Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 108C (2017) 2131–2140

Procedia Computer Science

International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland

Building a Community of Practice to Prepare the HPC Workforce

Katharine J. Cahill[1], Scott Lathrop[2], Steven Gordon[1]
[1]Ohio Supercomputer Center; [2]Shodor Education Foundation, Inc.
kcahill@osc.edu, lathrop@shodor.org, sgordon@osc.edu

Abstract
It has been well documented for more than 30 years, that significantly more effort is needed to prepare the HPC workforce needed today and well into the future. The Blue Waters Virtual School of Computational Science (VSCSE) provides an innovative model for addressing this critical need. The VSCSE uses a Small Private Online Course (SPOC) approach to providing graduate level credit courses to students at multiple institutions. In this paper, we describe the rationale for this approach, a description of the implementation, findings from external evaluations, and lessons learned. The paper concludes with recommendations for future strategies to build on this work to address the workforce needs of our global society.

Peer-review under responsibility of the scientific committee of the International Conference on Computational Science

Keywords: HPC workforce, petascale computing, on-line education, graduate education, SPOC course

1 Introduction

The Blue Waters project is funded by the National Science Foundation to provide computational science teams with an effective, scalable, sustained petascale computing platform. The system is configured to provide sustained petascale performance on a broad range of science and engineering applications, from applications that are compute-intensive to those that are data- and memory-intensive.

Achieving the full potential of the Blue Waters system, with its advanced technology components, requires well-educated and knowledgeable computational scientists and engineers. The Blue Waters project is committed to working with the community to educate current and future generations of scientists and engineers to make effective use of the extraordinary capabilities provided by Blue Waters and other extreme-scale computing environments.

In this paper, we share our experiences and lessons learned through the Blue Waters program to teach online graduate credit courses, called the Virtual School of Computational Science and Engineering (VSCSE).

1877-0509
Peer-review under responsibility of the scientific committee of the International Conference on Computational Science
10.1016/j.procs.2017.05.059

2 Background

There have been numerous reports (Rüde et. el., 2016) since 1985 highlighting the need for a larger and more diverse HPC workforce in all sectors of society. In a report on industry use of HPC, Intersect360 found that the gap in HPC skills is a significant barrier to implementation. As one respondent put it "The biggest gap I see with new graduates is that they generally understand what these tools are capable of but have no idea how to use them – or even which tool is appropriate for a certain type of job" (Intersect360, 2017).

Many different types of programs exist to address these issues using different methods, and they can all complement one another (Argonne National Laboratory, 2017; DOE Computational Science Graduate Fellowship, 2017; Lawrence Livermore National Lab, 2017. The virtual school model stands out for bringing these experiences to the students within their institutions. We are extending the old adage, "it takes a village to raise a child" to be "it takes a (inter/national) community to prepare an adequate (global) HPC workforce."

One mission of the Blue Waters outreach program is to energize additional educational innovations needed to prepare the HPC workforce. Accordingly, the Blue Waters team launched the Virtual School for Computational Science and Engineering (VSCSE) to address this need. The program was modeled on the Big10 Academic Alliance CourseShare program (Big 10 Academic Alliance, 2017) in which faculty from one institution teach an online course in which students from other CIC member institutions may enroll. The Big10 model works well administratively as the member institutions have established agreements to transfer course credits among the member institutions, and to accommodate balancing of the financial costs of students enrolling in courses at other member institutions. However, this model is not easily adaptable to interested academic institutions across the country.

Exclusively online courses often suffer from high dropout rates and low levels of performance by students. The literature shows that courses which blend online and face-to-face components are successful in improving those outcomes. Lopez-Perez et. al showed that students in an undergraduate business course using the blended model were more motivated, had higher course completion rates, and higher exam grades (Lopez-Perez, 2011). Blended learning provides a teaching presence to manage the learning process while online materials facilitates reflective learning to create a community of learning (Garrison, 2004). Ginns and Ellis provide a summary of the positive impacts of a blended learning approach to two years of a veterinary medicine program in Australia (Ginns, 2007). Picciano et.al. provide an edited on volume on blended learning research that includes a number of articles that echo these findings (Picciano, 2014).

The goal of the VSCSE was to establish a modified model for providing graduate-level petascale computing courses among academic institutions across the nation. The primary objective was to select faculty with petascale expertise to provide on-line petascale courses generally not available at academic institutions-through a blended approach of online delivery of content and the active participation of collaborating faculty at multiple institutions. A second and equally important objective was to allow participating students to earn graduate credit at their home institution, without having to worry about the transfer of credits among institutions.

3 Implementation

To pilot the program, the Blue Waters team initially recruited Dr. Wen-mei Hwu, Professor of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign, to teach a course titled "*Algorithmic Techniques for Scalable Many-core Computing*" during the spring of 2013. Dr. Hwu determined the course syllabus from a course he already planned to teach on the Illinois campus. With the Dr. Hwu's assistance, we recruited collaborating faculty, who were colleagues of Dr. Hwu. We initially received positive responses from faculty at four institutions. By the time the collaborating faculty were able to establish this special seminar graduate course at their institutions and enroll students, we ended up with two collaborating institutions at The Ohio State University and the University of Minnesota for a total of 54 graduate students.

Each of the campuses followed different academic schedules and there were multiple time zones to address. As a result, the course could not be delivered in real time. Therefore, the instructor recorded the lecture material for students to view at a later time. During his first course offering, Dr. Hwu created most of his recordings using the College of Engineering's high-end video classrooms. His course used several collaboration tools to share the course materials with participating faculty and students. In subsequent offerings, we leveraged the open source learning platform Moodle (Moodle Project, 2017) to host content and provide forums for discussion.

The collaborating faculty had the option to have their students view the lectures during their regularly scheduled class time, with the faculty member able to help answer their students' questions during this class time. More often we observed the faculty using a flipped classroom model, in which the students were tasked with viewing the lectures on their own time and using the regularly scheduled classroom time to discuss the video material. This provided for a richer learning experience for the students and improved the teaching experience for the faculty member.

In subsequent years, we also had to accommodate collaborating institutions following the quarter system, while the primary instructor followed a semester schedule. This was accomplished by shortening the content to be covered at the institution on the quarter system. A useful feature of this approach was the flexibility to adapt the material to meet the needs of the institutions.

The lead instructor had the option of recording the lectures in front of a live audience, or recording the material in a studio or even at their own desktop. Based on our experiences, we strongly advise the use of high quality HD video cameras for desktop recordings, as less expensive equipment provides "grainier images" that are not well received by students.

Recording the lectures without a live classroom has an added advantage for the lead instructor as well as the collaborating faculty of allowing the instructor(s) to travel and still have their class watch the video on their own. This approach accommodates the instructor's schedule without losing a valuable lecture session, although there is some doubling up discussions during the next regularly scheduled class period.

The extremely positive experiences and lessons learned from having Dr. Hwu teach this course in the spring of 2013 led to Blue Waters committing to support six additional VSCSE courses as shown in Table 1.

4 Course Contents and Structure

Each of the courses focused on a set of HPC software and hardware related skills. Recorded lectures introduce the course materials. Then a mixture of quizzes and assignments help students apply those skills. Early assignments do not require the use of an HPC system and were completed using laptops, small, local clusters, or virtual machines as appropriate. Later in the semester, students are given access to the Blue Waters system to allow them to complete assignments and a final project where scaling is required. A summary of the course contents below provides an overview of the technical content of each.

4.1 Algorithmic Techniques for Scalable Many-core Computing

This course is designed to allow students to master commonly used algorithm techniques and computational thinking skills for scalable, many-core/many-thread programming. It includes coursework in the following areas: many-core hardware limitations and constraints; desirable and undesirable computation patterns; practical algorithm techniques to convert undesirable computation patterns into desirable ones. The course includes regular quizzes and labs as well as a final group project implemented on Blue Waters. Familiarity with CUDA programming was a prerequisite.

4.2 Designing and Building Applications for Extreme Scale Systems

In this course, students learn how to design and implement applications for extreme scale systems, including analyzing and understanding the performance of applications, the primary causes of poor performance and scalability, and how both the choice of algorithm and programming system impact achievable performance. Dr. Gropp covers multi-and many-core processors, interconnects in HPC systems, and parallel I/O. Students who take this course are expected to have a strong knowledge of C, C++, or Fortran, including writing, debugging, and optimizing an application as well as some parallel programming experience. Weekly homeworks and final group projects are completed using Blue Waters. (Gropp, 2017).

4.3 High Performance Visualization for Large-Scale Scientific Data Analytics

The course is designed for graduate students in computer science or areas related to computational sciences who are interested in learning how to use visualization to analyze large-scale scientific data sets. To equip students with the ability to analyze very large-scale data sets, this course provides an in-depth discussion of the state-of-the-art in large scale scientific visualization algorithms and systems. In addition to the fundamental visualization techniques, parallel implementation of selected algorithms for high-performance architectures is covered. Hands-on visualization projects included work on Blue Waters.

Year	**Lead Instructor**	**Course Title**	**Collaborating Institutions**	**# of Students**
Fall 2014	Dr. Wen-mei Hwu	*Algorithmic Techniques for Scalable Many-core Computing*	UIUC, University of Tennessee, Knoxville, North Carolina State University, and the University of Oklahoma	74
Spring 2015	Dr. Bill Gropp	*Designing and Building Applications for Extreme Scale Systems*	UIUC, University of North Dakota and University of Wyoming	49
Fall 2015	Dr. Han-Wei Shen	*High Performance Visualization for Large-Scale Scientific Data Analytics*	Cyprus Institute, The Ohio State University, the University of Colorado Boulder, the University of Colorado Denver, the University of Illinois,	131

			Chicago, the University of Illinois, Urbana Champaign, and the University of Texas, El Paso	
Spring 2016	Dr. Bill Gropp	*Designing and Building Applications for Extreme Scale Systems*	Oklahoma State, Oregon State, University of Florida, Michigan State University, University of Illinois, Louisiana State University, University of Puerto Rico, University of Houston Clear Lake, Austin Peay State University, University of Washington, Prairie View A&M University	130
Fall 2016	Dr. Wen-mei Hwu	*Algorithmic Techniques for Scalable Many-core Computing*	University of Houston, Clear Lake, Clarkson University, Michigan State University, Perimeter Institute, Wake Forest University	56
Fall 2016	Dr. David Keyes	*Introduction to HPC*	University of Tulsa, Georgia Institute of Technology, Louisiana State University, Vanderbilt University, University of Wyoming, University of Puerto Rico Mayaguez, University of North Texas, Columbia University, Prairie View A&M University, King Abdullah University for Science and Technology	139

Table 1. Timeline of VSCSE courses, lead instructors, and participating institutions.

4.4 Introduction to HPC

This course focuses on high performance computing algorithms and software technology, with an emphasis on using distributed memory systems for scientific computing. The main topics covered in this course include computer architecture, state-of-the-art discretization techniques, solver libraries, and execution frameworks. The programming assignments use MPI and PETSc on local resources and Blue Waters culminating in an independent project leading to an in-class report.

5 Outcomes

An immediate outcome included providing 633 graduate students at 28 US institutions in 16 states-as well as institutions in Canada, Cyprus, Puerto Rico, and Saudi Arabia-with the knowledge and skills to advance computation and data-enabled discovery. Of particular note is that many of the collaborating institutions were what we would call tier 2 institutions, and includes 7 institutions in Experimental Program to Stimulate Competitive Research (EPSCoR) states/jurisdictions where the National Science Foundation seeks to build new research capacities. There was also participation by 2 Minority Serving Institutions. A secondary outcome was providing the collaborating faculty at 32 institutions with the knowledge, materials, resources and confidence to be able to teach the course content on their own. Subsequent to the courses being taught, we have made the materials available to any faculty member that requests access.

The approach to teaching in this manner readily accommodates access to learning by students at any and all institutions, regardless of the institution size or location. The courses are an exemplary model we believe can be replicated to accelerate the pace of adoption of new course content (HPC and petascale content in the case of these activities) on campuses nationally, and internationally, to prepare the needed workforce.

The Big10 model for offering shared courses among member institutions includes agreed upon cost-sharing among the member institutions to sustain the offering of their shared online courses. Over time, we hope to develop a comparable sustainable model for continuing to incorporate new content within the higher education community at the national/international level using either cost-sharing or exchange of services models.

6 External Evaluations

Each course was evaluated by external evaluators for the Blue Waters training and education activities. These evaluations have been instrumental for assessing the value of the offerings, identifying aspects that were well done, and aspects that need improvement, and for tracking the long-term impact of these courses on the instructors, the collaborating faculty, and the participating students.

The evaluators have identified the following findings from across multiple courses.

1. Overall, students had a very positive learning experience throughout the courses.
2. Students reported that one of the strengths of the courses was the extremely high-quality lectures from knowledgeable instructors.
3. The majority of the students said courses helped them experience the use of performance models to which they wouldn't have been exposed otherwise.
4. Respondents noted that the instructors were helpful at encouraging questions and discussions during the course.
5. Respondents reported that one of the primary strengths of the courses was providing students with hands-on experience.
6. Respondents reported that the courses had valuable, detailed contents and were well structured.
7. Respondents noted that TAs/local faculty were helpful and very supportive.

8. The students believed that the knowledge/skills they gained will contribute to their work/research and were interested in learning more about HPC as a result of their experience.

At the same time, the evaluators found a number of ways in which the courses could be improved. Students wished to have a clearer description of the course contents and prerequisites to better judge their preparation to take the courses. They also thought it would useful to have fewer assignments with more depth, supplementary materials which correspond to the required prerequisites, and more extensive online references to augment the lectures.

The participating faculty want to see clearer policies and extended support for the collaborating institutions. They would also like to have at least a two-year curriculum defined so that they could better plan for participation and integration into their curricula. Since some of the faculty were not themselves experts on some of the topics, they also wanted requested additional centralized TA grading and related support for their students.

The evaluation of the Extreme Scaling spring 2015 course identified that the majority of the students said this course helped them experience the use of performance models that they wouldn't have been exposed to otherwise. Students said that they were able to gain insight into the performance models by participating in this course that they would not have otherwise have gotten. They highly valued the practical examples among other course content. Some comments include the following:

- "I have never been taught how to practically go about measuring the real performance characteristics of large-scale systems in any formal way. This course has helped my research, and I am excited to apply the skills I have picked up here to my work in the future."
- "Insight into the use of performance models"
- "A detailed overview of key topics in extreme computing and performance modeling"
- "A broad understanding of HPC development."

The evaluators learned from evaluations of the GPU course in the fall 2014 that the students believed that the knowledge/skills they gained through the course will significantly contribute to their work/research, and they were interested in learning more about HPC as a result of this experience. Most (20 of 25) student respondents agreed that the knowledge/skills they gained through this course will significantly contribute to their research. More than 68% of the student respondents agreed or strongly agreed on "For the amount of time I invested in this course, I'm happy with what I've learned." Also, 76% of student agreed or strongly agreed on "I am interested in learning more about the resource and opportunities available through High Performance Computing as a result of this experience"

The student comments included:

- "For me, we don't have so much resources in our campus. Therefore, it is an opportunity we can learn something which is really help for our research and also study with other students"
- "I learned lot about CUDA and parallel programming for GPU's from this course. Given my limited knowledge on these before I am really happy about it now."
- "I have a greater understanding of not only CUDA now, but also using techniques to improve parallel algorithms. Additionally, I now have the foundation to explore and expand upon the knowledge gained from this course."
- "This course provided me a critical foundation in data/functional parallelism for HPC. Many of the techniques learned in class can also be applied to other types of processors such as coalescing, privatization, scatter/gather, tiling, etc."

- "I am a proficient programmer and I feel that this course gave me unique skills to tackle large scientific problems efficiently. I also had my first experience using GEM and its environment (always a plus to use other clusters)."

7 Lessons Learned

The single most challenging lesson learned is the difficulty recruiting faculty to teach high-end topics, such as petascale computing techniques. There are not enough faculty with these skills. Those with the appropriate expertise are often deeply involved in research and lack adequate time in their schedules to prepare such course content. One suggestion is to enlist postdocs who have recently learned many of these skills and who may be seeking the opportunity to expand their professional portfolio by teaching high-end petascale topics in such a novel approach. Another approach is to consider team teaching, thereby reducing the preparation time for any one faculty member and allowing more people to contribute their expertise to provide depth of coverage. Additionally, it may be beneficial to recruit professional staff, such as those at HPC Centers, to cover some of the topics.

There are numerous models for teaching online courses that are actively being pursued including SPOC and Massive Open Online Course (MOOC) styles of teaching. The VSCSE chose to use the Small Private Online Course (SPOC) model to ensure that the students gain in-depth understanding of the material with direct support and mentoring from their local faculty. The SPOC model has demonstrated much higher retention rates among students compared with the MOOC approach (Biemiller, 2013); (DeJong Peterson, 2014); (Gordon, 2015).

The initial lead faculty member, Dr. Wen-mei Hwu, Professor of Electrical and Computer Engineering at the University of Illinois at Urbana-Champaign, had been teaching a similar set of petascale topics through Coursera, which uses a MOOC style of teaching. He was excited to contribute to the VSCSE program to compare the different teaching and learning styles. After teaching the VSCSE course titled "*Algorithmic Techniques for Scalable Many-core Computing*" two times, Dr. Hwu's shared his experiences with these different teaching models at the annual Blue Waters Symposium in 2015. Dr. Hwu indicated that the students in the VSCSE course gained considerably more depth of knowledge since they had more frequent interactions with the course TAs and with their respective faculty.

Dr. William Gropp, Professor of Computer Science at the University of Illinois at Urbana-Champaign, taught the VSCSE course titled "*Designing and Building Applications for Extreme Scale Systems*" in the spring of 2015. He also shared his insights at the 2015 Symposium saying that he found the flipped classroom model to be far more effective for his students and himself. He plans to use this approach even when not participating in the VSCSE program. Dr. Gropp taught the same course again in the spring of 2016. He stated that "the material was also used for an undergraduate class (organized as independent study) that prepared an Illinois student team for the Student Cluster Competition at SC this year. I do think the reuse was a valuable outcome from the virtual school; that material wouldn't have been available to the students in the cluster competition class otherwise."

After completing his course on "*Introduction to HPC*" in the fall of 2016, Dr. David Keyes said, "The Virtual School has been a positive experience for this instructor. One result of it has been a much better version of my course than has ever been taught at Columbia or KAUST previously. The weekly feedback from the co-instructors and a few remote students has been useful, but the main has been the simple discipline of preparation of a lecture destined for broadcast and archive." He added, "there are other structurally very different attempts to address the same workforce training issue, but the Blue Waters SPOC provided something much closer in practicality and attainability for most students, while also having a transforming effect that is brought within the main educational institution, instead of taking place largely outside of the institution."

Key improvements to this program include improving recruitment of faculty both to develop courses and to offer them at their local institutions and working with the collaborating faculty on strategies to assist them in preparing to teach the material on their own after the VSCSE course ends. One way to achieve this would be to provide some pre-course training to the faculty so they feel more prepared to offer the class and facilitate discussions.

Lessons learned that have been implemented during the program:

- Textbooks/resources: Additional instructional resources (e.g., articles, chapters, and training materials) should be provided along with lecture slides to reduce the limitations of online-based courses
- Clear course prerequisites: Detailed information regarding course prerequisites would help students evaluate the course content prior to enrollment. Also, supplementary materials/resources which correspond to the prerequisite course would be helpful for students from different institutions who wish to prepare for the course.
- Regular (bi-weekly) calls among the instructor and collaborating faculty helps to ensure good communications, manage expectations, and handle unexpected problems in a timely fashion.
- Local implementation: Allowing local faculty to have flexibility in their grading system and polices would help them support their own students better in the local context.
- Discussion Quality: Some of the students claimed that when the discussion was led by the TAs, they were not well organized. The TAs need to be more prepared when they lead the discussion times.
- Assignments: Some of the assignments could have more depth/challenges for understanding the content better.
- It is easier for the lead instructor to teach the material a second time, as the videos required minor edits to update materials with new findings, and it helped improve the clarity of the material that caused the most questions by students during the previous course offering. Both Dr. Hwu and Dr. Gropp found that repeating the course was extremely valuable, and much less time consuming.

The external evaluators conducted a focus group with the collaborating faculty in December 2016, after all the courses were completed. There was unanimous agreement that the VSCSE program should continue. The faculty felt that the VSCSE added value to all collaborating institutions, that the courses helped build faculty capacity; and that the courses helped to jump start new programs at their institutions. The collaborating faculty felt that each course should be offered on a two-year cycle. The faculty also recommended additional topics that should be considered including big data (e.g. Hadoop), CUDA programming, and OpenACC.

8 Recommendations

Given the limitations of the availability of higher education faculty to teach these types of courses, the VSCSE model has been demonstrated to be an effective mechanism to provide course content to a larger audience. Future efforts would benefit from focusing on breaking down institutional barriers to allow for greater adoption by faculty and their institutions. As expressed by Dr. Hwu at the Blue Waters Symposium, this model may be a way to solve critical problems to enhance graduate education due to constraints on faculty time.

The VSCSE model of teaching is becoming much more common as SPOCs and MOOCs evolve over time to provide more students with greater access to a broader array of topics with the opportunity to earn course credits, certifications and badges.

We recommend funding additional research and development for VSCSE-type courses, as this is a model that deserves to be further improved. With time and experience, the VSCSE has strong potential to become a sustainable and scalable approach for preparing the future workforce.

9 Summary

The VSCSE experience was rewarding in providing the participating students with more advanced material on parallel computing than they would have otherwise been offered at their home institutions. The collaborating faculty and the lead instructors all learned from the experience, which will in turn improve their own teaching.

References

Argonne National Laboratory, 2017. *Argonne Training Program on Extreme-Scale Computing*. Retrieved April 3, 2017, from http://extremecomputingtraining.anl.gov/

Biemiller, L., 2013. MOOC Busters. *Chronicle of Higher Education, 60*(15), A23.

Big 10 Academic Alliance, 2017. *Big !0 Academic Alliance Course Share*. Retrieved April 3, 2017 from https://www.btaa.org/projects/shared-courses/courseshare/introduction

DeJong Peterson, R. , 2014. MOOC Fizzles. *Acad. Quest, 27*, 316–319. doi:DOI 10.1007/s12129-014-9434-3.

DOE Computational Science Graduate Fellowship, 2017. *DOE CSGF*. Retrieved April 3, 2017, from https://www.krellinst.org/csgf/

Garrison, D. R. and Heather Kanuka, 2004. Blended learning: uncovering its transformative potential in higher education. *The Internet and Higher Education,, 7*(2), 95–105.

Ginns, Paul and Robert Ellis, 2007. Quality in blended learning: Exploring the relationships between on-line and face-to-face teaching and learning. *Internet and Higher Education, 10*, 53-64.

Gordon, S. I. 2015. James Demmel, Lizanne Destefano, Lorna Rivera, “Extending Access to HPC Skills Through a Blended Online Course”, *XSEDE '15*, July 26 - 30, 2015, St. Louis, MO,

ISBN 978-1-4503-3720-5/15/07.

Gropp, W. 2017. *Designing and Building Applications for Extreme Scale Systems*. Retrieved April3, 2017 from http://wgropp.cs.illinois.edu/courses/cs598-s15/index.htm

Intersect360, 2017. *HPC in Industry: Recent Studies from Intersect360 Research and U.S. Council on Competitiveness.* Retrieved on April 7, 2017 from https://www.google.com/url?q=http://www.intersect360.com/industry/getPresentation.php%3Fid%3D17&sa=U&ved=0ahUKEwiviJXT74rTAhWmj1QKHRHRBywQFggGMAE&client=internal-uds-cse&usg=AFQjCNGL0_tacW5wScke3SgIx_b6TQck_g

Lawrence Livermore National Lab, 2017. *Institute for Scientific Computing Research*. Retrieved April 3, 2017, from http://computation.llnl.gov/about/organization/institute-scientific-computing-research

Lopez-Perez, M. V., Carmen Pérez-López, and Lázaro Rodríguez-Ariza, 2011. Blended Learning in Higher Education: Students' Perceptions and their Relations to Outcomes. *Computers in Education, 56*, 818-826.

Picciano, A. G., Charles D. Dziuban, Charles R. Graham (eds.), 2014. *Blended Learning Research Perspectives Volume 2.* New York: Routledge.

Project Moodle., 2017. *About Moodle*. Retrieved on April 3, 2017 from https://docs.moodle.org/32/en/About_Moodle

Rüde, Ullrich. et. al., 2016. Research and Education in Computational Science and Engineering. *Siam Review*. Retrieved from https://arxiv.org/abs/1610.02608