HPC Workforce Development of Undergraduates Outside the R1

Scott Feister
California State University Channel Islands
Camarillo, CA
scott.feister@csuci.edu

Elizabeth Blackwood
California State University Channel Islands
Camarillo, CA
elizabeth.blackwood@csuci.edu

ABSTRACT

Many Research-1 (R1) universities create investments in High Performance Computing (HPC) centers to facilitate grant-funded computing projects, leading to student training and outreach on campus. However, creating an HPC workforce pipeline for undergraduates at non-research-intensive universities requires creative, zero-cost education and exposure to HPC. We describe our approach to providing HPC education and opportunities for students at California State University Channel Islands, a four-year university / Hispanic-Serving Institution (HSI) with a primarily first-generation-to-college student population. We describe how we educate our university population in HPC without a dedicated HPC training budget. We achieve this by (1) integrating HPC topics and projects into non-HPC coursework, (2) organizing a campus-wide data analysis and visualization student competition with corporate sponsorship, (3) fielding undergraduate teams in an external, equity-focused supercomputing competition, (4) welcoming undergraduates into faculty HPC research, and (5) integrating research data management principles and practices into coursework. The net effect of this multifaceted approach is that our graduates are equipped with core competencies in HPC and are excited about entering HPC careers.

KEYWORDS

Undergraduate, Workforce, Education, HPC, HSI, Classroom, Student, Scientific computing, Data management

1 INTRODUCTION

When the authors began as new faculty members at California State University Channel Islands (CSU Channel Islands), each had a breadth of experience at multiple universities in bringing High-Performance Computing (HPC) education materials and training opportunities to undergraduates. However, their experience was entirely with Research-1 (R1) universities, and the transition to bringing HPC education to a campus without dedicated HPC on-site training resources was disorienting. In talking with peers who have made a similar transitions from being data researchers at R1 universities into faculty members at non-R1 institutions, our disorientation was not unique. This paper reflects the learning we have undergone in bringing HPC training opportunities to our students. Our goal is to share these experiences in order to help others undertaking similar efforts.

https://doi.org/10.22369/issn.2153-4136/13/2/2

The authors were accustomed to facilitating outreach and exposure for undergraduates by leveraging dedicated university centers. See Table 1 for a few examples of on-site research HPC resources associated with R1 universities, which also provide training support to their undergraduates. University-affiliated HPC resources such as these are created with the resources and research focus of R1 universities.

Undergraduate students at non-R1, more-teaching-focused universities can equally become valuable members of the HPC workforce. However, in the experience of the authors, opportunities for careers in HPC may not be known by some undergraduates at universities where HPC resources are not directly affiliated with the university. This lack of exposure to HPC education may be a barrier to some students who would otherwise pursue it as a career.

In this paper, we provide our approach to bringing undergraduate students into HPC at a university without dedicated, research-program-affiliated HPC resources. Our reflections will be valuable for other science, technology, engineering, and math (STEM) educators working in teaching-intensive universities who are also seeking to expose students to HPC and to support valuable learning experiences in this field.

2 APPROACH

2.1 Integrate HPC into Non-HPC Coursework

We have integrated HPC-relevant activities into everyday coursework for sophomore-level and junior-level undergraduate courses in Mechatronics Engineering and Computer Science. The first course in which HPC concepts are integrated is EMEC 231: Dynamics. This introductory Dynamics course (the physics of motion) is taught with tie-ins to Python numeric programming. Students use numeric programming tools in Python, including the NumPy library, to simulate the motion of a quadcopter in response to forces and torques along multiple axes. In the process, students are given a hands-on introduction to the idea of time-stepping through a simulation and at each step applying known physics equations in order to achieve a numeric solution to a physics problem. This foundation in scientific computing, specifically when it comes to a basic understanding of time-stepping solutions, translates well to many physics HPC applications running on today's supercomputing clusters.

We also have integrated HPC-relevant programming activities into COMP 262: Computer Organization and Architecture and COMP 362: Operating Systems. In both of these courses, students learn to program shared-memory parallel software using OpenMP. In COMP 262, the focus of these exercises is in scientific computing. For example, students in COMP 262 are instructed to parallelize a simple C program that performs a numeric integral to N threads and then create a scaling document to map the performance as a function of threads. Demonstrations of scaling are essential to

Table 1. Examples of On-Campus HPC Resources at R1 Universities. (Deliberately non-comprehensive. We are specifically highlighting a couple R1 institutions in which the co-authors were affiliated prior to joining CSU Channel Islands as faculty and a couple others for context.)

R1 University	Center Providing On-Campus HPC Education
University of Chicago	Research Computing Center
University of Pittsburgh	Office of Research Computing
University of California Los Angeles	Institute for Digital Research & Education
University of Michigan	Advanced Research Computing
Arizona State University	Research Computing

developing software for HPC applications, and this lab gives them their first experience with creating such graphs. We encourage critical thinking about the impact on a future career through a reflection essay on this activity with questions such as:

- Pick one of your C codes and explicitly walk through how your approach to migrating the code from serial to parallel followed each of these three steps: (1) identify parallelism, (2) express parallelism, and (3) express data locality.
- Explain how this lab might affect your own future choices of where to invest your time while developing parallel software in your career.
- From the perspective of a software developer, what did you learn about the difficulties of taking full advantage of the computing power within parallelized CPU architectures?

A second lab day in COMP 262 focuses students' attention on hardware acceleration and General Purpose Graphics Processing Units (GPGPUs). GPGPUs are an essential component of many cutting-edge national HPC resources, and fully utilizing the GPGPUs for scientific computing involves non-trivial programming. We give students their first exposure to NVIDIA CUDA programming through an exercise in which they offload the physics computation of an exploding bag of particles onto a GPGPU. Thousands of independent particles are modeled with kinematic variables stored in large arrays, and this exercise is an analog to physics simulations performed on supercomputers. In the process, students learn about the challenges of massive parallelization of existing computer codes as well as manage the challenges of memory transfer of large arrays of particles between processing units.

2.2 Plot-a-Thon: Campus-Wide Analysis and Visualization Competition

Visualizing large quantitative datasets is a major component of an HPC workflow. The authors co-founded a campus competition in quantitative data visualization that we call the Plot-a-Thon. We have partnered with several local businesses to provide funding for this event on our campus.

The Plot-a-Thon is an overnight hack-a-thon which engages about a hundred students at CSU Channel Islands from a variety of majors. We provide a large quantitative open dataset for students to analyze, drawn from an open online data repository. On the day of the competition, students in teams of four do their best to create compelling stories from the data. The teams are often interdisciplinary, bringing together students from computer science, business, and the humanities to work together and learn from one another. They learn to clean and analyze datasets using tools at various skill levels of expertise: Excel, Tableau, R, and Python. For the Python teams, we teach them quantitative data manipulation using Python packages that are relevant to HPC: NumPy, Pandas, and Matplotlib. Students who are new to the principles and tools have the option to attend faculty-led workshops where they learn using the competition dataset.

Most importantly, students are taught how to tell a story from their data rather than simply speaking about the numbers. The teams must submit a short video describing their process and the story that they intend to tell. This "soft skill" of communication about quantitative data analysis is a valuable one for a career in HPC and its adjacent STEM fields. The event also features a variety of guest speakers on topics of careers in data science from both local and national companies as well as networking opportunities for students who are seeking internships and jobs. The Plot-a-Thon has not only fostered students' skills with real-world HPC concepts but also created meaningful relationships between faculty and local businesses seeking to hire graduates.

2.3 2022 Winter Classic Invitational Student Cluster Competition

Providing students exposure to national-scale HPC resources and professional HPC workflows, as well as helping them develop a relevant professional networking in HPC contexts, will ultimately carry them into the HPC workforce after graduation. Our university was invited to a supercomputing competition targeted at HSI institutions called the 2022 Winter Classic Invitational Student Cluster Competition. According to the competition website, the competition in 2022 was comprised of "twelve teams of students from Historically Black Colleges and Universities (HBCUs) and Hispanic Serving Institutions (HSIs)." Ten of our undergraduate students competed, forming two of the twelve teams.

In the competition, our students learned about HPC workflows and were given access to HPC training crash-courses led by experts. Then, in a series of several weeks, they were given access to supercomputing facilities throughout the country to participate in bi-weekly competition events with other university teams. Every other week, students competed to compile and optimize computer codes for HPC applications. For example, in one event, the teams optimized a simulation of motorcycle air drag using the computational fluid dynamics code OpenFOAM. The winning team was able to coarsen the simulation mesh sufficiently to provide speedup

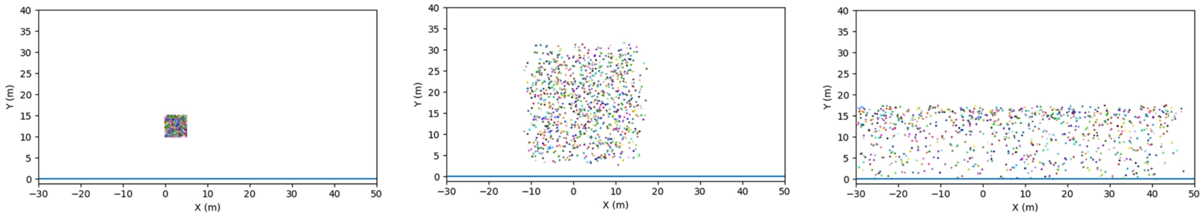

Figure 1. Three time frames from an assignment created by the author (Feister) simulated in C and visualized in Python: "An Exploding Brick of Bouncy Particles." In COMP 262, students offload the core physics time-stepping computation of this simulation onto a GPGPU.

without crossing a threshold in fidelity of the simulation. In addition to learning through doing, students at CSU Channel Islands learned from their competitors afterwards and analyzed each event's outcomes.

The Winter Classic Competition also provided students valuable networking with HPC experts. For example, they set up credentials on NASA computers and Oak Ridge National Laboratory computers. Several CSU Channel Islands students have reported that they have been contacted since the competition with opportunities in HPC and are now considering joining the HPC workforce. This was a high-impact event for our students and we are thankful to have been invited to participate.

2.4 Student Involvement in Faculty HPC Research Projects

Students at CSU Channel Islands can work with faculty members who have ties to the HPC community to develop their skills in undergraduate research projects. These one-on-one mentorship opportunities give students hands-on experience with national supercomputer resources and give them experience compiling and solving problems with advanced scientific codes. One student project worked in a collaboration between Computer Science and Physics to model the trajectory of particles through a proton spectrometer using the Geant4 software. Geant4 is the same software used in medical radiation applications and for high energy physics applications on HPC clusters. Another student project involved a collaboration with faculty in Environmental Sciences modeling sand erosion at beaches along the California coastline. Yet another involved a collaboration with Mathematics to model the cracks that form under rusty conditions in metals. The common link between these various student research projects is that each involved access to faculty member allocations at national-scale supercomputing resources. These projects are the first time any of these students have worked with an HPC resource allocation. Furthermore, each of the codes discussed are massively parallel Message-Passing Interface (MPI) codes written in C or Fortran.

2.5 Integrate Research Data Management Curriculum into Capstone Coursework

We have integrated Research Data Management Curriculum into required capstone coursework for a variety of HPC-adjacent projects. Completion of a capstone project is required for all students at CSU Channel Islands and is performed over one or two semesters during a student's senior year. These projects are mentored by a faculty member and are often based within a subset of the faculty member's research. Many of the aforementioned faculty HPC research projects become students' capstone work. At the beginning of the project, we require students to partake in a Research Data Management Workshop, led by a data librarian, which includes an introduction to the terminology, best practices, and implications of data management.

To solidify these learning objectives from the workshop, students are required to complete a data management plan using the DMPTool (an open online application) for their capstone project at the beginning of the term. Additionally, students are required to submit a final "repository" at the end of the project that includes the data that they plan to archive, relevant documentation, and a "readme" summary. These graded assignments build vital data management skills that students will need for HPC-related fields, especially those that rely heavily on grant funding.

3 ASSESSMENT

One way to assess the impact of these interventions is by examining these in the framework of skills and concepts learned by students that would be needed to enter the HPC workforce. Several were touched on and reinforced for our participating undergraduates through the various interventions.

We created a list of student learning outcomes related to HPC that were met by at least one of these interventions. Then, we compared the relative effectiveness at meeting these outcomes across our various interventions.

The student learning outcomes we assessed in Table 2 are:

(1) Professional Networking in HPC
(2) Remote SSH Access / File Transfer
(3) Multithreaded Programming
(4) GPU Programming
(5) Differentiating HPC Clusters
(6) Compiler Optimizations
(7) Implementing Math/Physics Equations
(8) Domain Decomposition
(9) Troubleshooting with HPC Staff
(10) Managing Compute Resource Allocations
(11) Teamwork in HPC
(12) Data Visualization

(13) Communication of Computational Analysis
(14) Data Management
(15) Writing Technical Documentation

Table 2. Various interventions had different effectiveness in meeting student learning outcomes relevant to HPC workforce development. PT = Plot-a-Thon, FR = Faculty Research, CI1 = Classroom Integration 1 (COMP262), CI2 = Classroom Integration 2 (EMEC 231), WC = 2022 Winter Classic Competition, DMW = Data Management Workshops.

	PT	FR	CI1	CI2	WC	DMW
SLO1		x			x	
SLO2		x	x		x	
SLO3		x	x		x	
SLO4			x			
SLO5			x		x	
SLO6					x	
SLO7		x	x	x		
SLO8		x			x	
SLO9		x			x	
SLO10		x	x		x	
SLO11	x	x			x	
SLO12	x	x	x	x	x	
SLO13	x	x	x		x	
SLO14	x	x			x	x
SLO15		x				x

4 REPRODUCIBILITY

The details of each course integration, campus event, student competition, and faculty project may not be identically reproduced. However, we argue that the overall structure can be reproduced at other universities. The following can be reproduced by faculty at non-R1 institutions using our work here, and the work of others, as example foundations:

- Assigning undergraduates to create computer simulations in non-HPC introductory science courses
- Assigning scientific computing projects, and GPU programming, in non-HPC computer science courses
- Creating campus events promoting communication and visualization of quantitative data
- Seeking out opportunities to engage students in external training and professional networking opportunities like the Winter Classic competition
- Partnering with librarians or other subject experts to build HPC-related training and skills into existing curricula

5 CONCLUSIONS

The approach described in this paper offers a model for exposing students in non-R1 universities to valuable HPC learning experiences. Despite a lack of on-site research-oriented HPC resources and infrastructure, we got students engaged with HPC and equipped them to pursue HPC careers after graduation. Some of the key successes of this effort were student understandings of the challenges and benefits of massively parallel computer programming, students' first access to national research computing facilities, and engagement of students with HPC professionals. The sustainability of such a model is key. We believe that by integrating HPC-focused assignments into existing courses and by organizing extracurricular HPC activities, universities without traditional dedicated computing centers can and will offer valuable HPC learning experiences to their students. The authors plan to further integrate HPC into non-HPC courses, advertise HPC student activities, provide professional networking for students, and seek external collaborations with the national HPC education community in the future to continue to implement these experiences. We also plan to create student feedback surveys on the effectiveness of these interventions and to begin to conduct meaningful assessments that can help us align our practices with the needs of our unique student body. With creativity, willingness, and persistence, educators at universities of all kinds can create a variety of HPC workforce development opportunities for their undergraduate students.

ACKNOWLEDGMENTS

The authors would like to thank Dan Olds for creating and inviting our university to the amazing HPC competition from which our students benefited: the 2022 Winter Classic Invitational Student Cluster Competition. The authors also thank Dr. Megan Kenny Feister for engineering education perspectives and writing support.