background Illumination Correction for Blue Screen Matting
296	Scalable software defined receivers running on desktop computers using General Purpose Graphics Processing Units
297	Approximate Dynamic Programming and Neural Networks on Game Hardware
298	A real time Breast Microwave Radar imaging reconstruction technique using simt based interpolation
299	OpenMPC: Extended OpenMP Programming and Tuning for GPUs
300	Accelerating Multi-Sensor Image Fusion Using Graphics Hardware
301	Performance Analysis of a New Real-Time Elastographic Time Constant Estimator
302	High-Speed Private Information Retrieval Computation on GPU
303	Frame-based parallelization of MPEG-4 on compute unified device architecture (CUDA)

304	GPU-S2S: A Compiler for Source-to-Source Translation on GPU
305	Research on ATI-CAL for accelerating FBP reconstruction