BarraCUDA Project: A Homebrew Supercomputer

Matt Redmond

1 February 2009

Contents

1 Main Proposal Questions

1.1 Objective and Motivation

What is the objective (goal or purpose) of your senior project and what do you hope to learn?

I'd like to build a supercomputer based on NVIDIA's CUDA (C Unified Device Architecture) platform, because CUDA represents the new direction that high performance computing is moving in: instead of waiting for a single fast CPU to perform heavy calculations sequentially, these calculations are farmed out in parallel to hundreds of individual GPU cores for computation. The actual performance speed-up (across diverse applications) is reported by various sources [1] to be anything from 5x to 2600x the time it would take a CPU-based system to do the same tasks. Furthermore, the CUDA architecture is alive and well-supported by the community, with various development tools available, and many new applications released frequently. Completing this project would provide me with a fair background in supercomputer design and construction, as well as an opportunity to learn the parallel programming skills that I'll need to be a successful researcher in the future. The programming API OpenMP is well-known and widely supported in the scientific computing community, and I'd like to gain further experience working with it on a platform that can unlock all of its power.

The BarraCUDA (Beginner's Analytic Research and Rendering Architecture) project will be a small-scale home brew supercomputer designed around the NVIDIA CUDA architecture. The project is meant to serve as a platform for student-researchers to develop parallel-programming skills, and design programs that showcase concepts in modern science research through real-time visualization. After building this supercomputer, I intend to develop a comprehensive and accurate three dimensional simulation of electrodynamics suitable for demonstrating the principles of electromagnetism to high school physics students. The supercomputer will be donated to Menlo upon the project's termination. I hope to further my skills in the areas of physics research and computer simulation.

1.2 Action Items and Resources

What will you actually be doing during the course of your project? What resources will you consult before and during your project?

The project essentially consists of two phases: the first phase entails construction and set-up of the computing resources, and the second phases entails software design and programming. For the first phase, I'll order computational components through the website www.newegg.com, and subsequently construct the computer platform. I hope to obtain the single most expensive component (a high-throughput graphics processor) through a grant provided by the NVIDIA foundation, and I've already contacted Dr. Kayvon Fatahalian of Stanford University and NVIDIA Corporation about this. I expect the construction to be completed relatively quickly, but the bulk of the first phase will lie in the configuration. Supercomputing differs dramatically from consumer-grade computing in the overhead required, and merely configuring all of the parts to interact with each other as quickly as possible will take a great deal of effort.

The second phase (programming the simulation) will extend on work I've undertaken in my Applied Science Research class. In this second half of the project, I will attempt to create a computational simulation of the behavior of objects under forces such as the Lorentz force and gravity. For this aspect of the project,

I will likely draw upon the expertise of Dr. Dann as well as Mr. Thibodeaux and Mr. Steinberg. I do not expect that I will need a lot of outside help for this project.

1.3   Personal Importance

Why does the project you've chosen matter to you?

Through my Applied Science Research and Computer Science classes, I've gradually become aware of the growing importance of high performance computational power in the fields of physical, chemical, and biological research. I'm also quite interested in learning more about fluid dynamics and visual representations of physical modeling problems. Current research in these computationally expensive fields includes [2], [3], and [4]. Last summer, I worked in Caltech's materials physics lab on the DANSE project. The DANSE project uses distributed computational analysis to model the theoretical results of neutron scattering experiments before actually conducting them. At Caltech, I was exposed to the elegant concept and power of a Beowulf cluster, and I realized that our high school could make good use of such a powerful tool (on a smaller scale) for instructional purposes, in the demonstration and visualization of scientific concepts as well as the programming process. I envision a class dedicated to high-performance computing, where students create programs for the science department to be used for various purposes. It would be an excellent way to prepare for a career as a science researcher.

1.4   New Challenges

How will this project challenge you? How does this project speak to your talents and limitations? If you do have extensive background knowledge, how are you going to push beyond it?

This project will combine my interests in physics and computer science, giving me an outlet to go further in depth in both fields. I've never actually had a chance to work with computational power on the teraflop scale, and this project should reach that benchmark with ease. I'm going to be working in areas I've never worked in before, like the synchronization of multi-threaded programs, and I should gain some more experience in multi-variable calculus as well as numerical analysis and optimization. Another aspect of the project's challenge lies in it's scope: the debug process on such a large project cannot be taken too lightly. From experience, I have found that the majority of the work in designing such a complicated systems lies in the optimization and debug processes.

1.5   Mission Statement

Tell us how your project embodies one or more of the values in Menlo's Mission Statement.

The BarraCUDA project will help me to "develop the skills necessary for success in college" as I learn about physics simulation and computational cluster management techniques. As stated above, these fields are drawing lots of funding and research time as the fruits of their labors become more readily applied to industrial and commercial situations. This project is absolutely governed by the mission statement's support for the process of discovery, as I expect that I will learn a lot of new techniques very rapidly throughout the four week process.

1.6 Budget

What is your budget? Does your project have any costs? How do you plan to pay for these costs?

Component	Description	Cost
Motherboard	ASRock X58 Extreme LGA1366	169.99
CPU	Intel Core i7-920 Quad Core Bloomfield 2.66 Ghz	288.99
CUDA Boards	2x Gigabyte GV-N295-181-B GTX295	2x 494.99
Memory	OCZ Gold 12GB (6 x 2gb) DDR3 1600	310.99
Hard Drive	Western Digital Velociraptor 300GB SATA @ 10,000 RPM	199.99
Power Supply	Tuniq Ripper 1000W	149.99
Total Cost		$2,109.93

Although I initially toyed with the idea of building a CPU-based 4-node cluster computer, after speaking with Dr. Kayvon Fatahalian (a post-doctoral researcher working at Stanford and NVIDIA on the CUDA project), I realized that I could actually get a lot more computational power for less money with a single-node, multi-GPU system. The NVIDIA cards like the GTX295 and the Tesla C1060 offer a compelling performance per dollar ratio, allowing this design to reach speeds greater than one TFLOP of processing power for substantially less money than an equivalent CPU-based system would cost. The motherboard was chosen because it utilizes the latest X58 chipset from Intel, with a LGA1366 socket that will support a quad core Bloomfield CPU. Furthermore, this particular motherboard was chosen because it supports double-wide PCI Express cards (the GTX-295 and the TeslaC1060 are both twice as wide as a standard PCI-E card, so any motherboard used for this project needs to meet this requirement). The motherboard also supports the current fastest DDR3 memory, so the computer will not bottleneck during intensive memory read-write operations. The Intel Core i7-920 CPU was chosen because it is provides the largest performance / price ratio of all quad-core CPUs on the market. Other options for this component include the Core i7-950, i7-960, and i7-975, but these processors offer only marginal performance gains in exchange for substantially higher costs. The CUDA processors can be found on the NVIDIA GeForce GTX-285 and -295, as well as the NVIDIA Tesla C1060. Unless the Tesla boards are donated, they are outside of the budget for this project. The GTX 295 offers twice the performance of the GTX-285, but it costs less than twice as much. Therefore, it was selected for this project. Supercomputing can put a high demand on the memory of any system, and I wanted to ensure that the BarraCUDA system wouldn't bottleneck here. 12 GB of memory ensures that the CUDA boards have enough working space (in addition to each of their 1.7GB on-board memory) to conduct any calculations easily. In addition to having high memory capacity, the system uses very high bandwith memory (DDR3 1600). This doesn't come cheaply, but it is worth the extra money to ensure the best possible performance. The hard drive is selected for its rapid data transfer rate. The "Velociraptor" drive does not provide a lot of storage space, but this is not a relevant concern given the applications that BarraCUDA will be used for. The biggest bottleneck in any supercomputing is generally I/O speed, and this drive helps to alleviate some of that pressure by providing rapid seek/write times for all sectors. An excellent introduction to the CUDA architecture (and indeed, a computer designed similarly to the one described above) can be found at [5].

I was hoping that Menlo School would be able to fund part of this project (as the final computer will be donated to the school), but I will need to meet with Bill Silver to establish the extent of available funding. Dr. Dann has agreed to fund $500 of the project, and Dr. Kayvon Fatahalian asserts that NVIDIA will be able to send me both GTX295s for free. I know that seeking funding for projects can take a long time,

so I will attempt to have all of the funding secured for the project no later than one month before it starts (funding by 3 April 2010).

1.7   Presentation

What do you plan to do for your public presentation? Please name any special equipment or facilities you may need.

I will present my simulations on a projected screen in real-time, demonstrating the full capabilities of the BarraCUDA supercomputer system. I will bring the computer, but I will need a projector system.

1.8   Transportation

Do you have transportation needs? If you are not driving yourself, please describe how you will meet your transportation needs. Please note that students are not allowed to drive each other for a senior project.

I have no expectations of traveling. If I do need to drive anywhere, I can drive myself.

2   Citations

[1] NVIDIA CUDA Project Repository. <http://www.nvidia.com/object/cuda_home.html#>

[2] CUDA-Based Incompressible Navier-Stokes Solver. Julien Thibault and Inanc Senocak.   <http://coen.

[3] TeraFlop Computing on a Desktop PC with GPUs for 3D CDF. J. Tolke and M. Krafczyk.   <http://www.

[4] CUDA Acceleration of Molecular Dynamics. David Kirk and Wen-mei W. Hwu.   <http://www.ks.uiuc.edu/

[5] Asrock Supercomputer Video. <http://www.youtube.com/watch?v=_87a6P0-Xjw&feature=related>

3   Calendar

Attached to this document is the proposed calendar for the senior project.