# Surviving the MPP Information Explosion

*Leslie Southern*, Ohio Supercomputer Center, *Mike Ess*, Arctic Region Supercomputing Center, and *Rich Raymond*, Pittsburgh Supercomputing Center

**ABSTRACT:** *The Pittsburgh Supercomputing Center, Arctic Regional Supercomputing Center, and Ohio Supercomputer Center have similar high performance computing environments and operational philosophies. Each center has a CRAY T3D with varying numbers of processors and system loads. In 1994, these centers formed the PhAROh Metacenter Regional Alliance (MRA) to support and outreach to a diverse parallel user community. I will focus on the advantages that this collaboration has provided to the staff and user communities.*

## Introduction

The transition from vector to parallel and heterogeneous computing environments presents both new opportunities and new challenges for state, regional and national supercomputer centers. Availability of massively parallel processing (MPP) computer systems and greater interoperabilitiy between supercomputer systems and supercomputing centers provide new opportunities for the industrial and academic users.

To address these challenges and facilitate effective collaboration, interoperabilitiy between systems and centers, and knowledge transfer to academic and industrial clients, the Pittsburgh (PSC), Arctic Region (ARSC) and Ohio (OSC) Supercomputer centers have formed the PhAROh Metacenter Regional Alliance (MRA). In a 1992 press release, the National Science Foundation defined the metacenter concept as "a coalescence of intellectual and physical resources unlimited by geographical constraint; a synthesis of individual centers that by combining resources creates a new resource greater than the sum of its parts." The PhAROh Alliance was formed in this spirit, the Alliance has the goal of creating an environment of collaboration that will foster the cross-fertilization of ideas, facilitate the sharing of human and physical resources, and stimulate the growth and development of expertise in high performance computing and communication (HPCC) technologies for the purpose of transferring that knowledge to a diverse community of local, state, and regional users.

The PhAROh Metacenter is a natural alliance between centers that have similar and compatible computational facilities and operational philosophies. The three centers have installed CRAY T3D systems with varying numbers of processors and system loads. Each center also has an installed base of CRAY Parallel Vector Processing (PVP) systems and has an established relationship of sharing documentation and other information resources through electronic mail and WorldWide Web (WWW) servers. By pooling resources, the Alliance centers can effectively share development expertise and understanding of the T3D environment, share code porting strategies and performance analysis tools, and co-develop outreach, training, and remote user service programs.

## Objectives

Project objectives are grouped into five categories representing the central challenges that Alliance centers face in HPCC development for parallel systems: 1) Knowledge Transfer, 2) Showcase Applications, 3) Alliance Resources, 4) Infrastructure Development, and 5) Project Management.

### *Knowledge Transfer*

PhAROh centers have extensive experience in conducting successful outreach programs for potential industrial and academic users and educational groups. The centers have traditionally conducted professional seminars and hands-on workshops to promote awareness of the benefits of HPCC technologies and to develop expertise in industrial and academic users. Workshop attendance reflects the increasing interest in parallel computing. In the 1994/1995 academic year, Alliance centers conducted 52 workshops for 722 attendees. Workshops on parallel processing concepts, usage, and techniques composed 40% of the workshops and attracted 54% of the attendees.

Each of the centers provides a workshop specifically designed for the CRAY T3D. The length of the workshops vary from site to site and range from 2 days of intensive programming and exercises to 5 days that include an introduction to parallel processing concepts and case studies describing conversion techniques and performance issues. Central to each of the workshops are fundamentals on hardware design, PVM, CRAFT, and SHMEM routines[1]. These workshops are constantly being restructured to meet the needs of the user community and include newly available parallel processing environments. For example, when the CRAY T3D was first introduced, the only message passing environment available was PVM. Shortly thereafter, Cray Research, Inc. released the shared memory library, referred to as SHMEM, and then a version of the message passing interface (MPI). The SHMEM library was developed by Cray Research, Inc. to achieve optimal performance on the CRAY T3D. While not intuitively obvious to use or well documented, it provides users the ability to hone their applications for the CRAY T3D. On the other hand, MPI, which encompasses many attractive features of a number of existing message passing systems, is apparently becoming the standard message passing environment for the parallel processing industry[2]. Unlike programs that implement SHMEM and that are specific to the CRAY T3D, programs that implement MPI are portable across a variety of systems.

Each Alliance center can draw upon the experiences of the other centers to select core topics and the amount of detail for all centers to include that follow the direction of high performance computing technologies and meet the users' needs within an appropriate timetable suitable for a given audience. They may share training materials, and, for special topics, the 'expert' center may provide that training.

The Parallel Processing Development Group at OSC developed LAM (Local Area Multicomputer), an MPI programming environment and development system for a message-passing parallel machine constituted with heterogeneous UNIX computers on a network. Through their experiences and underlying knowledge of the MPI environment, they can provide detailed information to support Alliance users and staff[3].

In addition to providing training workshops for a particular system, the Alliance centers host a number of high performance seminars and conferences that promote efficient design methods and inform business managers and decision makers of the benefits of parallel and heterogeneous computing.

To facilitate the transition from vector to parallel systems and to maintain effective use of HPCC technologies, each center also provides a variety of consultative and user support services. To reduce users' frustrations and to offer encouragement, each center provides information on problems, libraries, reports, techniques, and software changes. In addition to the normal methods used by each center to disseminate such information, ARSC has created a weekly newsletter[4]. The Alliance centers share in each others' users' groups. Through this forum, such issues as file I/O, file limits, random number routines, timing comparisons, CAM structures, and memory limitations have been resolved.

The goal of the Alliance knowledge transfer program is to leverage unique capabilities to provide access to a broader range of applications and related services to users at all centers and to reduce, when possible, the duplication of services.

### *Showcase Applications*

While knowledge transfer programs are valuable and critical components of the parallel and heterogeneous computing program, the lack of production quality applications running on MPP and heterogeneous systems has inhibited the adoption of these technologies in academia and industry. The PhAROh Alliance works with industrial partners and researchers to convert and develop key applications for the CRAY T3D and heterogeneous environments.

The PhAROh Alliance has selected showcase applications in the areas of medical rendering, computational chemistry scientific visualization, and coal combustion. These showcase applications take advantage of the expertise at each center and represent areas of broad interest and importance to industrial and scientific communities.

Researchers at OSC, The Ohio State University Advanced Computing Center for Arts and Design (ACCAD), and The Ohio State University College of Medicine have developed a volumetric rendering tool (PARAVOL), an intuitive multisensory user interface, that converts MRI or CT scan data into three-dimensional objects and scenes that allow physicians and surgeons to explore virtual bodies. This tool has tremendous potential for educational, diagnostic, preoperative planning, and radiation treatment planning applications. The CRAY T3D version implements a volumetric ray-casting parallel algorithm that exploits spatial coherency and relies on latency hiding for high performance rendering.

PSC has made remarkable progress in converting some of the most widely used computational chemistry packages. These programs are among the most difficult codes to convert, because sections of code require diverse architectural needs -- some are I/O intensive, some are CPU intensive, and others are mostly serial. These codes exploit one of the most attractive and useful features of a tightly-coupled heterogenous environment in that the codes can be distributed on both large-scale parallel vector systems and massively parallel systems according to their architectural needs. When PSC ported Gaussian using this approach, they identified 25,000 lines out of 300,000+ lines of code that accounted for up to 95% of most runs and modified those sections to run on a number of CRAY T3D processing elements. The high performance combination of the CRAY vector/T3D combination encourages an incremental approach to porting codes to a parallel environment. Users benefit during each porting phase as runtimes continue to decrease. Following similar porting procedures, PSC has ported Amber, designed for molecular modeling and simulations; CHARMM, designed to model and simulate molecular systems using an empirical potential energy function; GAMESS, designed to efficiently perform MCSCF calculations; and the Simulated Annealing code, designed to study large metal clusters using ab initio molecular orbital calculations.

ARSC is developing visualization capabilities that take full advantage of parallel processing using the AVS dataflow model. A well recognized advantage of high performance computing environments is the ability to fully utilize advanced visualization techniques. Visualization techniques enhance the understanding of physical phenomena by allowing the scientist or engineer to view important, but typically unseen, relationships in data. Furthermore, visualization can help in communicating information to others involved in the exploration process or who use the information to make important decisions. AVS employs a distributed dataflow architecture, which allows the user to integrate simulation models into visualization applications that can be executed on a wide range of workstations and supercomputers. The network paradigm employed by AVS lends itself extremely well to heterogeneous computations. For example, ARSC is running a complete medical imaging system in which the interface runs under AVS on the CRAY Y-MP while the volume processing engine resides on the CRAY T3D[5].

Researchers from OSC and The Ohio State University developed a computational reactive mechanics application (CReaM) and adapted it to simulate the behavior of pulverized coal in combustion under a range of operation conditions. CReaM allows scientists to study the complex processes that result in the production and emission of sulfur, nitrogen, and other pollutants. Like other fuels, coal combustion involves chemical reactions, two-phase flow transport phenomena, and heat exchange with the added complexities of particle pyrolysis, internal particle burning, and formation of ash and slag. Fluid flow in sudden expansion geometries, with or without solid particles, is an important technological process for many industrial and scientific applications, especially for combustion processes. This process, employing the Direct Numerical Simulation (DNS) technique, was the first to be ported to the CRAY T3D. The initial port employed the CRAFT data and work sharing parallel paradigms and showed somewhat discouraging performance measurements. To improve performance, researchers are restructuring the code to include references to SHMEM routines[6].

### *Alliance Resources*

Similarities in the installed CRAY parallel vector (e.g., Y-MP and C90) and massively parallel (e.g., T3D) high performance computing systems, operating systems, diagnostic tools, and common consultation and training services are the central reasons for establishing cooperative relationships for the PhAROh MRA. The continuity of resources and environments between centers and the diversity of configurations will serve as a magnet to attract a broader range of industrial and academic users to the MRA.

OSC maintains a 32 processing element (PE) T3D. The T3D host system is an 8 processor Y-MP with 128 megawords of memory. ARSC also has an 8 processor Y-MP, but their parallel vector system has 1 gigaword of memory. Their T3D contains 128 PEs. PSC maintains the largest combination of CRAY systems. Their large parallel vector system is a C90 with 16 processors and 512 megawords of memory. The T3D system that attaches to the C90 contains 512 PEs.

Together, the Alliance provides the opportunity for users to access a small environment, affordable to many industries, up to a large memory C90 with a 512 PE T3D system. Machines of this size are affordable only by a small number of industries and the National Science Foundation Centers. The common environment provides industries and academic users a true opportunity to match their computational needs to a variety of hardware configurations with a consistent, uniform support infrastructure. It allows the Alliance members to provide a more robust service than is available individually. In developing applications for parallel computing systems, Alliance centers have found that different applications have differing computational needs. Some applications use many processors, others use only a few. Some require extensive preprocessing, others require very little. Some use vast amounts of memory, while others use relatively little. No center can efficiently or economically support all possible combinations of resource utilization. Typically, applications begin as low-capacity projects, but as the complexity of research increases, demands for processing power grows. As computational complexity grows, high-capacity jobs requiring large amounts of CPU time are often delayed for days or weeks until adequate processing time is available. Through the Alliance, it is possible to increase turnaround time by moving high-capacity jobs to other centers.

For example, Dr. Charlotte Elster in the Physics and Astronomy Department at Ohio University and Dr. Pennuswamy Sadayappan in the Computer and Information Science Department at The Ohio State University initially developed parallel applications on the CRAY T3D at OSC. Dr. Elster's research application on proton scattering from nuclei involving the calculation of optical potential requires long running queues that are not available at OSC. On the other hand, Dr. Sadayappan's research in parallel sorting algorithms, requires additional PEs to study performance and scalability issues. Because the CRAY T3D supports both message passing and shared memory paradigms for interprocessor communication, it is an excellent target for their work.

The broad range of configurations available through PhAROh allows Alliance centers to help users evaluate their application needs and match those needs with the appropriate metacenter resources.

### *Infrastructure Development*

The goal of eliminating geographical boundaries and organizational barriers to provide users broad access to Alliance resources has significant implications for the intellectual and physical infrastructure needed to support industrial and scientific uses of parallel processing. Creating a consistent, unified image that allows users to easily access resources at any center is, perhaps, the most daunting challenge the PhAROh MRA will face. Despite obvious similarities between supercomputing systems, software systems, compilers, and archival systems, differences remain that must be studied and resolved before a unified image can emerge. To better understand infrastructure development issues and prepare for the long-term convergence of resources, the Alliance will

- analyze cost/benefits and performance issues regarding the industrial and scientific uses of loosely-coupled clusters of workstations and tightly-coupled parallel systems like the CRAY T3D.

- identify conversion, diagnostic, optimization, performance, visualization, and other software tools needed to provide consistent, comfortable access to Alliance-wide resources
- develop mechanisms for maintaining data security across Alliance centers
- explore linkages with other organizations, supercomputing centers, and metacenters that will expand Alliance resources and benefits.

The development and expansion of an infrastructure that fulfills the operational needs of industrial and scientific parallel processing is key to the long-term success of the PhAROh MRA. Through these infrastructure development activities, the Alliance will build upon initial successes and continue to enhance access to and use of Alliance resources.

***Project Management***

Alliance members utilize established meetings, e-mail, and conference calls to manage project activities. Each center is an active member in the Coalition of Academic Supercomputing Centers (CASC). Because each center has similar CRAY T3D systems, they are also members of the Cray Users Group (CUG). These forums provide a mechanism for one-to-one correspondence with site directors and staff to report on progress and exchange support activities.

## Conclusions

The NSF Metacenter was created to stimulate scientific and engineering research and to enhance competitiveness in our nation's universities and industries. Since its inception, the NSF Metacenter has been a vital breeding ground for research and development of MPP and heterogeneous computing technologies. By teaming with vendors and collaborating to integrate their resources, the NSF Metacenter is producing a transparent linkage between diverse supercomputer systems that will allow researchers to attack increasingly complex problems requiring simultaneous access to the resources of several centers.

By combining the resources of the Pittsburgh Supercomputing Center, the Arctic Region Supercomputing Center, and the Ohio Supercomputer Center, the PhAROh MRA provides significant benefits to industrial and academic researchers, the nation, and the members of the Regional Metacenter Alliance.

## Acknowledgments

This project is funded by the Advanced Scientific Computing Program of the National Science Foundation, Award number ASC-9418357.

## References

[1] Cray Research MPP Software Guide, SG-2508 1.2, Cray Research, Inc.

[2] The Message Passing Interface Forum (MPIF), MPI Report (http://www.mcs.anl.gov/mpi/mpi-report/mpi-report.html)

[3] G.D. Burns, R.B. Daoud, and J. R. Vaigl, "LAM: An Open Cluster Environment for MPI", Proceedings of Supercomputer Symposium 94, June 1994

[4] M. Ess, "Support Functions Influencing the Experience of T3D Users at AR-SC", Cray User Group, Inc. 1995 Spring Proceedings

[5] G. Johnson and J. Genetti, "Medical Diagnosis using the Cray T3D", Cray User Group, Inc. 1995 Spring Proceedings

[6] C.F. Bender, P.F. Buerger, M.L. Mittal, and T.J. Rozmajzl, "Fluid Flow in an Axisymmetric, Sudden-Expansion Geometry", Parallel CFD '95 Proceedings.