Application Profiling and Power Management for the Student Cluster Competition

Alex Ballmer*, David Ghiurco*, Ryan Mitchell*, Ryan Prendergast*, Hasan Rizvi*, Iva Veseli*, William Scullin+, Ben Allen+, Ioan Raicu*+

*Department of Computer Science, Illinois Institute of Technology, Chicago, IL
+Argonne Leadership Computing Facility, Argonne National Laboratory, Lemont, IL

Our Team

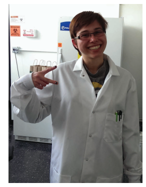
Alexander Ballmer is a 4th-year Computer Science undergraduate at IIT. He was a member of the IIT SCC teams in both 2014 and 2015. His focus is on LAMMPS, Born, and the cloud component. He is the team's system administrator.

David Ghiurco is a 4th-year Computer Science student at IIT. He has interned twice with HERE Technologies, LLC, and has also done research in the IIT DataSys lab. This is his first time as an SCC official contestant. He is the team's cloud component expert.

Ryan Mitchell is a senior at Adlai E. Stevenson High School intending to major in Computer Science. He has a background in Linux, server administration, and virtual machines. He is the team's system admin and benchmarking specialist, as well as being responsible for the mystery application.

Ryan Prendergast is a senior at Maine South High School, intending to major in Computer Science. Ryan has attended hackathons and worked extensively with Arduinos. His responsibilities include LAMMPS, Born, the mystery application, and visualization work.

Hasan Rizvi is a 3rd-year IIT Computer Science undergraduate with an Applied Math minor. Hasan is a computer science TA and frequents local hackathons. Over the summer, he performed an OS viability study for our hardware. His current task is integrating Mr Bayes with the cloud component.



Iva Veseli is a 4th-year undergraduate at IIT, majoring in Biology. She is a budding computational biologist and has spent over three years working in a genomics research lab on campus. Due to this background, she is responsible for MrBayes and visualization.

Our School

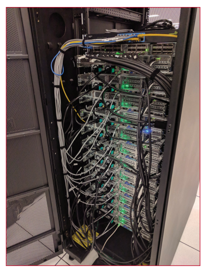
Illinois Institute of Technology (IIT) is a private, technology-focused research university offering undergraduates and graduate degrees in engineering, science, architecture, business, design, human sciences, applied technology, and law. IIT is centrally located in Chicago. For the past three years, IIT has worked closely with Argonne National Laboratory to send a team of students to the Student Cluster Competition at the annual Supercomputing Conference.

Our team is comprised of four IIT undergraduates and two Chicago-area high school seniors, as well as two backup members who are IIT undergraduate students. We have been working since the summer to explore and profile the competition applications in preparation for SCC17.

Hardware

Type	Hardware	Quantity
Chassis	Intel Wolf Pass with Calyos Heat Pipe Coolers	12 nodes
CPU	Intel Xeon Platinum 8180	2 sockets/node
Memory	384 GB DDR4-2666	384GB/node
Storage	Intel NVMe SSDs	2/node
Accelerators	NVIDIA Tesla V100	8
Interconnect	Intel OmniPath	

- Calyos coolers use liquid evaporation to efficiently cool each processor, reducing amount of fans in the system
- The Intel 8080's two AVX-512 FMA units/core gives us an advantage when running highly vectorized applications
- Our RAM fills all six memory channels for optimal memory bandwidth and performance
- NVMe SSDs using 3D XPoint memory allow for low-latency, high bandwidth IO
- OmniPath's on-load method of transferring data offers more performance



Software

Fedora 26
- More stable than other, more optimized versions of Linux
- Better package availability
- New enough to support modern features and instruction sets

Intel Compilers
Intel MPI
Intel MKL
- The Intel tools are better optimized for Intel processors and Intel OmniPath than other competing tools

Intel Data Center Manager (DCM)
- Integration with the Intel Management Engine on-die allows DCM to control power at a per-node level by changing processor power states
- Integration with SNMP and IPMI allows for cluster-wide hard power caps
- Easier to manage the power budget

IBM Spectrum Scale
- High-performance, low latency file system that is excellent for IO-intensive tasks

Slurm job scheduler
- SLURM schedules jobs on the cluster
- Extremely configurable but simple to use

Spack package manager
- Allows for rapid deployment of packages using different combinations of libraries and compilers

Preparation Strategy

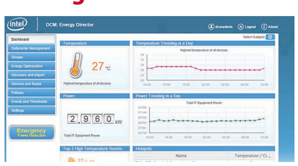
We started preparing as a team in June. Throughout the summer, we worked three days a week, with pairs of team members focusing on each of the competition's applications.
- Obtained a deep understanding of the applications, their structure, and their input files.
- This will especially benefit us with the application interviews and the reproducibility component of the competition.

We modeled the scalability of each application and determined the most efficient suite of parameters. Experiments included:
- Intra-node threading through OpenMP and MPI
- Inter-node scaling through MPI
- CPU frequency scaling
- Compiler optimization flags

Some team members focused on other competition aspects.
The cloud component:
- Determined which node types are the most efficient per dollar
- Determined which applications to run on the cloud

Visualization:
- Visualization tools for MrBayes, LAMMPS, and Born
- Ten 16x16 LED squares on the side of our cluster for system monitoring

Power Management



The primary constraint in the competition is a total cluster power consumption limit of 3000 Watts. To stay below this limit while maintaining high application performance, we have:
- Set a hard power limit of 3000 Watts over the entire cluster using Intel DCM
- Profiled application power consumption to inform scheduling decisions such that:
  - We can modify certain parameters to save energy without significantly impacting performance
  - Power-intensive applications can be run concurrently with those that use less energy
- Established system monitoring (including power usage)
- Determined contingency power-saving measures that can be taken, should the DCM power capping policy fail

Application Optimization

Pre-Compilation Optimizations
- Compiler usage (GNU vs Intel)
- Compiler optimization flags, such as O2, O3, fp-model, xHost, etc.
- Static vs dynamic compilation of binaries

Run Time Optimizations
Intra-node scalability:
- Determining the ideal number of cores per node to run on
- Whether to use pure MPI parallelism or a combination of MPI+threading (via OpenMP)

Inter-node scalability:
- Deciding whether each application scales well across multiple nodes

CPU frequency scalability, subject to power constraint (3 kW):
- Determining whether it is more efficient to maximize node count with a slower clock speed, or to maximize clock speed with fewer nodes

We used Allinea Map and Intel vTune to profile applications and identify resource bottlenecks.
- Allinea Performance Reports identified the primary bottleneck (CPU, MPI communications, or I/O)
- Profilers showed us which functions took up most of the runtime
- Profilers also confirmed the degree to which applications utilized vector instructions

Acknowledgments

This work would not have been possible without the generous support of Intel, Calyos, and NVIDIA, who provided requisite hardware and technical support throughout the project. In addition, this research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. We gratefully acknowledge the computing resources provided and operated by the Joint Laboratory for System Evaluation (JLSE) at Argonne National Laboratory. Special thanks to Alexandru Orhean for his assistance.

The Team. From top left: Alex B, Iva V, Ryan M, David G, Ryan P, Hasan R.
From bottom left: Ioan Raicu, William Scullin, Alexandru Orhean.

Why We Will Win

- Most of our team members have been working on the applications since June 2017.
- Each application has been studied extensively and profiled by at least two team members.
- We also have three team members responsible for the cloud component, and several working on visualization.
- We have practiced with an 8-hour mock competition including interviews and a 'mystery application' (WRF).
- We have had excellent support from our advisors, backup team members, and sponsors who have provided their experience and advice to help us prepare well.
- We have a steady supply of candy to fuel us throughout the competition