Cluster Expansion for Applied and Computational Mathematics
Final Report

December 17, 2007

PI Dr. Mark Stalzer
Center for Advanced Computing Research (CACR)

Co-PI Dr. Emmanuel Candes
Co-PI Dr. Oscar Bruno
Co-PI Dr. Thomas Hou
Engineering and Applied Science (EAS)

California Institute of Technology
Pasadena, CA 91125

NSF Award Number: DMS-0619860
1. Expansion of the Shared Heterogeneous Cluster

The nature of financial support for high-end computing resources has evolved given the widespread adoption of Beowulf clusters. Research groups that need computing often obtain funds for clusters as part of their grants. Caltech’s Center for Advanced Computing Research (CACR) participates in some of these efforts, and supports significant dedicated resources for high-energy physics, astronomy, geophysics, physics-based simulation, and many others. Unfortunately, the balkanization of computation by this model has created inefficiencies. The clusters do not take advantage of economies of scale, can be underutilized, and poorly administered.

To address these inefficiencies, CACR developed a shared cluster model and the first instance is the Shared Heterogeneous Cluster (SHC) that was commissioned in November of 2005 and subsequently expanded several times. SHC is a multi-teraflops class machine optimized for high precision parallel codes. The original partners were Professors Paul Dimotakis, Dan Meiron, and Kip Thorne with applications to simulating dynamic fluid-solid interactions, turbulent mixing, and relativistic systems. As of December 2007 SHC consists of 162 nodes, each with 4 AMD Opteron cores (mix of 2.2/2.4 GHz) for a total of 648 cores, ~1.2 TB of memory, all interconnected by an Infiniband (IB) networking fabric that can move 320+ GBps between the compute nodes. There are two head nodes for compilation and testing, each with the capacity of two compute nodes, and three high speed NFS file servers for a total of 85 TB of storage.

The machine is administered by CACR with funds from the partner groups, and each group has an allocation of time on the machine proportionate to its contribution. By sharing, the groups get better pricing from vendors, professional systems administration by experienced CACR staff, and the ability to use a much larger machine than each group could afford separately. Some of the partners are also supporting efforts at CACR in visualization and code tuning.

The purpose of the “Shared Cluster Expansion for Applied and Computational Mathematics” (DMS-0619860) project was to expand SHC to provide support for the research activities of Professors Oscar Bruno, Emmanuel Candes, and Tom Hou. All three are faculty members in Caltech’s Applied and Computational Mathematics (ACM) department. They work on some of the most intellectually challenging problems in mathematical algorithms. These problems are characterized by multiple length and time scales, inhomogeneity, and the sampling of very complex spaces. The applications are wide ranging, including electromagnetic modeling, optimal sampling for MRI biological imagery, and computing underground material flows for geophysics problems. In each case, progress in algorithmic development depends on access to large-scale computer resources. Their research is more fully described in Section 2.

The total award budget was $153,000 of which $152,716 was spent to acquire the cluster expansion components and installation support from Hewlett Packard. The period of performance was September 1, 2006 to August 31, 2007. Twenty four nodes were acquired, each containing two AMD Opteron 280 processors, dual core, running 2.4 Ghz, 8 GB memory, an IB adapter, and 250 GB local scratch disk space. A new 24-port line board for the Voltaire IB switch was also purchased.

The 24 ACM expansion nodes were delivered on December 11, 2006 and put into production on December 15, 2006. The quick introduction of new resources is due to the professionalism of CACR’s operations staff and HP technicians. These additional components increased the machine’s peak capacity by over 400 Gflops and 192 GB. The equipment is shown in Pictures 1 and 2 on the next two pages.
The expansion for ACM gives it about a 16% allocation of the machine. Many factors influence the scheduling priority, or weight, of a job on SHC, including but not limited to: CPU time consumed recently by a user or group; time spent waiting for resources to be available; a group's CPU allocation; requested runtime, and node count. These factors are used by a configurable fair share mechanism, with the resulting "weight" used to determine scheduling feasibility and priority on appropriate resources. The resulting operational model reflects a balanced mix of good job throughput, resource utilization and policies that reflect the contributions and priorities as set forth by the partners. SHC is presently running at about 85% utilization 24/7 with less than 2% downtime for administration and maintenance.
2. Research Objectives

This section contains examples of the research that is being done or planned by each of the faculty research groups.

In Prof. Bruno’s group, the cluster expansion has enhanced the ability to conduct research on high-performance highly accurate numerical solvers. Efforts currently being pursued include development of novel fast (linear time) unconditionally stable algorithms and corresponding parallel implementations for solution of high-frequency nonlinear wave equations in realistic anatomical configurations for medical ultrasound applications (with actual application to cancer treatment planning using High-Intensity Focused Ultrasound (HIFU)), solution of radar problems of very high frequency, and optimization of antenna arrays known as "reflectarrays".

Prof. Candes is using the cluster to address significant mathematical problems related to the detection of gravitational waves. These waves are predicted by the general theory of relativity and have never been observed directly. Gravitational waves may be detected by the evolving worldwide network of large-scale interferometers, but detection is difficult due to the lack of complete information about the expected signal. In addition, one has to deal with long streams of data for which the location and duration of the
signal, if there is any, are completely unknown. We have developed very sensitive statistical strategies to
detect these faint objects based on large-scale simulations. We have used the cluster to perform these
massively parallel computations to tune the algorithms to detect signals whenever their strengths are
detectable by any method whatsoever. Our findings are part of Hannes Helgason's PhD thesis and are
generating a great amount of interest.

Prof. Hou and his group are currently investigating the stabilizing effect of convection in 3D
incompressible Euler and Navier-Stokes equations. The convection term is the main source of
nonlinearity for these equations. It is often considered destabilizing although it conserves energy due to
the incompressibility condition. The preliminary study shows that the convection term together with the
incompressibility condition actually has a surprising stabilizing effect. This was demonstrated by
constructing a new 3D model for axisymmetric flows with swirl. This model preserves almost all the
properties of the full 3D Euler or Navier-Stokes equations except for convection that is neglected. The
group has performed some preliminary computations that strongly supports that the 3D model equations
develop finite time singularities, and the plan is to perform a series of large-scale computations for a class
of initial data using space resolution up to \(8000^3\). This level of resolution is necessary in order to resolve
the singular behavior of the solution. Access to the expanded cluster makes these computations much
more feasible in a research environment. The computational study will play an important role in
understanding the mechanism that leads to these singular events in the new 3D model and how the
convection term in the full Euler and Navier-Stokes equations would destroy such a mechanism. Prof.
Hou’s group expects that such study could have a significant impact in enhancing our understanding of
the global regularity of the 3D Navier-Stokes equations.

3. Concluding Remarks

The SHC expansion will have broader impacts in education and, potentially, public outreach. ACM has a
vital educational role at both the graduate and undergraduate level at Caltech. As computational science
emerges as the third pillar of scientific discovery, this role will increase. ACM teaches a required applied
mathematics course for all undergraduates, and its offerings in computational math and parallel
computing are increasingly popular. Access to SHC will give graduate students experience with high
performance computing, which is essential for workforce development.

CACR is active in public outreach, utilizing its visualization resources to produce compelling images of
numerical calculations. These include shake-maps to show ground motion on local television after an
earthquake, and a large composite astronomical image that is the centerpiece of the new main exhibit hall
at the Griffith Observatory in Los Angeles. Results of ACM computations will likely provide additional
material.

The SHC expansion for ACM leverages significant existing resources and will support several promising
research areas, provide graduate students with advanced computational capabilities, and help show the
public the value of computational mathematics. We believe the shared cluster model provides numerous
advantages, and based on the successes of SHC, this model is gaining significant support at Caltech.