High Performance Computing:Delivering Valuable and Valued Services at Colleges and Universities

Craig A. Stewart[1], Christopher S. Peebles[2], Mary Papakhian[1], John Samuel[1],
David Hart[1], Stephen Simms[1]
stewart@iu.edu, {peebles, mpapakhi, samuel, dhart, ssimms}@indiana.edu

ABSTRACT
Supercomputers were once regarded as being of very limited use – of interest to a very few national centers and used by a small fraction of researchers at any given university. As scientific research becomes more and more dependent upon management and analysis of massive amounts of data, advances in human knowledge will become increasingly dependent upon use of high performance computers and parallel programming techniques. Indiana University has undergone a transformation over the past four years, during which the capacity, use, and number of users of High Performance Computing (HPC) systems has dramatically increased. HPC systems are widely viewed as valuable to the scholarly community of Indiana University – even by those researchers who do not use parallel programming techniques. Economies of scale and vendor partnerships have enabled Indiana University to amass significant HPC systems. Carefully implemented strategies in delivery of consulting support have expanded the use of parallel programming techniques. Such techniques are of critical value to advancement of human knowledge in many disciplines, and it is now possible for any institution of higher education to provide some sort of parallel computing resource for education and research.

Categories and Subject Descriptors
D.3.3 [**Programming Languages**]: Parallel programming, High Performance Computing, HPC.

General Terms
Algorithms, Performance, Languages, Education.

Keywords
High performance computing, parallel computing, education.

1. INTRODUCTION
Twenty-five years ago supercomputers were used by a small and rarified group of experts – primarily physicists, researchers at weapons laboratories, and cryptographers. This began to change within the US in the 1980s when the government initiated the Advanced Scientific Computing Centers program and created five nationally shared supercomputer centers. These centers were given a mandate to provide access to supercomputer resources to scientists across the US. Still, the community making use of supercomputers – or high performance computing (HPC) systems as they are now more generally known – remained quite small, and was dominated by applications in physics and astronomy.

Today supercomputers are used in an increasingly diverse array of sciences as well as the arts. The use of parallel programming techniques – the application of many processors to a single problem simultaneously – is of fundamental importance to the advancement of knowledge and solution of some of the most important problems facing humankind today. Understanding the function of the human genetic sequence, creating new art forms based on virtual reality, and accurately modeling atmospheric change are enabled in unique ways by the application of parallel programming techniques. Parallel programming techniques are necessary to solve problems that would otherwise require impossibly long times if solved with uniprocessor, serial programming techniques. Parallel computing techniques are important in pushing back the boundaries of capability computing – the solution of the largest problems in simulation and analysis. Parallel computing techniques are also used in capacity computing – for example, analyzing the behavior of a function over a very large parameter space by calculating the value of a function at a very large number of points.

Indiana University acquired its own HPC system for the very first time in 1992 in a move praised by perhaps one dozen people – some of whom were the initial users of this system. This acquisition was at the same time decried by many more people who saw it as a waste of money. Indiana University (IU) now has three parallel computing systems with an aggregate processing capacity of hundreds of GFLOPS (millions of floating point operations per second) (Figure 1). The increase in capacity of IU HPC systems has been associated with a dramatic increase in the number of users of these systems, the diversity of the user community, total usage of computing cycles, and the extent to which parallel programming techniques are employed. According to a recent random survey of the university community, 6.4% of

[1] University Information Technology Services, Indiana University, Bloomington IN 47408
[2] Office of the Vice President for Information Technology, Indiana University, Bloomington IN 47408

all graduate students and faculty at IU Bloomington make use in some fashion of IU HPC systems [1]. On a 1 to 5 scale (with 1 being highly dissatisfied and 5 being very satisfied), the average evaluation provided by the users of IU HPC systems was a 3.73(± 0.09). Ninety-eight point six percent (98.6%) of all respondents rated these services at least satisfactory (giving a score of 3 or better). In this paper we will explain how Indiana University has been successful in gaining broad acceptance of HPC systems. We will discuss the ways in which the strategies employed by IU can be scaled to enhance intellectual accomplishment at a college of any size, including instructional settings in small colleges. We will end by discussing why it is critically important to the future of scientific progress for universities and colleges of all sizes to make greater use of HPC systems and parallel computing techniques in research and education.

2. DISCUSSION

2.1 Usage and acquisition of HPC systems at Indiana University

The growth in the aggregate peak theoretical capacity of IU's HPC systems is shown in Figure 1. Indiana University has adopted a policy of maintaining HPC systems in the three architectures most predominant in the US: distributed shared memory systems (an IBM RS/6000 SP, which comprises the largest system at IU); a shared memory system (an SGI Origin 2000 from 1997 to 2000, replaced by a Sun E10000 in 2000); and a cluster of PCs. Additional information about IU HPC systems is available [2]. The strategy of maintaining 'best of breed' of three architectures enables IU to support a variety of types of parallel applications, as well as adapt rapidly to changes in the HPC market.

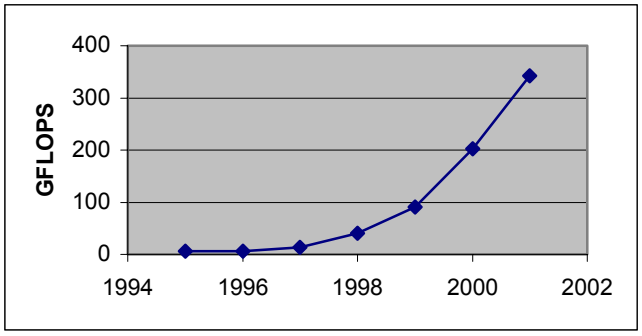

Figure 1. Growth in aggregate peak theoretical capacity of IU HPC systems (in millions of floating point operations per second – GFLOPS).

Merely expanding capacity might be done simply by spending large amounts of money, to no significant gain for the university community. However, as shown in Table 1, the number of users, the total number of GFLOPS-hours used, and the percentage of the GFLOPS-hours used devoted to a parallel program have also risen dramatically during this time.

Table 1. Trends in usage of Indiana University HPC systems.

Year	# users	% utilization	% of CPU hours used in parallel programs
1996/97	300		31.6%
1997/98	370		47.2%
1998/99	460	58%	63.2%
1999/2000	572	84%	56.8%
2000/01	675	64%	77.0%

2.2 Trends in usage over time

As one can see from Table 1, one of the limiting factors in usage of HPC systems is availability. An HPC system operating anywhere at or above 80% of capacity is operating very near its peak theoretical capacity. At a utilization level of 80% many users are waiting days for their jobs to be dispatched and the system feels "full" to the user. Part of the growth over time in GFLOPS-hours used is the result of more processing power becoming available. In any type of simulation that involves modeling a system by dividing it into a grid of points, the greater the number of grid points, the better the simulation. A scientist may well feel that s/he can afford to wait a certain amount of time and no longer to get the results of a computer simulation. This is often the factor limiting resolution of a simulation. Availability of faster and larger numbers of processors means that it becomes possible to perform a more thorough and better simulation. Thus there is more than a grain of truth in the maxim "time to solution will remain a constant forever."

But there is much more than this at work in the changes that have taken place at Indiana University. Recognizing the strategic importance of high performance computing in scientific and artistic endeavors, Indiana University placed the provision of HPC systems *and* support as a key component of a university-wide Information Technology Strategic Plan adopted formally in 1998 [3]. Funding by the State of Indiana for this IT strategic plan enabled the addition of parallel programming consulting staff and the formation of a team dedicated specifically to the support of HPC applications [2] within University Information Technology Services (UITS). The formation of this team has been just one part of an overall strategy to enhance the capabilities and productivity of researchers and artists at IU centered on IU's HPC systems.

2.3 Strategies in support of HPC resources

The growth in use of parallel programming techniques and HPC systems, and the resulting expansion in capabilities of IU researchers, has been made possible as a result of the following steps:

- Open accessibility to UITS HPC systems for all IU researchers
- Provision of widely used serial applications on IU's most robust HPC system
- Encouraging use of parallel programming techniques in instruction.
- Proactive engagement with researchers and delivery of assistance in parallelizing serial applications.

- Consulting to optimize existing parallel codes
- Constancy of purpose over time and across management units

It is quite common for Universities and national supercomputing centers to require some sort of application process before a researcher can gain access to a HPC computing system. In the case of extremely specialized systems this makes sense. However, in many other situations it creates perceived barriers to adoption of parallel computing techniques, and fosters a community perception that parallel computing is a tool for the select few. In contrast, IU has a policy of providing accounts on HPC systems to any member of the research community. There is no chargeback or allocation process for use of IU HPC systems, again as a step designed to encourage use of IU HPC systems. There are, however, limits on the number of simultaneous jobs that any user can run, and a fair share priority process, to ensure equitable access to IU HPC resources for all researchers using these systems.

Researchers ranging from beginning graduate students in statistical methods classes to IU's most distinguished computer scientists are all using the same system. This is encouraged by the provision of the most widely used serial research applications on IU's IBM RS/6000 SP, including SPSS, SAS, and Matlab. The consolidation of all research applications on IU HPC systems is made possible in part by astute use of batch management systems [4,5]. This makes possible the coexistence on IU HPC systems of very different types of users and use, and reduces the barriers to adoption of parallel computing techniques. For scientists whose computing needs never require adoption of parallel computing techniques, the large and constantly updated HPC systems still provide benefits in that serial programs run extremely quickly.

The wide variety of applications available on IU HPC systems assures their common use in instructional settings. UITS has also been cooperating with faculty members to enhance adoption of parallel programming techniques in instruction. IU is presently developing a new class focusing on HPC programming, aimed at upper level undergraduates and beginning graduate students.

Since many researchers use one of IU's HPC systems, the cost of adopting parallel programming techniques is exclusively in learning those techniques – researchers already know the operating system and batch management software used on the parallel computing systems. While the learning curve is still significant, the importance of this consistency in operating system cannot be overlooked. "The first thing you need to do to increase your computing capabilities as a scientist is learn a new operating system" is a statement sure to quash interest in use of parallel computing techniques in many potential users.

A critical part of consulting in parallel programming is the identification of researchers who could benefit from use of these techniques and subsequent steps to encourage adoption of parallel programs. The identification of researchers who could benefit from use of parallel programming techniques is dependent upon a great deal of communication with the University research community. This takes place in the form of seminars, special events, and informal discussions. The adoption of parallel programming techniques entails significant expenditure of effort on the part of the researcher. Thus UITS provides a fairly detailed analysis of the potential gains prior to an agreement to work together in parallelizing a program. Once a decision has been made to parallelize a code, a UITS consultant may do a very significant portion of the initial work. However, such work is always done in cooperation with a member of the research team, and the researcher is expected to have within her/his research team a person who understands the parallel code, and who is responsible for its ongoing maintenance. The process of seeking out researchers and encouraging them to adopt parallel programming techniques is one of the critical factors for the increases in HPC usage noted earlier, particularly the increase in the portion of the cycles used at IU that are consumed by parallel programs. Furthermore, this longstanding and ongoing effort has broadened the community of researchers who use HPC systems to include a great diversity of researchers in disciplines such as biology, medicine, fine arts, theatre, and others.

Consulting to improve the efficiency and scalability of existing parallel codes is important in two critical ways. First, any improvement in the efficiency of a parallel program has at least the same net effect as expansion of available HPC systems. Furthermore, there are many barriers to scaling parallel programs up to use large numbers of processors. Expert consulting to make this possible is important in expanding the capabilities of scientists using parallel programs.

Constancy of purpose over time and across management units has been important to the broad adoption of parallel computing techniques. The HPC environment has been a high priority consistently since the adoption of the IU IT Strategic Plan [3]. Each management unit that supports some aspect of research computation has in its responsibilities the promotion of some type of parallel computing. Instructional initiatives are important as well.

2.4 Strategies for funding HPC resources

Indiana University has seen a tremendous growth in the capacity and usage of its HPC systems, fueled in part by major new research activities such as the Indiana Genomics Initiative [6]. The cost of HPC systems is not to be overlooked. Such systems can be very expensive. However, IU has expanded its HPC resources very cost-effectively using strategies that are applicable to many institutions of higher education.

Consolidation of research computing activities into a small number of relatively large servers has been a significant part of the development of IU HPC resources, and this was particularly significant early on. The astute use open source batch management systems, such as PBS, was critical to effecting this consolidation while maintaining quality of service to all constituencies [4,5]. This approach, however, requires extremely skilled system administration staff.

Grants from national funding agencies, private organizations, and partnerships with computing vendors have contributed greatly to expansion of the University's HPC systems. Partnerships with computing vendors have been extremely important on an ongoing basis, and are a type of activity that many institutions of higher education can participate in. Collaborations between IU and industrial partners have been most effective when based upon ongoing collaborations between researchers at the computer vendor and researchers at IU. The identification of good collaboration opportunities, and the maintenance of ongoing

research relationships at times other than when acquisitions are being contemplated or hardware grants are being applied for, has been of critical importance in IU's relationships with vendors. At the same time, we have been careful to engage in such partnerships only when they are in the joint interest of the University and the vendor.

2.5 Scaling down – implications for smaller institutions

While there are many strategies that can reduce the cost barriers to acquisition of large HPC systems in large or medium-sized institutions of higher education, the cost of entry may make the use of a commercial parallel computing system beyond the means of a smaller institution. However, it is possible for research and educational usage to assemble a parallel computing system at a very low cost. It is quite feasible, for example, to assemble a parallel computing system from an assortment of older PCs and some networking equipment. Such a system could easily cost a very few thousand dollars and be of great use in education and research. The information and software needed to assemble a "Beowulf" cluster is now readily accessible [7].

3. IMPLICATIONS FOR THE FUTURE AND CONCLUSIONS

There is a tremendous need for graduate students, faculty members, and staff who are skilled in parallel computing techniques. Increasing use of parallel programming techniques in biomedical research will exacerbate this need. Advances in many disciplines now depend on use of HPC systems. The importance of increasing exposure to parallel computing techniques in the undergraduate curriculum cannot be overstated. The development of a thorough understanding of parallel computing techniques requires conceptual leaps on the part of the student not unlike those required to master calculus or differential equations. No one would consider an undergraduate student well prepared for graduate study in physics without a good grounding in these areas of mathematics. A good grounding in parallel computing concepts and techniques is similarly essential for graduate research in many scientific disciplines, but it is rare for students to enter graduate programs with even a rudimentary understanding of parallel programming.

Indiana University has engaged in an ongoing expansion of its HPC systems. Through partnerships between the central computing organization and researchers, parallel programming techniques have been widely adopted by an increasing number of researchers at IU, facilitating research endeavors that would otherwise have been impossible. Although it is difficult to measure scientific productivity, UITS attempts to track publications that make use of IU HPC systems. The hundreds of publications, conference presentations, and artistic works that were made possible, at least in part, by use of IU HPC systems [2] provides some measure of the value of these systems to IU's intellectual and artistic productivity.

The strategies developed by IU should be applicable at many mid-sized and large colleges and universities. PC Clusters provide an option for parallel computing environments within the reach of any institution of higher education. By providing and supporting such environments, colleges and universities can foster important new understanding in many areas of intellectual endeavor.

4. ACKNOWLEDGMENTS

The authors thank the many staff that run and support IU's high performance computing systems, including especially L. Grundhoefer, S. Burks, G. Turner, T. Todd, D.K. Berry, and S. Slavin. T. Todd assembled the usage statistics for this report. Malinda Lingwall deserves thanks for excellent editorial oversight. IU acknowledges multiple Shared University Research grants awarded by IBM, Inc.; assistance from Compaq Computing Corporation in the development of the IU Parallel PC Cluster; and its relationship as a Center of Excellence with Sun Microsystems, Inc. Acquisition of some of the HPC resources mentioned in this paper was made possible, in part, by NSF grant CDA-9601632 to Randall Bramley and D.F. McMullen.

5. REFERENCES

[1] UITS User Survey. http://www.indiana.edu/~uitssur/survey/

[2] Research and Academic Computing. http://www.indiana.edu/~uits/rac/

[3] Indiana University Information Technology Strategic Plan. 1998. http://www.indiana.edu/~ovpit/strategic/

[4] Papakhian, M. 1998. "Comparing Job-Management Systems: The User's Perspective", April-June 1998, IEEE's Computational Science and Engineering.

[5] Byun, C., C. Duncan, and S. Burks. 2000. A comparison of job management systems in supporting HPC ClusterTools. Presented at SUPerG, Vancouver.

[6] Indiana Genomics Initiative. http://www.ingen.iu.edu/

[7] Beowulf cluster information. http://www.uwsg.iu.edu/linux/LDP/HOWTO/Beowulf-HOWTO.html