# Designing a New High Performance Computing Education Strategy for Professional Scientists and Engineers

Julia S. Mullen, William Arcand, David Bestor, William Bergeron, Chansup Byun, Vijay Gadepally,
Michael Houle, Matthew Hubbell, Michael Jones, Anna Klein, Peter Michaleas, Lauren Milechin,
Andrew Prout, Antonio Rosa, Siddharth Samsi, Charles Yee, Jeremy Kepner and Albert Reuther
MIT Lincoln Laboratory
Lexington, MA 02420
Email: jsm@ll.mit.edu

***Abstract*—For decades the High Performance Computing (HPC) community has used web content, workshops and embedded HPC scientists to enable practitioners to harness the power of parallel and distributed computing. The most successful approaches, face-to-face tutorials and embedded professionals, don't scale. To create scalable, flexible, educational experiences for practitioners in all phases of a career, from student to professional, we turn to Massively Open Online Courses (MOOCs). We detail the conversion of personalized tutorials to a self-paced online course. In this demonstration, we highlight a course that mimics in-person tutorials by providing personalized paths through content that interleaves theory and practice, to help researchers learn key parallel computing concepts while developing familiarity with their HPC target system.**

## I. Introduction

Despite decades of effort to educate researchers on the use of HPC systems, developing scientific applications to run effectively on supercomputing systems remains difficult. The level of difficulty translates to a consistently smaller HPC community of practitioners relative to the breadth of active science and engineering research. A primary reason is the perceived level of effort required to achieve success using HPC resources. The trade-off between effort spent on research versus learning the details of parallel computing has deterred many investigators who could benefit from HPC resources.

To address this the MIT Lincoln Laboratory Supercomputing Center, LLSC, (formerly the Lincoln Laboratory Grid Team) took a different approach. The team flipped the parallel computing paradigm and designed an interactive system that brought supercomputing to the researcher. [1] Recognizing the power of interactive, high productivity software tools, the design seamlessly integrates the desktop with the computing cluster to create an interactive parallel environment.

To further flatten the HPC learning curve, the expert consulting team provides personalized training on how to use the HPC system. The expert consultant's role is analogous to a medical professional walking into an exam room, quickly diagnosing a condition and offering a remedy. Extending this analogy, the expert consultant team established a practice of "rounds", following up with users to evaluate their progress and assist as necessary. In general, users were operational and productive within days of the initial meeting.

While this form of consulting is highly effective, it is not scalable. To improve scalability the team leveraged Massively Open Online Courses, MOOCs, and built courseware to educate the newest generation of supercomputer users. The challenge lay in creating an asynchronous course that closely approximates the hands-on personalized approach of the expert consulting team.

## II. Design

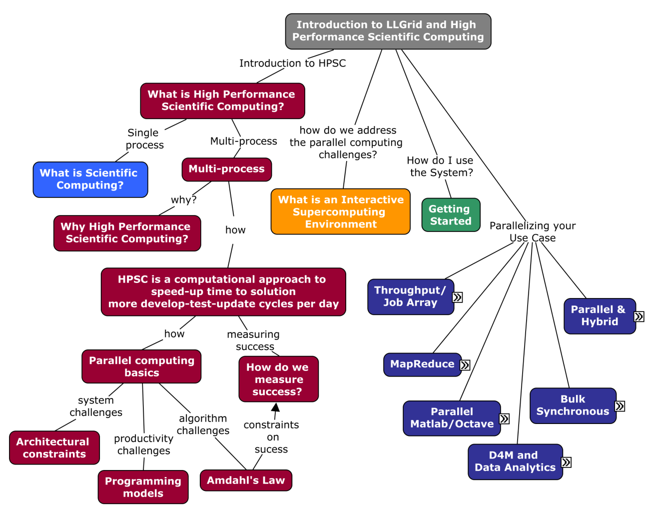


Fig. 1. "Introduction to High Performance Computing" Concept Map

The design centers on converting a personalized HPC computing tutorial into a self-paced online course. To better understand the challenge, consider the nature of the personalized tutorials. When meeting with a new user the expert

*This material is based upon work supported by the Assistant Secretary of Defense for Research and Engineering under Air Force Contract No. FA8721-05-C-0002. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Assistant Secretary of Defense for Research and Engineering.

978-1-5090-3525-0/16/$31.00 ©2016 IEEE

consultant needs to rapidly assess the user's level of computing experience, parallel computing experience, and the research application. Once assessed the consultant begins the tutorial from an appropriate starting point using a canonical example of an HPC use case matching the scientific application type.

Designing robust educational experiences requires understanding the student's starting point and goals. We identified three user types — novice, intermediate and advanced — and then identified the skills and knowledge each user needed to be successful. These were distilled into a set of concepts, illustrated in the high level knowledge map shown in Figure 1. A key idea is that every HPC application can be described in terms of a use case, *e.g.* throughput, loosely coupled, fully coupled, or data analytics. The basic assumption is that if we can teach the student how their application maps to an HPC use case, they can learn how to leverage the supercomputing system to decrease their time to solution. Once the use cases were identified, the remaining concepts were easily sorted to provide a framework for the supporting material.

The knowledge map is based on four primary questions:

- What is High Performance Scientific Computing?
- How do I address the challenge of High Performance Scientific Computing?
- How do I use the HPC system?
- How do I parallelize my use case?

Building mastery requires courseware that contains hands-on exercises in addition to simple questions. Simple questions reinforce the concepts and definitions to scaffold material for subsequent units or confirm that students have read the basic policies. To develop hands-on experiments we build on the canonical examples demonstrated in the personalized tutorials. For example, to experiment with the serial to parallel conversion process, the student applies the steps to a parameter sweep example using the LLSC system. [2] The LLSC is the computing environment used for course demonstrations, hands-on exercises and the researcher's workflow. Completing the course exercises has a two-fold purpose: prepare students for parallel processing and familiarize them with their new work environment.

## III. Deployment

Online courses and programs have existed for decades, but recently Massively Open Online Courses (MOOCs) have demonstrated the ability to scale education to hundreds of thousands of students. Among the leading MOOC consortia, only edX [3] has open sourced their learning platform. The platform used by edX, and adopted by hundreds of businesses and academic institutions around the world, is available through Open edX [4].

We built the first version of the "Introduction to High Performance Computing" course in 2015, and are currently revising it, based on user feedback. The course layout maps to the concepts and major questions described in Section II, with the addition of the Welcome Module to orient the student. To augment the sections associated with determining and building parallel use cases, we are developing a new module to highlight and explore the mapping of real world applications to use cases. The applications are drawn from our user base and offer opportunities to evaluate the trade-offs between parallelization strategies.


Fig. 2. LLx High Performance Computing Course

Building on pedagogical research demonstrating the value of segmenting content and providing frequent assessments to build mastery learning [5], assessment questions are auto-graded and interspersed between units of video content. The auto-grading provides students with immediate feedback to quickly resolve misconceptions and reaffirm learning. To provide immediate feedback on the hands-on exercises we crafted a set of questions for each step of the assignment. The questions are formed by reviewing the process required to reach a correct result. Identifying the steps is akin to creating a rubric for partial credit. Using each element of the partial credit rubric we can create a full set of questions for the programming experiments. The benefit of this approach is that students can adjust their solution approach at each stage of the process, based on their understanding of the exercise as reflected by their performance.

## IV. Summary

We demonstrated a course deployed on the Open edX MOOC platform designed to teach High Performance Computing (HPC) to professional scientists and engineers. Starting from a highly successful personalized tutorial system, we described the challenges and approaches involved in converting the in-person experience to a self-paced asynchronous online course. The design required rethinking traditional approaches to HPC and online education in order to create courseware with personalized paths, interleaved concepts and assessments and hands-on exercises on the student's target system. We selected the Open edX platform because it integrates interleaved content and assessments to produce an easy to navigate user experience that supports mastery based learning. Though the conversion of face-to-face workshops to online courseware is an active area of discussion in the educational arena, our experience shows that careful attention to courseware design and the use of a supportive online platform yields an online learning experience equivalent to a workshop.

## Acknowledgment

The authors would like to thank Robert Bond, David Martinez, and all the students who tested the course material.

978-1-5090-3525-0/16/$31.00 ©2016 IEEE

# REFERENCES

[1] N. Bliss, R. Bond, J. Kepner, H. Kim, and A. Reuther, "Interactive grid computing at lincoln laboratory," *MIT Lincoln Laboratory Journal*, vol. 16, no. 1, pp. 165–216, 2006.

[2] H. Kim, A. Reuther, and J. Kepner, "Writing parallel parameter sweep applications with pmatlab," 2005. [Online]. Available: http://www.ll.mit.edu/mission/cybersec/softwaretools/pmatlab/pmatlab.html

[3] [Online]. Available: http://www.edx.org

[4] [Online]. Available: http://www.open.edx.org/

[5] R. C. Clark and R. Mayer, *E-Learning and the Science of Instruction, Third Edition*. San Francisco, CA: Pfeiffer, A Wiley Imprint, 2011.

978-1-5090-3525-0/16/$31.00 ©2016 IEEE