

New approaches for advanced computing and data processing in climate modeling

Marek Błażewicz, <u>Milosz Ciżnicki</u>, Krzysztof Kurowski, Michał Kulczewski, Piotr Kopta

Applications Department



PSNC overview



Center for Development of e-Infrastructure

- Research Metropolitan Area Network POZMAN
- HPC Center
- National Research and Education Network PIONIER
- Digital Libraries Federation

Center for R & D

- New Generation Networks
- Clouds & Grids
- Digital Libraries and Portals
- Technology, Applications and Services
- Cyber Security



PIONIER – Polish Optical Network



- Area
 312k sq km
- Population 38M
- Main academic centers 22
- State universities 165+
- Students
 2M+
- R&D institutions and Univ. interconnected via PIONIER network 700+

6494 km of fiber infrastructure in Poland

763 km of fiber in Germany (IRU)

7257 km of fiber in total

22 MANs and 5 HPC Centers (Gdańsk included) in PIONIER Consortium



European optical networks





International optical networks





Networking + Storage + Computing





Key services for e-Science



POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER Application Department @ PSNC

- 28 people + 14 students
- Advanced applications and computing, including accelerated computing, masive data analysis & high-resolution advanced visualization



• R&D acitivities in areas:

- Computational science
- Energy & environment science
- Medicine
- Material science
- Biology
- Chemistry
- Numerical weather prediction





HPC & key services



HPC Center

•GEANT 2/3 (10 x 10Gb/s)

- National Research and Education Network PIONIER
- •10 Gb/s dedicated optical fiber connection to USA
- High speed •HD videoconference services



uster (Tier-1 resource)
224.3 Tflops
5448 x AMD Opteron
6234
334 x Nvidia Tesla
M2050
10.64 TB
55.42 TB

Chimera - SMP						
Performance	21.8 Tflops					
CPU	2048 x Intel Xeon E7-8837					
RAM	16 TB					
HDD	1.2 TB					

Reef - cluster					
Performance	54.27 Tflops				
CPU	6176				
RAM	10.208 TB				
HDD	163.668 TB				

 •Total peak performance: 304 TFlops (increasing)
 •European Grid Initiative / National Grid Initiative (PL-GRID)

• European HPC infrastructure (PRACE)

HPC Center • Applications on demand

- National Data Storage
- Disk storage 4.6 PB (increasing)
- Archiving and backup on demand

Data Center

network

POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER (POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER (POZNAŃ Advanced visualization (up to 16K)

Vitrall

- Distributed web-based visualization system
- Multi-GPUs support
- Remote (web browser) or local (attached screens) rendering
- Support for sensor fusion and multi-touch based user interfaces





Advanced computing



New approaches for advanced computing





CUDA Research Center

- PSNC is CUDA research center from 2011
- CaKernel exa-scale GPGPU application
- CAVE visualization





MultiGPU Applications

Motivation:

- Common PC are often equipped with two or more GPUs
- HPC centers offer access to supercomputers with a large number of GPUs

Communication issues:

- High parallelism in multi-GPU systems requires effective dealing with communication on many levels, e.g. between threads, blocks of threads, GPUs etc.
- Different memory types with its properties (bandwidth, latency) should be taken into account when planning communication
- Data distribution patterns have different pros and cons



Computational Fluid Dynamics (1)

Algorithm:

- Two and three-dimensional CFD simulation with a free surface approximation (VOF method);
- Navier-Stokes equations discretized using finite-difference method;
- Fluid propagation over an Eulerian grid;



Volume of Fluid (VOF) Method for the Dynamics of Free Boundaries, C.W. Hirt, B.D. Nichols, 1979.

POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER CONTACT CONDUCTION CONTINUE AND NETWORKING CENTER (2)

Results:

- MPI-based, Multi-GPU heterogeneous application;
- 80x speedup comparing to sequential CPU implementation;
- Good scaling on 32 GPUs.





Motion Tracking

Motivation

• A need to track object efficiently in high resolution movies in real-time

Approach:

- Good points to track are selected by Harris corner detector
- Lucas-Kanade method is used to track points
- All steps are performed on GPU

Results:

Real-time motion tracking in movies up to 4K





JPEG 2000

Motivation

- New image compression algorithm
- Better quality than JPEG
- Digital cinema standard
- Lossy hyperspectral image compressic
- Benchmarking

Approach

- Parallel implementation on GPU
- The fastest available implementation







3D Tomography images (1)

Motivation:

- Existing implementations are capable of processing small domains only
- There's no application dealing with real-life models efficiently

PSNC's software:

- Merging 2D slices of CT/MRI images into a 3D model
- Application: medical tomography (body scans), petroleum engineering (rock scans)
- Algorithm: Marching Tetrahedra creating isosurface from neighbouring pixels representing the same density



POZNAŃ SUPERCOMPUTING AND NETWORKING CENTER (3D Tomography images (2)

Results:

- ~ 50-fold speedup over single threaded CPU version
- Multi-GPU support scaling up well
- isorock command line version designed to create rock structure models (computes open and closed porosity)







Quantum chemistry

- Code rewrite Fortran -> C
- Use of parallel numerical libraries (BLAS/LaPACK)
- Memory usage minimization
- Hybrid parallel implementation (for large number of cores optimization)
- Scalability up to 500 cores (limitations of the algorithm)
- Up to 6x speedup
- PRACE grant given



	hs r001hs r002hs 📧					
Ξ.	lotspots - Hotspots 🥕 🧿					Intel VTune Amplifier XE 20
4	🕹 Analysis Target 📄 Analysis Type 🖬 Summar	y 🙈 Bottom-up 😽 Top-down	Tree			
Group	ping: Function / Call Stack.				•	Thread create stack.
	Function / Call Stack	CPU Time	* Module	Function (Full)		1 stack(s) selected. Viewing <1 of 1
>mkl	blas_xsdot	4.208s	lbmkl_mc3.so	mkl_blas_xsdot		Current stack is 100.0% of selection
>[libis	iomp5.so]	0.539\$	libiomp5.so	[libiomp5.so]		100.0% (5.620s of 5.620s)
≻uti_	_random_fmatrix	0.440s	kit	util_random_fmatrix		[Uninews]
>cov	_matrix_omp	0.200s	kit	cov_matrix_omp		[Granowij
≻mid,	_serv_cpu_detect	0.063s	libmid_core.so	mld_serv_cpu_detect		
>[Imp	port thunk rand_r]	0.040s	kit	[Import thunk rand_r]		
>mair	in	0.030s	kit	main		
>[[m	port thunkkmp_execute_tasks]	0.030s	libiomp5.so	[Import thunkimp_execute_tasks]		
>[Imp	port thunkkmp_yield]	0.020s	lbiomp5.so	[Import thunkkmp_yield]		
>mkl,	L_serv_load_lapack_dll	0.020s	libmid_core.so	mkl_serv_load_lapack_dll		
≥mid_	Lserv_mkl_free	0.0195	libmid_core.so	mid_serv_mid_free		
>[Imp	port thunkkmp_x86_pause]	0.010s	libiomp5.so	[Import thunkkmp_x86_pause]		
≯mid,	_blas_diyr2_nb	0.000s	libmid_mc3.so	mid_blas_dsyr2_nb		
	Sele					
_	Selo	cted 0 row(s):				
-	Selo	0.2c 0.25c 0.3c 0.35	0.44 0.454	052 055 062 0	45e 0.7e 0.7	Threads
(T=L	Sele	cted 0 row(s):	; 0.4s 0.45s	0.5s 0.55s 0.6s 0	.65s 0.7s 0.7	75s 0.0s + V Threads
(市)	Selo	o.2s 0.25s 0.3s 0.35	: 0.4s 0.45s	0.55 0.555 0.65 0	65s 0.7s 0.7	755 0.05 + V Threads
0 1	Selo 2°2+2-2* read (0x5590) MP Worksr Thr	o.2s 0.25s 0.3s 0.35	s 0.4s 0.45s	0.55 0.555 0.65 0	655 0.75 0.7	755 0.05 + ♥ Threads ♥ ■ Running ♥ ▲▲ CPU Thre
0.01	Selo	cted 0 row(s):	s 0.4s 0.45s	0.55 0.555 0.65 0	655 0.75 0.7	T55 0.05 + ♥ Threads ♥ ■ Running ♥ ↓ CPU Time ♥ ♥ ♥ Open/MP Re ♥ ♥ ♥ Open/MP Re
1000	Sele 	cted 0 row(5)	s 0.4s 0.45s	0.55 0.555 0.66 0	65 0.7 0.7	TES 0.05 1 + ♥ Threads ♥ Automing ♥ Automing ♥ CPU Usage ♥ CPU Usage ♥ CPU Usage
0 8 8 8 3 3	Sele 	cted 0 row(s):	s 0.4s 0.45s	0.55 0.555 0.66 0	655 0.75 0.7	75 0.65 + ♥ Threads ♥ ■ Rurning ♥ ■ CPUTme ♥ CPUTwage ₩at CPUTme
* 81818181∄ ∧	Seb CCCC+C+C+ wead (0x550) MP Worker Thr MP Worker Thr MP Worker Thr	cted 0 rcw(s)	2 0.42 0.455	0.55 0.55 0.65 0	.655 0.75 0.7	Participation of the second s
nexts 일 입 입 입 입 입 걸	Sele	cted 0 row(5)	5 0.45 0.455		665 0.75 0.7	55 0.05 W ■ Rurring W ■ CPUTme W CPUThese CPUThese W CPUThese W CPUThese
Threads 응 [응 응 응 응 공]	Concept 0.65 0.15 0.15 Present (0.5559) 0.65 0.15 0.15 Present (0.5599) 0.55 0.15 0.15 Present (0.5599) 0.15 0.15 0.15 Present (cted 0 row(5)	2 0.45 0.45s		655 0.75 0.7	Constant
Threads 8 8 8 8 8 8 8 3 ∄ <	Selection of the select	0,25 0,25 0,35 0,35	5 0. 4 5 0.455	0.5s 0.55s 0.0c 0		m 0 0 m 4 V Threads V Rurring V CPU Usage A CPU The
Threads 입 입 입 입 입 입 입 입 입 권		ctel 0 row(3)	0.45 0.455	0.55 0.755 0.45 0	655 0.75 0.7	15. 0.0. The mass of the set of
Threads 의 의 의 의 의 의 의 의 의 의 크)	Control (Control (Contro) (Contro) (Contro) (Contro) (Contro) (Contro) (Contro) (Contro)	0.25 0.25 0.35 0.75	5 0.45 0.455	0.55 0.755 0.65 0	0.75 0.7	main (1) and
Threads 회 81 81 81 81 81 81 81 81 81 31		0.25 0.25 0.25 0.35	0.45 0.455	0.55 0.575 0.58 0	655 0.75 0.7	10 0.00 + ♥ Threads ♥ threads ♥ Grutes ♥ G
	Set 0.05 0.15 0.15 Version 0.05 0.15 0.15 Version 0.05 0.15 0.15 Version 0.05 0.15 0.05 Version 0.05 0.05 0.15 Version 0.05 0.05 0.05 Version 0.05	0.25 0.25 0.35 0.35	: 0.45 0.45s	0.55 0.556 0.46 0		S 0.5 + ♥ Threads D
Treess 1 (3 (3 (3 (3 (3 (3 (3 (3 (3 (3 (3 (3 (3	State	0.25 0.25 0.35 0.35	0.45 0.455	0.55 0.575 0.56 0	665 0.75 0.7	15 0.5 1 + ♥ Theods ♥ Avress ♥ Av
Treasc 888888888888888888888888888888888888		odo ove(o) ods ozes ods oze	0.45 0.45		, , , , , , , , , , , , , , , , , , ,	(C) Theorem (C) Theorem (C) Constraints (C) Constraints (C
10000000000000000000000000000000000000	C C C C C C C C C C C C C C C C C C C	ααδ στοφότι δ.2. δ.24 δ.3 δ.3 δ.2 δ.25 δ.3	0.49 0.45			15 0.5 + 2 Hereads D Correspondences D Correspon



CaKernel/Cactus/Kranc

Cactus

- Framework for developing portable, modular applications solving 3D PDE time evolutions
- HPC
- Hardware architecture independence
- Programming language independence

CaKernel

- Integrated with Cactus, fully compatible with legacy applications
- Execution of routines on GPU and/or CPU without modifications to applications
- Keeps track of recent data changes and copies minimum amount required for consisted computations

Kranc

- Mathematica application
- Translates tensorial description of a time dependent partial differential equation into a Cactus module
- Very efficient time-to-solution factor
- Mathematica translates and simplifies equations for their more efficient compilation/execution



Chemora = CaKernel +Cactus + Kranc

Results:

- Define problem using a **domainspecific language**: EDL (Equation Description Language)
- Compile this specification into a C/C+ +/CaKernel source code module Cactus thorn) – Kranc
- Launching the same kernel definition in different environment (CUDA, OpenCL, CPU, OpenMP) by just changing templates - **CaKernel**
- Cactus code compiled and run on laptop/ workstation/ supercomputer /heterogeneous environment
- Very short time to solution



- Simple CFD problem
- Chemora vs. Physis
- GPU for computations
- CPU for data transfer
- Single precision





Eulag activities





Airflow and heat transfer in data center (2)

- Realistic wind flow
- Heat transfer
- Temperature
- Power consumption optimization
- Racks placement optimization





Poznan urban flow



- Simulation of ideal and realistic wind flow
- Contamination & dispersion of hazardous substances





Future activities





Backup slides



CaKernel Optimizations

- Dynamic tile size selection (JIT compilation)
 - The stencil size
 - The number of grid variables
 - The shape of the local grid
- Lightweight kernel generation
 - Program parameters and tile shape seen as constants by compiler
 - Fixed load offset
- Fat kernel detection
- Profiling information
 - NVIDIA Cupta API



Thank you

Poznań Supercomputing and Networking Center affiliated to the Institute of Bioorganic Chemistry of the Polish Academy of Sciences, ul. Noskowskiego 12/14, 61-704 Poznań, POLAND, Office: phone center: (+48 61) 858-20-00, fax: (+48 61) 852-59-54, e-mail: office@man.poznan.pl, http://www.man.poznan.pl