team swarm Georgia Institute of Technology

learn more at andf.us/team-swarm

hello world!

We are Team Swarm, Georgia Tech's inaugural team at the Student Cluster Competition!

Each year, Georgia Tech's presence at the Supercomputing conference is felt: from *papers in the main conference and workshops to faculty participations in panels and keynotes to students participating in the multiple poster sessions*.

For this year's Supercomputing conference, we aim to further expand our presence by competing in the Student Cluster Competition. We are a team of seven talented undergraduate students mentored by HPC experts, staff and doctoral students in Georgia Tech.

Statement On 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.

Our team is a direct reflection of this commitment. Although we are all either Computer Science or Electrical Engineering majors with prior interest in HPC, our team members' geographic backgrounds span four states, three countries, and three continents. Including our advisors, even reach five countries

our team

Andy Fang *Junior*
Team Captain
Focus Areas: Software Infra
Power Management, Born

Petros Eskinder *Senior*
Team Captain
Focus Areas: HPCG
Reproducibility, Proposals

Jess Rosenfield *Senior*
Computational Biology
Focus Areas: MrBayes
Born Seismic Imaging

Alok Tripathy *Junior*
Graph Analytics and GPUs
Focus Areas: Linpack
HPCG, Power Management

Nick Fahrenkrog *Senior*
Cloud Computing
Focus Areas: Reproducibility
MrBayes and Cloud

David Meyer *Senior*
Operating Systems Expert
Focus Areas: Born Tuning
Mystery Application

How We Formed Our Team

Early January, our advisors posted an announcement on our college's mailing list inviting students to join GT's inaugural team. Soon after, they held an information session. From the attendees, each member was selected based on their enthusiasm and prior HPC experience.

How We Prepared

Weekly Meetings: We meet each week as a team to discuss the applications, our experiences tuning them, as well as general HPC topics relevant to competition

***Specializations*:** Each team member has taken ownership of at least one component of the competition. Alongside general learning, *they become our local experts* for that specific focus area.

***Vendor Support*:** We have benefitted greatly from our vendor's assistance. IBM and NVIDIA have provided optimized binaries for HPL and HPCG. They have also assisted us with setting up our file system.

***Coursework*:** Alongside fundamental CS courses, several members have also taken advanced courses in Computational Biology, Numerical Linear Algebra, Compilers, Operating Systems, and Parallel Computing Architecture.

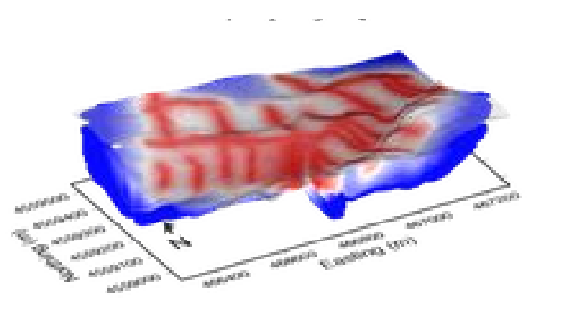
Although this will be our first time competing in the Student Cluster Competition, we believe we've assembled a *competitive* team of passionate and capable students.

our optimizations and strategy

Benchmarks (HPL & HPCG): We are using optimized binaries for both HPL and HPCG. These have been provided by our sponsors NVIDIA and IBM. This has allowed us to consistently achieve ~750 GFLOP/s on HPCG and ~22 TFLOP/s with HPL, while staying within the 3000 Watt power limit.

Since without optimizations our machine regularly hits over 3800 Watts, much of our focus has been staying within the power limit. We have employed various optimizations like CPU frequency scaling and setting a GPU power limit.

MrBayes: Our biggest optimization comes from tuning the parameters and using the Beagle library to run our bayesian and maximum likelihood calculations on GPUs. Additionally, on larger files, we can speed convergence by frequently running independent runs with more chains. This means that calculations take place on the CPU rather than GPU.



Born Seismic Imaging: Some of the optimizations we have explored include vectorization, exploiting cache locality and employing the GPU's. Depending on the mystery application, we may run Born in the cloud. We have found it much easier optimizing Born for an x86 cloud instance rather than our Power machine. Additionally, this would allow us to avoid the challenges of the Power Shutoff Activity.

Reproducibility Task: We will be reproducing the results on our Power machine. After the author ported the paper's optimizations to Power, we worked with the author through Github to ensure we successfully applied them to our machine. This challenge is exciting for us, as the original paper claimed performance portability but did not test on our vector ISA.

Mystery Task: To prepare for the mystery task, we have examined previous competition's applications by compiling and running them. If advantageous, we intend to leverage the cloud. One challenge we may have is running the application on our Power machine. If running the application requires considerable effort and we have any remaining compute hours, we intend to call up our friends at Cycle Cloud.

TURN IT OFF

Power Shut Off: We explored using general checkpointing for the power shutoff activity. However, we concluded that, provided we run Born in the cloud, the application durations were not long enough for a power loss to be particularly damaging. We have found moderate success using Mr. Bayes built-in checkpointing, and may use it if the power is shut off during a Bayes run.

our system

Hardware Configuration

Two node IBM Minsky system (Power S822LC).

CPU	2 x IBM POWER8+	10 core 64 threads,
GPU	8 x NVIDIA P100	with NVLink
MEM	2 x 256 GB	DDR4 223
HDD	2 x 7200rpm	1TB drive

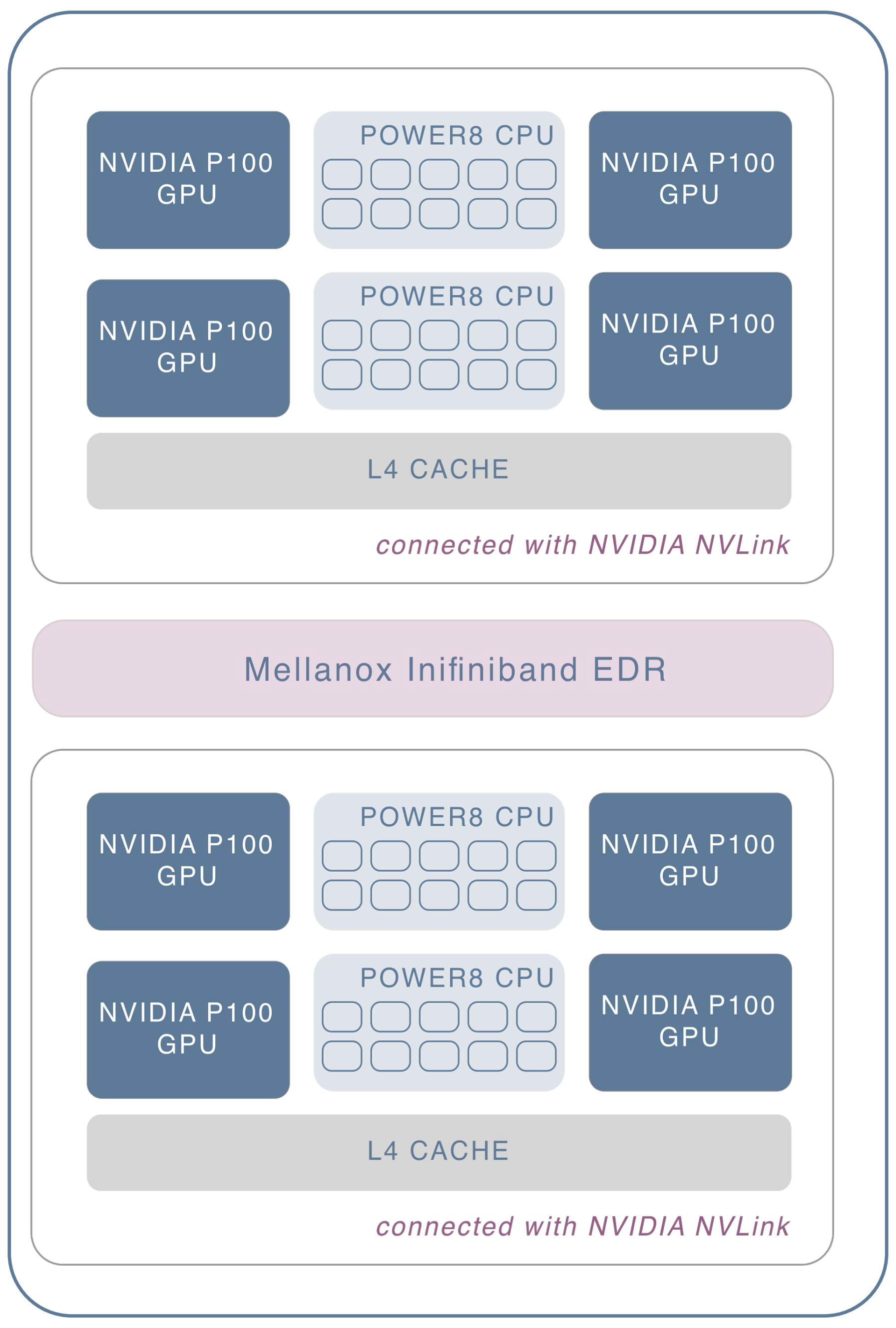
For our **interconnect**, we are using *Mellanox ConnectX-4 InfiniBand EDR*, the fastest interconnect on the market.

How We Chose Our System

Why Power8: The POWER8 architecture is the first processor to come out of the OpenPOWER consortium. We found the following features most attractive:

- Embedded NVLink
- Support for Mellanox Infiniband
- Fast Memory Subsystem

GPU Advantage: We chose to build a cluster consisting of a *small number of nodes* with multiple GPUs because GPUs provide significant performance improvements at a lower financial cost and a lower power consumption than adding another identical purely CPU-based node.



Software Configuration

Operating System	CentOS 7.4
Scheduler	Slurm
File System	NFS v4
Compilers	GCC, IBM's XL C/C++ CUDA C/C++
MPI	OpenMPI compiled with Infiniband supportive libraries
Math Library	ESSL (IBM version of BLAS)
System Monitoring	NVIDIA-smi and Datadog
Profiling	Allinea, Flame Graphs

In general, we let our hardware dictate our selection of system software, so we typically chose IBM versions of tools. For example, CentOS 7.4 was the only OS that both the Power8 system AND the ESSL libraries supported.

Also, we found that MPI compiled for Infiniband outperformed IBM's Spectrum MPI and other reference implementations.

Where this hardware first approach failed was with GPFS, IBM's high performing file system. We explored using it. However, given our limited number of nodes and limited time, we found using NFS a better option.

We would like to thank

Dr. Oded Green
Advisor, School of CSE

Chirag Jain
Advisor, School of CSE

Will Powell
Advisor, School of CSE

our advisors

our sponsors