THE HARRISON HIGH SCHOOL
SUDO WRESTLERS

MEET THE TEAM

Everyone on the team is from William Henry Harrison High School in West Lafayette, Indiana in the United States. We are all excited to be the first full high school team to compete at this level of the Student Cluster Competition. The Sudo Wrestlers were all originally a part of **F.I.R.S.T.** Robotics Team 1747, making up most of the team's programming subteam. We became aware of SCC when two volunteers from Purdue University came to a meeting and gave a presentation on the competition. It got our attention and we have been preparing since.

There were certain challenges that faced the team as preparation began. One issue being that none of us had any experience regarding cluster computing and only very limited knowledge of Unix. Another issue of the team being diversity, an issue that many high school STEM activities face. Through months of meeting and the guidance of Lev Gorenstein and Chuck Schwarz, the technical challenges were gradually overcome and as they did the team's knowledge grew. Now, after starting from nothing, we are able to run our scientific applications and use the our newfound technical skills. We are excited to participate in the year's competition and show what we have learned!

The team is a group of primarily juniors, with 1 senior and 1 sophomore. The members are interested in fields of Math, Science, Engineering, Programming, Machine Learning, and Cyber Security. Among the field of STEM the interests of each member are diverse. Members are (Clockwise from left to right) Eduardo Wedekind, Nick Nolte, Jeremy Meyer, Chris Page, Jacob Sharp, Callum Gundlach, and Nick Frooninckx. Advisors are Chuck Schwarz (left) and Lev Gorenstein (right).

CYCLECLOUD

CycleCloud, the newly introduced Cloud computing component of this years competition has proven to be very valuable tool. It is perfect for serving as a second place to run processes that are too resource intensive for the local cluster. Born is an excellent example of such an application, and therefore will be run exclusively on a CycleCloud cluster. Not only is it a second environment for computing, but we have also often used it for testing system administration processes, such as installing and running torque before installing it on Raider. The cloud clusters have shown the be a great place for providing information without the risk of breaking our real machine. All around, it has been a great tool for the team to experiment with.

BORN

CALLUM & JACOB

Born is a seismic analysis tool used for surveying for oil and gas deposits below the surface of the Earth. Experiments are run and the data collected by each experiment is organized into a 'shot'. When Born is run it processes the data shot by shot. This makes the application 'embarrassingly parallel' and scale very effectively.

Scaling Studies showed that Born was multi-threaded, but not MPI-capable, and thus capable to fully utilize multiple CPU cores in one node but not scale to multiple nodes. The application also appeared to use up all of the available resources on a single node, and run for hours at a time. From these discoveries it was inferred that Born was an excellent candidate for running on CycleCloud. The application simply uses too much of the limited local CPU resources to work on other apps at the same time, and while not being MPI-capable it would able to run on cheap instances without premium interconnect. Born was discovered to use approximately 34GB of RAM, and instances were tested accordingly. The tests showed that an instance with 16 cores gave the best performance/cost, and said instances will be used to run the application at the competition.

Cores	Total Memory (gb)	Shot Time	Cost ($/hour)	Cost Per Shot ($)
4	28	N/A*	$1.22	N/A*
8	56	32581	$0.59	5.37
16	112	15214	$0.94	3.96
20	140	19726	$1.85	10.15

*The 4-core instance lacked enough RAM to complete the shot.

MEET RAIDER

Architecture:

- 2 HPE ProLiant XL270 nodes
- Apollo D6500 chassis.
- Dual-socket, 12-core Intel Xeon E5-2650v4 (Broadwell) processors
- 512GB DDR4 2400MHz RAM
- 4x480GB Intel SSDs
- HP InfiniBand EDR 2-port 840QSFP28 Adapter (Mellanox ConnectX-4)
- 4 NVIDIA V100s per node (with 16GB of memory and 5120 CUDA cores each)

Software

- CentOS v7.3
- Torque Resource Manager
- BLCR Checkpointing
- OpenMPI & IntelMPI
- GCC & Intel Compilers

NVIDIA V100 GPU

LAMMPS/TERSOFF

NICK F. & JEREMY

LAMMPS is molecular dynamics simulator that computes interactions and movements of many atomic particles to simulate physical properties of molecular systems. The Tersoff Multi-Body Potential is one of the force fields used to model covalent systems. When computing pairwise atomic interactions, Tersoff potential uses information about current atoms' bonds with other neighbors. While improving the simulation accuracy, this approach is computationally expensive.

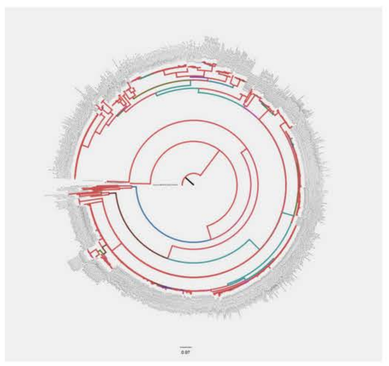
Unlike other applications, for the LAMMPS/Tersoff task, the goal was not to demonstrate raw performance, but rather to reproduce the findings of the SC16 Gordon Bell Award finalist paper by M. Höhnerbach and coauthors [1]. The paper proposed an efficient, scalable and portable vectorized implementation of the Tersoff potential for LAMMPS. A digital artefact (in the form of Github repository [2] with source code and computational experiments) was provided by authors. We were able to successfully reproduce the experiments and confirm original observations on our hardware platform.

[1] M. Höhnerbach, A. Ismail, and P. Bientinesi. 2016. "The Vectorization of the Tersoff Multi-Body Potential: An Exercise in Performance Portability". In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis (SC '16)
[2] https://github.com/HPAC/lammps-tersoff-vector

MRBAYES

NICK N. & CHRIS

MrBayes is a program that uses Bayesian inference to evaluate genetic trees. It uses the Metropolis-Hastings algorithm to find the best tree by randomly changing the tree and comparing it to the original. It was used to analyze whitefly DNA in order to make the cassava plant in Africa resistant to a disease carried by whiteflies. We ran tests to determine the best values for nchains, nruns, and nswaps for speed and performance. We also experimented with the beagle library but found that it did not speed up the runs or improve convergence. As it turns out, for MrBayes the beagle library is most useful with codon or amino acid models and not DNA.



HPL & HPCG

NICK N. & CHRIS — JACOB & JEREMY

High Performance Linpack is a very common benchmark that tests the raw computing power of a system and its rank on the TOP500 list, while the High Performance Conjugates Gradient benchmark is a complement to the HPL benchmark designed to represent a diverse range of applications. Through the series of scaling studies and power monitoring, the team learned how to run HPL and HPCG efficiently and with optimal parameters. Power usage has been profiled throughout the runs, and GPU frequencies were adjusted to stay within power budget.

MATRIX INTEGRATION

NVIDIA

Hewlett Packard Enterprise