Building Zero Latency Matrix Multiplication Engines using FPGAs
SAAHPC
14th July 2010

Craig Petrie – Product Manager

www.nallatech.com
Nallatech at a glance

- Leading supplier of COTS high performance computing solutions
- Serving Embedded Defense and High Performance Computing markets
- Founded 1993
- Wholly owned subsidiary of Interconnect Systems Inc
- Headquartered Camarillo, CA.
- 200 employees worldwide
Capabilities

- Privately owned, stable
- Advanced SMT
- Bare die assembly
- 3D memory packaging

- US design center
- ITAR registered
- HP approved supplier
- Miniaturization

- Xilinx Alliance Partner
- UK design centres
- Security cleared staff
- Dedicated Engineering support desk manned by UK nationals
- IBM Integrated Solutions Reseller (ISR)

- US design center
- ITAR registered
- Dedicated Engineering support desks manned by US nationals
  - California
  - Maryland

Copyright © 2010, Nallatech.
Nallatech and ISI offer a wide range of FPGA accelerator products and design tools.
What is an FPGA?

» Essentially a “reconfigurable” ASIC
» Each FPGA contains thousands of “computing building blocks”
  » Registers
  » Adders
  » Multipliers
» Can efficiently solve specific computing problems
» Hundreds of GPIO pins:
  » Memory – DDR3 SDRAM, QDR-II+ SRAM
  » Sensors - ADCs, DACs, Video etc
» High speed serial I/O supporting:
  » PCI Express (8-lane Gen3)
  » Ethernet (up to 100GbE)
  » Serial Rapid I/O

“Xilinx” - FPGA market leader
- www.xilinx.com
- ~$2Bn revenue per year
- “Virtex” and “Spartan” brands
- 40nm silicon process
- Used extensively in embedded computing applications, particularly signal processing
- Typical power consumption of up to 25 Watts
Application Areas

### Signal Intelligence
- Data acquisition and transmission
- Sophisticated ADC and DAC front end
- Usually real time data processing
- Host bandwidth and latency important

### Network Processing
- Monitoring, Interception and Searching of internet/network traffic
- Ethernet (1GbE & 10GbE)
- Mostly real time data processing
- Host bandwidth and latency critical
- Rapid deployment of new technologies

### Accelerated Computing
- FPGA co-processing
- FPGA/Memory architecture important
- Mostly streaming applications that do not require real time data processing
- Platform CPU/FPGA density extremely important
European Southern Observatory (ESO)

» ESO is the intergovernmental science and technology organisation in astronomy. It is carrying out an ambitious programme focussed on the design, construction and operation of powerful ground-based observing facilities for astronomy to enable important scientific discoveries.

» ESO is currently designing a new class of ground-based telescopes scheduled to begin operations in 2018

» The “European Extremely Large Telescope” (E-ELT) will help track down Earth-like planets around other stars in the “habitable zones” where life could exist
E-ELT

» The E-ELT dome will be similar in size to a football stadium
  » Diameter at its base of over 100 m
  » Height of over 80 m.

» The primary mirror will be a record setting 42 m in diameter

» Impressive… however the instrumentation precision and compute power required to successfully detect an exoplanet is exceptional
Let’s put things in perspective…
Let’s put things in perspective…
Let’s put things in perspective…
Let’s put things in perspective…

Sun  Sirius  Pollux  Arcturus
Jupiter is about 1 pixel in size
Earth is invisible at this scale
Let’s put things in perspective…
Transit Method

» If a planet crosses in front of its parent star's disk, then the observed brightness of the star drops by a small amount. The amount by which the star dims depends on its size and on the size of the planet.

» The effect is akin to watching a mosquito flying in front of a searchlight two hundred miles away.
Detecting exoplanets using ground-based telescopes is especially challenging since the light you are sampling is constantly being distorted by Earth’s atmosphere.

Adaptive Optics (AO) provides a means of compensating for atmospheric turbulence.
Real Time Box - Critical Requirements

1. Acquire large amount of data and distribute them to processing unit in an extremely short time
2. Collect results and distribute them to the next stage or to the actuators in an extremely short time
3. Deliver constant and reliable performance as far as execution time is concerned (jitter)
4. Pack enough computing power in a reasonable space with a real-time low latency fabric connecting them
At the heart of the AO processing is a Matrix Vector Multiplication (MVM) operation. Here, input data arrives over a 3kHz frame in blocks, such that the first element of the input vector is transmitted at the beginning of the frame and the last element arrives at the end of the frame. Blocks are of pre-determined size and their arrival is deterministic. Maximum latency is half the update rate.
Which technology?

**Multi Core**
- Good all rounder, built to run sequential code
- Difficult to distribute applications across multiple servers
- Data communication overhead kills system latency

**Many Core**
- Single and Double precision floating point
- Can be used as co-processors for the supervisor, where real time is not a requirement

**Many, Many Core!**
- Bit manipulation, variable bit-width, integer & FP
- Ideal for direct interfacing with sensor I/O (ethernet)
- Architecture is inherently parallel
- Deterministic, real time capable
System Complexity – Moore’s Law

![Complexity vs time graph]

- Complexity (logscale in MAC/s) vs Year
- Log scale for both axes
- Data points for various initiatives (NAOS, MACAO, SPHERE, AOF, SCAO, NGSGLAO, LGSGLAO, ATLAS, MAORY, EAGLE, EPICS)
- Trend lines indicating growth over time
System Complexity – Moore’s Law Corrected
Adaptive Optics Telescope Requirements

- Microprocessor based platforms will not be able to meet the 6 TMAC ultimately required by EPICS
- A platform featuring tightly-coupled FPGAs could deliver this demanding level of performance

<table>
<thead>
<tr>
<th>AO Class</th>
<th>AO Module</th>
<th>WFS</th>
<th>Size</th>
<th>DM size</th>
<th>Freq</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>EAGLE</td>
<td>MOAO</td>
<td>6+4</td>
<td>84x84</td>
<td>1 x IFU</td>
<td>84x84</td>
<td>250</td>
</tr>
<tr>
<td>EPICS</td>
<td>XAO</td>
<td>1</td>
<td>210x210</td>
<td>211x211</td>
<td>2500</td>
<td>6 TMAC</td>
</tr>
<tr>
<td>MICADO</td>
<td>SCAO</td>
<td>1</td>
<td>84x84</td>
<td>1</td>
<td>7200</td>
<td>1000</td>
</tr>
<tr>
<td></td>
<td>MCAO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80 GMAC</td>
</tr>
<tr>
<td>METIS</td>
<td>LTAO</td>
<td>1</td>
<td>84x84</td>
<td>1</td>
<td>7200</td>
<td>1000</td>
</tr>
<tr>
<td></td>
<td>SCAO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>80 GMAC</td>
</tr>
<tr>
<td></td>
<td>MAORY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAORY</td>
<td>MCAO</td>
<td>6</td>
<td>84x84</td>
<td>3</td>
<td>7200</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>370 GMAC</td>
</tr>
<tr>
<td>ATLAS</td>
<td>LTAO</td>
<td>6</td>
<td>84x84</td>
<td>1</td>
<td>7200</td>
<td>500</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>240 GMAC</td>
</tr>
<tr>
<td>Telescope1</td>
<td>N-GLAO</td>
<td>3</td>
<td>68x68</td>
<td>1</td>
<td>7200</td>
<td>500</td>
</tr>
<tr>
<td>Telescope2</td>
<td>L-GLAO</td>
<td>4</td>
<td>68x68</td>
<td>1</td>
<td>7200</td>
<td>500</td>
</tr>
</tbody>
</table>
FPGA-based Real Time Computer (RTC)

» 10GbE acts as interface to sensors
  » Solid roadmap
  » Ubiquitous
  » Competitive cost

» Used as a point-to-point connection between FPGAs, 10GbE can be used for real time, deterministic comms

» PCIe cards with multiple FPGAs + Memory per server

» Low latency communications fabric between FPGAs

» PCIe for control and matrix updates
PCIe-180 Overview

PCIe-180 Low Profile 10G Ethernet FPGA Accelerator Card

- 8-lane PCI Express 1.1
- Up to 2.5GB/s total host bandwidth
- Half height, Half length
- XFP supporting 10GE/SONET

- Xilinx Virtex-5 LX155-1 user FPGA
- 1 bank of DDR2 SDRAM memory
- 5 banks of DDR-II SRAM memory
- 10GbE MAC IP core
PCle-180
Results so far…

» PCIe host interface
  » Total bandwidth of 2.5GBytes/s sustained
  » 1.2GBytes/s WRITE, 600ns latency
  » 1.3GBytes/s READ, 600ns latency

» UDP IP 10GbE
  » Complete UDP IP framer/deframer
  » FPGA implementation can achieve 99.97% of maximum bandwidth
  » Total trip latency = 118 clock cycles @ 156.25MHz
    = 775ns (387.5ns each direction)
  » FPGA implementation can play tricks a NIC cannot
    » Avoid buffer packets and waiting for packet completions
    » Start distributing and processing incoming data before the packet is complete
  » Adjust frame sizes on the fly
Results so far…

» FPGA external memory accesses
  » 5 independent banks of DDR-II SRAM
    » 32-bit @ 250MHz DDR
    » 10GBytes/s
  » SDRAM
    » 72-bit @ 250MHz DDR
    » 4GBytes/s

» FPGA-to-FPGA latency
  » Direct connectivity via LVDS links
  » 14 clock cycles @ 324MHz = 43ns
FPGA in Memory Module (FIMM)

» Accelerator products will utilize novel FPGA and Memory packaging techniques to maximize compute density

» Module footprint will bring out standard I/O with support for:
  » LVDS
  » 4x Gigabit Serial I/Os

» Module variants with common footprint. E.g:
  » FPGA only
  » FPGA + 2x memories
  » FPGA + 4x memories

» Memory types:
  » DDR-II and QDR-II SRAMS
  » DDR3, RLDRAM SDRAMs
  » FLASH etc
Concept

PCle-287
PCI Express 2.0 FPGA Accelerator Card

- 8-lane PCI Express 2.0
- Full height, 241.3mm length
- Up to 5GB/s total host bandwidth
- Quad SFP+ I/O via PCI backplate

- 24 user FPGAs
- Up to 2 banks of SDRAM per FPGA
  - LPDDR or DDR3
- Dedicated 1-lane PCI Express 1.1 Host interface per FPGA
Software Design Process

Develop software only application
• Determine key accelerator functions

Convert key functions to HDL
• Can use any tool e.g. DIME-C, Impulse C, System Generator or simply code in HDL
• Easy access to Node capabilities using standard HSD interfaces

Create accelerated system
• Wrap function using Framework Builder
• Automatically includes PCIe interface instantiation & ID generator
• Add autogenerated router code to standard software function
• Fit hardware and go….
Nallatech Integrated Solution

1U Hybrid Computing Platform

» IBM x3550-M3 1U server
» PCIe-180 FPGA card
» Finnisar Optical Transceiver
» Reference Design
» Integration, Support and Warranty services
Thank you

c.petrie@nallatech.com
www.nallatech.com