A First Look At the Cell: Performance Estimates of a STAP Benchmark on the IBM/Sony/Toshiba Cell Processor

Mr. Luke Cico (lcico@mc.com)
Mr. Jon Greene (jgreene@mc.com)
Dr. Robert Cooper (rcooper@mc.com)
Mercury Computer Systems
Chelmsford, MA 01824

The Ultimate Performance Machine

© 2002 Mercury Computer Systems, Inc.
What do we want to learn from this exercise?

- Examine Cell architecture under the light of a representative Radar processing chain.
- Analyze data flow patterns between SIMD engines and main memory.
- Identify Cell processor’s strengths and constraints for these applications.
- Estimate computational efficiency of the Cell.
- Estimate balance of IO and computation.
- To gain insights into the Cell’s programming model and productivity and performance tools that will be needed.
- Not intending in this exercise to examine balance of Cell’s external IO to computational throughput. The goal of this exercise is to examine the mapping of a Radar mode to the cell architecture and examine some metrics for a single chip.
Cell Architecture

- PowerPC Processor Element (PPE)
  - PowerPC Processor Unit (PPU)
  - 64 Bit PPC core
  - 128-Bit Vector Multimedia Extension (VME) SIMD unit
  - 32K Instruction + 32K Data L1 cache
  - 512K L2 cache
  - 2 way hardware Hyper-Threading

- Synergistic Processor Element (SPE)
  - 8 Per Chip
  - 128-Bit SIMD Synergistic Processing Unit (SPU) is VME-like instruction set and architecture
  - 256K Local Storage (LS) for data and instructions
  - Memory Flow Control (MFC) unit with DMA controller
  - 32-Bit single precision FP. Also supports 64-Bit double precision numerical operations but with less efficiency than single precision

Note: 3GHz is Mercury operating point, CBE can be operated at frequencies greater than 3GHz.

Peak Computational Throughput @ 3GHz
3.0 GHz X 8 FLOPS (re. MAC) = 24 GFLOPS per SPE or 192 GFLOPS per chip
This paper presented an analytical model and performance estimates for a STAP processing chain on the IBM/Sony/Toshiba Cell processor. Follow on work will measure the computational and data flow metrics presented here.

This paper has estimated that a single cell is likely to produce an order of magnitude increase in sustained performance per chip for these types of applications.

- Total CPU loading for this application is 44% on single cell compared to 100% loading on 16 current generation Altivec Processors.

Mapping the application required lots of fine grained data strip-mining and careful layout of data sets in SPE Local Store memory in order to maintain throughput of chip.

A DRI programming model has been proposed with a sub-routine engine to allow application designers a higher level of abstraction for managing the data and work flow of the application to the distributed computational resources.

<table>
<thead>
<tr>
<th>Stage of Processing</th>
<th>Throughput (MFLOPS)</th>
<th>Compute time to IO time</th>
<th>CPU Loading (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Processing (video I/Q, pulse compression)</td>
<td>83736.9</td>
<td>2.7</td>
<td>7.2%</td>
</tr>
<tr>
<td>Doppler Filtering and data reorganization</td>
<td>32938.1</td>
<td>&lt; 1</td>
<td>2.0%</td>
</tr>
<tr>
<td>Adaptive Weight Computation</td>
<td>106912.8</td>
<td>7.5</td>
<td>31.2%</td>
</tr>
<tr>
<td>Adaptive Weight Application</td>
<td>11139.5</td>
<td>0.22</td>
<td>4.5%</td>
</tr>
</tbody>
</table>