

Real-time Visualisation and Analysis of Tera-scale Datasets

Christopher J. Fluke[†]

Centre for Astrophysics & Supercomputing, Swinburne University of Technology, Hawthorn, Victoria, 3122, Australia ,
email: cfluke@swin.edu.au

Abstract. As we move ever closer to the Square Kilometre Array era, support for real-time, interactive visualisation and analysis of tera-scale (and beyond) data cubes will be crucial for on-going knowledge discovery. However, the data-on-the-desktop approach to analysis and visualisation that most astronomers are comfortable with will no longer be feasible: tera-scale data volumes exceed the memory and processing capabilities of standard desktop computing environments. Instead, there will be an increasing need for astronomers to utilise remote high performance computing (HPC) resources. In recent years, the graphics processing unit (GPU) has emerged as a credible, low cost option for HPC. A growing number of supercomputing centres are now investing heavily in GPU technologies to provide O(100) TeraFlop/s processing. I describe how a GPU-powered computing cluster allows us to overcome the analysis and visualisation challenges of tera-scale data. With a GPU-based architecture, we have moved the bottleneck from processing-limited to bandwidth-limited, achieving exceptional real-time performance for common visualisation and data analysis tasks.

Keywords. methods:data analysis, techniques:image processing

1. Introduction

Next generation facilities such as the Large Synoptic Survey Telescope and the Square Kilometre Array, and its pathfinders, will survey more of the sky, more often, with more pixels covering more wavelengths – leading to more data that needs to be processed, analysed and visualised before more science will result. This petascale astronomy data era brings with it many new computational challenges, as existing desktop-based solutions are unlikely to scale. Instead, a move away from analysis of data on desktop computers will need to occur, with greater reliance on remote data processing. A critical ingredient for the success of remote processing solutions will be their ability to support real-time, interactive data analysis and visualisation tasks for individual data products measured in terabytes. High performance computing facilities that leverage the processing power of graphics processing units (GPUs) are showing great promise as a solution to the interactive data analysis and visualisation of terabyte-scale data cubes.

2. Graphics processing units

GPUs have emerged as low-cost computational accelerators, capable of providing 10×–100× speed-ups for algorithms with “massive data parallelism” (e.g. single instruction multiple data) and high arithmetic intensity (ratio of floating point operations to memory access). While existing codes cannot be used directly on a GPU, there are a growing number of programming options including: the CUDA and OpenCL application programming

[†] Present address: Mail H30, PO Box 218, Hawthorn, Victoria, 3122, Australia.

interfaces; code libraries including cuFFT, cuBLAS, and Thrust (C++ template library with a focus on algorithms); PyCUDA for Python; support within IDL and Matlab; and the OpenACC compiler-based directives. Prior to choosing a specific implementation, it is essential to determine whether an algorithm is suitable for a GPU – see Barsdell *et al.* (2010) and Fluke *et al.* (2011) for practical GPU-adoption strategies.

3. GPU-accelerated data analysis and visualisation

The WALLABY all-sky, extragalactic HI survey, to be performed with the Australian Square Kilometre Array Pathfinder (ASKAP), will produce over 1000 spectral data cubes (right ascension, declination and frequency or line-of-sight velocity) with a likely resolution of $4096 \times 4096 \times 16384$ voxels, requiring ~ 1 TB of storage each. Compare this to the 12 GB data volume of the entire HI Parkes All-Sky Survey [HIPASS; Barnes *et al.* (2001)] where the 387 individual cubes contribute to a single $1721 \times 1721 \times 1024$ voxel super-cube.

Using Australia's gSTAR national facility, Hassan *et al.* (submitted) have benchmarked several common data analysis and visualisation tasks for a 542 GB spectral data cube (replication of the HIPASS supercube 48 times = $6884 \times 6884 \times 3072$ voxels). Once in GPU memory of the cluster (576 GB total memory available using $90 \times$ Tesla C2070 and $6 \times$ Tesla C2090 GPUs, each with 6 GB/GPU) it takes approximately:

- 0.14 seconds to provide a three-dimensional volume rendering;
- 2 seconds to calculate the global mean and standard deviation;
- 4 seconds to calculate a histogram;
- 45 seconds to iteratively calculate the global median; and
- 20 milliseconds to extract a one-dimensional spectrum from an arbitrary orientation of the cube, a step towards interactive quantitative investigation.

4. Conclusions

Graphics processing units are a practical solution to the real-time data analysis and visualisation needs of the petascale astronomy era. They provide a low-cost option to $O(100)$ Tflop/s high performance computing, compared to the equivalent number of CPUs. The majority of the benchmarked data analysis and visualisation tasks were bandwidth limited – the GPUs have more than enough processing capability. Future advances in GPU cluster computing will benefit from more memory per GPU, more GPUs per compute node (currently limited by the PCI Express bus), and faster inter-node communication. Astronomy is headed towards an exciting, massively-parallel future.

Acknowledgements

CJF thanks his collaborators Amr Hassan (Swinburne), David Barnes (Monash), and Virginia Kilborn (Swinburne) for their contributions. This work was performed on the gSTAR national facility at Swinburne University of Technology. gSTAR is funded by Swinburne and the Australian Governments Education Investment Fund.

References

- Barsdell, B. R., Barnes, D. G., & Fluke, C. J., 2010, *MNRAS*, 408, 1936
Barnes, D. G., *et al.* 2001, *MNRAS*, 322, 486
Fluke, C. J., Barnes, D. G., Barsdell, B. R., & Hassan, A. H., 2011, *PASA*, 28, 15
Hassan, A. H., Fluke, C. J., Barnes, D. G., & Kilborn, V. A., *MNRAS*, under review