Astrophysical Algorithms on Novel HPC Systems

Robert J. Brunner, Volodymyr V. Kindratenko
University of Illinois at Urbana-Champaign
rb@astro.uiuc.edu, kindr@ncsa.uiuc.edu
Objectives

• Demonstrate the practical use of novel computing technologies, such as those based on Field-Programmable Gate Arrays (FPGAs) and Graphics Processing Units (GPUs), for advanced astrophysical algorithms and applications involving very large data sets

• Make the developed data analysis tools available to NASA research community
Digit{ized|al} Sky Surveys

From Data Drought to Data Flood

1977-1982
First CfA Redshift Survey
spectroscopic observations of 1,100 galaxies

1985-1995
Second CfA Redshift Survey
spectroscopic observations of 18,000 galaxies

2000-2005
Sloan Digital Sky Survey I
spectroscopic observations of 675,000 galaxies

2005-2008
Sloan Digital Sky Survey II
spectroscopic observations of 869,000 galaxies

Sources: http://www.cfa.harvard.edu/~huchra/zcat/
http://zebu.uoregon.edu/~imamura/123/images/
http://www.sdss.org/
Example Analysis: Angular Correlation

- **TPACF** ($\omega(\theta)$) is the frequency distribution of angular separations $\theta$ between celestial objects in the interval $(\theta, \theta + \delta\theta)$
  - $\theta$ is the angular distance between two points
- **Blue points** (random data) are, on average, randomly distributed, **red points** (observed data) are clustered
  - Random (blue) points: $\omega(\theta) = 0$
  - Observed (red) points: $\omega(\theta) > 0$
- Can vary as a function of angular distance, $\theta$ (yellow circles)
  - Blue: $\omega(\theta) = 0$ on all scales
  - Red: $\omega(\theta)$ is larger on smaller scales
- Computed as
  \[
  \omega(\theta) = \frac{1}{n^2_D} \cdot DD(\theta) - \frac{2}{n_D n_R} \sum DR_i(\theta) + \frac{1}{n^2_R} \sum RR_i(\theta)
  \]
Special-Purpose Processors

- Field-Programmable Gate Arrays (FPGAs)
  - Digital signal processing, embedded computing
- Graphics Processing Units (GPUs)
  - Desktop graphics accelerators
- Physics Processing Units (PPUs)
  - Desktop games accelerators
- Sony/Toshiba/IBM Cell Broadband Engine
  - Game console and digital content delivery systems
- ClearSpeed accelerator
  - Floating-point accelerator board for compute-intensive applications
- Stream Processor
  - Digital signal processing
Why not HPC Systems?

• The gap between the application performance and the peak system performance increases
  • Few applications can utilize high percentage of microprocessor peak performance, but even fewer applications can utilize high percentage of the peak performance of a multiprocessor system

• Computational complexity of scientific applications increases faster than the hardware capabilities used to run the applications
  • Science and engineering teams are requesting more cycles than HPC centers can provide

• I/O bandwidth and clock wall put limits on computing speed
  • Computational speed increasing faster than memory or network latency is decreasing
  • Computational speed is increasing faster than memory bandwidth
  • The processor speed is limited due to leakage current
  • Storage capacities increasing faster than I/O bandwidths

• Building and using larger machines becomes more and more challenging
  • Increased space, power, and cooling requirements
    • ~$1M+ per year in cooling and power costs for moderate sized systems
  • Application fault-tolerance becomes a major concern
Summary of Year 1 Progress

- **Two-point angular correlation algorithm implemented on SRC-6 reconfigurable computer**
  - 2 GFLOPS on an FPGA vs. 80 MFLOPS on a CPU
  - 24x speedup over a 2.8 GHz Intel Xeon
  - 3.2% of power of the CPU-only based system

- **Two-point angular correlation algorithm implemented on SGI RASC RC100 reconfigurable module**

- **Instance based classification algorithm**
  - Reference implementation of the n-nearest neighbor kd-tree based classification algorithm
Conclusions from Year 1

- Novel ways of computing, such as reconfigurable computing, offer a possibility to accelerate astrophysical algorithms beyond what is possible on today’s mainstream systems, but
  - Such systems are expensive and
  - Are not easy to program

- We should look at architectures based on commodity accelerators, e.g., GPUs
NCSA’s Heterogeneous Cluster

- 16 compute nodes
  - 2 dual-core 2.4 GHz AMD Opterons, 8 GB of memory
  - 4 NVIDIA Quadro 5600 GPUs, each with 1.5 GB of memory
  - Nallatech H101-PCIX FPGA accelerator, 16 MB SRAM, 512 MB SDRAM
Summary of Year 2 Progress

- Extended two-point angular correlation function implementation from previous year to work on a cluster consisting of multi-core SMP nodes using Message Passing Interface.

- Implemented compute kernel of the cluster application on a Nallatech H101 FPGA application accelerator board using DIME-C language and DIMEtalk API and expanded the application to utilize FPGA accelerators available in all cluster nodes.

- Experimented with the two-point angular correlation compute kernel on the NVIDIA GPU G80 platform using CUDA development tools.

- Extended our reference $n$-nearest neighbor $kd$-tree based implementation of the instance based classification code to work on a multi-core SMP system via pthreads and tested it with multi-million point datasets.
GPU Results

- **Single Node Performance**
  - Dataset
    - 32K observed points
    - 100 x 32K random points
  - Analysis parameters
    - no jackknifes re-sampling
    - Min angular distance: 1°
    - Max angular distance: 100°
    - Bins per decade of scale: 5
  - GPU vs. CPU speedup
    - 25x for 32K dataset
    - 22x for 8K dataset
    - 60x for optimized kernel that works only with small datasets

- **Observations**
  - Single-precision floating-point
    - Cannot perform calculations for angular separations below 1 degree
  - 32-bit integers
    - Overflow in bin counts
    - Requires additional storage and code to deal with overflow
  - Read-after-write hazard is very costly to work around
FPGA Results

- **Single Node**
  - Dataset
    - 97K observed points
    - 100 x 97K random points
  - Analysis parameters
    - 10 jackknifes re-sampling
    - Min angular distance: 0.01 arcmin
    - Max angular distance: 10000 arcmin
    - Bins per decade of scale: 5
  - One CPU core
    - 44,259 seconds
  - Four CPU cores per node
    - 11,159 seconds (3.9x speedup)
  - One FPGA
    - 7,166 seconds (6.2x, 1.6x)

- **8-node Cluster**
  - Dataset
    - 97K observed points
    - 100 x 97K random points
  - Analysis parameters
    - 10 jackknifes re-sampling
    - Min angular distance: 0.01 arcmin
    - Max angular distance: 10000 arcmin
    - Bins per decade of scale: 5
  - One CPU core per node
    - 5,428 seconds
  - Four CPU cores per node
    - 1,449 seconds (3.8x speedup)
  - One FPGA per node
    - 881 seconds (6.2x, 1.6x)
Conclusions from Year 2

• As architectures based on commodity accelerators are becoming readily available, they too offer a possibility to accelerate astrophysical algorithms beyond of what is possible on today’s mainstream systems

  • At a substantially smaller cost as compared to highly tuned and specialized systems such as SRC-6
  • Still suffer from some of the hardware limitations and difficulties with programming
Year 2 Outreach Highlights

- NSF STCI grant: *Investigating Application Analysis and Design Methodologies for Computational Accelerators*


- Reconfigurable Systems Summer Institute (RSSI), July 2007, NCSA, Urbana, IL
Future Work

• With the introduction of double-precision floating-point GPU chips later this year, we will research and implement the two-point angular correlation kernel on double-precision GPUs.

• Extend our existing cluster application to simultaneously take advantage of the multi-core chips as well as the Nallatech H101 FPGA accelerators and NVIDIA GPUs.

• Investigate the use of FPGAs and GPUs to accelerate the $kd$-tree based range search algorithm used in the $n$-nearest neighbor classifier.
Reconfigurable Systems Summer Institute (RSSI) 2008

- July 7-10, 2008
- National Center for Supercomputing Applications (NCSA), Urbana, Illinois
- Organized by