Swarming SCC17: Georgia Institute of Technology

Petros Eskinder, Nicholas Fahrenkrog, Dezhi Fang, Manas George, David Meyer, Jessica Rosenfield, Alok Tripathy
Chirag Jain, Will Powell, Oded Green
{peskinder3, nick.fahrenkrog, dezhifang, xkcd, dmeyer31, jrosenfield3, atripathy8, cjain, will.powell, ogreen}@gatech.edu

***Abstract*—Each year, Georgia Tech's presence at the Supercomputing conference is felt: from papers in the main conference and workshops, faculty participations in panels and keynotes, students participating in the multiple poster sessions and the exhibition booth. For this year's Supercomputing'17 at Colorado, we aim to further expand our presence by competing in the Student Cluster Competition (SCC). We are a team of seven talented undergraduate students mentored by HPC experts, staff and doctoral students in Georgia Tech. In this proposal, we introduce the student members in our team, our advisors, as well as our background in HPC. We outline our strategy to win the competition and discuss the software/hardware details of our IBM Power8 compute cluster that we plan to use to benchmark SCC applications. Reproducibility is now an essential component of the HPC research process. Besides addressing numerous challenges that are already part of SCC, we propose to make all our software optimization efforts in this competition reproducible for future teams by utilizing Docker in our software architecture.**

I. Our Team

We believe we have assembled an excellent team of passionate and capable students and mentors. Our team consists of seven undergraduate members, six competitors and one alternate. We are all Computer Science or Electrical Engineering majors. Half of our team is in their junior year of college, with the rest either in their sophomore or senior year.

We each possess prior interest and knowledge of High Performance Computing (HPC), through research projects, internships and coursework. We view the competition as an excellent opportunity to develop a more systematic and practical understanding of HPC. *Besides, pushing numbers is its own reward.*

We are mentored by Dr. Oded Green, Chirag Jain, and Will Powell, who are all affiliated with the School of Computational Science and Engineering at Georgia Tech. Dr. Green is a research scientist; Chirag a doctoral student; Will is a research technologist.

A. Strength of Team

Although this will be our first time competing, we believe the qualities described below make our team competitive for this year's competition and many more to come.

- **Strong Academic Background.** Through our coursework, we each share knowledge of computer architecture, operating systems, data structures and algorithms, and parallel programming. Members on our team have also taken graduate level courses in Numerical Linear Algebra, Operating Systems, Distributed Systems, and Parallel Computing Architecture.
- **Experienced with Linux.** All team members have prior experience using Linux through courses, internships, and personal usage. Some members have even worked as administrators. For instance, Andy administers the clusters for his research lab, and for the company he previously led. Other members are active competitors of capture-the-flag (CTF) competitions, where they exploit vulnerabilities in Linux ELFs or binaries.
- **Extensive Research and Industry Experience.** We have a broad range of research and internship experiences relevant for the competition. Our shared research experience ranges from working on parallel streaming graph algorithms, to accelerating genomic applications using GPUs, to integrated circuit fabrication, to data analytics and parallel programming instrumentation. Additionally, we have also interned at a number of companies, providing us the opportunity to further improve our software engineering skills and face a wider breadth of programming and engineering problems. These companies include Apple, Bloomberg, Facebook, Google, Sandia National Labs, Square, Symantec, and Texas Instruments.
- **Leaders in Our Community.** Every one of us has loved some course or club enough that we wanted to take a leadership position to make a positive impact on it. For example, 5 of our 7 members have served as undergraduate teaching assistants for our computing courses. Jessica has served as the president of College of Computing's Undergraduate Council, and has been recognized as Outstanding Senior in Computing. Manas leads our school's Computer Security club, where he organizes weekly CTF competitions.
- **Passionate Programmers.** We all also program extensively outside of school, largely for fun. Several of our team members have competed in and won at hackathons. Others have competed in algorithmic programming competitions. Some members have also started and ran their own technology companies.

We believe that by coupling our shared experiences and talents with duly applied hard work, we will be competitive at this year's competition as well as many future ones to come.

B. Further Information About Members

We have placed short biographies of each member in the appendix of this proposal. These further discuss our relevant experiences, and interests in HPC.

C. Strength of Diversity

At Georgia Tech, we pride ourselves on our institution-wide commitment to support and foster diversity in all of its manifestations. Each year, we are consistently rated among the top universities in the nation for graduation of women and underrepresented minorities in engineering, and computer science. More recently, in 2015, we received a $5 million grant from Intel to expand our existing initiatives to recruit and retain qualified underrepresented students in STEM. These initiatives include summer research programs for undergraduates, mentorship programs at the K-12 level as well as at the graduate and undergraduate levels, and outreach programs attracting bright women and underrepresented minorities to attend graduate school.

Our team is a direct reflection of this commitment to inclusive excellence. Early January, our advisors posted an announcement on our college's news and events mailing list inviting students to join GT's inaugural Student Cluster Competition team. Soon after, they held an information session describing the competition and expected level of commitment. From the attendees, seven members, six competitors and one alternate, were then selected based on their enthusiasm and prior experience with HPC. What resulted was a very diverse group of students.

Geographically, our team members' backgrounds span four states, six countries, and three continents. Manas grew up in Dubai. Chirag is from Haryana, India. Dr. Green stems from Haifa, Israel. Andy is from Ürümqi, China. Petros was born in Ethiopia. Jessica was raised in Farmington, Connecticut, Alok grew up in Princeton, New Jersey. Nick is from Boulder, Colorado, and David is from Buford, Georgia. Beyond this variety in background, we also share an eclectic mix of interests outside of our work in computing. Members of our team can be found running marathons, playing sports such as squash and baseball, designing fashionable clothing, and hacking gaming consoles.

Additionally, two team members are members of underrepresented communities, and have each held leadership positions within our overall undergraduate computing community. One of these two members also organized the creation and submission of this proposal. We believe we can leverage our diversity in background and talents to not only succeed in the competition, but to also attract a broader range of students in the future, *expanding the pipeline into HPC*.

II. Team Preparation

A. Coursework

As apart of our degrees, we have each taken coursework on computer architecture, operating systems, data structures and algorithms, and parallel programming. Several members, through their own choosing, have also taken advanced courses in Numerical Linear Algebra, Operating Systems, Distributed Systems, Processor Design, Compilers, High Performance Computing Algorithms and Parallel Computing Architecture. Note, we not only *enrolled* in these courses, we also *excelled* in them. Most notably, 5 of our 7 members have served as teaching assistants for relevant courses like Data Structures and Algorithms, Introductory Computer Architecture, Introductory Computer Systems and Networks, and Algorithm Design. Other members also tutor privately on the above coursework.

B. Available HPC Resources

Our plan is to assemble a two node Power8 cluster. We expect the first node to arrive in early August, right before the start of the fall semester. Then, between mid-September and early October, we expect our second node. This gives us at most two months with access to our competition machine.

In the ensuing time, we possess access to a number of clusters. Georgia Tech's College of Computing has provided us with two Intel Xeon based clusters, PACE and Jinx. PACE is a four node research cluster with each node outfitted with 24 core Intel Xeon E5-2680 CPU. Jinx is a 24 node instructional cluster with each node containing two Intel Xeon X5650 processors and two NVIDIA Tesla GPUs. PACE is shared amongst members of a research group at Georgia Tech and Jinx is shared with various classes. Additionally, we have been given access to Top500 machine Comet@SDSC. These servers will be used for practicing our skills and developing an intuition for the applications.

In addition to the aforementioned clusters, we currently have a Power8 server, on loan to us from IBM. This server is an older generation of the Minsky server we plan to use at the competition. We are using this system to get accustomed to the PowerPC architecture and to properly tune each application for our future server. This way, during the fall, when our server arrives, we will be productive. Note, IBM loaned us this cluster explicitly to prepare for the Student Cluster Competition. Consequently, the team has exclusive access to the server, and each member has sudo access.

C. Method of Preparation

After forming our team in early January, we have held weekly meetings throughout the spring semester. Since many of our members had strong fundamental HPC knowledge, our primary focus during the spring was to familiarize everyone with the applications, and with the practice of tuning an application. Consequently, our meetings have largely consisted of discussions on applications and our experiences tuning them. On a number of occasions, faculty and research scientists at Georgia Tech have given talks on topics like the mathematics behind the LINPACK algorithm and the basics of distributed memory parallel computing using Message Passing Interface (MPI).

Outside of meetings, we communicate using Slack, an instant messaging and collaboration tool. We have segmented our Slack into different channels such as proposal writing, HPL benchmark results, and general announcements. Within each channel, we communicate topically relevant questions, progress and concerns.

Additionally, we have split our team into groups of two or three members. Each group is responsible for a specific aspect of an application. During the week, they conduct investigations together, and then present their results and learnings during our weekly team meeting. We believe this weekly communication practice is an excellent preparation for impressing the judges during our application and team interviews.

D. Plans Moving Forward

During the summer, many of our members will be away at either internships or summer research programs. We will use this time to conduct a literature review on the various applications, as well as our hardware and software. Meetings will be held biweekly through Google hangouts. In addition, since we will have remote access to our servers, we will individually investigate any existing

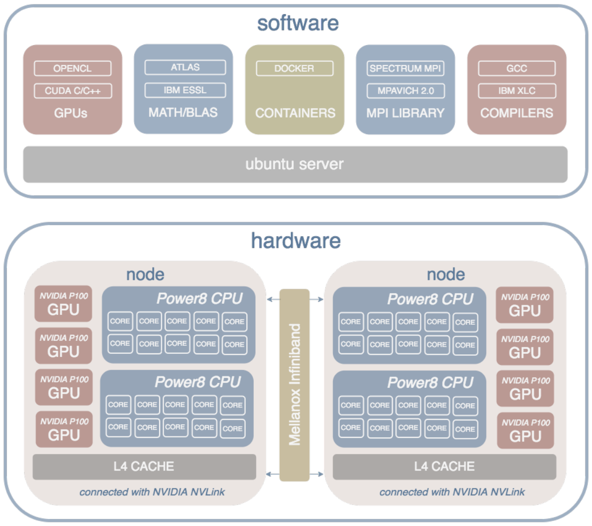


Fig. 1. Our Hardware and Software system architecture

interests or problems. This will allow us to hit the ground running in the fall. In August, at the start of our fall semester, our first server node will arrive. Our focus will be getting our system together and optimizing the applications for our system. To best prepare for the competition, we plan to hold longer meetings and meet twice as frequently.

E. Reproducibility through Containers

We propose making all of our software optimization efforts reproducible for future Georgia Tech teams, by utilizing *containers* in our software architecture. Last year, the Student Cluster Competition introduced the Reproducibility Initiative, using our advisor's application ParConnect, with the intent of promoting the practice of reproducibility [1]. We consider using *containers* a big step towards better reproducibility. Containers are isolated sandbox environments for system-level virtualization in Linux. They can be thought of as lightweight virtual machines with minimal performance loss. Because a container specifies the system right down to every file in the file system, there's little room for misconfiguration on the software side. Moreover, containers can be easily shared across team, ensuring every team member has the exact same configuration. Although nascent, there is a growing interest in using containers in HPC environments [2]. Moreover, this method of ensuring reproducibility is already been actively employed in computational biology [3]. Additionally, we believe our containers can serve great cornerstones for Georgia Tech's SCC 2018 team. Using our containers, they would be able to start their conquest based on a collection of trial and error of ours. Containers could also serve as baselines for competition. A container could be provided to teams that is guaranteed to work, and further optimizations of different teams can have a single source.

III. Hardware Design

Although we are in the preliminary stages of deciding our cluster design, the hardware currently under consideration is heavily based on the two node IBM Minsky system (Power S822LC). Each node is outfitted with two 10-core Power8 CPUs and four NVIDIA P100 GPUs. We intend to use Infiniband EDR's as our interconnect between the two nodes, and NVIDIA's NVLink interconnect within each node for GPU-GPU and GPU-CPU communication.

The POWER8 architecture is the first processor to come out of the OpenPOWER consortium, and has been designed for computationally intensive applications. We see this in part with the NVLink protocol, showing how the POWER8 system is optimized to work well with NVIDIA GPUs. A similar synergy is seen in the choice of interconnect which supports some of the fastest interconnects, like Mellanox InfiniBands.

A. Why POWER8?

At the time of writing this proposal, only two applications have been released, HPCG and HPL. Consequently, we will make general arguments for using the Power8, irrespective of one's application. Here is a listing of the advantages we believe the Power8 possesses:

- **Fast Memory Subsystem**: The Power8's biggest attractor's is its very fast memory subsystem. The POWER8 memory subsystem series shows consistently high memory bandwidth ($\sim$90GB/s) on the Stream bandwidth benchmark, staying in the high nineties even as the number of threads varies from 20 to 80 [4]. These numbers were obtained using the GCC compiler, which is free and much easier for most developers to obtain and integrate into their toolchains, as opposed to vendor-specific compilers, and is therefore more representative of real world performance. The POWER8 processor also has big, high-bandwidth caches at each level of abstraction, further strengthening the memory subsystem.
- **SMT**: The per-core 8-way simultaneous multi-threading provides an incredible boost to memory-heavy applications, allowing the processor to hide memory latency over many threads. This means improved performance in applications like graph traversal. Thread scaling is further helped by support for hardware transactional memory, which removes the overhead of locking mechanisms in software.
- **Embedded NVLink**: Currently, POWER is the only architecture that supports the NVLink protocol, providing fast data transfer between GPUs. NVLink supports transfer rates 5 and 12 times faster than the more common PCIe3 protocol. Combined, the four NVLink ports on the Minsky system provide a 80GB/s full duplex link between the GPUs, minimizing the bottleneck presented by CPU to GPU transfers.
- **CAPI**: The Coherent Accelerator Processor Interface allows for tighter integration between the POWER8 processors and the P100 accelerators, allowing for better heterogeneous performance. It accomplishes this through two means: (1) providing a unified virtual memory space; meaning that the accelerators and the CPUs can use the same memory addresses, reducing device driver overhead, (2) permitting

accelerators to behave like normal threads, reducing overhead with respect to cache coherency, as that is now taken care of in hardware.

B. The GPU Advantage

We chose to build a cluster consisting of a small number of nodes based on a combined CPU/GPU platform as opposed to a larger cluster of smaller purely CPU based nodes. A survey of the literature suggested that using GPUs to accelerate computation would provide significant performance improvements at a lower cost and power than adding another identical purely CPU-based node. For instance, Matsuoka et al. [5] showed that 66% of their performance on the LINPACK benchmark came from their GPUs which only accounted for 15% of their total power consumption. We hope to replicate their results with our LINPACK execution, showing that the GPUs yield higher performance returns than would be expected from their power budget.

Getting down to numbers, Iqbal et al. [6] obtained 9X speedup on HPL by adding four K80 cards to a system that originally only had two Intel Xeon E5 processors. The quad-card setup also showed a 2.7X improvement in power efficiency across the entire system, as measured in GFLOPS/W. An older benchmark run by Phillips and Fatica [7] from NVIDIA shows an 8.2X improvement in raw performance, a 5.45X improvement in performance per dollar, and a 4.5X improvement in performance per watt when adding two Tesla C2050 GPUs to a dual Intel Xeon X5550-based system. These results support our decision to add more GPUs as a way to achieve higher performance with a lower financial cost and lower power consumption, as opposed to building a cluster with a larger number of nodes.

Given the P100's massive 10.6 TeraFLOPS peak single-precision performance, fed by 720GB/s memory bandwidth and backed by 16GB of high-bandwidth memory, choosing NVLink allows us to minimize the bottleneck presented by CPU to GPU data transfers. Forgoing the incredible transfer speeds NVLink affords us would mean sacrificing performance to shuttling data between the host and the device, which is something we want to avoid at all costs.

IV. Software Selection

A. Operating System

We plan on using Ubuntu Server 16.10 for our system. The Power8 system only supports two operating systems: Red Hat Enterprise Linux (RHEL) 7.1+ and Ubuntu 16.04+.[8] We have chosen Ubuntu for two primary reasons. The first reason is that Ubuntu is free while RHEL requires a subscription, and we prefer being cost effective. The second reason is that we have more experience administering Debian systems.

B. File system

After investigating numerous file systems, we have narrowed our choice down to four options: Parallel Virtual File System (PVFS), Lustre, IBM Generation Parallel File System (GPFS), or BeeGFS. Although the IBM GPFS matches the vendor for our hardware, GPFS is mostly designed for extremely high-performance machines as opposed to a 2-node cluster [9]. The same applies for the Lustre filesystem, and the concern we have by including such a massive file system would be the performance loss incurred due to overhead, although Lustre does benefit from being a completely kernel based file system [10]. This results in a lowered amount of context switches to user space, which could benefit performance. On the other hand, PVFS is a userspace process [11]. This would help with errors as it would be much easier to restart PVFS if it is killed than Lustre, for instance. PVFS also has some performance benefits by having implementations for a non-blocking message interface for many different networks, including Infiniband which we will use. BeeGFS is also a fairly popular and performant file system [12]. However, it emphasizes its scalability which will not help us greatly as our cluster has only two nodes.

C. Toolchain

Our hardware guided our toolchain selection. In general, we elected to use vendor specific tools, as they are clearly fine-tuned for our machines. Since our machine consists of IBM Power8 CPUs with NVIDIA P100 GPUs, this translated to IBM's proprietary toolkit and NVIDIA's CUDA toolkit.

- **Compilers**: We will be using the standard GNU compilers as well as IBM's XL C/C++ compiler and CUDA C/C++.
- **BLAS**: We will use the Engineering and Scientific Subroutine Library (ESSL), IBM's BLAS library, for math applications like LINPACK. We were motivated to use a vendor-specific BLAS after comparing the performance differences between them and the standard HPL implementation from the University of Tennessee, Knoxville on appropriate systems.
- **Message Passing Interface**: We intend to use IBM's Spectrum MPI, since it supports CUDA-aware MPI, which allows the GPUs to pass messages directly to each other, reducing unnecessary copies through the host or the network. This allows us to minimize message passing overhead.
- **Profilers and System Monitors**: We will use a variety of miscellaneous tools to fine tune the performance of our system. For instance, to diagnose performance bottlenecks (e.g. communication latency, thread creation) in MPI and OpenMP applications, we will use *Allinea*, an OpenMP and MPI profiler. We will also use *Ganglia* to monitor our system, and ensure our application runs have good load balancing. To monitor our GPUs' performance, we will use the NVIDIA visual profiler.

D. Scheduler

We are currently investigating whether or not to use a scheduler during the competition. We hypothesize we will *not* receive a significant hit on productivity by manually scheduling our applications. Because our cluster only contains two nodes and six users, we could use system monitors to manually schedule jobs such that no two competing applications run concurrently. This provides us more fine-grained control of our applications' execution, which we *may* find beneficial.

V. Power Consumption Concerns

For our Power8 system to support its four NVIDIA P100 GPUs, it requires a 220-volt power supply. Although the competition's requirements state systems be powered through “110-volt range, 20-amp circuits [13],” the stance on 220-volt supplies was unclear to us. In particular, we were unsure whether we were permitted to transform the 110-volt power lines to 220-volts using a voltage converter, or whether we required to use the 110-volt supply directly.

We contacted John Cazes, the Deputy Chair of the Student Cluster Competition for clarification. He confirmed we were permitted to use a voltage transformer, provided we were able to plug our transformer into the Geist Power Distribution Units.

VI. Sponsorship

A. Vendor Relationship

Our company sponsors are IBM, Avnet, and Flagship. They will provide our cluster hardware and fund hardware shipment to Georgia Tech. We expect to receive the first node in August and the second in late September. In the ensuing time, IBM has loaned us a Power8 cluster until June. We have been using this machine to investigate optimizations for HPL and HPCG.

During the months preceding the competition, we will work with IBM to ensure we utilize our systems to peak capacity. In particular, we will work with two researchers in IBM’s High Performance Analytics department at the T. J. Watson Research Center, Dr. Xing Liu and Dr. Jee Choi. Both are close colleagues with our advisor, Dr. Oded Green.

Their individual expertise align with the demands of the competition. Dr. Liu’s research interests lie in sparse linear algebra, large scale molecular simulations, and analyzing fundamental kernels like SpMV. Moreover, he is an early adopter of the Power8 system. Dr. Choi has written extensively on energy and power modeling for HPC applications [14]. We plan to leverage their extensive experience during our preparation.

B. Institutional Support

Our team’s travel expenses will be supported by Georgia Tech’s School of Computational Science and Engineering (CSE) as well as it’s newly formed Institute of Data Engineering and Science (IDEaS). This includes the cost of shipping our machine to the conference. We provide our projected budget in Table I.

C. Contact Information

For questions regarding our system’s architecture, the best points of contact are:

- **Team**: Oded Green oded.green@gatech.edu
- **Team**: Chirag Jain cjain@gatech.edu
- **Team**: Will Powell will.powell@cc.gatech.edu
- **Avnet**: Pat Fleming pat.fleming@avnet.com
- **IBM**: David Carpin dacarpin@us.ibm.com
- **IBM**: Clifton Jones cejones@us.ibm.com
- **Flagship**: Ed Turetzky eturetzky@flagshipsg.com

VII. Acknowledgments

We thank our advisors for challenging us to write this proposal, as well as for mentoring us throughout this process.

TABLE I
SUPERCOMPUTING 2017 TRAVEL BUDGET

Description	Estimated Cost	Quantity	Frequency	Total Cost
Airport Transport	$10	7-people	4-trips	$280
Airfare	$550	7-people	1-flight	$3850
Per diem	$70	7-people	6-days	$2940
Hotel Room (provided by SCC)	$0	7-people	7-nights	$0
Conference Registration (provided by SCC)	$0	7-people	7-nights	$0
Hardware Transport	$5000	1	1	$5000
Onsite expenses	$1500	1	1	$1500
Total Cost				**$13,570**

References

[1] (2016). Reproducibility application for the student cluster competition at SC16.

[2] D. M. Jacobsen and R. S. Canon, “Contain this, unleashing docker for HPC,” *Proceedings of the Cray User Group*, 2015.

[3] F. Moreews, O. Sallou, Y. Bras, *et al.*, “A curated domain centric shared docker registry linked to the galaxy toolshed,” in *Galaxy Community Conference 2015*, 2015.

[4] J. D. Gelas, “Assessing IBM’s power8, part 1: A low level look at little endian,” Tech. Rep.

[5] S. Matsuoka, T. Aoki, T. Endo, A. Nukada, T. Kato, and A. Hasegawa, “GPU accelerated computing—from hype to mainstream, the rebirth of vector computing,” *Journal of Physics: Conference Series*, vol. 180, no. 1, p. 012 043,

[6] S. Iqbal, M. Deshmukh, and N. Dandapanthula, “Game-changing extreme gpu computing with the dell poweredge C4130,” Dell HPC Engineering, Tech. Rep., 2015.

[7] E. Phillips and M. Fatica, “CUDA accelerated linpack on clusters,” GPU Technology Conference, 2010.

[8] (2015). Supported linux distributions for power8 linux on power systems.

[9] F. Schmuck and R. Haskin, “GPFS: A shared-disk file system for large computing clusters,” *Proceedings of the Conferenced on File and Storage Technologies*, 2002.

[10] T. K. Petersen, “Inside the lustre file system,” Seagate, Tech. Rep., 2015.

[11] P. H. Carns, W. B. L. III, R. B. Ross, and R. Thakur, “PVFS: A parallel file system for linux clusters,” *Proceedings of the 4th Annual Linux Showcase and Conference*, 2000.

[12] J. Heichler, “An introduction to BeeGFS,” ThinkParQ, Tech. Rep., 2014.

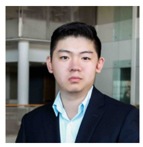
[13] (2017). Student cluster competition 2017 rules.

[14] G. Lawson, M. Sosonkina, and Y. Shen, “Energy evaluation for applications with different thread affinities on the intel xeon phi,” in *Computer Architecture and High Performance Computing Workshop (SBAC-PADW), 2014 International Symposium on*, IEEE, 2014, pp. 54–59.

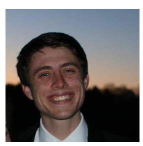
[15] D. Fang and D. Chau, “M3: Scaling up machine learning via memory mapping.,” *Proceedings of the 2016 International Conference on Management of Data.*, 2016.

Appendix I. Our Members

Petros Eskinder is a fourth year computer science major whose interests lie in graph algorithms and high performance computer architecture. As an undergrad at Georgia Tech, he served as a teaching assistant for courses on algorithms and computer systems. He previously interned at Google and Facebook, where he developed tools to automate build-file creation and made *pretty* mobile screens. Last year, he launched a viral mobile application that was featured on various media publications like Business Insider, CNN, and NBC News.



Dezhi "Andy" Fang is a second year computer science major. He was first exposed to HPC through a data analytics research project where he reimplemented the lower-level communication of a big data framework to use memory-mapped I/O. The results of his work were published in SIGMOD'16 [15]. Andy has interned with Symantec where he created a malware visualization pipeline using Hive and Hadoop. He has also served as technical lead for a Chinese startup. He will be responsible for managing our usage of Docker.



Nick Fahrenkrog is a fourth year student and the only electrical engineer on the team. He possesses a wide breadth of experiences, ranging from conducting integrated circuit fabrication research, squashing security issues at Texas Instruments, and developing full stack web applications. He is an avid programmer, frequently attending and winning hackathons. He was first introduced to HPC through a class on MPI, pthreads, CUDA, and OpenGL. He intends to leverage his experiences by focusing on power management and providing a holistic perspective from the transistor level upwards.

Manas George is a third year computer science student. He is a member of Georgia Tech's High Performance Computing Lab, where he works on cuStinger, a GPU-based dynamic graph processing framework. He was a research intern at Apple where he built a data processing pipeline for the iTunes app submission process. He leads the computer security club at Tech, giving presentations on security vulnerabilities, as well as setting up competitions that involve administering multiple systems under attack over a period of 24-48 hours.

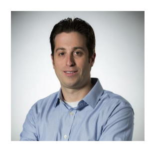
David Meyer is a third year CS major, who became interested in HPC through reading about Beowulf clusters. For the past two years, he has worked as a teaching assistant for courses on Object Oriented Programming and Computer Organization. He is currently developing a web application to elegantly visualize genealogy data. Ever since, he has been searching for a ways to expand his knowledge on HPC.

Jessica Rosenfield is a third year computer science major who has served as the president of the College of Computing's Undergraduate Council and was recently awarded Outstanding Senior in Computing. This semester she is researching high performance computing algorithms to improve DNA alignment and assembly software. Jessica has interned twice with Square where she worked in distributed systems and customer insights. She plans to work at Snapchat this summer in data engineering.

Alok Tripathy is a second year CS major. His research focuses on designing and implementing parallel streaming graph analytics for massively multi-threaded systems. He has interned at Sandia National Labs where he implemented a distributed cache coherency protocol in Go, and at Bloomberg designing and implementing machine learning features to extract tables from PDF files. He is a recipient of the Sidney Goldin scholarship for outstanding leadership abilities. In his free time, he likes to play capture-the-flag (CTF) computer security and competitive programming competitions.

VIII. Appendix II. Our Advisors

Chirag Jain is a third year PhD student in the School of Computational Science and Engineering at Georgia Tech. His research interest lies in designing scalable serial and parallel algorithms for combinatorial problems in genomics. In 2016, he was a summer fellow at the National Institutes of Health. He is the recipient of the Reproducibility Initiative Award at Supercomputing'16.



Oded Green is a research scientist in the School of Computational Science and Engineering at Georgia Institute of Technology, where he also received his PhD. Oded received his MSc in electrical engineering and his BSc in computer engineering, both from the Technion, Israel Institute of Technology. Oded's research focuses on improving performance and increasing scalability for large-scale data analytics using a wide range of high performance computing platforms.

Will Powell is a research technologist in the School of Computational Science and Engineering at Georgia Institute of Technology. Will has over 20 years experience in the IT industry including almost 12 years at IBM. Will is an expert in designing high performance servers and clusters. He'll assist the team with designing the team's cluster and teaching the students good system administration techniques.