

# Teaching Supercomputing via Videoconferencing

H. Neeman,<sup>a</sup> Member, IEEE, H. Severini,<sup>a</sup> D. Wu,<sup>b</sup> Member, IEEE, and K. Kantardjiev<sup>c</sup>

<sup>a</sup>University of Oklahoma, Norman OK 73019, USA

<sup>b</sup>University of Oklahoma Health Sciences Center, Oklahoma City OK 73104, USA

<sup>c</sup>California State University, Fullerton CA 92831, USA

Lead author contact: hneeman@ou.edu



## 1 INTRODUCTION

Computing technologies are rapidly evolving not only in speed and size, but also in structure, particularly with the rise of multicore and the anticipation of many-core. Concomitantly, the teaching of High End Computing (HEC) topics has become increasingly crucial, not only for computer scientists but especially for domain scientists and engineers, who are the primary consumers of large scale computing. This broad target audience has often been inadequately served by computing educators, whose reward structures typically emphasize education of CS students rather than delivering content to other disciplines. The sizable population of domain investigators who are deeply sophisticated about their own research topics but substantially underexperienced in computing, especially in advanced computing, demonstrates a strong and growing need for opportunities to learn these materials.

Because of the considerable increase in the population of Computational Science & Engineering (CSE) consumers, and the strong linkage between CSE and HEC, the need for advanced computing education and training has never been greater. Delivery of these materials remotely via conferencing technologies is critical to the advancement of investigations in hundreds of topics, in dozens of disciplines, in academia, industry and government.

Taking note of this need, the National Science Foundation (NSF) established the Cyberinfrastructure Training, Education, Advancement and Mentoring (CI-TEAM) program in 2005, with workforce development as a key goal. In 2006, a CI-TEAM grant was awarded to a multi-institution team led by the University of Oklahoma (OU), specifically by the OU Supercomputing Center for Education & Research (OSCER). The project began as a collaboration of 43 faculty and staff at 14 institutions in 6 states, and has since grown to 90 faculty and staff at 32 institutions in 18 states.

One of the objectives of this CI-TEAM project is to provide advanced computing education to the entire national community. Unfortunately, achieving this objective via face to face interactions is impractical at the national scale, because of the growing disparity between supply and demand. Therefore, conferencing technologies are essential to addressing the needs of the community as a whole.

## 2 SUPERCOMPUTING IN PLAIN ENGLISH

Education, Outreach and Training (EOT) have been shown to be crucial to workforce development under initiatives such as the TeraGrid. In this paper, the following definitions are used:

- *training*: teaching the particulars of how to use a specific technology;
- *education*: teaching the fundamental concepts underlying a class of technologies;
- *outreach*: teaching the reasons why a class of technologies is valuable to a particular person or community.

In fall 2001, OSCER began teaching a series of hourlong educational workshops titled “Supercomputing in Plain English” (SiPE) [1,2,3,4]. These lectures, which have been presented several times over the past several years, focus on the fundamental issues of High End Computing, with an emphasis on practicalities rather than on the specifics of various technologies. The prerequisite for these workshops is one semester of programming experience with any relevant language (for example, Fortran, C, C++, Java); in particular, participants need familiarity with variables, loops and arrays, but with little else (the first workshop, an overview of HEC, requires only a basic familiarity with PCs; anyone who has used e-mail or a web browser can follow this material). The pedagogical approach is to eschew technological details, instead using analogies, storytelling and play to illustrate principles rather than particulars.

The workshop series has grown to 9 presentations, described below. Some have whimsical titles, as a means of setting a relaxed mood for the participants.

### 2.1 Overview: What the Heck is Supercomputing?

This workshop provides a broad overview of HEC, and is split into several sections: what is supercomputing?; a brief overview of OSCER; the fundamental issues of supercomputing (i.e., storage hierarchy and parallelism); a quick hardware primer; introduction to storage hierarchy issues; introduction to parallelism via an analogy (multiple people putting together a jigsaw puzzle); an introduction to Moore’s Law; the motivation for using HEC.

This overview workshop has been presented far more often than any of the other workshops, because its content is accessible to a far broader and more diverse audience than the subsequent presentations. In-person recipients have included not only academic institutions (at 17 colleges and universities in 5 states, as well as a high school), but also a professional society (the Oklahoma City chapter of the American Society of Mechanical Engineers) and civic organizations (the Oklahoma State Chamber of Commerce, the Norman OK Lions Club). Ages of participants have ranged from elementary school to retirees. In addition, this overview has been presented via videoconferencing to a university in another state and to a university in South America (with the slides pre-translated into Spanish and with live simultaneous translation into Spanish of words spoken in English); this latter presentation was so successful that the lively question-and-answer period that followed ran longer than the presentation itself.

## 2.2 The Tyranny of the Storage Hierarchy

Fundamentally, just two issues govern performance: the storage hierarchy, and parallelism. This presentation focuses on the implications of a fundamental reality: fast implies expensive implies small, and slow implies cheap implies large. Topics include: what is the storage hierarchy?; registers; cache; RAM; the relationship between cache and RAM; cache hits and misses; cache lines; cache mapping strategies (direct, fully associative, set associative); cache conflicts; write-through vs. write-back; locality; tiling; hard disk; I/O strategies; virtual memory. A key point in the broader context of HEC is that parallel performance can be difficult to predict or understand without a thorough grounding in the storage hierarchy.

## 2.3 Instruction Level Parallelism

This presentation is the first in-depth introduction to parallelism, and as such is a relatively gentle way to ease the participants into understanding how parallel computing behaves in practice. Participants are regularly urged not to panic, which helps lighten the tone of material that is otherwise somewhat technical. Topics include: recap of parallelism; what is Instruction Level Parallelism (ILP)?; kinds of ILP; instructions and cycles; scalar (non-ILP) operation; independence, dependence and order; superscalar; loops; pipelining; loop performance; pipeline inhibitors; superpipelining; vectors.

## 2.4 Stupid Compiler Tricks

Although this presentation is ostensibly about high performance compilers, the first section is actually an introduction to dependencies, to reinforce the participants' thinking about parallelism and its implications. Topics include: dependency analysis (control vs. data dependencies, branch dependencies, loop carried dependencies, call dependencies, I/O dependencies, reductions, data dependencies, output dependencies, loop carried dependencies); tricks compilers play (scalar optimizations such as copy propagation, constant folding, dead code removal, strength reduction, common subexpression elimination, variable renaming; loop optimizations such as hoisting loop invariant code,

unswitching, iteration peeling, index set splitting, loop interchange, unrolling, fusion, fission; inlining); tricks to play with compilers (command line options for optimization; profiling).

## 2.5 Shared Memory Multithreading

This presentation is the participants' first introduction to using multiple, independent flows of execution. Topics include: parallelism basics (definition, threads vs. processes, Amdahl's Law, speedup, scalability, granularity, parallel overhead); recap of the jigsaw puzzle analogy from the Overview presentation; the fork/join model; OpenMP (compiler directives, hello world, parallel do/for, chunks, private vs. shared data, static vs. dynamic vs. guided scheduling, synchronization, barriers, critical sections, race conditions explained via the analogy of "The Pen Games," reductions, how to parallelize a serial code).

## 2.6 Distributed Parallelism

By the time the participants reach this presentation, they have a fairly good grasp of how to think about parallelism, but no experience with distributed parallelism or multiprocessing. Topics include: an analogy for understanding distributed parallelism (desert islands), which covers distributed operation, communication, message passing, independence, privacy, latency vs. bandwidth; recap of parallelism issues; parallel strategies (client-server, task parallelism, data parallelism, pipelining); MPI (structure of MPI calls, MPI program structure, Single Program/Multiple Data strategy, hello world, MPI runs, compiling for MPI, rank, determinism vs. indeterminism, MPI data types, tags, communicators, broadcasting, reductions, non-blocking vs. blocking communication, communication hiding).

## 2.7 Multicore Madness

The purpose of this presentation is to frighten the participants, because multicore (and, soon, many-core) are highly disruptive technologies that will require substantial redesign of many existing software applications and will make more difficult the design of new software. Topics include: implications of Moore's Law; recap of the storage hierarchy, including a practical example of the disparity between CPU speed and RAM bandwidth; recap of tiling; multicore/many-core basics (definition, RAM challenges, network challenges); weather forecasting example (Cartesian mesh, finite difference, ghost boundaries); software strategies for weather forecasting (tiling won't work because of inadequate calculations per byte, strategies for improving cache reuse, multiple subdomains per process, expanded ghost stencil to improve both cache reuse and communication hiding, higher order numerical schemes to increase the number of calculations per mesh zone per timestep, parallelization in Z to improve the size of each subdomain, cache size limitations).

## 2.8 High Throughput Computing

This presentation focuses on Condor [5] and similar technologies. Topics include: what is HTC? (definition, throughput vs. performance, throughput on a cluster); tightly vs. loosely coupled; Monte Carlo methods;

opportunistic computing (supercomputing at night, BOINC [6]); Condor (Condor vs. BOINC, features, limitations, running a job, batch script, Linux Condor on Windows via coLinux[7]); grid computing; OU's CI-TEAM project.

### 2.9 Grab Bag: Scientific Libraries, I/O, Visualization

This presentation concludes the series with some smaller issues of interest. Topics include: scientific computing pipeline; scientific libraries; I/O libraries (native binary vs. text, portable binary); visualization (contour lines, slice planes, isosurfaces, streamlines, volume rendering).

## 3 REMOTE DELIVERY OF WORKSHOPS

OSKER delivered its first national-scale instance of these advanced computing materials in fall 2007, to over 200 participants – students, faculty, staff and non-academics – at 43 institutions in 22 states, Puerto Rico and Mexico (of this group, approximately a quarter were at OU and three quarters were remote). Institution types varied substantially, including PhD-granting, masters-granting and bachelors-granting academic institutions, government laboratories and non-governmental organizations. (Participants included industry, but industrial participation wasn't local to the participant's firm.) Likewise, the group was broadly distributed geographically, and there was substantial variation in networking capability.

### 3.1 Conferencing Technologies

From the beginning, the authors recognized that selecting a single conferencing technology for delivery of these presentations would be a serious mistake, for several reasons:

- **Reliability:** Conferencing technologies are notorious for breaking down during use, potentially rendering an event non-operational.
- **Cost:** Some conferencing technologies are too expensive and/or too difficult to maintain for smaller institutions.
- **Scalability:** Most individual conferencing technologies have significant constraints on the number of simultaneous connections, and respond to overuse by either degrading quality of service or collapsing entirely.
- **Robustness:** A power or network outage at the presenter's institution could devastate a presentation.

Given these considerations, it quickly became clear that the presentations would need to employ multiple conferencing technologies simultaneously. OU was already prepared to provide two such technologies: an Access Grid (AG) [8] node located in the same room as the presentations were to occur, and QuickTime Broadcaster [9]. In addition, a now-defunct open source conferencing technology, VRVS [10], was designed to be compatible with a subset of AG venues, so as long as the venues were chosen carefully, people using VRVS could participate as well. However, these three technologies were deemed inadequate.

In summer 2007, the NSF had conducted a workshop about the CI-TEAM program, for the benefit of both current and prospective CI-TEAM PIs. This workshop provided

contact information for approximately 150 participants. The OU CI-TEAM PI contacted most of the NSF workshop participants to invite them to participate in the remotely delivered SiPE workshops.

Among those contacted was one of this paper's co-authors, who at her home institution had already acquired 50 licenses for iLinc [11], a proprietary videoconferencing technology. In general, licenses of this kind are only occasionally in use, because of the need to schedule specific events to use them. As such, she was in a perfect position to offer this capability to the SiPE workshops, at no additional cost to herself and at no cost to the presenters nor to the recipients. Furthermore, this technology is broadcast from a company-owned facility rather than from OU, which meant that OU didn't need to consume multicast capability to reach participants who connected via iLinc, substantially increasing the total carrying capacity for the project as a whole without requiring a costly increase in capacity at OU.

At approximately the same time, the PI realized that a power or network outage at OU would guarantee failure. Understanding that the recipients of these workshops are busy people who have neither the time nor the inclination to wait for service to be restored, the PI determined that a land line phone bridge, which would use neither OU's network nor OU's power system (land line phone cables carry a small power capacity of their own), would substantially enhance the reliability and robustness of the endeavor. However, crucial to the land line's value to the recipients would be a lack of long distance charges; that is, the phone bridge would have to be toll free, and would need to have adequate capacity to handle all participants, in the event that OU's network or power failed catastrophically. Unfortunately, OU doesn't currently provide toll free capability; fortunately, Prof. Amy Apon of the University of Arkansas arranged to provide her institution's toll free land line phone bridge, with a capacity of up to 100 simultaneous connections, to this endeavor. In practice, the number of simultaneous connections was very low, typically under 10.

With all of these conferencing technologies operating simultaneously, each with its own audio feed, and some with no echo cancellation capability, it became clear that having the presenter able to hear the recipients would be extremely impractical. Therefore, it was necessary to mute audio from all remote sites for the duration of each presentation.

However, interaction is a key component to the education process. To address these conflicting needs, the PIs created both a Google Talk session and a Gmail account for the workshops, and had them monitored during the presentations, along with the text messaging capability provided in iLinc for those participants who chose iLinc as their conferencing technology. Questions were submitted by text and e-mail, read out during the presentations, and answered either immediately or at the end of each presentation. Thus, even though interaction was impractical via audio, there was no barrier to question-and-answer.

Regardless of how the materials were delivered live, the slidesets needed to be available prior to the start of each

session, both in case of OU network or power failure, and because some of the videoconferencing technologies have inadequate video resolution to show the slides directly. Therefore, it was necessary that participants be able to download the slides and view them alongside the audio/video feeds. Thus, these materials were provided on the OSCER education webpage [4] in advance of each session. For completeness, each slideset included, at the beginning, the mechanisms for connecting via each of the conferencing technologies available, in order to ensure that participants didn't need to pore through their e-mail archives to find the weekly e-mails describing how to connect. This section of each slideset pointed out that, by definition, every large scale videoconference is an experiment, and is bound to fail at least partially. This set the right mood among the participants, since failures were then looked upon as minor setbacks rather than as catastrophes. And as predicted, every technology failed at least once, some of them failed regularly, and many participants experienced one or even frequent failures over the course of the semester

Finally, the audio/video feed was captured, by both QuickTime and iLinc, and is now available for streaming on the OSCER education webpage [4]. This approach served both participants who had to miss some of the presentations, and non-participants who wanted to view the workshops afterward.

### 3.2 Constraints on Presenter Behavior

The mix of conferencing technologies, each with its advantages, disadvantages and idiosyncrasies, placed significant constraints on the behavior of the presenter. Setup required considerable preparation, but more importantly, the presenter employed several tactics to provide the best possible experience for the participants.

First, a professional camera operator (a full time OU Information Technology staff member, who therefore added no increase in funding burden to the endeavor) provided video feed for both AG/VRVS and QuickTime Broadcaster. Unfortunately, the project team was never able to find a mechanism to get iLinc to accept the professional camera feed, so iLinc video used a standard webcam. Thus, while the AG/VRVS and QuickTime video followed the presenter around the room, the iLinc camera was locked down.

As such, the presenter was required to stay within a fairly small area near the projection screen, which was bounded by chairs to ensure compliance. Leaving this small region meant leaving the iLinc image, though on AG/VRVS and QuickTime the camera followed the presenter.

Second, because of the possibility that some participants would be connecting via audio only, it was necessary to call out slide numbers as each presentation progressed. This approach helped keep everyone synchronized during the session, and allowed people who experienced connection dropouts to find their place quickly upon reconnection.

Third, each session required, at OU, not only the presenter, but also an AG technician to run the local AG session, a question wrangler to transfer questions to the

presenter from the various text and e-mail sources, and an iLinc page turner. In addition, at least one participant was required at the University of Arkansas, to initiate the land line phone bridge session. Finally, an iLinc session chair was also required, which often was the iLinc licensee (co-author), but sometimes was a technician at OU who the licensee had trained.

## 4 CONCLUSION

Providing High End Computing education to large numbers of remote participants via conferencing is non-trivial, but doesn't have to be overwhelming. Keys to success include: careful planning and preparation; use of multiple technologies deployed in multiple, geographically distributed locations; availability of slidesets in advance; ability to stream recordings of the sessions; flexibility in approach and attitude by all involved. The experience of this team provides a model for the national Cyberinfrastructure EOT community, one that the authors recommend be broadly adopted. Costs are low, and the return on investment is incredible.

## ACKNOWLEDGMENTS

The authors wish to thank the University of Arkansas for providing the land line toll free phone bridge at the behest of Prof. Amy Apon; John Chapman and Jeff Pummill of the University of Arkansas, Kathy Smith-Miller of the Desert Research Institute, and Patrick Calhoun of OU for managing Access Grid conferencing; Kevin Blake of OU for videotaping, managing QuickTime Broadcast, and managing streaming of both live and recorded presentations; Brandon George and Brett Zimmerman of OU for running the in-person sessions and fielding text questions; many of the remote workshop participants for serving as workshop evangelists at their institutions. Also, several of the local participants at OU acted as page-turners for OU's iLinc session.

## REFERENCES

- [1] H. Neeman, J. Mullen, L. Lee and G. Newman: "Supercomputing in Plain English: Teaching High Performance Computing to Inexperienced Programmers." Proceedings of the 3rd LCI International Conference on Linux Clusters: The HPC Revolution 2002.
- [2] H. Neeman, L. Lee, J. Mullen, G. Newman: "Analogies for Teaching Parallel Computing to Inexperienced Programmers", inroads: SIGCSE Bulletin, 38 (4), 64-67, 2006.
- [3] H. Neeman, H. Severini, D. Wu: "Supercomputing in Plain English: Teaching Cyberinfrastructure to Computing Novices", inroads: SIGCSE Bulletin, 2008, to appear.
- [4] OSCER education webpage. <http://www.oscer.ou.edu/education.php>
- [5] BOINC website. <http://boinc.berkeley.edu/>
- [6] Condor website. <http://www.cs.wisc.edu/condor/>
- [7] coLinux website. <http://www.colinux.org/>
- [8] Access Grid website. <http://www.accessgrid.org/>
- [9] Quicktime Broadcaster webpage.

<http://www.apple.com/quicktime/broadcaster/>

[10] VRVS website. <http://www.vrvs.org/>

[11] iLinc website. <http://www.ilinc.com/>