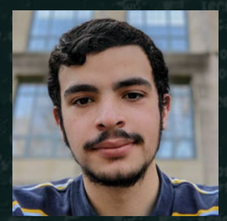
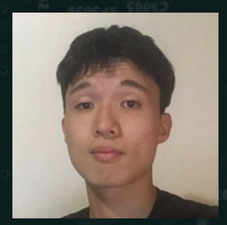
MASSACHUSETTS GREEN TEAM @ SC21

PROFILES



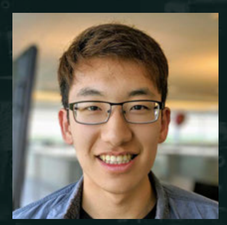
Carlton Knox
Class of '23

Mike Klein
Class of '24



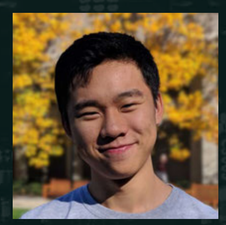
Howie Chen
Class of '23

Richard Kumahia
Class of '22



Benjamin Li
Class of '22

Reproducibility



David Shen
Class of '22

OUR TEAM

TEAM INTRODUCTIONS

We are the Massachusetts Green Team, formed from a partnership between three Universities in the Boston Area: Boston College, Boston University, and University of Massachusetts Lowell.

Coming from diverse backgrounds around the world, we are bonded by the same enthusiasm for high speed computing and algorithm optimization. In forming this team, we reached out to underrepresented groups within the student bodies at our colleges to make sure that we could put together the most accessible and equitable team this year. Richard is a first generation African American, a group which is vastly underrepresented in computing. Carlton is a second generation Mexican American, so by participating in the SCC, he wants to make his voice heard in a space where Hispanic representation is low.

Our advisor this year is MIT's Kurt Keville, who has been providing support to Boston-area teams for as long as we can remember. He is frequent speaker at the Boston HPC + GPU group and was the chair for IEEE HPEC this year. Kurt has spent the past few months connecting our team with industry partners and providing feedback for our cluster designs.

We would like to acknowledge the Boston University SPARK! Ignite Council for funding our club and providing resources to facilitate our growth. We would also like to give thanks to our competition advisor Kurt Keville and our club advisor Martin Herbordt from BU. We could not have made it this far without their guidance!

OUR SYSTEM

We will be bringing a single node computer consisting of two AMD EPYC 7763 CPUs as well as 4 Nvidia V100s. Each CPU will have 256 GB of DDR4 ECC RAM, for a total of 512 GB of RAM in the system. Our storage will be NVMe SSDs. Because we are running a single node, we will not need any type of communication card. Our estimated power consumption will be 2000W under load.

Our system will be running CentOS 7.6, which has proven in the past to be well suited for HPC related applications and benchmarks. The 7.6 version specifically provides the necessary management tools and driver support for our system. Slurm is our scheduler of choice, and it also doubles as our job monitor. Members of the team have used Slurm in the past, and so are familiar with it. We will be using the Linux utility CRIU to take process snapshots/checkpoints.

OUR SPONSORS

WHY WE'LL WIN

OUR PREPARATION

Our team prepared extensively during the Summer and Fall semesters, leading workshops on Parallelization, Optimization, Cluster Configurations, Benchmarking, and Large-scale Simulations. Each member of our team has specialized in one of the competition applications, leading research and optimization efforts in accordance with their skills & background. Each team member is an expert in their benchmark/application, which can be found on the top of the poster.

OUR STRENGTHS

The Massachusetts Green Team brings together a diverse group of students with a variety of skills developed through our computer science and engineering backgrounds. As a team, we have shared our knowledge to build a well rounded and competitive group. With our combined expertise, we know we can solve any HPC problem thrown at us efficiently and optimally.

OUR STRATEGY

One the ways our team will stand out from other competitors is by using custom compilers for our workloads. We plan to use the AMD Optimizing C/C++ Compiler for the Reproducibility challenge as it contains libraries that are optimized for AMD processors. We also plan to leverage CUDA-X, a collection of GPU library implementations like cuBLAS and NCCL for inter-GPU communication.

Moreover, our team has been investigating the potential of using a distributed filesystem in the event that the competition goes completely virtual.

During the leadup to the competition, we were analyzing how best to use our power budget. Due to our system falling under the power limit at load at all times, our main focus was on application scheduling. We determined an algorithmically optimal knapsack-based strategy to run our applications.

APPS

OPTIMIZATION TECHNIQUES

The initial plan is to perform some preliminary analysis of the executions of the competition applications. This would determine where various bottlenecks occur and how best to solve them. Afterwards, performance tuning can be done to find out exactly what kind of software stack gives the best performance. Finally, validation will need to be done to ensure that the system chosen works well against the budget constraint and will produce good results.

In terms of optimization, each application lends itself to being split across multiple executions units. This is what the team is most comfortable with, and it will be the focus for the one-month period leading up to the competition. This will be the hardest part of warming up to the competition, but it also where there is the most room for improvement.
In preparation for the mystery application, we have peeked into the past mystery applications to try to gain some insight there.

APPS & BENCHMARKS

HPL: The standard benchmark for dense linear equations. Our usage of the NVIDIA Tesla V100 should give us good results with HPL. Additionally, we plan to profile and optimize our system with Nvidia Profiler to improve our score as we prepare for the competition.

HPCG: A broad test on Conjugate Gradient for ranking top systems. The GPU focus in our system should yield strong results. HPCG 3.1 supports the newest Volta and Ampere cards.
Quantum Expresso: Quantum Espresso is a nanoscale electronic structure simulator. Because we are using CUDA GPUs, we are able to leverage the GPU version of the code.

Cardioid: Cardioid is a cardiac simulation suite developed at LLNL. We can again leverage our V100 GPUs to achieve better results.

IO500: The IO500 benchmark is a benchmark suite for High-Performance IO. We have worked with IO500 in past competitions and are confident in our system to achieve a competitive result.

ramBLe (Reproducibility): The Reproducibility challenge runs exclusively on the CPU. Our system should provide good results to base our paper on. We have been exploring the code and testing different configurations to better our submission.