Hardware Support for GPU Computing from IT

Fons Rademakers, CERN openlab Chief Research Officer
GPU Computing Event, CERN, 12-1-2016

Intel Modern Code Developer Challenge

The Challenge - Speedup Brain Development Simulation Code

- Original code is 14000 lines of Java
- Recoded in C++
- CERN openlab provided a summer student to start this task
- Intel provided tools and hardware
- A 500 line kernel from this program was used for the Challenge
- This kernel took 45 hours to run with the target set of parameters

The Prizes

- 1 Grand Prize: CERN openlab fellowship
- 3 First Prizes: visit to CERN
- 3 Second Prizes: visit to SC’16

Contestant Engagement

- 17000 students reached
- 2077 students registered for the challenge
  - 130 universities
  - 19 countries
- Over 1200 code downloads
- 1000 students accessed free training

Grand Prize Winner

Mathieu Gravey

Alès School of Engineering

France

- Original code C/C++ running on single core single thread Xeon Phi 7120A
- Final optimized code runs on Xeon Phi 7120A taking advantage of all cores and threads

From 45 hours to…..

8 minutes 24 seconds

320x Increase

Mathieu’s Optimisations

- Change from AoS to SoA to allow vectorisation and improved cache layout
- Custom memory allocator, reuse memory for many small memory allocations
- Use OpenMP for parallelisation over all Xeon-Phi cores
- Use icc scatter/gather intrinsics

Code Modernisation Can Payoff Big Time

Idea: Create CERN Modern Code Developer Challenge

- Find critical pieces of code in CERN programs
- Put them up for the acceleration challenge
- Keep running scores of fastest times to create competition
- Allow students to refine their submissions till end of challenge
- Thinks of some nice prizes
- Also a perfect requirement tool ;-)

IT GPU and Co-Processor Hardware Support - TechLab

TechLab

- TechLab is an IT project providing a test environment aimed at improving the efficiency and performance that can be obtained from modern hardware
- TechLab systems are operated, as evolving test systems, on a best effort basis
  - Not to be used for production work
  - Host are reinstalled on a regular basis and user data is deleted
  - Machines can be booked for a limited amount of time
- For more on TechLab see: https://twiki.cern.ch/twiki/bin/viewauth/IT/TechLab

Current TechLab Systems

Hardware type	Specs summary	HEP-SPEC06	OS & Kernel	Expected availability
iWARP 10Gb	60 nodes with 10 Gb iWARP		SLC 6.4 2.6.32-358.el6	Now
Quad Socket	4 nodes SandyBridge and 4 nodes Westmere-EX	499.38	SLC 6.5	Now
Intel Xeon Phi	4 nodes, each with dual socket 8 cores SandyBridge + Xeon Phi 7120P		SLC 6.5 + Intel MPSS 3.1.2	Now
Nvidia K20X GPU	4 nodes, each with dual socket 8 cores SandyBridge		SLC 6.5 + CUDA 5.5 Prod Release	Now
AMD GPU	1 node, dual socket 8 cores SandyBridge + AMD GPU			Now
Intel Atom S1260	45 cartridges	10.43	SLC6 2.6.32-358.18.1.el6.x86_64 gcc 4.4.7 20120313	On demand
Intel Atom C2000 "Avoton"	45 cartridges	53.40	SLC6 2.6.32-358.23.2.el6.x86_64 gcc 4.4.7 20120313	Now
ARM A9 Calxeda SOCs	4 independent ARM A9-based SoCs cluster	13.75	Fedora 20 (3.11.10-301.fc20.armv7hl)	Now
ARM 64	X-Gene 1, 8 cores @ 2.4 GHz, 64 GB of RAM	56.52	Ubuntu 14.04, kernel 3.13, gcc 4.9.2	Now
Maxeler Data Flow Engine	Galava PCI-e DFE card		Maxeler OS	Now
OpenPOWER8	IBM Turismo, 4 physical cores (32 logical) @ 3 GHz, 64 GB of RAM	112.29	Fedora 21, kernel 3.18.7-200.fc21.ppc64le, gcc 4.9.2	Now

Conclusions

- Code modernisation, vectorisation and parallelisation, can result in big performance increases
- TechLab has a number of GPU and co-processor systems available
- We would like to extend TechLab and make it a real supported service
- Need user feedback and requests for more support and hardware
- Please use it