

LA-UR-15-27495

Approved for public release; distribution is unlimited.

Title: Student Cluster Competition: A Multi-disciplinary Undergraduate HPC Educational Tool

Author(s): Harrell, Stephen L
Nam, Hai Ah
Vergara Larrea, Veronica G.
Keville, Kurt
Kamalic, Dan

Intended for: EduHPC-15 Workshop on Education for High-Performance Computing, 2015-11-16 (Austin, Texas, United States)

Issued: 2015-12-01 (Rev.1) (Draft)

Disclaimer:

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the Los Alamos National Security, LLC for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

Student Cluster Competition: A Multi-disciplinary Undergraduate HPC Educational Tool

Stephen Lien Harrell
Purdue University
West Lafayette, IN
sharrell@purdue.edu

Hai Ah Nam
Los Alamos National
Laboratory
Los Alamos, NM
hnam@lanl.gov

Verónica G. Vergara
Larrea
Oak Ridge National
Laboratory
Oak Ridge, TN
vergaravg@ornl.gov

Kurt Keville
Massachusetts Institute of
Technology
Cambridge, MA
klk@mit.edu

Dan Kamalic
GNS Healthcare
Cambridge, MA
dkamalic@gnshealthcare.com

ABSTRACT

National labs, academic institutions and industry have a strong need for scientists and staff that understand high performance computing (HPC) and the complex interconnections across individual topics in HPC. However, domain science and computer science undergraduate programs are not providing sufficient educational resources, and are far from conveying the interdisciplinary and collaborative nature of the HPC environment. The Student Cluster Competition (SCC) was created as an educational tool to immerse undergraduates in HPC. It is a microcosm of professional HPC centers that teaches and inspires students to pursue careers in the field. The SCC's impact is reflected in new undergraduate curricula and through the experience of the students themselves. The SCC can complement a strong parallel and distributed computing (PDC) curriculum through experiential learning and engagement with the HPC community as a whole, which will prepare graduates for the interdisciplinary nature of work in HPC fields.

Categories and Subject Descriptors

K.3 [Computing Milieux]: Computers and Education;
C.5.1 [Computer Systems Organization]: Computer System Implementation—*Super (very large) computers*

Keywords

HPC, Student Cluster Competition, education, undergraduate, curriculum, PDC

1. INTRODUCTION

High performance computing (HPC) has become an essential tool to advance scientific discovery over the last twenty years [12], and is itself, an area of active research to create larger systems to accommodate new modes of scientific discovery with complex workflows [17]. With the push to Exascale, large-scale scientific programming and HPC are becoming more important to achieving national goals, further increasing the importance of adequately educating the next generation [18]. On July 29, 2015, the White House issued an executive order [19] establishing the National Strategic Computing Initiative (NSCI) to ensure the US continues its leadership role in HPC and its use in solving complex scientific and national problems. The NSCI will make strategic investments toward exascale computing and data-driven HPC. However, there is a significant deficiency in current academic curricula, to prepare the next generation to use, develop, and innovate in HPC at the undergraduate level or even, often times, at the graduate level.

In a domain science degree program (physics, biology, chemistry, etc.), computational science is, at best, offered as an elective and exposure to parallel computing is only attained through a computer science degree program. Furthermore, the majority of computer science degree programs only provide a basic parallel programming course, which fails to expose students to the complex and inter-disciplinary nature of the HPC environment. Efforts are underway on a national level to address this educational deficiency through the *Curriculum Initiative on Parallel and Distributed Computing* [20] led by the National Science Foundation and the IEEE Computing Society. They provide a more comprehensive HPC curriculum encompassing HPC architecture, programming and algorithms, and incentivize early adopters through additional support and stipends for instructors. This early effort is targeted toward integrating HPC into computer science (CS) and computer engineering (CE) programs.

These efforts still leave a deficiency in HPC exposure for students in domain science degree programs, who often become the applications developers of the future. Science application developers are often poorly equipped to develop massively parallel applications without considerable on-the-job training. Code design choices, both methods and algo-

©2015 Association for Computing Machinery. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of the United States government. As such, the United States Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only.

EduHPC2015, November 15-20, 2015, Austin, TX, USA

©2015 ACM. ISBN 978-1-4503-3961-2/15/11\$15.00

DOI: <http://dx.doi.org/10.1145/2831425.2831428>.

gorithms, are increasingly impacted by HPC architecture and require a shared language to facilitate close collaborations between domain scientists and CS/CE researchers to bring it to large-scale HPC systems. In addition, no curriculum exposes the complexity of a production HPC ecosystem, which includes, not only a massively parallel HPC system, but is also comprised of networks, file systems, archival storage, programming tools, and administrative tools, which can significantly impact scientific productivity.

The Student Cluster Competition (SCC) was developed in 2007 as a mechanism to expose undergraduate and high school students to HPC through immersion. With the help of a vendor partner student teams design and build the hardware and software for a small cluster, learn designated scientific applications, apply optimization techniques for their chosen architecture, and compete in a non-stop 48-hour challenge to correctly complete a real-world scientific workload at the SC conference, all while impressing conference attendees and interview judges with their HPC knowledge.

The SCC was developed prior to efforts to integrate HPC education into undergraduate curriculum. Identifying this deficiency early, the original intent of the competition was to provide students with early exposure to HPC and motivate undergraduate curriculum development. Going forward, despite undergraduate HPC curriculum slowly becoming main stream in CS/CE programs, the SCC remains an invaluable tool in HPC education. The SCC reaches a multi-disciplinary student population and provides domain scientists with, often times, their first exposure to HPC. The competition also exposes teams to an HPC work environment, requiring team work and broad understanding of systems, software and applications by all team members.

Some unique impacts of the SCC to HPC education include providing students the opportunity to work with cutting-edge hardware and software seen in the Top500 HPC systems. In addition, SCC applications and rules are chosen so that teams are aware of the interconnections between system architecture and applications, and competition applications are chosen so that teams can see the societal impacts of HPC (e.g. simulating weather and catastrophic events can impact urban development, simulating zombie invasions is similar to understanding disease spread, etc.). The competition also enables students to become inspired by HPC and use the playful nature of a competition to take risks and drive their own education. And, by staging the final event during the SC conference, the SCC allows students to see the vibrancy of the HPC community, identify role models, network, establish mentorship opportunities, and expose students to career possibilities in HPC.

This paper describes the SCC in detail and its continued impact in HPC education. In Section 2 we describe current HPC workforce needs. Sections 3 and 4 provide background on the Student Cluster Competition and describe qualitative and quantitative impacts of the SCC. We conclude with suggestions for broadening the SCC and increasing its potential impact on HPC education.

2. HPC WORKFORCE NEEDS

High performance computing is pervasive in research, academia, industry and government facilities. According to the Top 500 List for June 2015 [9], the top 500 highest performing HPC systems were distributed among the following market segments, presented in Fig. 1: Industry 42%, Research

(e.g. National Labs) 26.2%, Academic 23%, Government (e.g. Army, Air Force, NASA, etc.) 5.6%, and Vendor 3%.

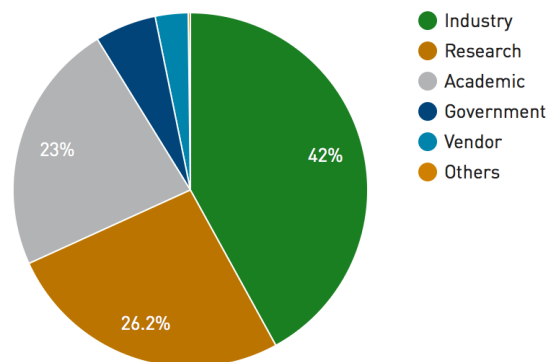


Figure 1: Percentage of the Top 500 systems in each HPC segment for June 2015 [9]

In recent years, there has been a significant uptick in demand for applicants who are versed in HPC methodologies. This trend has been spurred by increased ability to take advantage of inexpensive computational resources to achieve faster or more accurate solutions to problems that were formerly solved serially on single workstations, or deemed unfeasible given the time and budget constraints of the organization.

A 2010 International Data Corporation (IDC) study [16] of talent and skill impacting HPC data centers found the HPC workforce aging and retiring, and 93% of HPC centers having difficulty hiring staff with the requisite skills as a result. A 2013 Networking and Information Technology Research and Development Program (NITRD) publication [14] concluded, “current approaches to HEC (high-end computing) workforce development and education are inadequate to address today’s needs in HEC centers and scientific disciplines that depend upon HEC; as demands upon HEC increase, this gap will widen.”

HPC penetration is increasing across the board in industry. A 2014 IDC forecast [15] predicted the greatest increase in HPC systems revenue from 2013 to 2018 in electronic design automation, IT, and independent software vendors (9.5%). Government labs took the second place prediction at (9.1%), with the economic and financial industry in third place at (8.7%), and university and academic increases further down at 6.7%. It is significant to note the pace of HPC growth in industry keeping step or outpacing that of traditional sources of HPC workforce employment such as government labs and academia.

2.1 HPC workforce roles at national labs

The collaborative nature of the work at national laboratories requires not only a strong academic and technical background but also excellent interpersonal, communication, organization, and problem solving skills. In the high performance computing area, national laboratories often look for different types of professionals including: computational scientists, data scientists, system engineers, system administrators, and user assistance staff. It is often difficult to find professionals with the desired HPC background, described in Table 1, to fill these roles. In addition, the staff listed

need knowledge beyond any specific skill set but also need to understand the interconnectedness of all the technologies and roles involved.

Table 1: Roles in HPC and desired skill set

HPC Role	Skill Set
Computational Scientist	Computer science, including but not limited to computer architecture, operating systems, and algorithms, and expertise in a specific scientific domain and its applications
Data Scientist	Data analytics, machine learning or statistical modeling with a strong understanding of big data technologies such as Hadoop/MapReduce and NoSQL databases
System Engineers	HPC architecture in systems, storage and/or networking as well as systems programming for monitoring, change management and repair
System Administrators	Cluster or HPC system, network or storage administration and a deep understanding of how each systems influences the others
User Assistance Specialist	Use of batch systems, workflow, compilers, programming languages used in scientific computing, profilers and debuggers, and great communication skills

2.2 New skills for emerging architectures

To further complicate the problem, the fast-paced evolution of HPC systems, for example heterogeneous systems with accelerators, requires all HPC staff to be agile and constantly educate themselves to drive the next evolution. The new skills required include:

Computational scientists are required to have experience with modern programming models including CUDA, OpenCL, OpenMP and OpenACC compiler directives, vectorization, and message-passing interfaces (MPI). Furthermore, computational scientists must demonstrate ability developing, profiling, and optimizing scientific applications for a variety of architectures. Computational scientists collaborate with users and work closely with the user support staff to identify and disseminate best practices.

Data scientists must have in-depth knowledge of computer science, as well as extensive knowledge or practical experience with “big data” technologies such as MapReduce, Hadoop, Spark, and NoSQL databases. Data scientists often work with researchers in many different fields to provide assistance understanding their data needs, as well as analyzing and improving their workflow. In order to efficiently operate in HPC environments, data scientists must also have experience with modern programming languages and analysis tools. In addition, data scientists are also often required to have experience with the latest visualization techniques and tools.

System engineers often have a computer science or computer engineering background. In industry this may be referred to as DevOps. The position usually requires a balance between theoretical and technical knowledge. System engineers have in-depth technical knowledge of the architecture

of HPC systems. In addition, they must gain an understanding of current and upcoming technologies in order to improve and develop the next generation of tools. Generally, system engineers have a particular specialization (e.g. storage, networking, compilers) and experience with modern programming and scripting languages. Their technical expertise in combination with their programming and analytical skills allows them to develop tools and systems that can monitor/diagnose the HPC ecosystem, manage change within the system, as well as automate some repairs. System engineers work closely with user support staff in order to develop tools that can improve the overall user experience.

System administrators for HPC systems are required to have practical experience managing large-scale systems. Previous experience deploying, configuring, and managing HPC Linux clusters is also generally required. In addition, system administrators must have a thorough understanding of Linux, the different layers in the software stack, networking, and storage. Due to the rapid evolution of technology, systems administrators must stay up to date with the latest technologies. Furthermore, system administrators must be able to quickly diagnose and troubleshoot operational issues which often involves knowledge on many levels of the stack. They are also responsible for software updates and upgrades to maintain the availability and security of the systems at the center. The systems administration group works closely with several groups inside and outside the organization, and therefore, the role requires excellent organizational as well as interpersonal skills.

User assistance specialists provide support to scientists that use and rely on large-scale HPC systems for their research. Because users enter with different backgrounds, a user assistance specialist must be capable of tailoring technical information to specific audiences and must be able to explain complex systems clearly and succinctly. This type of role requires adaptability, and excellent problem solving skills. From a technical standpoint, the position requires practical experience with multiple programming and scripting languages including but not limited to Fortran, C, C++, Perl, Python, and shell scripting. In addition, knowledge of parallel programming models and previous experience with Linux clusters, batch schedulers, and parallel programming tools (profilers and debuggers) is beneficial.

2.3 HPC workforce in academia

Organizational structure varies from university to university much more than national labs and as such there may be many different types of organizational structures for research computing. A common structure used is a central research computing organization that is attached to the IT or Research administrative department.

If the need at national labs is depth then the need in academia is breadth. The academic needs focus less on development of cutting edge HPC resources and architectures and instead focus on deployment and use of existing architectures. While research on these new and upcoming architectures is done at many universities, this type of research does not typically get deployed into production at universities first. HPC centers at universities are typically helping to push faculty’s research forward. While there is certainly HPC application development at many universities, most users of academic HPC resources will be focused on using existing codes. Although all of the jobs enumerated for the

national labs are relevant in academia, the different focus translates into a different distribution of need and responsibilities among the jobs.

For instance, the lab’s computational scientist role responsibilities are distributed between user assistance specialists and the faculty/graduate students that are users. Specialists are expected to know some of the domain science and they often focus on educating users who often have never done a simulation or used a command line interface before. Data scientists are similarly distributed. While there may be a data scientist role in an HPC center, it can also take the form of a technologist (systems engineer or user assistance specialist) who understands the “big data” tools and a statistics faculty. Systems engineers and systems administrators are very similar although they may be more focused on deployment than their lab counterparts.

3. STUDENT CLUSTER COMPETITION

The Student Cluster Competition (SCC) debuted at the Supercomputing Conference in 2007 in Reno, NV. Since then it has been a growing component of the SC Conference and will celebrate its ninth year at SC15. The overarching goal of the SCC is to introduce the next generation of students to HPC and integrate them into the HPC community. This is achieved through a friendly yet spirited competition, where teams of six undergraduate and/or high school students assemble small clusters at the conference, and race to complete a real-world HPC workload, across a series of applications, and to impress HPC-industry judges. In the competition, teams partner with vendors to design and build cutting-edge clusters, from commercially-available components, and work with application experts to prepare for the competition. All with a catch, clusters must not exceed the 3,120-watt power limits (26 amps at 120 volts).

3.1 Microcosm of the HPC Community

The steps the teams undergo to prepare and compete resembles an HPC systems procurement and acceptance testing (e.g. Trinity-NERSC8 [2], CORAL [3]); whereby HPC centers describe a set of requirements through a list of applications and performance metrics, vendors propose a system to meet some or all of the requirements, and upon delivery of the system, a comprehensive acceptance test is performed to ensure the HPC system meets the contractual functionality, performance and stability to accept the machine in exchange for payment.

The SCC culminates in a real-time, non-stop, 48-hour challenge during the SC conference in November that is similar to an acceptance test. However, the competition really starts in the spring when the competition applications are announced. Typically 4-6 large-scale scientific applications are the basis for the competition, where teams will need to learn how to run and optimize the applications for their chosen architecture. Ideally, the competition applications should inform their decisions on their system architecture (e.g. if a GPU version of an application does not exist, does it make sense to bring a system with GPUs). Based on the competition workload, teams must find vendor partners and submit a proposal describing their anticipated system and their strategies to achieve optimal performance for the workload. See Harrell et al. [13] for a description from past advisors on team creation and the preparation required to participate in the competition.

Accepted teams then work with their vendor partners and advisors to build a suitable cluster for the competition workload, setup the programming and tools environment (schedulers, compilers, MPI, libraries, profilers/debuggers, etc.), decide on storage needs (file system, SSDs, etc.). In addition, teams are encouraged to find domain scientists and applications developers to assist them with understanding the competition applications (i.e. what are the use cases for the application, what are the impacts of changing parameters to the science outcome and the performance, what input and output are expected, etc.).

During the SC conference, the winning team is decided based on the highest aggregate score when combining their correctly completed workload across the competition applications, best HPL run, application interviews and HPC interviews. The interview portion assesses whether teams understood the applications, their system and engaged with the other conference activities. Often, teams go above-and-beyond through intricate visualizations of the application output to impress conference attendees and interview judges.

Although the competition doesn’t require students to go in depth into parallel computing techniques, typically the team that wins the overall competition has taken the extra step towards understanding the parallel applications and the programming model to understand how best to execute the codes for their chosen architecture.

3.2 SCC and HPC Education

The SCC was developed prior to emphasis on integrating HPC into undergraduate curriculum. Its original goal was two-fold, one, to expose and excite students about HPC and two, to motivate and influence curriculum development. Competing teams require considerable training to prepare for the competition at the SC conference. The topic areas students need to prepare for the competition could be used toward guiding HPC curriculum.

Organizers of the SCC are equally focused on creating a high-quality competition and increasing the exposure and integration for each student participant in the HPC community. Over the years, the SCC has demonstrated and continues to refine its unique impact on HPC education and building the next generation of experts to fill the HPC pipeline. Some of these benefits include:

Multi-disciplinary reach.

The premise of the SCC, where domain applications are run and optimized for cutting-edge HPC systems, attracts CS/CE students as well as domain science students, who are often recruited on teams for their science expertise. During SCC interviews, teams are required to explain their knowledge of competition applications and its use. Often, this is the first exposure domain science students receive in HPC and the applicability of their field of study into a broader landscape.

Team work.

The scope of the competition is so broad that even a seasoned HPC expert could not single-handedly run a team and must rely on team work, similar to a real HPC production environment. In a team of six students, the work to understand the 4-6 applications is typically distributed among team members, as well as the cluster architecture and programming environment setup. The SCC also pro-

vides students with early exposure to communicating across domains. Often in HPC, scientific experts from different domains must find a common language to collaborate in the HPC environment.

Student initiative.

The competitive yet playful nature of the SCC helps to drive student initiative, versus course requirements, and attracts students to the field through a fun activity. The target age, undergraduates, are drawn to fields for their ‘cool’ factor, such as using drones and robots to spark interest in embedded systems. The SCC gives students a competition framework to drive their participation but does not impose strict rules on how to compete. They are encouraged to make design decisions with the guidance of their vendor partners and advisers, and they are provided cutting-edge hardware, ‘real-world toys,’ on which to be creative.

Impacts of HPC.

In addition, competition applications and data sets are chosen to spark their imagination and allow students to see the broad impact and applicability of HPC in society. Applications have included password cracking important to cybersecurity, WRF [10] to simulate weather patterns, such as the Katrina hurricane, with impacts to urban development, and NAMD [8] to simulate biomolecular systems to accelerate new drug design.

Integration in the community.

One of the attractions to compete in the SCC is for students to travel to the SC conference and experience one of the largest HPC conferences. Student participants are encouraged to attend technical talks, the job fair, the mentor-protégé program, and network with other attendees. Students are also constantly presenting their team, architecture and know-how to attendees visiting the competition area. This exposure provides students with a glimpse at the diversity of career paths in HPC and gives them role models and future mentors.

4. SCC IMPACT

There has not been a formal mechanism to quantify the impacts of the SCC, in large part, because the organizing committee is comprised of volunteers and changes from year to year. For this study, we analyzed the team proposals starting from 2010, when consistency in the submission guidelines was implemented. Team proposals are applications to participate in the SCC. The contents of the proposals include the expected members of the team, the expected hardware configuration, the current HPC curriculum at the institution, and a range of other information requested by the SCC committee [1]. This data is self reported and can be considered biased towards getting accepted into the competition. All proposals are used in this study regardless of the status of acceptance. Data was gathered from 87 applications from 40 distinct teams to the SCC from 2010 to 2015.

4.1 SCC Statistics

4.1.1 Interest in the competition

Continued interest in the competition from organizations

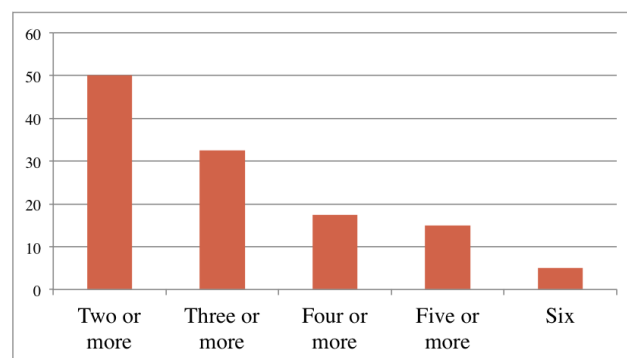


Figure 2: Percentage of teams with recurring proposal submissions from 2010 to 2015.

is a measure of usefulness of the SCC to the HPC community since it requires institutional commitment in staff, funding and other types of support. From all teams within this survey 51% have applied to the SCC twice or more, and over one third have applied more than 3 times within the last six years as seen in Figure 2. This high recurrence rate indicates that the SCC has a perceived value to undergraduates and institutions who have participated before.

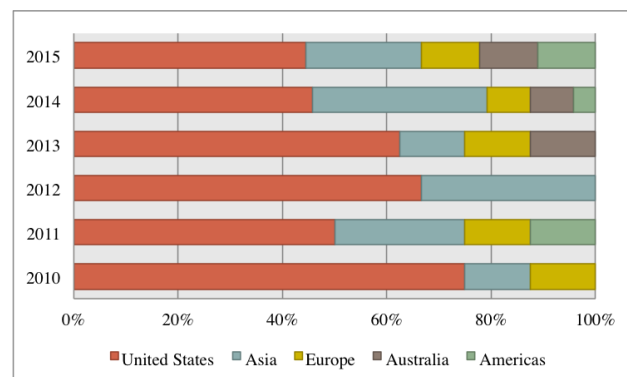


Figure 3: Percentage of applications broken down by continent and the US.

In addition, Figure 3 shows increased interest internationally to participate in the SCC. In 2010, 75% of all team proposals came from schools within the United States. In 2015, the increase in international team submissions resulted in less than 50% submissions from the US. Between 8 to 12 teams have been accepted into the competition each year, with total submissions ranging from 11 to 17. Submissions from Europe, Australia and the Americas is still limited, resulting in a 100% acceptance rate seen in Figure 4, with the exception of 2012. The majority of submissions come from the US and Asia, resulting in lower acceptance rates.

The growing international interest in HPC is also reflected in the percent of system share of the Top500 list, when analyzed by geographical region [9]. In 2010, North America represented 56.4% of system share and Eastern Asia at 14.2%, whereas in 2015, North America had 47.8% and Eastern Asia increased to 17.6%.

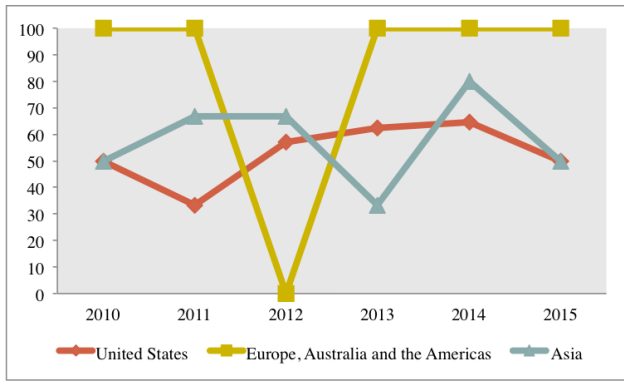


Figure 4: Acceptance rates for US, Asia and Other.

4.1.2 Curriculum Development

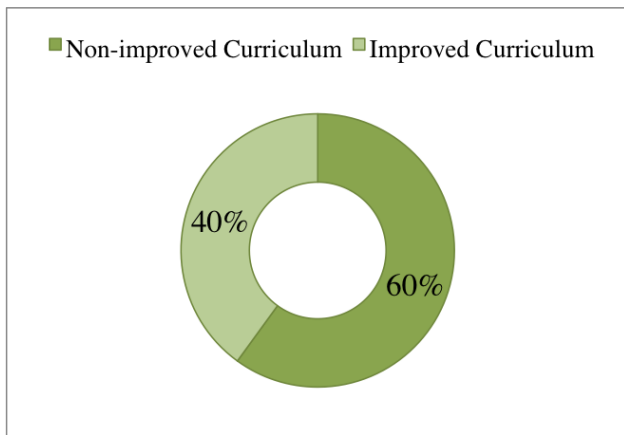


Figure 5: Percentage of repeat teams that have improved their HPC curriculum from 2010 to 2015.

Information about the team’s existing HPC coursework is also collected. The call for applications asks for respondents to “Explain the commitment of the institution to educating the broader student community about the usefulness and the accessibility of High Performance Computing at your institution; explain how cluster computing is integrated in the educational curriculum of the proposing institution.” [1] The answers are graded on a three point scale. Some respondents choose not to answer the prompt. These responses were scored as zero. The lowest non-zero score is given to institutions that only have a parallel programming course. The middle score is given for having a parallel computing course but also are offering parallel architectures and/or offering domain specific parallel computing courses. The top score is given to those that describe holistic HPC programs that include HPC system deployment and use as well as parallel programming classes in domain sciences.

During the first few years of the competition, teams received their training for the competition as an extra-curricular activity or a special project. Advisers and team members had to see the benefit of spending additional time preparing for the competition in exchange for the opportunity to learn about HPC and experience the SC conference. Since then, many teams from returning academic institutions have

adopted more formal coursework to prepare students for the HPC workforce and, by proxy, the competition. Figure 5 shows the percentage of repeat teams that have improved their curriculum offering in HPC.

4.1.3 Domain Science Majors in the SCC

Multi-disciplinary work is important in many of the roles in the HPC workforce, yet it is rarely integrated into undergraduate curriculum. The SCC gives many students a first look into the inter-disciplinary nature of HPC. The majors of the students were gathered from each application. These percentages are averaged per year and then averaged together to correct for teams who submit subsequent applications. As seen in Figure 6, we see close to 1/6th of students having a non-computing major or an average of one student per team that is not a major in computer science, computer engineering or information systems. This means that a majority of the SCC applicants are exposed to domain science majors not only in the competition but also on their team. The non-computing majors are also similarly exposed to a new area that they may never have been exposed to otherwise.

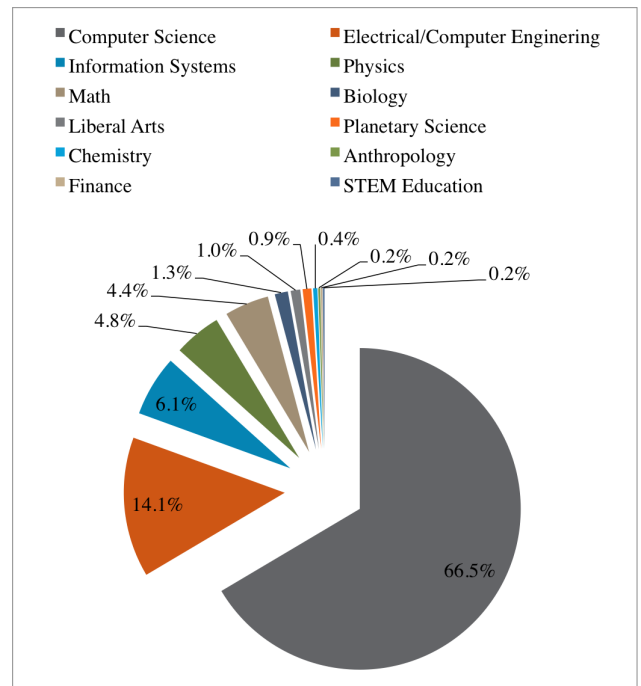


Figure 6: Diversity of undergraduate majors in SCC applicants from 2010 to 2015.

4.2 Participant Experience

In 2013, the SCC committee distributed a survey to past participants to attempt to gauge the competition’s impact on students. A common theme emerged: every one of the survey respondents expressed that many of their teams’ students have moved on to graduate or industry work in HPC, and over half of the respondents expressed that HPC education was not commonplace at their university, and the competition exposed them to this field which they would not have had an opportunity to explore otherwise. In addition, several respondents cited the career networking op-

portunities they received from being part of the larger SC conference and being able to attend other parts of the technical program.

The impact of the competition can perhaps be best summed up in quotes from past students and advisors:

“The Cluster Competition was a great way to learn skills that are not taught whatsoever at school. I’m a Computer Science major, but [my university’s] CS curriculum is highly theory-based, and it was a relief to be able to learn the concrete skills needed to set up, manage, and maintain a cluster, as well as the work it took to learn the scientific applications. The conference as a whole was memorable—combining the work of the competition during the day with the networking at night—it was one of the high points of the school year, and one that opened up new career possibilities to me.”

This is an area of education that is lacking formally... and having this competition opens up the eyes of many undergrads who would have never considered HPC as an option.”

“Our team concluded that SC12 itself was the BIG experience. That is, coming from a small liberal arts college setting, it was horizon-expanding to see what the ‘big boys’ are doing. The competition itself had all the ‘normal’ highlights: teamwork, all-nighters, collegiality with other teams, bug-fighting, etc. But it was the experience of being with the HPC community that made the big difference.”

“Huge amounts of student interaction between teams - friendly competition is awesome!”

“Almost all of the students go to graduate schools. Some of them become my masters and PhD students. They are all doing research in HPC. The SCC is a great event for undergraduate students who never have HPC experiences during their undergraduate study... This event guides them into the HPC community.”

“As only sublimely confident undergraduates can do, my team decided last night to scrap the design we’ve been working for the whole summer and take a radical left hand turn using hardware we don’t actually own yet. It made me so proud! These 5 guys are mastering so much material and digging into technical spec’s in a way I’d never get if they were just taking a class.”

4.3 Outcomes

Anecdotally, many students who are past participants at the SCC go into HPC related jobs. One advisor for a regularly participating team said that about 30% of students from their teams go into HPC for a career. Additionally many students intern in HPC related jobs even before they finish their degree. Although there is not data for every student who has participated in the SCC there is a growing group of SCC alumni who participate in the Supercomputing Conference, have jobs at national labs, in HPC industry, and are in graduate programs related to HPC.

5. CONCLUSIONS

The growing importance of high performance computing to solve complex problems, as well as the fast-paced evolution of the HPC environment, has heightened the need for a trained and experienced workforce to meet these challenges. The HPC community needs to address this challenge by providing the next generation with the necessary education, mentorship and inspiration to join this exciting field. The complexity of the HPC landscape calls for the need for

both structured undergraduate educational curriculum and more multi-disciplinary educational experiences. Additionally, the community needs to remain agile so when a new technology like Hadoop/MapReduce gains wide acceptance it can be quickly reflected in job descriptions and syllabi. The Student Cluster Competition can complement the current push toward integrating HPC curriculum into undergraduate education by providing breadth and context to a depth-wise program. Activities like the SCC provide a vital context to the education to see how they can fit in and get students, that are otherwise uninterested in HPC, inspired to pursue a career in the field.

The Student Cluster Competition has shown that it can inspire undergraduates to pursue HPC. Students that have participated in the SCC are joining the HPC workforce through industry, national labs and academia. The value of the SCC framework has been recognized internationally. There are two other major student cluster competition events [4] [5] that invite teams from around the world and at least one other [6] that is a local variant on the SCC. The international HPC community recognizes the importance of these type of activities to augment undergraduate education and inspire the next generation of HPC professionals.

5.1 Broadening the reach

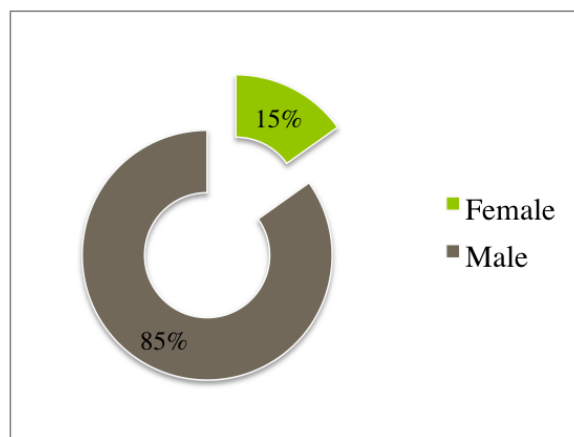


Figure 7: Gender distribution of SCC applicants from 2010 to 2015.

While the SCC has seen many successes, the competition strives to increase its overall impact and reach of HPC. This can be achieved through increasing its current scale and broadening its reach to include greater diversity (e.g. gender, underserved minority, science discipline, institution size, etc.). Analysis of the current composition of participants and their institutions shows that the SCC needs to make targeted recruiting efforts. Figure 7 shows the gender distribution of SCC applicants from 2010 to 2015, with only 15% female participation. In addition, smaller schools with an undergraduate population less than 5,000 are not represented at all in the SCC as seen in Figure 8. These figures show that more work can be done to further increase the diversity in SCC participation.

The SCC is continually trying to grow the competition, to allow more teams to compete each year. This is often constrained by financial limitations as well as space limitations in the competing venue. Through ACM/IEEE sponsorship

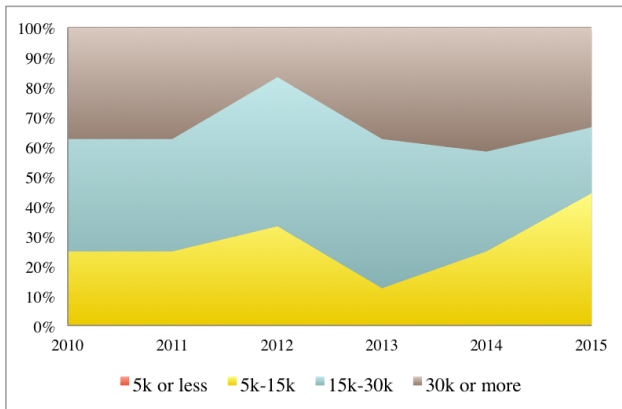


Figure 8: Acceptance rates based on school undergraduate population by year.

from the SC conference and support from HPC industry supporters, the SCC provides competing teams with local accommodations during the conference, registration, and meals. The SCC requires each team to provide their travel costs and system expenses (hardware, software, shipping), provided through their institution and/or vendor partner. Thus far, the SCC has accepted a maximum of 12 teams due to funding limitations.

With increased and stable funding the SCC would be able to expand its reach in size as well as to more underserved institutions. Specifically, the SCC would be able to bring more teams to the competition. The competition could also broadly advertise through targeted and aligned conferences such as the Grace Hopper Celebration of Women in Computing conference [7] and the ACM Richard Tapia Celebration of Diversity in Computing conference [11], as well as visits to underrepresented institutions. Lastly, additional funding would enable the SCC to identify, mentor, and support smaller academic institutions to establish an SCC team and provide additional financial support for travel and resources to teams comprising underrepresented groups.

Within the current framework, the SCC plans to implement team matching and mentorship to expose students from smaller academic institutions to HPC. In this model, past competing teams with existing HPC resources partner with teams that have little access to HPC. Currently, the SCC continues to provide ‘matchmaking services’ for any institution that needs a vendor partner to provide hardware, software and HPC mentorship. Although broader and earlier advertising of the competition is achievable, intensified targeting and advertising requires more time and effort than a volunteer workforce can provide.

The SCC committee plans to continue quantified assessment of the impact of the SCC on students and the HPC community. A comprehensive “where are they now” study will be conducted, which is crucial for the growth and stewardship of the competition.

Acknowledgment

The authors would like to thank Brent Gorda and Ricky Kendall for creating the Student Cluster Competition, the Supercomputing conference series and ACM/IEEE for continuing to host and provide financial support, and the HPC

vendors and supporters who provide systems/software, mentorship, and financial support to make the competition possible each year. A portion of this work was performed at the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725 and Los Alamos National Laboratory, supported by the U.S. Department of Energy contract DE-FC02-06ER25750. The publication has been assigned the LANL identifier LA-UR-15-27495.

6. REFERENCES

- [1] Sc12 student cluster competition. <http://sc12.supercomputing.org/content/student-cluster-competition.html>, 2012.
- [2] Trinity / nersc-8 request for proposal. <http://www.nersc.gov/users/computational-systems/cori/nersc-8-procurement/trinity-nersc-8-rfp/>, 2013.
- [3] Coral request for proposal b604142. <https://asc.llnl.gov/CORAL/>, 2014.
- [4] Isc 15 student cluster competition. <http://www.hpcadvisorycouncil.com/events/2015/isc15-student-cluster-competition/>, 2014.
- [5] Asc 15. <http://www.asc-events.org/ASC15/index15.php>, 2015.
- [6] Chpc student cluster competition. <http://www.chpc.ac.za/index.php/training-workshops/scc>, 2015.
- [7] Grace hopper celebration of women in computing. <http://gracehopper.anitaborg.org/>, 2015.
- [8] Namd: Scalable molecular dynamics. <http://www.ks.uiuc.edu/Research/namd/>, 2015.
- [9] Top 500 list statistics. <http://top500.org/statistics/list/>, June 2015.
- [10] The weather research & forecasting model. www.wrf-model.org, 2015.
- [11] Acm richard tapia celebration of diversity in computing. <http://tapiaconference.org/>, 2016.
- [12] S. Graham, M. Snir, and C. Patterson. Getting up to speed: The future of supercomputing. Technical report, National Research Council of the National Academies, 2005.
- [13] S. Harrell, P. Smith, T. Hoefler, A. Labutina, and T. Overmyer. Methods of creating student cluster competition teams. In *Proceedings of the 6th TeraGrid Conference*, June 2011.
- [14] High End Computing Interagency Working Group. Education and workforce development in the high end computing community. Technical report, NITRD.
- [15] E. Joseph, S. Conway, C. Dekate, and L. Cohen. Idc hpc update at isc’14. http://www.hpctoday.com/files/docs/pdfs/IDC_ISC14_Breakfast_Briefing_slides.pdf, 2014.
- [16] E. Joseph, S. Conway, and J. Wu. A study of the talent and skill set issues impacting hpc data centers conducted on behalf of the us department of energy. Technical report, IDC, July 2010.
- [17] P. Kogge. Exascale computing study: Technology challenges in achieving exascale systems. Technical report, DARPA, 2008.
- [18] R. Lucas. Top ten exascale research challenges. Technical report, U.S. Department of Energy, 2014.
- [19] B. Obama. Executive order – creating a national strategic computing initiative, 2015. The White House Office of the Press Secretary.
- [20] S. Prasad. Nsf/ieee-tcpp curriculum initiative. <http://grid.cs.gsu.edu/~tcpp/curriculum/?q=home>, 2012.