Productively Scaling I/O Bound Streaming Applications with a Cluster of FPGAs

Andrew Schmidt, Siddhartha Datta, Ashwin Mendon, and Ron Sass

University of North Carolina at Charlotte

July 13, 2010
Introduction

- use FPGAs as compute accelerators
- Question:
  - how well can we replicate cores across a cluster of FPGAs?
- FPGA Cluster and Existing Components
- Experimental Set-Up & Results
- The Next Step
Reconfigurable Computing Cluster (RCC) Project

- investigating use of all-FPGA cluster for HPC
- no discrete microprocessors
- custom integrated high-speed network
- filesystem implemented in a hardware core
- application specific hardware cores
- effectively one *large* FPGA

NSF CRI Grant CNS 06-52468
RCC project’s 64 node Spirit cluster
single node (Xilinx ML410) in Spirit cluster
custom network card (8 bidirectional channels at 4.0 Gb/s)
Design

Three key components:

- AIREN Network
- Hardware Filesystem
- Hardware Core (Application): BLASTn
Architecture Independent Reconfigurable Network

AIREN Performance

AIREN Network Bandwidth for Both On-Chip and Off-Chip Transfers

<table>
<thead>
<tr>
<th>Hops</th>
<th>Latency</th>
<th>Latency/Hop</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.83 µs</td>
<td>0.83 µs</td>
</tr>
<tr>
<td>2</td>
<td>1.60 µs</td>
<td>0.80 µs</td>
</tr>
<tr>
<td>3</td>
<td>2.37 µs</td>
<td>0.79 µs</td>
</tr>
<tr>
<td>4</td>
<td>3.10 µs</td>
<td>0.78 µs</td>
</tr>
</tbody>
</table>

on-chip and off-chip bandwidth and latency of AIREN
HWFS Performance

HWFS sequential Read/Write efficiency with four RAM disks
Basic Local Alignment Search Tool

BLASTn:
- nucleotide based queries over existing databases
- calculates statistical significances of matches
- I/O bound streaming application
- our approach:
  - accelerate Scan and Ungapped Extension functions
  - based on NCBI software implementation
  - produces same results as NCBI
BLASTn Hardware Core

Design

Four node types created to test scalability:

- Head Node
- Disk Node
- BLASTn Intermediate Node
- BLASTn Leaf Node
Experimental Setup

- each node running Linux 2.6.25
- PowerPC 405 running at 300 MHz
- AIREN running at 100 MHz (4.0 Gbps/channel)
- HWFS running at 100 MHz Block Size 1024 Bytes
- BLASTn running at 100 MHz (1 query/core)
- Sub-cluster of 25 total node assembled from Spirit
- 3 configurations tested (each with 4 disk nodes):
  - 1 Node
  - 5 Nodes
  - 21 Nodes
1 Node Configuration

Head Node

1 Node
10 BLASTn Cores:
10 - Head Node

Disk Node 0
Disk Node 1
Disk Node 2
Disk Node 3
5 Node Configuration

Head Node

5 Nodes

54 BLASTn Cores:
6 - Head Node
12 - Each Leaf Node

BLASTn Leaf Node 12 Cores

Disk Node 0

Disk Node 2

Disk Node 1

Disk Node 3
21 Node Configuration

Head Node

**21 Nodes**

230 BLASTn Cores:
- 6 - Head Node
- 8 - Each Intermediate Node
- 12 - Each Leaf Node

Disk Node 0
Disk Node 1
Disk Node 2
Disk Node 3

BLASTn Intermediate Node 8 Cores

BLASTn Leaf Node 12 Cores
### BLASTn Databases and Queries

<table>
<thead>
<tr>
<th>Database Name</th>
<th>Size (MB)</th>
<th>Number of Sequences</th>
<th>Num. Hits Q1 248 Bytes</th>
<th>Numb. Hits Q2 4,292 Bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>env.nr</td>
<td>71</td>
<td>6,027,398</td>
<td>226,576</td>
<td>5,262,389</td>
</tr>
<tr>
<td>month.nt</td>
<td>172</td>
<td>167,450</td>
<td>831,680</td>
<td>15,572,380</td>
</tr>
<tr>
<td>env.nt</td>
<td>677</td>
<td>2,799,207</td>
<td>3,650,764</td>
<td>63,179,552</td>
</tr>
<tr>
<td>human.nt</td>
<td>1,536</td>
<td>30,565</td>
<td>6,732,670</td>
<td>158,369,822</td>
</tr>
</tbody>
</table>
### BLASTn Performance

<table>
<thead>
<tr>
<th>DB name</th>
<th>248 Byte Query Size</th>
<th>4,292 Byte Query Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 Core</td>
<td>10 Cores</td>
</tr>
<tr>
<td>env.nr</td>
<td>0.99×</td>
<td>8.60×</td>
</tr>
<tr>
<td>month.nt</td>
<td>3.92×</td>
<td>18.80×</td>
</tr>
<tr>
<td>env.nt</td>
<td>3.05×</td>
<td>18.83×</td>
</tr>
<tr>
<td>human.nt</td>
<td>7.63×</td>
<td>22.30×</td>
</tr>
<tr>
<td><strong>Geometric Mean</strong></td>
<td><strong>3.09×</strong></td>
<td><strong>16.14×</strong></td>
</tr>
</tbody>
</table>

Speedup of a single BLASTn node with 1 core and 10 cores over 2.1 GHz AMD Opteron 2352
Scalability of BLASTn Cores Across the Cluster

Speedup over Variable Number of Cores with Query Set Size 248 Bytes

Speedup over Variable Number of Cores with Query Set Size 4,292 Bytes
Scalability of Nodes Across the Cluster

Speedup over Number of Nodes with Query Set Size 248 and 4,292 Bytes

- linear speedup
- env.nr (248)
- month.nt (248)
- env.nt (248)
- human.nt (248)
- geometric mean (248)
- env.nr (4,292)
- month.nt (4,292)
- env.nt (4,292)
- human.nt (4,292)
- geometric mean (4,292)

SAAHPC 2010
Conclusion

Question

How well does replicating cores scale across multiple FPGAs?

- no special effort to replicate core
- multiple unmodified BLASTn cores instantiated and system successfully synthesized
- limitation on speedup (1 to 230 cores) due to memory bandwidth bottleneck in original BLASTn implementation
- report linear speedup for 1, 5 and 21 FPGA nodes due to very low end-to-end network latency ($\approx 1.6\mu s$)
Scalability

1 Head Node
4 Intermediate Level 1 Nodes
32 Intermediate Level 2 Nodes
256 Intermediate Level 3 Nodes
2048 Leaf Nodes

theoretical configuration of 2,341 nodes and 26,918 BLASTn cores
Future Work

heterogeneous cluster with 64 BLASTn nodes
Future Work

- Automate Design Process Across the Cluster
- Support Heterogeneous Nodes
  FPGAs, CPUs, GPUs, Cell, etc...
- Implement Other Applications:
  - Smith-Waterman
  - (NAMD) Nanoscale Molecular Dynamics
  - (FDTD) Finite-Difference Time-Domain
Questions?