Benchmark Evaluation of Radar Processing Algorithms on GPUs

HPEC 2010
15 September 2010

Lockheed Martin MS2
199 Borton Landing Road
Mooresstown, NJ 08057

Scott Sawyer
R. Bennett, R. Pancoast, M. Iaquinto,
R. Putatunda, N. Doss, J. Broadbent
Agenda

• Background
• GeForce GTX285 Architecture
• Pulse Compression Implementation
• Benchmark Analysis
• Shared Memory Study
• Path Forward
Background

• Next-generation radar architectures will need significantly more processing capacity
• Architecture options include multi-core CPUs, FPGAs and GPUs
• Algorithms of interest include:
  – Pulse compression
  – Pulse integration
  – Adaptive beam forming
• Pulse compression benchmarked on GeForce GTX285
NVIDIA GeForce GTX285

- CUDA Compute Level 1.3
- Multiprocessors: 30
- Total cores: 240
- Global Memory: 1 GB
- Shared Memory: 16 KB per multiprocessor
- Processor Clock: 1476 MHz
- Theoretical peak throughput: 1063 GFLOPS
Benchmark System Architecture

- Intel Host CPU
- System Memory
- Northbridge
- Graphics Memory (1 GB)
- PCIe 1x (4 GB/s)
- GPU Processor
- Red Hat Linux Host OS
Pulse Compression Implementation

• C for CUDA
  – CUFFT for FFT and IFFT
  – Custom kernel for point-by-point multiplication

• MATLAB reference implementation
  – Input vectors
  – Result verification

• Assumptions
  – Pulsed radar
  – LFM Waveform
  – Frequency domain matched filter

• Best of ‘N’ runs taken
  – Assume least host OS interference
Input Vectors

• Simulated radar returns
  – Complex signal (interleaved real/imaginary)
  – Single-precision floating point
  – Constant delay/range
  – AWGN

• Variable pulse interval (FFT size)
  – 1024, 2048, 4096, ..., 65536 samples

• Variable pulse count
  – 1, 2, 4, ..., 64 pulses
Performance Metrics

Algorithm Timing

Total Latency

Host-to-Device Transfer

FFT
Mult.
IFFT
Device-to-Host Transfer

Algorithm Run Time

• Timing does not include file I/O
• Used CUDA event API to time asynchronous kernel calls

Operation Counts

• Assume radix-2 FFTs
• Count all floating point adds and multiplies in custom kernel
Performance vs. FFT Size

![Performance vs. FFT Size Graph]

- 1 Pulse
- 4 Pulses
- 16 Pulses
- 64 Pulses

GFLOPS vs. FFT Size

Copyright © 2010 by Lockheed Martin Corporation
Timing Breakdown

Example: 16K, 16 Pulses

- 15% run time, 85% transfer time
- 1696 MB/s PCIe (1x) transfer rate

All times in ms

<table>
<thead>
<tr>
<th></th>
<th>4096, 4 Pulses</th>
<th>4096, 16 Pulses</th>
<th>16K, 4 Pulses</th>
<th>16K, 16 Pulses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Host-to-Device Copy</td>
<td>0.089</td>
<td>0.319</td>
<td>0.314</td>
<td>1.179</td>
</tr>
<tr>
<td>FFT</td>
<td>0.029</td>
<td>0.050</td>
<td>0.077</td>
<td>0.194</td>
</tr>
<tr>
<td>Multiply</td>
<td>0.011</td>
<td>0.023</td>
<td>0.023</td>
<td>0.056</td>
</tr>
<tr>
<td>IFFT</td>
<td>0.028</td>
<td>0.049</td>
<td>0.077</td>
<td>0.176</td>
</tr>
<tr>
<td>Device-to-Host Copy</td>
<td>0.116</td>
<td>0.399</td>
<td>0.400</td>
<td>1.306</td>
</tr>
<tr>
<td>Run Time</td>
<td>0.068</td>
<td>0.123</td>
<td>0.177</td>
<td>0.427</td>
</tr>
<tr>
<td>Latency</td>
<td>0.273</td>
<td>0.840</td>
<td>0.891</td>
<td>2.912</td>
</tr>
</tbody>
</table>
Real Time Operation

**GTX285, Single DMA Channel**

For run time less than one-way transfer, real-time operation requires HTD+DTH time less than pulse interval.

**Fermi, Dual DMA Channels**

Separate HTD and DTH channels allow tighter timelines. Real-time operation requires HTD + algorithm run time less than pulse interval.
Shared Memory Study

Investigated optimizing multiplication step by using shared memory.

Reduce reads from global to on-chip registers by sharing filter coefficients:
Coalesced Reads

- Global memory accessed via 32-, 64- or 128-byte reads
- Threads accessing aligned memory may use coalesced reads (strict rules based on compute capability)
• Shared memory implementation lowers multiplication block performance.
Path Forward for Tactical Applications

- Update algorithms for Fermi architecture
- Benchmark adaptive beam forming and other algorithms

**Desktop System**

- Graphics Card
- Device Memory
- GPU
- Host Memory
- PCIe
- Host CPU

**Future Tactical Systems**

- Sensor Data
- Dual-Port Memory
- GPU
- Digital Processing Board