Theme: Towards Reconfigurable HPC
Lecture **7**

Reconfigurable HPC I – Introduction

Iris Christadler

Leibniz Supercomputing Centre

Inverted CERN School of Computing, 3-5 March 2008

Introduction

- **Objectives:**
  - Define reconfigurable high-performance computing (RHPC)
  - Give an overview of RHPC platforms
  - Give an overview of accelerated platforms
  - Explain drawbacks of RHPC

Definition

- **Reconfigurable Computing (RC)**
  - Idea of reconfiguring a computer to your current needs
  - Use FPGAs for the reconfiguration
- **Concept exists since 1960s (Paper by Gerald Estrin)**
  "Unfortunately this idea was far ahead of its time in needed electronic technology."
- **Renaissance in the 80s/90s**
  "The world's first commercial reconfigurable computer, the Algotronix CHS2X4, was completed in 1991. It was not a commercial success."
- **Reconfigurable HPC (RHPC)**
  "Currently there are a number of vendors with commercially available reconfigurable computers aimed at the high performance computing market."
  Quotes are taken from [http://en.wikipedia.org/wiki/Reconfigurable_computing]

FPGAs

- **Field-programmable gate array**
- **"Adjust the architecture to the needs of your algorithm"**
- **Invented 1984**
- **Used heavily in embedded and real-time systems**
- **Used in supercomputers like Cray XD1, SGI RASC Blades**
- **Programmability!**
- An overview can be found at:
  [http://en.wikipedia.org/wiki/Field-programmable_gate_array]

Where is RC important today?

- **Cryptography**
- **Digital Signal Processing**
- **Medical Imaging**
- **… and other highly specialized embedded systems**

➢ **All of them have special requirements that make RC a necessity**

➢ **These domains invest in hand-coded VHDL**

➢ **One can learn from them, but RC must become easier if it should be widely used**

Why HPC is …

BECOMING RECONFIGURABLE

Why HPC is not yet reconfigurable

Very interesting idea, but

- **“As long as there is no Fortran compiler…”**
- **FPGA programming is too cumbersome**
- **SDKs don’t meet the needs of HPC Programmers**
- **FPGAs were not big enough for scientific kernels**

Why HPC might become reconfigurable

- **“The free lunch is over”**
- **Stability**
  Thousands of cores are no longer manageable
- **Energy Consumption**
  - Power
  - Cooling
- **Footprint**
- **Performance gain is promising**

SC07

"**GPGPUs and FPGAs are now fully implanted in our brains**
The server and chip industries have felt the rise of the accelerator coming for some time. Last year's conference, for example, had the hardware heads pitted against the coders. The software set wondered if enough developers would ever exist to write custom code for tricky FPGAs and GPGPUs. The hardware crowd countered such skepticism by forcing pitches about their super silicon down the throats of anyone who showed even a faint sign of interest in the technology.
Much of the reticence around the server accelerators vanished at Supercomputing '07. Yes, the software folks still have their concerns. And, yes, grizzled veterans complain that they've seen accelerator fads come and go. Ultimately, however, it's clear to us that enough big vendors, ISVs, start-ups and customers have coalesced around the accelerator idea to push the technology forward in a profound way."

[http://www.theregister.co.uk/2007/11/20/accelerators_fpga_gpu_sc07/]

THE GOOD, THE BAD AND THE UGLY
(NOT NECESSARILY IN THIS ORDER ☺)

Reconfigurable High Performance Computing

EXAMPLES OF RHPC SYSTEMS

Cray XD1

- **Per Chassis:**
  - 12 AMD Opterons (single or dual-core)
  - 6 Xilinx Virtex II Pro (Virtex-4) FPGAs
  - 3.2 GB/s interconnect
- **Released Oct. 2004**
- **RapidArray Interconnect**
- **Development kit:**
  - Specialized libraries
  - Xilinx Tools (HDL)
  - Mitrion-C
- **"Affordable power"** (advertisement claim)

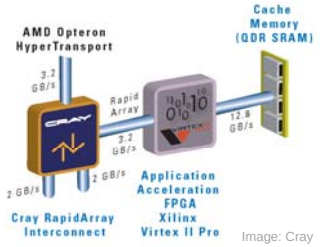


Image: Cray

Cray XD1 Installations

- **FZJ, Juelich, Germany**
  - Not as much development as planned
  - Mainly use of codes from other institutes instead
  - "FPGA programming needs electrical engineers"
  - Pricing models of many FPGA development tools are not suitable for research institutes
- **The George Washington University, USA**
  - RC-Tutorial at Supercomputing Conference
  - Are very happy with their machine(s)
  - Part of CHREC
- **Many more sites:**
  CINECA (Italy), MHPCC (Maui), ASA (Alabama), Zuse Institute Berlin (Germany), …

SGI RASC Blades

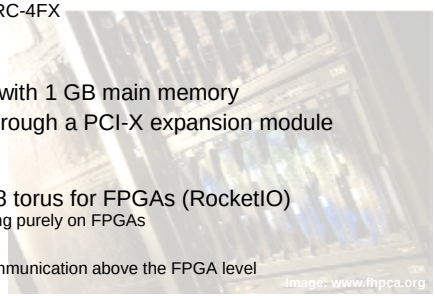
- **Reconfigurable Application-Specific Computing (RASC)**
  - Allows hybrid programming
  - Numalink Interconnect
- **RC100 Blade:**
  - Accelerates SGI Altix Itanium series
  - Dual Xilinx Virtex 4 LX200 FPGA
- **RC200 Blade:**
  - Available for SGI Altix Xeon series
  - Multiple Altera Stratix III FPGAs
- **Development kit:**
  - RASCLib, RASC API
  - Mitrion-C, Handel-C, Xilinx Synthesis Technology

SGI RASC Installations

- **SGI built the world's largest FPGA supercomputer:**
  (Nov. 2007)
  They ran a broadly used bioinformatics application more than 900 times faster than the same application would run on a traditional cluster. They used the Mitrion Accelerated BLAST-n "off-the-shelf" code.
  SGI's reconfigurable supercomputer featured 70 FPGAs, more than any single system built to date. SGI's FPGA supercomputer accelerated the performance of a complex BLAST-n query by more than 900 times, completing in less than 33 minutes what took a 68-node Opteron-based cluster approximately three weeks to finish.
  [http://www.sgi.com/company_info/newsroom/press_releases/2007/november/fpga.html]
- **Many more:**
  The George Washington University, ZIH Dresden (Germany), …

Maxwell

- **FPGA supercomputer**
  - 32 IBM Intel Xeon Blades
  - 64 Xilinx Virtex-4 FPGAs mounted in two card types
    - Nallatech H101
    - Alpha Data ADM-XRC-4FX
- **Each Blade**
  - 2.8 GHz Intel Xeon with 1 GB main memory
  - hosts two FPGAs through a PCI-X expansion module
- **Two networks**
  - two-dimensional 8x8 torus for FPGAs (RocketIO)
    allows parallel programming purely on FPGAs
  - GigE for the Xeons
    supports inter-process communication above the FPGA level



Maxwell: Porting 3 Demo Codes

- **Financial Engineering**
  - Monte Carlo simulation of stock option pricing
  - Classic Black-Scholes model
  - runs 320 times faster
- **Medical Imaging**
  - 3 and 4D facial image reconstruction codes
  - batch process video images on FPGAs
  - runs 2.5 times faster (sustained)
- **Oil & Gas**
  - 3D controlled source electromagnetic (CSEM) code
  - pretty typical physic simulation code
  - runs 5.5 times faster (but scales only up to 8 nodes)

Maxwell: Findings

- **FPGAs can be used as main processors**
- **Porting will yield good results for HPC Applications with a compact & well-defined kernel**
- **Complexity & compilation overhead makes development slow**
- **"Cost of an effective port is still too high"**
- **Better tools are necessary**

Breaking Petaflops…

EXAMPLES OF ACCELERATED SYSTEMS

TSUBAME (ClearSpeed)

- **Tokyo Institute of Technology, Japan**
  fastest computer in Japan
- **single accelerated system in the top500**
  currently ranked 16, highest rank was 9 (11/2006)
- **TSUBAME Grid Cluster Sun Fire x4600 Cluster**
  - 11664 Opteron cores (2.4/2.6 GHz)
  - 360 ClearSpeed Advance Accelerator Boards
    - 96 GF theoretical peak per board
    - 50 GF sustained double-prec. DGEMM
  - Peak Performance: 102 TF
  - Linpack Performance: 56 TF

ClearSpeed: Technical Details

- **Offer two versions:**
  - Advance X620 (PCI version)
  - Advance e620 (PCIe version)
  - Two CSX600 chips on a board
- **CSX600 chip:**
  - IEEE 754 standard conform
  - Clocked at 250 MHz
  - 1GB DDR2 SDRAM (500 MHz)
  - 96 compute engines:
    - Each has a **64 bit floating point** adder and multiplier
    - Each has 6KB of high-speed local storage

ClearSpeed

- **Advertisement claims:**
  - "World's highest performance processor" (96 GF per board)
  - "World's highest performance per watt" (<25 W/Board, 2 GF/Watt)
- **Programming environment:**
  - Compiler with extension (Cn)
  - Assembler, debugger, execution profiler
  - Pre-written performance libraries: Level 3 BLAS, LAPACK, FFT
  - Ported versions of some codes: Amber, Molpro, …

RIKEN MD-GRAPE 3

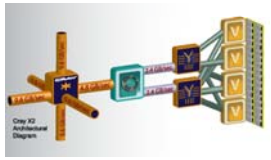
- **RIKEN Yokohama Institute, Japan**
- **Special-purpose system** for molecular dynamics simulation (protein structure prediction)
- **Consists of:**
  - 201 unit of 24 custom MDGrape-3 chips (4808 total)
  - plus several Xeon servers as hosts
- **Performance:**
  - 1 Petaflop (June 2006)
  - but: not capable of running Linpack

IBMs Roadrunner (Cell)

- **First petaflop system worldwide** Peak Performance will be 1.33 PF
- **A hybrid Opteron-Cell system** Cells used as accelerates
- **Cluster of:**
  - 6,912 AMD dual-core Opterons (1.8 GHz)
  - 12,960 IBM Cell eDP accelerators (~ 100 GF peak each)
- **Power Consumption:**
  - 3.9 MW Power
  - 0.35 GF/Watt (peak)
- **due in Q3/2008**

Cray XT5$_h$

- **"A milestone on the path to Adaptive Supercomputing"**
- **An integrated hybrid supercomputer**
  - XT5 Blade: scalar processing (Opterons)
    - Dual- or quad-core Opterons
  - X2 Blade: vector processing
    - more than 100 GF peak performance
    - support Unified Parallel C **(UPC)**
    - support Co-Array Fortran **(CAF)**
  - XR1 Blade: FPGA acceleration
    - 1 Opteron
    - 2 DRC Computer's RPUs
    - HyperTransport



Images: Cray

Cray XT5$_h$ Installation: HECToR

- High-End Computing Terascale Resources (HECToR)
- Edinburgh Parallel Computing Centre (EPCC)
- XT4 system is already installed
- X2 vector “Black Widow” system will be delivered in September 2008
- Upgrade to XT5$_h$ in 2009

Shortcomings of RHPC

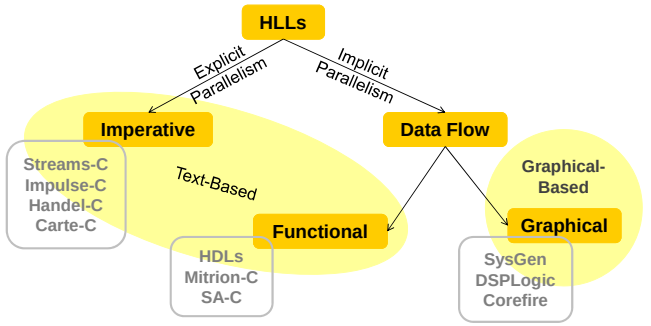
DEVELOPMENT KITS FOR RHPC

Development Kits for RHPC

- Many (different!) SDKs are available
- They should facilitate FPGA software development
- Software development for FPGAs is still cumbersome

➢ We do need new ideas for programming FPGAs!

Classification of HLLs



taken from
"Productivity of High-Level Languages on Reconfigurable Computers: An HPC Perspective" (see further readings at the end of the talk)

FPGA COMMUNITIES

FPGA Communities

- **OpenFPGA**
  [www.openfpga.org]
  - consortium to foster reconfigurable computing
  - established 2004
- **FPGA High Performance Computing Alliance (FHPCA)**
  [www.fhpca.org]
  - established 2004
- **Center for High Performance Reconfigurable Computing (CHREC)**
  [www.chrec.org]
  - U.S. center for research in reconfigurable computing
  - operational since January 2007

Conclusion

- **At the moment, RHPC is only used for very special applications, and very special computers (breaking Cryptography algorithms, more military use?)**
- **It is way too difficult to port an existing 1000+ lines of Fortran/MPI code to FPGAs**
- **Tools and DK for FPGAs exist but are still hard to use**
- **HPC people started to look at other accelerators**
- **Many codes are not well suited for acceleration**
- **The performance gain, the energy savings and the smaller footprint are compelling**
- **We (Manfred & I, anyone else?) believe that the lessons learned with different accelerators (and their programming paradigms) will be useful and lead, in the end, to the widely use of FPGAs in HPC.**

Future Work

- **Which codes are able to scale to 10000+ processors?**
- **Which codes can benefit from special accelerators?**
- **Who determines if a code is scalable?**
  - The research area,
  - the problem itself,
  - the chosen algorithm or
  - the written code ?
- **What are we going to do with codes that are already hitting scalability limits?**
- **How can we convince people to use better programming languages?**

Further Reading

- **Productivity of High-Level Languages on Reconfigurable Computers: An HPC Perspective**
  [IEEE International Conf. on Field-Programmable Technology]
- **RAT: A Methodology for Predicting Performance in Application Design Migration to FPGAs**
  [http://www.ncsa.uiuc.edu/Conferences/HPRCTA07/papers/Brian_Holland.pdf]
- **A Preliminary Investigation of a Neocortex Model Implementation on the Cray XD1**
  [http://sc07.supercomputing.org/schedule/pdf/pap321.pdf]
- **Parting Shots at 2007**
  [http://www.hpcwire.com/hpc/1967844.html]

"SOFTWARE STANDARDS, ANYONE?" [ABOUT FPGAS]