Why FPGAs will win the Accelerator Battle: Building Computers that Minimize Data Movement

Allan Cantle – President & Founder

www.nallatech.com
Overview

» Commercial Realities For HPC
» A View From Berkeley
» Two Fundamental Computing Architectures
  » Processor Centric – Data is brought to the Processor
  » Data Centric – Processors are built around the Data
» Mixing Processor/Data Centric Architectures
» Data Centric Computing Architectures
» Integrating Processor & Data Centric Architectures
» Summary
Commercial Realities for HPC

- Code base readily Available and portable between Architecture generations
- Code base NOT readily Available and NOT portable between Architecture generations

$M Market Size*
0.1 1 10 100 1,000 10,000

- GPGPUs
- Cell
- Clearspeed
- x86
- GPUs (as graphics accelerator)
- FPGAs
- = Single Source Vendor
- = Multi Source Vendors

*Market size based on processor chip revenues only
A processor that is capable of morphing itself into the most efficient implementation for any given computing problem.
Commercial Realities - Technology Battleground

FPGA Technologies building software on an inherently parallel architecture

Incumbent Technologies building parallel architectures that can still support inherently serial software

Growing FPGA Families
- Distributed Memory
- Multipliers/DSPs
- Gb Serial I/O

Reconfigurable Processors
- Cell/Multi-Core

Application Data Types
- Symbolic
- Vector/Streaming
- Bit Level

SWEEP Efficiency
- Size, Weight, Energy, Performance, Time
» HPC’s traditional markets
   » Scientific Modelling – Largest in CFD / FEA
   » Typically Symmetrical problems
   » Non Real Time (Non Deterministic)

» Growing HPC Markets
   » Triple Play – Voice, Video & Data
   » Internet
   » Centralisation of Application Servers – Back to the Mainframe
   » Virtualisation
   » Deterministic & Real Time – e.g. VOIP

» Some of the new HPC challenges have already been solved in HPEC world
A View from Berkeley

The Landscape of Parallel Computing Research: A View from Berkeley

Krsti Ananovic
Ras Bodik
Bryan Christopher Catanzaro
Joseph James Gebis
Parr Husbands
Kurt Keutzer
David A. Patterson
William Lester Plisker
John Shalf
Samuel Webb Williams
Katherine A. Yelick

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2006-183
http://www.eecs.berkeley.edu/Pubs/TechRpts/2006/EECS-2006-183.html

December 18, 2006
Two Fundamental Computing Models

» Processor Centric – HPC World
  » Von Neumann & Harvard Architectures
  » Invented when the processor was the bottleneck
  » Hub and Spoke approach

» Data Centric – HPEC World
  » System Level Data Centric
  » Chip Level Data Centric
  » Distributed processing approach
  » Minimise Data Movement
Today’s Processor Centric Architecture (Intel)

Quad Core Xeon Processor
Quad Core Xeon Processor
Quad Core Xeon Processor
Quad Core Xeon Processor

SDRAM Memory DIMMs

MCH / Northbridge Chipset

21 GB/s
2+2 GB/s

PCIe x8

Northbridge Chipset

2+2 GB/s

PC

OS

Storage I/O
USB I/O
1394 I/O
Graphics I/O
Wireless I/O
Network I/O
Accelerators

Dependent & Non Deterministic

Leads to High Latency & hence the need for fully decoupled accelerators & I/O

8.5 GB/s
8.5 GB/s
8.5 GB/s
8.5 GB/s
21 GB/s
2+2 GB/s
2+2 GB/s
2+2 GB/s
2+2 GB/s

Commercial In Confidence. Copyright ©2007, Nallatech.
Mixing Processor/Data Centric Architectures

Are Data Centric Architectures the future of Computing?!

Application Data Types

Symbolic
Vector/Streaming
Bit Level

Parallelised
Bit level processors
Parallelised
Vector/Streaming processors
Parellised
Symbolic processors

Bit level processor
Vector/Streaming processors
Symbolic processor

Size, Weight, Energy, Performance, Time
SWEPT Efficiency

= Processor Centric = 33%
= Data Centric = 66%
Mixing Processor/Data Centric Architectures

» Cell, ClearSpeed, others have a 1-2 year code base

» Complex Heterogeneous Architectures will provide significant challenges on code portability
  » FPGA code very portable
  » X86 Architectures where we are inherently porting from

![Diagram showing application data types and efficiency]

- **SWEPT Efficiency**
  - Size, Weight, Energy, Performance, Time

- **Application Data Types**
  - Bit Level
  - Vector/Streaming
  - Symbolic

- **Parallelised**
  - Bit level processor
  - Vector/Streaming processors
  - Symbolic processors

- **FPGA VHDL Code Base 10+ Years**
- **X86 Code Base 25+ Years**
The FPGA can effectively concatenate functions

» Minimises Memory Thrashing
  » Critical Bottleneck in today’s Architectures

» Requires Data to be ordered

» The Microprocessor almost becomes the Coprocessor
  » When viewed from a “Data Centric Perspective”
A Generic Data Centric Computing Architecture

System Partitioned to minimise comms bandwidth

FPGA Data Centric Processor

X86 Symbolic Processor

FPGA Data Centric Processor

X86 Symbolic Processor

FPGA Data Centric Processor

X86 Symbolic Processor

Internet

Main Processor & Co-Processor?!
or Peer Processors?

Multimedia
We are entering a Gigabit serial comms world!

- All Processor I/O will be Gigabit serial
- Phy layer will be compatible across memory, system interconnect and I/O
- Software Configurable
- Possibility to choose your architecture that best suits your application.

= 4 x 5Gb bidirectional serial links
X86 Architecture Migration

Quad Core Xeon Processor

Quad Core Xeon Processor

Quad Core Xeon Processor

Quad Core Xeon Processor

SDRAM Memory DIMMs

Interfaces already Gbit Serial

Storage I/O

USB I/O

1394 I/O

Graphics I/O

Wireless I/O

Network I/O

Accelerators

MCH / Northbridge Chipset

PCIe x8

2+2 GB/s

2+2 GB/s

2+2 GB/s

21 GB/s

8.5 GB/s

8.5 GB/s

FSB

FSB

Geneseo PCIe extensions resolving Latency, Determinism, small & large packet Bandwidth issues (2010)
Intel & Nallatech – Bringing processor & Data Centric Architectures together

Nallatech’s Data Centric FPGA Computing Platform – DIME-II

Intel’s Processor Centric MP Xeon Platform
Summary

» Only Two Fundamental Computing Approaches
  » Processor Centric – x86 Microprocessor
  » Data Centric – FPGAs

» HPC Can Learn from HPEC

» FPGAs are by far the best all round accelerator
  » Commercial viability
  » Multi Vendor
  » application Flexibility
  » extremely complementary Microprocessors

» Nallatech & Intel working together to deliver balanced FPGA/x86 architectures in early Q1/2008