SAND REPORT
SAND2008-2189
Unlimited Release
Printed April 2008

Computer Science Research Institute 2005 Annual Report of Activities

David E. Womble, S. Scott Collis, Bernadette M. Watts, and Deanna R. Ceballos

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

Issued by Sandia National Laboratories, operated for the United States

Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.doe.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2008-2189
Unlimited Release
Printed April 2008

Computer Science Research Institute 2005 Annual Report of Activities

David E. Womble, S. Scott Collis, Bernadette M. Watts, and
Deanna R. Ceballos
Computer Science Research Institute
Sandia National Laboratories
P. O. Box 5800
Albuquerque, NM 87185-1319

Abstract

This report summarizes the activities of the Computer Science Research Institute (CSRI) at Sandia National Laboratories during the period January 1, 2005 to December 31, 2005. During this period, the CSRI hosted 182 visitors representing 83 universities, companies and laboratories. Of these, 60 were summer students or faculty. The CSRI partially sponsored 2 workshops and also organized and was the primary host for 3 workshops. These 3 CSRI sponsored workshops had 105 participants, 78 from universities, companies and laboratories, and 27 from Sandia. Finally, the CSRI sponsored 12 long-term collaborative research projects and 3 Sabbaticals.

Table of Contents

Chapter 1. CSRI Overview

1.1. Introduction

The Computer Science Research Institute (CSRI) at Sandia National Laboratories brings together researchers from universities and the national laboratories to solve problems in computer science, computational science and mathematics and to provide new capabilities in modeling and simulation. Participants are also encouraged to develop long-term relationships with laboratory scientists and researchers.

Through the inclusion of university faculty, the CSRI expands the range of expertise and research capabilities that can be applied to problems in modeling and simulation at the national laboratories. Through the interactions with laboratory scientists, researchers from universities and industry are exposed to computational problems that arise at the laboratories in connection with their DOE stockpile stewardship mission.

The Computer Science Research Institute also includes both graduate and undergraduate student programs. These include post-doctoral positions, summer jobs and graduate fellowships. The CSRI encourages students to choose careers in computer science, computational science and mathematics that support directly the challenges of national security programs.

The Computer Science Research Institute complements existing laboratory research programs and university alliances. It provides both a physical and technical focal point for identifying problems, for conducting research and for developing and strengthening interactions between the university and laboratory researchers.

This report presents an overview of the CSRI and describes the projects, visitor programs, and other activities conducted by the CSRI during the period January 1, 2005 to December 31, 2005.

1.2. Technical Focus of the CSRI

A number of potential long-term focus areas for the Sandia program in FY05 are listed and described below. This list represents key technologies in the high-performance massively parallel computing area with the potential to provide substantial benefit in efficiency and accuracy for the Sandia engineering analysis codes and other codes used or being developed for stockpile stewardship applications. Indeed, in some cases the focus areas include breakthrough technology which, when developed, will enable entirely new classes of simulations to be performed. The focus areas are divided into a small number of overarching technical areas, specifically, Algorithms and Computational Mathematics, Enabling Technologies, and System Software.

1.2.1 Algorithms and Computational Mathematics Focus Area:

1.2.1.1 Design and Optimization: As the ability to do "forward" simulations increases, the ability to do the "inverse" problem needs to be developed, e.g., parameter identification and system design, as well as the traditional inverse problems of applied mathematics. Optimization tends to be very application-specific, although some toolkits have been developed that can be generally applied. Current research efforts include work on large-scale optimization, global optimization, and discrete optimization.

1.2.1.2 Linear Solvers: Linear solvers are at the heart of many engineering simulations. There are many algorithms available; however, significant challenges remain. These challenges include the

development of scalable preconditioners and preconditioners designed for the specific needs of various applications. Much attention is currently focused on "multiscale" methods and preconditioners as the hope for truly scalable solvers, but a lot of work remains to be done, especially for unstructured adaptive grids, systems of equations, and complex boundary conditions. Additional work is also needed in many other related areas, including algebraic preconditioners, coupling direct methods for better or more robust convergence, ways to improve performance for machines with deep memory hierarchies, and developing solvers for matrices without the traditional finite-element structure, e.g., in circuit simulation.

1.2.1.3 Nonlinear Solvers: Nonlinear solvers often depend on repeated linear solvers, but there are additional research questions. For example, it will be necessary to solve systems with hundreds of variables for 3-D high-fidelity simulations. Present technology is expected to achieve tens of variables within the next year, falling far short of the ultimate requirement. Newton methods and their use in conjunction with preconditioned Krylov methods for specific problems, are of particular interest.

1.2.1.4 Eigensolvers: Many scientific and engineering problems require the eigenvalues and eigenvectors of extremely large matrices. Examples of particular interest include modal analysis for structural dynamics, minimum energy eigenfunction calculations in quantum chemistry models, and detecting the onset of turbulence in fluid flow. A common feature of these eigenvalue problems is that the number of eigenvalues required is small relative to the size of the matrices, the matrix systems are often very sparse, and only the action of the matrix on a vector (or several of them) is available. Standard techniques that involve directly factoring the matrix (including sparse direct methods) are often impractical for these problems because of excessive memory and computational requirements. Algorithmic work is needed on scalable eigensolvers, reduced accuracy algorithms, parallel implementations and application-focused algorithmic research.

1.2.1.5 Algorithms for Differential and Integral Equations: Differential or integral equations lie at the heart of most engineering simulations. A mathematical analysis of these equations can often reduce the amount of computing needed by simplifying or improving models, choosing better algorithms, or designing better computational experiments. Research topics of interest include coupling or de-coupling of scales, subgrid modeling, asymptotics, bifurcation, and stability analysis.

1.2.1.6 Complex Phenomena: This is a very large area, but general goals include identifying and quantifying the effects of uncertainty, developing a predictive capability for complex systems and processes based on computational "experiments," and algorithms that reduce fundamental computational complexity. Topics of interest include stochastic finite elements, sensitivity analysis, experimental design, stability analysis, summability methods, and general methods for handling multiscale (time and space) phenomena.

1.2.1.7 Adaptivity: The purpose of the adaptivity area is to develop the methodologies and algorithms for finite element error estimation and adaptive computing, with the general goal being to reduce the cost of computing by increasing the mesh resolution only in areas where needed. Finite element error estimation addresses the discretization error of the finite element solution for some (local) quantity of interest. The goal is to obtain tight bounds or estimates of the error in a way that is relatively cheap to compute (compared to the cost of solving the original problem).

1.2.2 Enabling Technologies Focus Area:

1.2.2.1 Meshing: Meshing is a time consuming and difficult part of any engineering simulation, yet the quality of the simulation is highly dependent on the quality of the mesh. Of particular interest are hexahedral meshes and high-quality hex-tet meshes. Research questions here include mesh connectivity, mesh optimization, and mesh refinement. Fully automatic methods and the ability to mesh large complex geometries are of particular interest. The general issue of a robust parallel meshing toolkit remains a high-priority goal of the high-performance computing (HPC) programs at the laboratories.

1.2.2.2 Automatic Mesh Refinement and Dynamic Load Balancing: More and more simulation codes include the ability to handle multiple meshes or to automatically refine meshes, and the efficient parallel implementation of these codes will require dynamic load balancing algorithms. Much of the current work is on design and implementation, but as the implementations become available, many new research questions will be raised. The need for dynamic load balancing will be more acute in heterogeneous environments such as will be developed under DisCom2. There will also be the need for "online" load balancing algorithms.

1.2.2.3 Visualization: The visualization needs at Sandia have outstripped the abilities of the commercially available tools. New algorithms are needed, and there are many questions to be answered about the appropriate and optimal visualization algorithms that should be used for SSP applications. Also, there is the question of where and when to do the visualization in a large simulation, e.g., as a post-processing operation or as a runtime process, on a stand-alone platform or on the MP machine directly, etc. The answer to these questions will have a major impact on the type of algorithms that are developed for visualization applications. Emphasis in this area will be on scalable visualization tools and algorithms for very large data sets. Distributed, commodity visualization platforms are being developed as an alternative to the costly, nonscalable platforms currently available.

1.2.3 System Software Focus Area:

1.2.3.1 Operating Systems: The operating system is a critical component in the effective and efficient use of massively parallel processing (MPP) computers. Current research topics include the use of commodity operating systems (primarily Linux) with modifications and extensions for MPP computers and distributed, cluster-based, virtual MPP computers. As in other areas, a key focus is on scalability. Projects include adding simple memory management and process management to Linux to improve performance while preserving Linux's portability and expandability, improving communication and connectivity, and fault tolerance. The efficient use of SMP nodes within the MPP computing environment is also being considered; this includes the development and implementation of efficient thread and virtual node capabilities and the efficient utilization of resources that are un-partitionable, such as the network interface.

1.2.3.2 Environments: An effective environment must address several issues. First, it must provide a fast and "user friendly" environment that allows designers to access easily all of the modeling tools, the data comprehension tools, the problem setup tools and the resources required. Second, it must provide a robust and efficient environment for developers to prototype new methods, algorithms and physics, without redoing major portions of the existing codes. Examples exist of application problem solving environments aimed at designers, but these are all "one-of-a-kind" products that are developed for a specific physics code. Examples also exist of component interfaces that allow specific methods to be rapidly prototyped, but again these are not general-purpose, nor are they in common use. Finally, new software tools are needed to model and predict the performance of code and algorithms on MPP computers. The development of tools that combine object-based, Web-centric, client-server technology with high-performance parallel server technology, made available on demand, will also be pursued.

1.2.3.3 I/O: Large-scale, simulation-based analysis requires efficient transfer of data among simulation, visualization, and data management applications. Current efforts seek to improve I/O performance of parallel codes by facilitating I/O operations from multiple nodes in parallel through highly portable user-level programming interfaces. This work will involve design, implementation, and testing of a portable parallel file system. Ideally, the parallel file system should include a server side, which may require a particular hardware configuration, and a client side, which is appropriate for use on any ASCI platform. This is not a replacement for MPI-IO. Just as the MPI data movement standard relies on an underlying message-passing or remote memory access protocol, the MPI-IO standard relies on an underlying file system. The goal is to produce at least a prototype of such a system and, if possible, a product that is appropriate for any future (or current) machine.

1.2.3.4 Heterogeneous and Distributed Systems: Parallel computers based on heterogeneous clusters of commodity workstations are starting to appear and will become common. Yet the effective use of these machines presents many research problems. For example, resources such as processors must be scheduled and managed, systems must be fault-tolerant, operating systems must be compatible, protocols for communication must be established, environments must be developed, and the integrated system must be latency-tolerant. The distinguishing feature for work in this area will be scalability to terascale and larger distributed systems.

1.2.3.5 Architecture: Our basic architecture is influenced by the highly successful ASCI Red. Cplant™ follows this architecture in spirit if not in details. This project will consider new architectures that will scale to 100 TF, petaflops, and beyond. Among other things is the need for research into interconnect technologies (hardware and software). In addition, for many current and future supercomputing applications, the enormity of the data in processing or post-processing for visualization is a major consideration. This project will consider such questions as how this should affect the architecture of future machines.

1.3. Research opportunities at the CSRI

The CSRI presents many opportunities for collaborations between university researchers and laboratory scientists in the areas of computer science, computational science and mathematics. These include the following

1.3.1 Collaborative research projects. The CSRI accepts proposals for collaborative research projects lasting from one to three years. Projects must have a principle investigator and a Sandia collaborator. Projects should address one of the technical areas listed above and the work must be performed on-site at Sandia. Proposals may be submitted to the CSRI director at any time and must be approved by the CSRI executive board.

1.3.2 Postdoctoral appointments. The CSRI offers several postdoctoral positions each year. Postdoctoral positions are for one year and are renewable for one additional year. Applications should include a statement of research interests, a resume, and a list of references.

1.3.3 Summer faculty positions and long-term research visits. Faculty are invited to consider the CSRI for summer employment or for extended visits. Salaries are generally commensurate with academic year salaries. Proposals to hire research groups including both faculty and graduate students for the summer are also encouraged.

1.3.4 Faculty sabbaticals. Faculty may spend all or part of a sabbatical year at the CSRI. Proposals for sabbatical visits are accepted at any time and the salary depends on the normal academic year salary and the sabbatical salary.

1.3.5 Summer student positions. Students are encouraged to apply for summer positions at the CSRI. Employment is generally for eight to twelve weeks. Students may be associated with a research group (see Summer Faculty Positions above) or may apply independently.

1.3.6 Graduate Fellowships. The CSRI sponsors graduate fellowships through the Krell Institute and the National Physical Sciences Consortium (NPSC). For more information, students can contact the Krell Institute or the NPSC directly, or they may contact the CSRI.

1.3.7 Short term visits. The CSRI hosts approximately 100 research visits lasting from several days to weeks. The CSRI generally reimburses visitors for travel expenses.

1.3.8 Workshops. The CSRI hosts one to five workshops per year. Workshops are generally co-organized by Sandia staff and university researchers. Workshop proposals are accepted at any time.

Chapter 2. Research Projects

1. Efficient Implementation for Overlapping File Access in MPI-IO
   Alok Choudhary, ECE Department, Northwestern University
   Wei-keng Liao, ECE Department, Northwestern University
   Kenin Coloma, Northwestern University

2. The impact three-dimensional unstructured finite element meshes have on preconditioned iterative solvers
   Eric de Sturler, Virginia Tech

3. Novel *CUR*-type Decompositions of Matrices and Applications to Preconditioned Iterative Regularization of Ill-Posed Problems and Structuring of Multi-mode Data Sets
   Petros Drineas, Rensselaer Polytechnic Institute

4. Reconfigurable Architectures for Floating-Point Computations
   Scott Hauck, University of Washington

5. Multiscale Methods in Science and Engineering
   Dr. Thomas J. R. Hughes, The Texas Institute of Computational and Applied Mathematics, The University of Texas at Austin

6. CSRI Executive Board Member
   Deepak Kapur, University of New Mexico

7. CSRI
   Professor Robert C. Kirby, University of Chicago

8. Simulating SuperComputer Interconnects
   Dr. George F. Riley, Georgia Institute of Technology

9. Polyhedra Matter in Nonlinear Equation Solving
   J. Maurice Rojas, Computational Biology Group, and Texas A&M

10. Large Scale Eigenvalue Methods and Model Reduction of Second Order Systems
    Dan C. Sorensen, Rice University

11. Combinatorial Optimization: Algorithms and Applications
    K. Subramani, LDCSEE, West Virginia University

12. Using Linearly Constrained Generating Set Search to Solve Nonlinear Programming Problems
    Robert Michael Lewis and Virginia Torczon, College of William and Mary

Title: Efficient Implementation for Overlapping File Access in MPI-IO (Years Two and Three)

PI: Alok Choudhary, Northwestern University
Wei-keng Liao, Northwestern University

Dates: October 1, 2003 - August 15, 2005

CSRI POC: Eric Russell, (505) 844-3679

Project Summary: Numerous studies of the I/O characteristics of parallel applications have shown that in most cases multiple processors access shared data objects. However, the partitioning and layout of the shared data objects to be stored in the memory can be different from its physical layout on disks, in which case the I/O performance can significantly degrade. In order to solve such problem, collective I/O was proposed in which each participated processor performs I/O on behalf of other processors and, then, all processors use available interconnection network to exchange the data so that each processor obtains the desired data. This technique has been adopted by MPI-IO, the I/O part of the MPI-2 standard, whose goal is to design a high-performance I/O library for parallel and distributed I/O. Collective I/O operations may have the situations that multiple processors issue concurrent read/write requests to overlapped regions in the shared file. The results of writing to the overlapped regions can be defined as written by one of the processors, an aggregation of all processors, or undefined. The mechanism of solving this problem, called atomicity, is implemented differently across file systems, which may involve locking shared files to guarantee the desired results. However, file locking reduces the parallelism of performing concurrent I/O and becomes the bottleneck of the collective operations. We propose to develop techniques to solve this problem. We plan to design a mechanism that automatically detects overlapping region accesses in the collective I/O operations in order to reduce the number of file locking, pass proper parameters to file locking mechanism, or even remove the locking.

ROMIO, a portable MPI-IO implementation, provides uniform parallel I/O APIs to access files on different file systems. Internally, ROMIO is built on top of ADIO, which is implemented separately on each file system using its native machine-dependent I/O library. When the underlying file system is the Network File System (NFS), ADIO use the file lock mechanism, fcntl, to perform non-coherent client-side caching on local processor memory by NFS default. This effect can prevent the file consistency problems occurred in other processors. ADIO disables client-side caching by locking the portion of the file being accessed so that the updated data can be flushed to the disk and viewable by other processors. This implementation results a pair of lock and unlock wraparound every native NFS read/write calls, even for those collective I/O performing only non-overlapping region access.

Client-side caching policy is a well-known technique to reduce the communication cost between clients and servers. However, in order to avoid the cache coherence problem, many file systems choose not perform this policy or lack of efficient solutions to it. We have designed a Persistent File Domain (PFD) approach, at the MPI-IO level, to solve the cache coherence problem while maintaining the caching policy performed on the clients. The idea of PFD is to partition a file into exclusive domains and assign each domain to a process. Once a process is assigned a file domain, it is responsible for delivering the data to the requesting processes for any successive I/O requests to this domain until the file is closed. Since a file domain is read/written by one process only, the client-side cache coherence problem is solved. We compared the performance of using different striping factors for the domain assignment as well as with the intuitive approach that explicitly invalidating/flushing the cache at every read/write call. By using a sliding window access pattern, the PFD approach shows a significant performance improvement over the invalidating-flushing approach. This implementation is also submitted to MPICH development team for potential software incorporation into the MPICH distribution.

Benefits to Sandia

It should be noted that this research would directly benefit Cplant clusters' applications. Currently there is no mechanism that controls application nodes from overwriting results from nodes within applications. Also, it is possible to retrieve obsolete data on one node due to the caching issues mentioned above. This research will provide software that will provide controls on output and eliminate retrieving obsolete data.

It is expected that this research will result in the publication and presentation of papers to juried symposia. Scalable Implementations of MPI Atomicity for Concurrent Overlapping I/O, which resulted from Year 1 efforts, was accepted at ICPP 2003. It should be noted that only 37% of the submitted papers were accepted.

FY05 Objectives:

Objective 1: User-Space Cache Coherence Solution for MPI Non-collective I/O

As mentioned, the objectives that non-collective I/O poses a tremendous challenge to optimize I/O because different processes can perform I/O at any time and do not synchronize their requests with each other. It is important to design a mechanism for individual process to manage its own cache data in order to maintain the coherent file access. Traditionally and currently in all file systems, client-side cache is managed at the file system level in which users can only access to it through regular read-write calls. System managed client-side caching can easily cause the non-coherent file access since there is no mechanism to deliver more useful application-level access information to the file system. On the other hand, MPI-IO provides several interfaces, such as MPI_File_set_view(), MPI_File_set_info(), MPI_File_get_info() and MPI_Info that potentially can be used to describe such information. Contractor will investigate the use of MPI-level cache buffering for the file data in each individual client process such that caching can be more efficiently manipulated and consistent with the desired access patterns. Contractor will also evaluate the performance of this software extensively using various applications.

Objective 2: Design and Evaluate the MPI-IO software for Lustre File System

Lustre file system is being considered by SNL (in fact the tri-labs) as a scalable file system for their platforms. Contractor will first port their MPI-IO software to Lustre utilizing its features. Contractor will evaluate performance of MPI-IO on Lustre using the various optimizations already developed. Contractor will also determine specific optimizations that may be important for the Lustre interface. For this task, first version of Lustre will be used.

Objective 3: Design and Evaluate Group Pipelined Collective I/O

In most of existing high-performance computing systems, the number of I/O servers is usually much smaller than the number of the compute nodes. Therefore, it is easily to cause traffic bottlenecks at I/O servers when all compute nodes simultaneously access to the shared files in a parallel I/O operation. The basic idea behind group pipelined collective I/O is to explicitly overlap I/O with data reorganization and communication in order to obtain better performance. The I/O is partitioned dynamically among groups of processors, where at any given time, only one group performs I/O, reducing bottleneck in the network and at the storage nodes, while keeping them busy. The contractor will design techniques to implement group pipelined collective I/O techniques within the MPI-IO infrastructure.

FY05 Tasks:

First quarter:

1. Design MPI-level caching management policy for non-collective MPI-IO optimizations
2. Set up Lustre environment and study its native interface
3. Design the group pipelined collective I/O algorithms for MPI-IO

Second quarter:

4. Implement MPI-level caching management mechanism
5. Design alternate strategies for MPI-IO port to Luster
6. Create hints for Luster file systems to use access patterns from MPI-IO
7. Implement first version of group pipelined collective I/O for MPI-IO

Third quarter:

8. Implement first MPI-IO version on top of Luster file system
9. Evaluate MPI-level caching for non-collective I/O
10. Evaluate group pipelined collective I/O using different access patterns

Fourth quarter:

11. Perform initial evaluation of the MPI-IO on top of Luster
12. Identify designs for implementing MPI-IO efficiently on future versions of Luster, which will have more sophisticated features and interface
13. Extensively benchmark and evaluate the group pipelined collective I/O performance and identify future directions for designs
14. Prepare report.
15. Identify future research directions

FYO5 Deliverables:

Quarterly Status Report and a Final Report

Title: The impact three-dimensional unstructured finite element meshes have on preconditioned iterative solvers

PI: Eric de Sturler, Virginia Tech

Dates: May 16 – August 12, 2005

CSRI POC: Richard Lehoucq (505) 845-8929

Project Summary: This project will be carried out as summer faculty from May 16 until August 12. We propose the following research to be carried out at Sandia National Laboratories (Albuquerque) in collaboration with Richard Lehoucq. We also plan to collaborate with Patrick Knupp on the relevant (re)meshing issues and to discuss/obtain good meshes for analysis and evaluation.

We will study the impact three-dimensional unstructured finite element meshes have on preconditioned iterative solvers. This involves the following major research questions. First, how do various mesh features translate into matrix properties, in particular those properties that are important for convergence (theory) of iterative methods. Second, how can we use understanding of these features to generate meshes that lead to faster convergence or improve existing meshes in this respect, while maintaining other mesh quality metrics. Third, can we generate preconditioners, either by purely algebraic means or by employing mesh information, that mitigate potentially negative effects of mesh features. These preconditioners could be further complemented by more traditional preconditioning methods.

For the case of self-adjoint, coercive PDEs, theory exists relating relative element size and element shape to bounds on the smallest and largest eigenvalues of the assembled matrix [Axelsson & Barker (SIAM 2001), Fried'73 and Fried'79, Shewchuk (paper in progress)]. This can be generalized to relating the maximum and minimum singular values for the Jacobians of the element transformations to a reference element to smallest and largest singular values of the matrix. Such a generalization may be quite useful in considering the effect that methods like mesh smoothing, which locally improve the conditioning of those Jacobians, would have on the overall condition number of the matrix and on iterative method convergence. On the other hand, those bounds on maximum and minimum eigenvalues of the matrix tend to be wide for actual

meshes. Moreover, the theory does not provide information about the eigenvalue distribution, which governs the actual convergence of Krylov methods. Taking the existing theory as a starting point, we are interested in deriving more accurate information about the eigenvalue distribution. Clearly, this is related to the eigenvalues and eigenvectors associated with the bilinear form and function spaces derived from the weak form of the PDE and the trial and test space(s) as determined by the mesh. The more accurately an eigenvector can be approximated the closer a matrix eigenvalue is to the corresponding eigenvalue associated with the weak form of the PDE. It is interesting to consider how the smallest eigenvalues of the PDE (weak form) are related to the eigenvalues of the matrix. The smallest eigenvalues of the matrix can be made arbitrarily small since an algebraic Cartesian basis vector corresponds to a nodal basis function and not to a unit vector in L2 norm. Moreover, it will be interesting to analyze the local or more global influence of a small volume element, or region with multiple such elements, on eigenvectors and hence eigenvalues. Similarly, it will be interesting to see how a poorly-meshed local area impacts the approximation of the eigenvectors (global) and hence the approximation of the eigenvalues. For example, following Shewchuk, anisotropic coefficients in the PDE should be accompanied by anisotropic meshes or elements. Analysis along these lines may lead to interesting results regarding meshing for geometrically complex objects, objects with large variations (and/or discontinuities) in the coefficient functions, meshes resulting from ALE formulations and other moving meshes, AMR-type meshes, and meshes resulting from mesh adaptation based on a posteriori error estimators.

The next step in terms of analysis will be to extend results on eigenvalue bounds from the self-adjoint case to the nonself-adjoint case. Little theory seems to exist for meshes here and the convergence theory for Krylov methods is much less conclusive as well. Extending the eigenvalue bounds mentioned above to singular value bounds should be easy (as already indicated), but obtaining detailed information on the eigenvalues will be harder. Decomposing the nonsymmetric bilinear form into its symmetric and antisymmetric part and applying Bendixson-type theorems [Householder, Dover 1975] will allow us in many cases to define regions containing all the eigenvalues. More detailed results on the eigenvalues again would require information about the eigenvectors of the infinite dimensional problem and its finite dimensional approximations. It is well known that for nonsymmetric problems the eigenvalue perturbations may be larger and more complicated than for the symmetric case. Moreover, now the angles between the approximate eigenvectors play a fundamental role in the convergence. Unless certain regularity conditions are observed, convergence cannot be guaranteed even in the limit. Nevertheless, with some regularity constraints, the main principles outlined above, to relate the quality of eigenvalue approximation to the accuracy with which eigenvectors can be approximated, apply also to the non-selfadjoint case. Some relevant questions to start with might be whether or under which conditions the spectrum of the matrix is positive real for PDEs that have this property. Analogous questions can be formulated for other desirable properties of the spectrum of the matrix. For example, can we bound the eigenvalues inside an ellipse or family of ellipses, give the distance to origin, etc.? What are the relations to convergence theories for GMRES and other methods? If useful, how should we extend convergence theory for this purpose?

In collaboration with Patrick Knupp, in addition to more theoretical analysis, we plan to empirically analyze the topics mentioned above on a collection of test meshes. We plan to consider meshes arising from ALE (with or without mesh smoothing), meshes defined on a deforming geometry, meshes with various types of 'bad' elements, and meshes where the singular values and eigenvalues of the Jacobians of the element transformations vary significantly.

As the second part of the project we study how to improve meshes for iterative solution while maintaining other appropriate mesh quality metrics. This could involve changing elements while maintaining the node locations or local remeshing. Another interesting topic might be to put certain criteria on the mesh density, possibly in relation to the coefficient functions. It might be interesting to consider (re)meshing problems that occur for moving meshes, such as ALE meshes and dynamic/adaptive meshes. Since the change of eigenvalues due to meshing issues is related to the approximation of the eigenvectors on a mesh, it will be interesting to consider how quantitative information of this kind can be obtained and used cheaply. We plan to collaborate with Patrick Knupp on these issues. In particular, we plan to study the effects of the mesh smoothing methods proposed by Patrick Knupp on the convergence of iterative methods and study the trade-offs with mesh smoothing for accuracy.

The third part of the project would be to address convergence problems that are related to meshing issues in a preconditioning step or in the solver. This could be done with or without information about the mesh. Note that under certain regularity constraints the triangular or tetrahedral mesh can actually be computed from the matrix and the PDE (Stathopoulos and Teng 2002). Since the Jacobian of the transformation from the reference element to actual elements plays an important role in the eigenvalue bounds it would be interesting to consider preconditioners that involve rescaling dependent on these Jacobians. For meshes that have a large variation in element size (area or volume) this may significantly reduce the condition number. Note that it would also bring unit algebraic vectors more in line with unit nodal basis vectors in the L2 norm. We should also analyze the results of this for elements that are not small in edge length but are small in area or volume (large aspect ratios). In this respect we should also consider the coefficient functions of the PDE.

Other research opportunities

In addition to the main project discussed above I would like to explore the following possible collaborations. (1) Collaboration with John Shadid and Mike Parks on solving long sequences of difficult linear systems arising in various applications. For the common case of slowly changing linear systems arising in a sequence we have recently developed several techniques for reusing Krylov subspaces to significantly accelerate the solution of such problems. (2) Collaboration with Mike Heroux and Mike Parks to explore the possibilities and benefits of contributing solver software that we have developed to Trilinos (such as solver techniques for sequences of linear systems).

General research interests

- Iterative methods for linear systems, including preconditioning, in particular solving sequences of linear systems, preconditioners for (generalized) saddle-point problems, and multilevel preconditioners and multigrid for (highly dynamic) AMR meshes,
- Large-scale parallel computing, parallel solvers, frameworks,
- Eigenvalue problems,
- Nonlinear systems of equations and optimization,
- Meshing and graphics,
- Applications in computational mechanics (fluids, solids), materials science, physics, chemistry, and biology.

Title: Novel *CUR*-type Decompositions of Matrices and Applications to Preconditioned Iterative Regularization of Ill-Posed Problems and Structuring of Multi-mode Data Sets

PI: Petros Drineas
Rensselaer Polytechnic Institute

Dates: August 1 – December 2, 2005

CSRI POC: Bruce Hendrickson, (505) 845-7599

1 Proposal Summary

Extremely large matrices (two-mode arrays), and, more generally, extremely large tensors (multi-mode arrays) arise in numerous applications, e.g., engineering applications, Information Retrieval and Data Mining, etc. Algorithms that efficiently operate on linear algebraic objects such as matrices and tensors are of particular interest and importance in today's highly networked and cooperative environments. We propose research on two topics, building upon prior work of the author on matrix decompositions of the form *CUR* (see Section 2) where, given a matrix *A*, *C* consists of a few columns of *A*, *R* consists of a few rows of *A*, and *U* is a carefully constructed matrix such that the difference *A ¡ CUR* is *provably* small. First (see Section 3), we propose to investigate whether *CUR* type decompositions are useful in the context of preconditioned iterative regularization for ill-posed problems. Towards that end, we intend to collaborate with Bruce Hendrickson and his colleagues at SNL, and use their expertise in the field of numerical preconditioning to identify specific applications where novel, *CUR*-based preconditioning techniques might improve existing results. Second (see Section 4), we propose to investigate extensions of our *CUR*-type decompositions to tensors, with the explicit goal of developing new results for structuring multi-mode tensors and analyzing multi-mode datasets. Towards that end, we intend to collaborate with Tamara Kolda and her colleagues at SNL. Length of visit. During my visit (if approved) at Sandia National Laboratories, I will be partly supported by Rensselaer Polytechnic Institute. I would like to spend anywhere between 3 to 5 months at SNL (in either location), starting early or middle August 2005, and (potentially) up to December 2005.

2 Prior Results

In the past, the author and collaborators have developed and analyzed fast Monte Carlo algorithms for performing useful computations on large matrices [2, 3, 4, 5, 9, 6, 7, 8]. Examples of such computations include matrix multiplication, the computation of the Singular Value Decomposition of a matrix, and the computation of compressed approximate decompositions of a matrix. The Singular Value Decomposition (SVD) in particular is a central result from Linear Algebra that is used in numerous data applications. In addition, many common matrix problems, such as the computation of the pseudoinverse, least squares calculations, and matrix compression, reduce to the computation or approximation of the SVD. Our work pioneered (both theoretically and experimentally) the development of Monte Carlo methods to perform approximate computations for matrix operations.

Since such computations generally require time which is superlinear in the number of nonzero elements of the matrix, our algorithms are particularly useful in applications where the input matrices are extremely large. One of the main contributions of our work is to demonstrate that a "sketch" consisting of a small, judiciously chosen random sample of rows and/or columns of the input matrix (or matrices) is adequate for provably rapid and efficient approximation of several common matrix operations. An important implementation issue is how to form the random sample. We sample according to a judiciously chosen (and data-dependent) set of nonuniform sampling probabilities. This nonuniform sampling, in which in the first pass through the data we compute sampling probabilities (e.g., we may keep rows or columns of a data matrix with probability proportional to the square of their lengths), and in the second pass we draw the sample, offers substantial gains. For example, it allows us to approximately solve problems in sparse matrices as well as dense matrices with significant heterogeneous structure.

Creating "sketches" of matrices: In [5, 8] we presented two algorithms which, when given an $m \pounds n$ matrix *A*, compute an approximation to *A* which is the product of three smaller matrices, *C*, *U*, and *R*, each of which may be computed rapidly. In particular, *C* consists of a few columns of *A* and *R* consists of a few

rows of A. Let CUR be the computed approximate decomposition. Both algorithms have provable bounds for the error matrix A-CUR. Both algorithms may be implemented without storing the matrix A in RAM, provided they can make two (or three) passes over the matrix stored in external memory, and use $O(m+n)$ ($O(1)$) additional RAM memory. In more recent work in progress with M. W. Mahoney, we have suggested new constructions for the matrix U, as well as improved sampling schemes for picking columns and rows of A to be included in C and R respectively, in order to improve the error bounds.

3 *CUR*-type decompositions and Preconditioning

We briefly illustrate how and why a *CUR*-type decomposition satisfying bounds similar to (1) might be a useful tool in the context of preconditioned iterative regularization for ill-posed problems. Let A *be* a large matrix. We seek to solve $Ax = b$, given A and b where ´ models the measurement noise for the "ideal", noise-free vector b. If the matrix A is ill-conditioned, solving $Ax = \tilde{}b$ exactly will not compute an accurate approximation to x, since the noise term ´ might become amplified in the solution; see [10] and references therein for a detailed discussion. Preconditioning techniques post-multiply the matrix A, and solve a different system using the Conjugate Gradient (CNGR) algorithm. The overall algorithm is commonly called Preconditioned CNGR or PCNGR. The convergence rate of PCNGR depends on the singular values. Also, it is commonly assumed that the uncontaminated data vector b satisfies the discrete Picard conditions, namely that the spectral coefficients of b decay in absolute value like the singular values of the matrix A, and that the signal vector b dominates the noise vector ´. The observation on the convergence rate of PCNGR and the aforementioned assumptions justify (and we do not expand on this) choosing preconditioners which only capture information from the top k right singular vectors and corresponding top k singular values of the matrix A, for some value of k typically much less than n. Whereas it is practically infeasible (due to memory and time constraints) to compute the Singular Value Decomposition of A and subsequently define the preconditioner as the inverse of Ak (the "optimal" rank k approximation to A), our *CUR*-type decompositions *approximate* both the left and the right subspace of Ak efficiently. In particular, C may be used to approximate the top k left singular vectors of A, R may be used to approximate the top k right singular vectors of A, and U may be used to approximate the top k singular values of A. We propose to investigate – both theoretically and experimentally – various potential directions for constructing matrices from the *CUR*-type decomposition of A.

4 Extending *CUR*-type decompositions to tensors

We propose to extend the *CUR*-type decompositions of Section 2 to tensors and develop a basis for dealing with complex, heterogeneous, multi-mode data in a principled manner. The basis of our approach is the extraction of the linear structure of tensors by sampling in order to infer global properties of the data sets. There are numerous theoretical and practical applications of such developments.

The first step towards handling large tensors is the development of tensor decompositions that are (i) intuitively simple, (ii) capture the (linear) structure of a tensor efficiently, and (iii) may be computed – exactly or approximately – in limited RAM space and computational time. A possible strategy is to attempt to generalize our approximation techniques to the "analog" of SVD for tensors. Unfortunately, we are confronted with a major hurdle: there is no known – widely accepted – "analog" of the SVD for tensors! We emphasize that, even though there are many candidate decompositions [12, 15, 13, 11, 14, 1], they all suffer from various deficiencies. The most common issues are: it is not clear that the decompositions can be computed in polynomial time, the "rank one" components that comprise the tensor are not pairwise orthogonal in a "strong" sense, none of the proposed "SVD analogs" satisfies the so called *Eckart-Young theorem*, namely that (for a certain fixed rank) the computed decomposition has the minimal number of "rank one" components among all possible fixed rank approximations to the given tensor, and such decompositions are usually not unique in a "strong" sense.

Although the above discussion implies that extending previous tools in order to deal with tensors is difficult, we are not interested in exact results. Simple and approximate extensions of the two-mode *CUR* decomposition seem promising for our goals for two main reasons. First, they are intuitive and easily computable (thus satisfying points (i) and (ii) as described in the beginning of the section), and second they seem to provide approximation bounds similar to the ones of equation (1). Thus, we hope to satisfy (iii) as well and identify the structure in large, multidimensional databases by accessing only as small, judiciously sampled, fraction of the database.

References

[1] R. Coppi and S. Bolasco. *Multiway data analysis*. Elsevier Science Publishers, North Holland, Amsterdam, 1989.

[2] P. Drineas. *Randomized Algorithms for Matrix Operations*. PhD thesis, Yale University, 2002.

[3] P. Drineas, A. Frieze, R. Kannan, S. Vempala, and V. Vinay. Clustering in large graphs and matrices. In *Proceedings of the 10th Annual ACM-SIAM Symposium on DiscreteAlgorithms*, pages 291–299, 1999.

[4] P. Drineas and R. Kannan. Fast Monte-Carlo algorithms for approximate matrix multiplication. In *Proceedings of the 42nd Annual IEEE Symposium on Foundations of Computer Science*, pages 452–459, 2001.

[5] P. Drineas and R. Kannan. Pass efficient algorithms for approximating large matrices. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 223–232, 2003.

[6] P. Drineas, R. Kannan, and M.W. Mahoney. Fast Monte Carlo algorithms for matrices I: Approximating matrix multiplication. Technical Report YALEU/DCS/TR-1269, Yale University Department of Computer Science, New Haven, CT, February 2004.

[7] P. Drineas, R. Kannan, and M.W. Mahoney. Fast Monte Carlo algorithms for matrices II: Computing a low-rank approximation to a matrix. Technical Report YALEU/DCS/TR-1270, Yale University Department of Computer Science, New Haven, CT, February 2004.

[8] P. Drineas, R. Kannan, and M.W. Mahoney. Fast Monte Carlo algorithms for matrices III: Computing a compressed approximate matrix decomposition. Technical Report YALEU/DCS/TR-1271, Yale University Department of Computer Science, New Haven, CT, February 2004.

[9] P. Drineas, I. Kerenidis, and P. Raghavan. Competitive recommendation systems. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 82–90, 2002.

[10] M. Hanke, J. Nagy, and R. Plemmons. Preconditioned iterative regularization for illposed problems. *Numerical Linear Algebra, L. Reichel, A. Ruttan, and R. S. Varga, eds., de Gruyter, Berlin*, pages 141 – 163, 1993.

[11] A. Kapteyn, H. Neudecker, and T. Wansbeek. An approach to *n*-mode components analysis. *Psychometrica*, 51(1):269–275, 1986.

[12] T. G. Kolda. Orthogonal tensor decompositions. *SIAM Journal of Matrix Analysis and Applications*, 23(1):243–255, 2001.

[13] J. B. Kruskal. Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and Applications*, 18:95–138, 1977.

[14] D. Leibovici and R. Sabatier. A singular value decomposition of a *k*-way array for principal component analysis of multiway data, pta-*k*. *Linear Algebra and Applications*, 169:307–329, 1998.

[15] T. Zhang and G. H. Golub. Rank-one approximation to high order tensors. *SIAM Journal of Matrix Analysis and Applications*, 23(2):534–550, 2001.

Title: Reconfigurable Architectures for Floating-Point Computations

PI: Scott Hauck, University of Washington

Dates: January 1, 2004 – December 31, 2005

CSRI POC: Keith Underwood, (505) 284-9283

Project Summary: ASCI supercomputers are notorious for achieving a low percentage of peak performance. Programmable hardware has been shown to provide much higher realized performance than commodity processors for many algorithms that can use integer or fixed point arithmetic, yet currently provides poor support for floating-point computation. In order to accelerate these computations, we will develop techniques for augmenting general-purpose FPGAs to better support floating point computations, as well as research to transition this technology to commercial devices. The impact will be a significant acceleration of floating-point intensive pipelined computations in ASCI applications.

Background

ASCI platforms have traditionally been based on commodity microprocessor technology; however, as ASCI systems approach the petaflop scale, there is great concern that commodity processor based systems will no longer continue to deliver a sufficient percentage of that performance to real applications. Already, ASCI platforms achieve as little as five percent of peak performance on some real codes. Numerous efforts are underway to investigate ways to augment or replace commodity processors so that systems of the future will realize a higher percentage of peak performance. One such effort funded by LDRD at Sandia National Laboratories (PI: Keith Underwood) is considering ways in which reconfigurable logic based on Field Programmable Gate Arrays (FPGAs) can be integrated in future ASCI platforms to deliver acceleration to applications. This proposal seeks to complement that effort by improving the ability of FPGAs to perform floating-point computations.

Reconfigurable logic has become a dominant technology for signal processing because of its high realized performance via spatial parallelism and post-fabrication customization. It has outstripped both commodity processors and special purpose DSPs (digital signal processors) in terms of absolute sustained performance, price/performance, and power/performance. However, FPGAs have generally been considered unsuitable for floating-point computations because of the high complexity of floating point operations. This manifests itself both in the computations, which require expensive multiplication and variable length shifters, and in interconnect, which requires wide buses for transferring data values. This means many applications must suffer from lower performance processor based solutions, or significantly higher cost ASICs (Application-Specific Integrated Circuits).

FPGAs do not reach their full potential on double precision floating point operations because the industry is not currently focused on optimizing FPGAs for scientific computing. The industry instead focuses on replacing floating-point with fixed point computations, a technique which is adequate for DSP (Digital Signal Processing), but woefully inadequate for more intensive floating-point applications. However, at the University of Washington we have an ongoing effort in domain-specific FPGA design, architecture-adaptive mapping tools, and floating-point optimization tools.

We have the opportunity to dramatically increase the performance of scientific computing by incorporating floating-point-specific optimizations into general-purpose FPGAs, and harnessing the techniques of reconfigurable computing. This involves the integration of new functionalities and interconnect strategies into commodity FPGA devices. To facilitate a long term transition to commercial technology, we will study improvements to FPGAs that will not decrease their commercial viability, but will significantly increase their floating point performance. This can give rise to a set of floating-point optimized FPGAs in tandem with traditional integer/random logic FPGAs, just as DSP processors have bifurcated into integer and floating point architectures.

Objectives
We will develop new architectures for floating-point computations. This includes extensions to standard general-purpose FPGAs to efficiently support floating-point computations, and the supporting CAD flow.

Work Description

General-Purpose FPGAs
One of the most powerful capabilities of FPGAs is the ability to optimize the computations to the needs of a given application, yet still benefit from economies of scale from a commodity part. To support this type of operation, we will augment standard island-style FPGAs with floating-point specific features such as:

- **Embedded multipliers:** FPGAs currently have embedded multipliers, which can also be used as shifters. We will investigate variations that will improve their floating-point capabilities.

- **Embedded floating-point units:** We will augment groups of logic blocks (CLBs) with a floating point functional unit. A mode bit will be provided that allows the CLBs to be used for their normal computations, or to redirect the CLB's I/O connections to the floating point unit.

- **Nibble-oriented floating-point units:** To improve their flexibility, we will consider smaller floating-point units, which can be cascaded to form wide operations. This will both allow small-range functions to use more restricted bit-widths, as well as support larger dynamic ranges – perhaps even greater than double-precision – for computations that need even greater accuracy.

Applications of Floating-Point Oriented Reconfigurable Devices
Guiding all of these efforts, we will focus on floating-point operations from common libraries that could be provided to application developers to minimize the impact on existing code and eliminate the hardware design barrier to using the technology while still providing significant acceleration. Examples of these include the N-dimensional FFT (Fast Fourier Transform), sparse matrix solvers, and BLAS (Basic Linear Algebra Subprograms) implementations. We will work with Sandia to select a representative set of library routines to study to insure that the results of the effort are generally applicable.

Expected Outcomes
In order to demonstrate the feasibility of these techniques, we will provide the following results:
- Quantification of the benefits of these technologies vs. standard FPGAs.

Key deliverables will be copies of the layouts and analysis routines, and a final report summarizing the results.

Title: Multiscale Methods in Science and Engineering

PI: Dr. Thomas J. R. Hughes
The Texas Institute of Computational and Applied Mathematics
The University of Texas at Austin

Dates: October 1, 2005 – September 30, 2006

CSRI POC: John Shadid, (505) 845-7876

Project Summary: The numerical solution of partial differential equation systems (PDEs) arising in engineering and the applied sciences implicitly involves the elimination of spatial and/or temporal scales. The appropriate solution manifolds of PDEs are infinite-dimensional whereas numerical methods, such as finite elements, wavelets, finite volumes, finite differences, and spectral methods, necessarily employ finite-dimensional approximations. Typically, the eliminated scales are "small" and the retained ones are "large." The eliminated scales essentially correspond to unresolved sub-grid scale physics that are often necessary for proper approximation of the large-scale behavior of the physical system. This can be most easily understood in a spectral context in which low wave number and/or frequency Fourier modes are retained and higher ones neglected. In many applications of physical interest, standard numerical methods, such as Galerkin approximations or central difference methods, fail under these circumstances because important interactions between small and large scales are precluded. Examples are shock wave propagation and turbulence in which various somewhat ad hoc procedures have been developed. These methods have been used over the years to account for missing or unresolved effects, namely, proper entropy production in the former case and reestablishment of the energy cascade in the latter. These ad hoc modifications, while attempting to provide reasonable physical behavior of the large scales, modify the basic underlying equations and can often produce a mathematically inconsistent formulation of the problem. Many other examples can be mentioned.

The Variational Multiscale Method was devised to provide a framework for the development of numerical methods in which the effects of small scales are identified and accounted for ab initio. The first study focused on the finite element method and showed how so-called "stabilized methods" could be derived from fundamental principles. Stabilized methods had been shown to be effective in various circumstances previously, but were derived by ad hoc means. The Variational Multiscale Method demonstrated that there was a solid foundation to these methods and provided a way for systematically developing them for more complex applications. In fact, it may be said that most all stabilized methods to date do not appropriately take account of nonlinear effects. Consequently, there is an opportunity to systematically explore the variational multiscale concept in a broad spectrum of important nonlinear applications. This is the intent of the proposed research. Potential benefits include upgrading the performance of existing production codes based on stabilized methods, correcting deficiencies noted in current codes, providing a rationale for and refining ad hoc techniques currently used in practice, and developing new systematic and fundamentally sound methods for applications which have not been successfully dealt with heretofore.

Tasks:

1) Continue multiscale work on incompressible and compressible Navier-Stokes equations.
2) Continue multiscale work on compressible Euler equations with particular reference to artificial viscosity.
3) Continue development of stabilized mixed formulations for the drift diffusion model of semiconductor devices.

I) Shock Hydrodynamics

Investigate: 1) the physical and numerical aspects of the von Neumann-Richtmeyer viscosity in a multi-dimensional Lagrangian framework, and 2) develop suitable multi-dimensional models for the closure terms that derive from the multiscale variational formulation, but that are not currently included in the von Neumann viscosity. The proposed work shall be demonstrated initially on suitable Lagrangian strong-shock problems and shall be documented in an archival publication. Contractor

shall develop extensions of the formulation to materials with strength and incompressible materials, and shall initiate development of an Eulerian formulation.

II) Compressible and Incompressible Navier-Stokes Equations and Turbulence

Pursue the development, implementation and analysis of a sequence of multiscale methods for transport/reaction systems. The major component systems shall be: 1) Incompressible Navier-Stokes, 2) Navier-Stokes, thermal energy with Boussinesq interaction model, 3) Compressible Navier-Stokes with a low Mach number assumption, 4) Navier-Stokes compressible flow. Contractor shall derive the appropriate multiscale interaction models for these systems and provide consistent turbulence models to include relevant physical mechanisms.

III) Additional tasks to be performed

Address extensions of the multiscale variational formulation to control-volume schemes, discontinuous Galerkin methods, semi-conductor device modeling with drift diffusion and hydrodynamic models, stabilized mixed discontinuous Galerkin Methods for Darcy Flow, and plasma simulations. Contractor shall apply the multiscale discontinuous Galerkin method to a stabilized mixed formulation of Darcy Flow and to the drift-diffusion, semi-conductor device equations.

Deliverables:

1) Short Course sequel to multiscale analysis and stabilized methods lectures at Sandia National Laboratories, Albuquerque.

2) Course notes.

3) Reference bibliography.

4) PI shall spend two weeks on site at Sandia.

5) Consultation/collaborations with Sandia technical personnel on specific research topics.

6) End of year report and compilation of any technical reports and articles produced by supported work.

A final report is due after completion of work. The final report shall cover work accomplished, results obtained, problem areas, data obtained and recommended solutions for actions. This report shall be a summary of technical activities during the entire performance and a comprehensive evaluation of progress in the areas of research, study, or development supported by this work.

Title: CSRI Executive Board Member

PI: Deepak Kapur, University of New Mexico

Dates: October 1, 2004 – September 30, 2005

CSRI POC: David Womble, (505) 845-7471

Project Summary: The CSRI Executive Board will consist of 6-8 people representing both technical contributors and programmatic stakeholders of the Institute. This will be on a part-time basis. The responsibilities of the Executive Board member activity for Dr. Kapur are as follows:

1. Executive Board Participation

Attend regular Executive Board Meetings.

Attend any special Board meetings required to meet ad hoc activities.

Provide programmatic guidance and advice in regards to selection of CSRI applicants for the CSRI activities.

Actively seek staff, research programs and facilities that are highly visible and respected in computer and computational science community. Build on this reputation by actively recruiting recognized and respected university researchers to participate in the CSRI.

2. Student Research

Continue to Investigate the use of Dixon resultant based method for solving nonlinear polynomial equations.

Parallelize the code developed for solving nonlinear polynomial equations.

Study how homotopy and eigen-value based methods can be adapted to exploit algebraic information available through Dixon resultant formulation.

Deliverables:

Periodic reports in the form of publications will be provided.

Annual one page report summarizing the accomplishments, benefits to Sandia National Laboratories and publications that have resulted from work performed.

Title: Automatically Generated Code for Finite Elements

PI: Prof. Robert C. Kirby, University of Chicago

Dates: August 2005

CSRI POC: Kevin Long, (925) 294-4910

1. Introduction

Supported by the DOE Early Career PI program, Robert Kirby is currently investigating techniques for automatically generating optimal code for computing arbitrary finite element basis functions, and working with Kevin Long to integrate these techniques with Sundance, Sandia's high-level tool for rapidly deploying finite element simulations and PDE-constrained optimization.

As indicated in the proposal for the Early Career program, we request support for Kirby to visit Sandia Livermore for one month during summer 2005 to expedite this collaboration. His primary work will be with Kevin Long on the integration of his tools into Sundance, however, he will also interact with Jonathan Hu, Ray Tuminaro, and Victoria Howle on fast solvers for the high-order elements made practical by Kirby's techniques, and with Philippe Pebay and Kevin Long on the application of algebraic symmetries to automatic generation of efficient code for scientific computation.

2. Background

Though initially conceived to enable development of the auxiliary operators required in high-performance PDE-constrained optimization, Sundance's high-level interface provides an effective \equation laboratory" in which researchers may rapidly develop different models for a physical system by changing the variational formulation. It is in use in numerous research projects at Sandia and elsewhere, including several ASCI-supported projects such as Long's work on applying PDE-constrained optimization to robust decision-making, Boggs' work on large-scale inversion, and Howle's work on physics-based preconditioners for penetration problems. Enhancements and generalizations to Sundance will thus have impact on many research problems in optimization, fast solvers, and applications.

Like other codes, Sundance is limited in the kind of approximating functions that it provides. For example, Sundance does not currently support the curl-conforming elements for electromagnetic fields, although these elements are known to outperform the typical continuous elements. Adding support for these elements would allow other Sandia investigators to apply the advanced PDE-constrained optimization methods enabled by Sundance to electromagnetic problems. Research by Hu and Tuminaro on fast solvers for electromagnetics will be facilitated by the use of Sundance for rapid generation of large-scale parallel simulations. Furthermore, adding general order instances of these elements will allow researchers the novel opportunity to study whether higher order approximations are suitable for core DOE problems, and to begin investigation of fast solvers for high-order curl-conforming elements. In summary, our vision is to make Sundance an \element laboratory" as well as an \equation laboratory" and a tool for PDE-constrained optimization.

Kirby has developed a tool, FIAT, that generates computable representations for general finite elements from a high-level specification. He is developing a new version that will be ready to support general three-dimensional elements by the summer. We will use FIAT to generate C++ code for evaluating basis functions that will be compiled into Sundance. Doing so will require extensions to Sundance; in particular, it is necessary to implement contravariant element transformations. We will abstract the notion of element transformations, creating an abstract class from which we will derive classes to perform the affine and Piola (contravariant) transformations. Furthermore, when we move beyond Lagrange elements, it is not always the case that the trace of a basis function to a face or edge of an element belongs to the finite element family of the lower dimensional space, as is currently assumed in Sundance. We will replace this assumption (valid for Lagrange elements of any order and certain hierarchical elements) with the notion of a \support relation" in which we specify which basis functions are nonvanishing on each component of the mesh, along with rules for evaluating those basis functions.

A month of face-to-face collaboration will provide a great opportunity for progress, as it will be relatively easy to ask and answer questions about finite elements and Sundance details. Additionally, we hope to use this time for Kirby to interact with solver researchers regarding various finite element formulations and what challenges they would present, both algorithmically and to Sundance implementation.

3. Deliverables

Between work through the spring and the summer visit, we will produce several deliverables, in the form of concrete extensions to Sundance and papers written describing and using these extensions. The results of the summer work will be summarized by Kirby in a CSRI seminar at the end of the visit.

To begin, we will integrate FIAT-generated elements into Sundance. This will be an off-line process by which we generate C++ code for evaluating Lagrange, nonconforming, Raviart-Thomas, Nedelec, and other elements for several approximating degrees, wrap it into new Sundance classes, and include it in the build. We will also make the further modifications of adding an element transformation interface and support relations to Sundance.

These concrete extensions to Sundance will allow us to produce drafts for at least one paper during the summer. We will write a paper describing what information the PDE solver needs from FIAT. Such a paper will be of general interest to the computational finite element community, as several projects that seek to write general PDE solvers often claim to leave hooks for higher order elements, but do not really understand what is required to implement them. We will submit this paper for publication at ACM Transactions on Mathematical Software.

Title: Simulating SuperComputer Interconnects

PI: Dr. George F. Riley, Georgia Institute of Technology

Investigator: Elizabeth Whitaker, Student

Dates: Spring 2005 – Fall 2006

CSRI POC: Keith Underwood (505) 284-9283

Background. Modern supercomputer platforms typically consist of a large number of computing elements (CPU's) that communicate using some type of an interconnecting switch network. As applications run on the supercomputer, each of the CPU's communicates with some subset of the other CPU's by sending messages through the interconnect network. It is well known that the design and implementation of the interconnect network can have significant effect on the overall application performance. Differing methods for interconnect topology design, message routing techniques, message blocking and backpressure, and message latencies (just to name a few) can result in wide variations in the run–time of a typical supercomputer application.

Since the purchase price of such platforms can exceed $100M, it is clear that some method for performance prediction, using *message workloads from existing applications*, is needed. Vendors are continually improving the performance of supercomputers by re–designing and re–engineering interconnect networks in a variety of ways. For example, newer technology can lead to higher bandwidth between pairs of nodes due to faster transistor rise times. A new topology design might lead to a reduction in maximum latency between pairs of nodes in the system due to a smaller diameter network. Better packaging and manufacturing technology can lead to smaller single–hop latencies between neighbors due to closer physical proximity. All of these improvements can be shown to improve overall network capacity using analytical methods for performance prediction. However, such analytical models require somewhat restrictive assumptions on network workload, such as a *Uniform* workload with equal traffic between all

pairs of nodes. There is no indication that improved network performance under the restrictive traffic models used by the analytical results will lead to improved performance for actual applications.

Discrete event simulation has historically been the method of choice to evaluate such systems when analytical models fail to provide useful results. A simulation tool can be designed to model closely the behavior of an arbitrary interconnect network with parameters specifying link bandwidth, per hop latency, routing decisions, flow control, and the like. Further, the simulation can use realistic message traffic models, either by using a traffic matrix determined from analysis of actual program workload, or by instrumenting an existing program and logging all message traffic. However, given the size of new and proposed supercomputer hardware, consisting of 65,000 nodes or more, the performance of the simulation environment can become an issue, both in terms of running time (CPU requirements) and memory constraints.

Description of Research and Goals. The goal of the proposed research is to provide a scalable, configurable, and easy–to–use simulation environment that can be used to predict the performance of an arbitrary interconnect switch network when applied to an existing distributed application. The simulation tool will be based on our existing *Georgia Tech Network Simulator (GTNetS)* network simulation tool. *GTNetS* has been designed from the outset for scalability, extensibility, distributed simulation, and ease–of–use. It has been demonstrated to support network topologies exceeding 1 million network elements and data flows using distributed simulation on a large–scale computing platform. In this research, we will investigate efficient methods for simulating the specific needs of supercomputer interconnect networks. Specifically, we will design:

1. Configurable TopologyModel for the *k-ary-n-cube* torus topology. Parameters will include the *k* (number nodes in each dimension) and *n* (number of dimensions), link bandwidth, and one–hop delay (which may be constant across all pairs of nodes, or unique to each pair, depending on physical layout).

2. Configurable Topology for *FAT-Tree*. Another popular topology model is the so–called *FAT–Tree,* which is designed to make more bandwidth available near the root of the tree, to alleviate potential congestion at possible convergence points. The FAT–Tree is parameterized with the number of leaf nodes, fan–out at each level, as well as the bandwidth and delay parameters as above.

3. Routing and Flow–Control Models. The routing methods used in these interconnect networks differ considerably from typical wide–area network routing. There are typically no routing protocols or routing tables, but rather routing decisions are made independently at each switch node based on information found in small *flow–control digits,* or *flits*. Flow control using *backpressure* is a common approach to reacting to congestion on parts of the network. Other methods will be explored.

4. Model of *Network Interface Card (NIC)* delays. When investigating interconnect networks with potentially very small message latencies (nanoseconds), the delays between the application and the switch network can add considerably to the overall message latency. Detailed models for a variety of *NIC* cards will be developed and included in the *GTNetS* interconnect library.

5. Application Model. Utilizing instrumented MPI libraries, the message traffic of any existing MPI application can be captured and logged. Utilizing these log files, we will incorporate models of message traffic into the simulation environment, and account for reduced or extended message latency observed by the simulated interconnect model.

6. Integration of the interconnect network model with the detailed processor models currently under development at Sandia.

7. Creation of a parameterizable multi-level network model to capture network behavior and multiple degrees of fidelity. A common methodology to improve simulator performance is to utilize more abstract, less detailed models where appropriate. In this research, we will identify those areas in

the supercomputer simulation models where omission of computationally expensive details can be used without significant loss of accuracy.

8. Validation of our newly created models against an actual execution of the same application on an existing supercomputer interconnect system. Validation of simulation models can be a difficult issue, since the simulations are frequently modeling hardware or systems that do not yet exist. Fortunately, in this case our models can be configured to simulate existing interconnects, as well as newer and non–deployed hardware. Further, our application models are based on traffic traces from existing applications. These two facts should lead to a high degree of confidence in the accuracy of the simulation results.

Research Plan. The principal investigator, Dr. Riley, will identify a Georgia Tech Ph.D. student who is a U.S. citizen to begin working on this effort beginning in the Spring semester of 2005. The student will start with an extensive literature survey of existing work on supercomputer interconnect design, and analytical methods for performance analysis. The student will concurrently begin a design for the new simulation models needed for this research. The student will spend the Summer of 2005 on–site at Sandia, and will continue working upon his return to Georgia Tech, Fall 2005. Depending on availability of funding, the student will continue through 2006 with a second summer at
Sandia.

Further, Dr. Riley will plan two extended on–site visits to Sandia, once in the Spring of 2005 and a second in Summer of 2005. Each visit will be approximately 2 weeks and will result in a close collaboration between Dr. Riley and researchers at Sandia. We will leverage existing research by our group and others, including:

- The *Georgia Tech Network Simulator* is an existing tool designed for scalability and distributed simulation. All models developed in this research will be included in the *GTNetS* library, and will consistent with the design philosophy of **GTNetS**. Further, all our interconnectmodels will be suitable for distributed simulation methods leading to larger scale simulations as needed.

- Existing work by Dr. Keith Underwood's group at Sandia in modeling Network interface Cards (NIC's) will be incorporated into our models.

- The instrumented MPI library developed by Dr. Jeffery Vetter at Oak Ridge National Laboratory will provide the basis for the application models used in our work. This library is available to others, and will be provided to Sandia.

Expected Outcome. The successful conclusion of this research is expected to result in simulation models that can be used to predict performance of existing supercomputer applications on hardware platforms not yet purchased or not yet designed. The simulation tools will be delivered to Sandia, including full source code and indefinite right–to–use. We expect that this effort will lead to longer–term research collaborations between Sandia and Georgia Tech.

Title: Polyhedra Matter in Nonlinear Equation Solving

PI: J. Maurice Rojas
Computational Biology Group, and Texas A&M

Dates: October 17 – November 18, 2005

CSRI POC: Danny Rintoul, (505) 844-9592

Project Summary: The intent of my sabbatical is to assist Sandia personnel to help understand the structure and solutions of large sets of linear Diophantine equations. The goal is to improve the ability of Sandia to design molecules for applications relating to materials design and counter-bioterrorism. My work will potentially include both mathematical studies and computational implementation.

Tasks:
Typical tasks shall include the following:

1. Understanding how the structure of different sets of integer equations affects the time for the basis set of all solutions.
2. Explaining how different algorithms compare in terms of solution time and memory as the parameters of the input sets of Diophantine equations vary.
3. Outlining what is available in the literature with respect to solutions of Diophantine equations, and how this work could be leveraged by Sandia National Laboratories.

Deliverables:
1. During this sabbatical I intend to make improvements to the code that is currently used to solve these equations and potential publications submitted to refereed journals.
2. At the completion of this sabbatical, I will provide a summary of the visit, including documentation of all research results in the form of a publication or technical report, to Danny Rintoul, Org. 01412 and David Womble, Org. 01410.

Title: Large Scale Eigenvalue Methods and Model Reduction of Second Order Systems

PI: Dan C. Sorensen, Rice University

Dates: October 1, 2004 – September 30, 2005

CSRI POC: David Womble, (505) 845-7471

Project Summary:

Large Scale Eigenvalue Problems - We shall continue to develop techniques for improving the performance of ARPACK. This software for large eigenvalue problems is in wide use at Sandia and is based upon our implicitly restarted Arnoldi method. We hope to develop pre-conditioning techniques appropriate for stability and bifurcation analysis of dynamical systems. These will be closely related to the use of iterative methods for solving the equations required to implement a Cayley transformation. However, they will construct a fixed (preconditioned) polynomial operator as an approximation to the shift-invert operator.

Model Reduction of Second Order Dynamical - Systems Direct numerical simulation of dynamical systems has been an extremely successful means for studying complex physical phenomena. However, as more detail is included, the dimensionality of such simulations may increase to unmanageable levels of storage and computational requirements. One approach to overcoming this is through model reduction. The goal is to produce a low dimensional system that has the same response characteristics as the original system with far less storage requirements and much lower evaluation time. The resulting reduced model might be used to replace the original system as a component in a larger simulation or it might be used to develop a low dimensional controller suitable for real time applications.

In the past year, we have made considerable progress on the fundamental model reduction of systems of the form $\dot{x} = Ax + Bu$, $y = Cx$ where *A, B, C* are real $n \times n$, $n \times m$ and $p \times m$ matrices, while *u, y, x* are vector valued functions of time. Large systems of this form arise in many applications, for example in circuit simulation and in the simulation of PDEs. We have developed balanced model reduction techniques for large-scale systems through low rank approximation of certain system Grammians. These techniques are matrix-free in the same sense as Krylov methods for eigenvalue computation and solution of linear systems. Balanced reduction is an excellent candidate for the development of robust and widely applicable software because of the existence of *a-priori* error bounds and the preservation of important system properties. We intend to extend these results and also develop new techniques for model reduction of second order systems $M\ddot{x} + G\dot{x} + Kx = Bu$, $y = Px + Q\dot{x}$.

Such systems are far more challenging to work with but they have many more applications. Moreover, even though a second order system can be reduced to a first order system through standard techniques, a reduced model obtained from the first order formulation is usually not valid when converted back to the second order setting. We hope to develop balanced model reduction techniques that work directly with the second order system.

Title: Combinatorial Optimization: Algorithms and Applications

PI: K. Subramani, LDCSEE
West Virginia University

Dates: May 17 – August 13, 2005

CSRI POC: Bruce Hendrickson (505) 845-7955

1 Background

I am a fifth-year Assistant Professor at the Lane Department of Computer Science and Electrical Engineering, West Virginia University. I was invited by Bruce Hendrickson to spend the summer of 2005 as a Summer Faculty at CSRI with the goal of collaborative cooperation between my expertise in logic, combinatorial optimization, and graphs and some of the more pressing defense-oriented concerns of Sandia.

Graphs and combinatorial optimization arise in many Sandia applications, ranging from linear solvers and parallel tools to sensor placement and data mining. The wide range of high-impact problems were a source of continual inspiration for me during my time at Sandia.

2 Progress

One of the more interesting problems that I encountered at Sandia is the Red-Blue Shortest Path Tree problem. In this problem, we are given a graph with two types of edges, viz., blue (light-weight) and red (heavy-weight). The goal here is to find a linear-time algorithm for the Single-Source shortest path problem on this graph. The characterization of the red-blue graph structure arose out of my discussion with J. Berry about his work on the *El Dorado* project.

A second problem in which I am heavily invested is Chain Programming. A Chain program is a linear program in which a total ordering is known on the variables and this total ordering is part of the constraint system. Chain Programming problems arise naturally in the verification of sequential programs and job-shop scheduling.

Additionally, as Section 5 and Section 3 indicate, I am involved in a number of projects involving CSRI researchers. I fully expect that our collaboration will continue over the foreseeable future.

Title: Using Linearly Constrained Generating Set Search to Solve Nonlinear Programming Problems

PI: Robert Michael Lewis and Virginia Torczon
College of William and Mary

Dates: May 16 – September 30, 2005

CSRI POC: Tammy Kolda (925) 294-4769

Overview
We propose to develop techniques for handling nonlinear constraints in pattern search methods because many optimization problems are formulated with nonlinear constraints. This will be an extension of a successful ongoing collaboration that has produced numerous papers and influenced both codes and applications at Sandia.

Goals

- Use the results in Technical Report SAND2003–8550, "Stationarity results for generating set search for linearly constrained optimization," written under prior CSRI contracts and accepted for publication in SIAM Journal on Optimization [7], to derive results for general nonlinear constraints using an augmented Lagrangian approach of the form proposed by Conn, Gould, Sartenaer, and Toint [2]. The goals are both to develop a general algorithmic form and the accompanying convergence analysis as well as to write this up for submission to SIAM Journal on Optimization.

- Initiate an investigation of efficient ways to extend extant software (APPSPACK [3], in particular) to handle general nonlinear constraints.

- Begin an investigation of alternatives to the augmented Lagrangian approach to handling general nonlinear constraints (e.g., filter methods and/or an interior point approach).

Collaborators
We plan to spend one month during the summer of 2005 in residence at Sandia National Laboratories, Livermore, collaborating with the following individuals:

- Tammy Kolda (our host and primary contact),
- Bill Hart (consulting on various pattern search issues),
- Genetha Gray (consulting on the use of oracles within GSS methods, as motivated by Joe Castro's multi-fidelity optimization LDRD on which she is working),
- Josh Griffin (newly hired member of 8962 who may be a collaborator on the implementation of extensions to APPSPACK to handle nonlinear constraints),
- and student interns in residence at Sandia, as appropriate.

We also will consult on various applications at Sandia, as appropriate.

Context
We propose to spend a month in residence at Sandia, Livermore, as part of an ongoing collaboration to develop pattern search/generating set search (GSS) algorithms for solving the nonlinear programming problem

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } x \in S \end{aligned}$$

where $f: \mathrm{R}^n \rightarrow \mathrm{R}$ and $S \in \mathrm{R}^n$.

The goal is to combine the results from our previous CSRI-funded work [7] with work by Conn, Gould, Sartenaer, and Toint [2], to develop generating set search methods for nonlinear constraints. This is a further, and more flexible, generalization of the results in [10].

APPSPACK [3] currently handles nonlinear constraints by simply rejecting any trial point that violates the constraints (i.e., any x 62 S simply is not considered). This simple strategy ensures that the final answer produced by APPSPACK is at least feasible, but there is no way to ensure that the solution produced is optimal; in fact, we have a test problem for which APPSPACK produces a variety of answers (since it is an asynchronous method), none of which satisfy the Karush-Kuhn-Tucker (KKT) conditions for a constrained minimizer. Since nonlinear constraints are a critical part of the definition for many practical applications, clearly this behavior is not acceptable. It also is not necessary if we can craft algorithmic mechanisms for ensuring that the KKT conditions are satisfied, which is precisely the goal of the work we propose.

Past work

During our visit to CSRI last summer, we completely revised our paper "Stationarity results for generating set search for linearly constrained optimization" [7] in response to the first round of referee reports we had received. This proved to be an unusually productive exercise as we came up with a far more elegant—and more practical—variant of the algorithms that had been proposed for handling linear constraints [9, 11, 6].

We now have a prototype C/C++ implementation for handling the special case of linear constraints (which can occur in practice) that correctly handles all the degenerate problems we have tested. From the start, we have known that our analysis covers the case of degeneracy [9, 6, 7], but it is well known from linear programming that degeneracy is inherently difficult to handle in practice. This work was done in conjunction with Anne Shepherd, a DOE High-Performance Computer Science Fellow who spent last summer in residence at Sandia National Laboratories, Albuquerque. Michael Lewis, Anne Shepherd, and Virginia Torczon are in the process of writing up a description of the implementation for submission to SIAM Journal on Scientific Computing [8].

While in residence last summer, Michael Lewis worked with Robert Darwin, a student intern, to integrate our prototype implementation for handling degenerate linear constraints into APPSPACK. Tammy Kolda has the immediate goal of overseeing the integration of our latest procedures for handling degenerate constraints into the next release of APPSPACK [3].

Impact

The work undertaken during our visits to Sandia, Livermore is largely developmental in nature, but it has a practical impact on other work undertaken at Sandia. For instance, Tammy and Genetha Gray used prior results, as implemented in APPSPACK, in a collaboration with Kenneth Sale and Malin Young (both members of the Biosystems Research group at Sandia, Livermore), to examine the problem of transmembrane protein structure determination [4]. Tammy also worked with Michael Chiesa, Reese Jones, and Kenneth Perano to use APPSPACK to solve ASC problems involving forging processes for optimal material properties [1]. Another application includes optimization for circuit simulation using ChiliSpice and Xyce.

Further, after the collaborations of the past few summers, Tammy, with the help of Genetha Gray, thoroughly revised APPSPACK [3, 5]. Our understanding is that the new revised version of APPSPACK matched or bested all competition in the Community Problems and Solutions showdown being run by Tim Kelley, Department of Mathematics, North Carolina State University, as part of an NSF ITR project on sampling methods for optimization and control of subsurface flows (see http://www4.ncsu.edu/~ctk/community.html).

In all these applications, a question that remains is how to deal effectively with constraints, which arise quite naturally in the definition of such application problems. The success of pattern search has led to its incorporation in several software projects at Sandia National Laboratories, including OPT++ and APPSPACK, both of which are part of the DAKOTA software toolkit. Throughout the development of our software for handling constraints, we will continue to consult with all the parties who have already made use of our pattern search technology to ensure that any new software can be incorporated as quickly and easily as possible into their work and this can be applied as soon as possible to applications at Sandia National Laboratories.

References

[1] Michael L. Chiesa, Reese E. Jones, Kenneth J. Perano, and Tamara G. Kolda. Parallel optimization of forging processes for optimal material properties. In Proceedings of NUMIFORM 2004: The 8th International Conference on Numerical Methods in Industrial Forming Processes, Columbus, Ohio, June 13–17 2004.

[2] A. R. Conn, N. Gould, A. Sartenaer, and Ph. L. Toint. Convergence properties of an augmented Lagrangian algorithm for optimization with a combination of general equality and linear constraints. SIAM Journal on Optimization, 6(3):674–703, August 1996.

[3] Genetha Anne Gray and Tamara G. Kolda. APPSPACK 4.0: Asynchronous parallel pattern search for derivative-free optimization. Technical Report SAND2004-6391, Sandia National Laboratories, Livermore, California, December 2004.

[4] Genetha Anne Gray, Tamara G. Kolda, Kenneth L. Sale, and Malin M. Young. Optimizing an empirical scoring function for transmembrane protein structure determination. Special Issue on Computational Molecular Biology/Bioinformatics, INFORMS Journal on Computing, 16(4):406–418, 2004.

[5] Tamara G. Kolda. Revisiting asynchronous parallel pattern search for nonlinear optimization. Technical Report SAND2004–8055, Sandia National Laboratories, Livermore, CA 94551–9217, 2004.

[6] Tamara G. Kolda, Robert Michael Lewis, and Virginia Torczon. Optimization by direct search: New perspectives on some classical and modern methods. SIAM Review, 45(3):385–482, 2003.

[7] Tamara G. Kolda, Robert Michael Lewis, and Virginia Torczon. Stationarity results for generating set search for linearly constrained optimization. Technical Report SAND2003-8550, Sandia National Laboratories, Livermore, California, 2003. Accepted for publication in SIAM Journal on Optimization, subject to suitable revision.

[8] Robert Michael Lewis, Anne Shepherd, and Virginia Torczon. Implementing generating set search methods for linearly constrained minimization. In preparation, 2005.

[9] Robert Michael Lewis and Virginia Torczon. Pattern search methods for linearly constrained minimization. SIAM Journal on Optimization, 10(3):917–941, 2000.

[10] Robert Michael Lewis and Virginia Torczon. A globally convergent augmented Lagrangian pattern search algorithm for optimization with general constraints and simple bounds. SIAM Journal on Optimization, 12(4):1075–1089, 2002.

[11] Stefano Lucidi, Marco Sciandrone, and Paul Tseng. Objective-derivative-free methods for constrained optimization. Mathematical Programming, 92(1):37–59, 2002.

Chapter 3. Sabbaticals

The following university faculty did all or part of a sabbatical stay at Sandia National Laboratories during calendar year 2005.

Dr. Richard Byrd,
University of Colorado

Dr. Stephen Guattery,
Bucknell University

Dr. Fredrik Manne,
University of Bergen, Norway

Title: Developing Limited-Memory Methods for Nonlinear Optimization Problems

PI: Richard Byrd, University of Colorado

Dates: September 1, 2004 – February 28, 2005

CSRI POC: Paul Boggs, (925) 294-4630

Project Summary:

1. Goals

- To develop improved strategies for employing limited-memory quasi-Newton methods for general nonlinear optimization problems.
- To design and implement software to allow general use of these strategies in existing SNL codes, including MOOCHO and Split/03D.
- To explore applications at SNL, including optimization problems constrained by partial differential equations, e.g., topology optimization inverse problems, design, and control, etc. Possible specific applications include the incorporation of these ideas into Sierra codes such as Salinas and Premo with application to structural members of AFF assemblies.
- To draft a paper, or papers, on advances in limited-memory technologies for submission to a referred journal.
- To investigate other topics in collaboration with SNL personnel, including nonlinearly constrained search methods.

2. Background

Limited memory (LM) quasi-Newton methods constitute one of the two most effective approaches currently existing for solving large scale unconstrained optimization problems (the other being truncated Newton). LM methods approximate the Hessian of the objective (or Lagrangian for constrained problems) by storing the quasi-Newton update vectors for, say, 3-50 of the most recent iterations. Then at each iteration, these updates are applied to a nominal initial matrix to construct the Hessian approximation. These methods have minimal computational overhead, and can be used regardless of the problem structure. However, they have difficulty dealing with ill-conditioning and with attaining high accuracy. [4, 1] We are interested in ways to make these methods more effective.

3. Proposed Research

- **Initial Approximations (Preconditioners)**
  In almost all cases up to now the initial matrix, to which the LM updates are applied, is chosen to be a multiple of the identify, multiplied by a suitable scalar. However, in many situations there is a part of the Hessian that is important, and which can be computed and factorized cheaply. This is analogous to preconditioning in Krylov methods. Preliminary experiments on protein structure prediction problems, using the Hessian of the bonded terms for the initial matrix, indicate that the number of LM steps can be reduced by a factor of six or more by this approach. (We should point out that our approach involves a scaled initial matrix for the remaining part of the Hessian, which makes the effective much more effective than in the experiments reported by Xie and Schlick [5].) We believe this technique can be very effective in many applications. For example, in computing a Hessian approximation in least squares formulations arising from inverse problems e.g., data matching for validation in ASCI and source-inversion problems, one can sometimes easily compute the diagonal of the J^TJ term and/or readily compute the contribution from the regularization term. In the context of multi-physics applications, there are often parts of the Hessian that can be easily computed.

- **Use with AD**
  An important aspect of this procedure is to efficiently compute the preconditioning part of the initial Hessian. It is clear how to use AD software or finite differencing to do this in a time related to the sparsity of the preconditioner. However, in the protein context, implementations of the objective function with modeling language such as AMPL indicate that repeated expressions can be saved and reused in other parts of the preconditioner, resulting in significant computational savings.

- **Guidelines for adjusting memory as iterations proceed**
  Currently the number of directions to be saved in a limited memory method is chosen in an ad hoc fashion with little guidance available on how to choose this number. We believe it should be possible to dynamically estimate the relevance of a past direction at the current point by maintaining a measure of the deviation of the objective function from a quadratic. This would lead to an adaptive criterion for removing a step that occurred more than certain distance from the current point and allow a variable number of previous steps to be used.

4. Related Optimization Research

- Nonlinearly constrained direct search methods
  This group at Sandia is known for the development of practical direct search methods and software. Most of this work has been for unconstrained optimization, although an approach for linear constraints has recently been developed by Kolda, Lewis and Torczon [3]. We have broad experience in the handling of nonlinear constraints in optimization problems. Therefore, I propose to work with Tammy Kolda and Patty Hough to develop a direct search method for nonlinearly constrained optimization.

- **PDE-constrained optimization**
  Large-scale nonlinear optimization problems with partial differential equations (PDE) as constraints are becoming increasingly important at SNL. Although much of the work in limited-memory methods has been directed toward unconstrained problems, these techniques can, in principle, be used in this more difficult context. I propose to work with Paul Boggs and Kevin Long in California, and with Bart van Bloemen Waanders and Ross Bartlett in Albuquerque, on various aspects of these problems. In particular, the topology optimization approaches seem very interesting.

5. Software Development

Our goal would be to develop preconditioned limited memory software, that given a specification of a preconditioning part of the objective, would automatically compute an approximation to the Hessian of this term. It would then apply limited-memory quasi-Newton with that as a preconditioner. Any adaptive memory size strategies that are developed would also be included.

6. Applications

Limited-memory methods are currently being used in several SNL codes including MOOCHO (a reduced sequential quadratic programming (SQP) code), Split/03D (a full space SQP code), as well as in a geophysical inverse code and a prototype topology optimization code. Improvements in the efficiency of the limited-memory aspects of these code would have a dramatic effect on the overall performance. The above codes are being used in a number of ASCI-related applications.

7. Participation

The proposal is that Richard Byrd spend approximately three months at Sandia National Laboratories, California as part of his sabbatical. Richard's primary contact will be Paul Boggs. Others in the department, including Tammy Kolda, Kevin Long, Patty Hough, and Genetha Gray are also interested in working on related optimization projects with Richard. In addition, Bart van Bloemen Waanders,

Ross Bartlett, David gay and Bill Hart in Albuquerque are also interesting in collaborations with Richard; thus, Richard will spend approximately two weeks in Albuquerque.

8. Deliverables

There are three major deliverables:

- The development of advanced strategies for using limited-memory ideas in very large scale nonlinear optimization problems;
- Draft of research paper describing results; and
- Prototype implementation of Limited-Memory software

9. Impact

As noted above, LM methods are among the most effective methods for solving large-scale nonlinear optimization problems. Nevertheless, many questions remain as to how to most effectively use these techniques. If successful, the research here will shed light on the very promising ideas of preconditioning the LM methods and of keeping a variable number of previous steps. Increased understanding and the development of appropriate software would allow the incorporation of new, more powerful, variants of LM into many existing SNL codes.

References

[1] R. Byrd, J. Nocedal and R. B. Schnabel, "Representations of quasi-Newton matrices and their use in limited memory methods", *Mathematical Programming 63* (1994): 129-157.

[2] J. E. Dennis, H. J. Martinez, R. A. Tapia, "Convergence theory for the structured BFGS secant method with application to nonlinear least squares", JOTA 61 (1989): 161-177.

[3] T. G. Kolda, R. M. Lewis, and V. Torczon, "Stationarity Results for Generating Set Search for Linearly Constrained Optimization", Technical Report SAND2003-8550, Sandia National Laboratories, Livermore, California, October 2003.

[4] D. C. Liu and J. Nocedal, "On the limited memory BGFS method for large scale optimization methods", *Mathematical Programming 45* (1989): 502=3:528.

[5] D. Xie and T. Schlick, "Efficient Implementations of the truncated-Newton algorithm for large-scale chemistry applications", *SIAM Journal on Optimization 10* (1999): 132-154.

Title: Support Theory for Analyzing Preconditioners

PI: Stephen Guattery, Bucknell University

Dates: January – June 2005

CSRI POC: Bruce Hendrickson, (505) 845-7599

Project Summary: I will be on sabbatical from the Department of Computer Science at Bucknell University for the 2004-2005 academic year. I am applying for a sabbatical appointment at the computer Science Research Institute at Sandia National Laboratories in Albuquerque, for the period mid-January through mid-June 2005. During that time I hope to extend current collaborations with Sandia researchers and develop new collaborations. My main area of interest is in connections between algebraic properties of matrices and combinatorial properties of graphs representing those matrices, and in applying those connections in the area of large scientific computations. I am more generally interested in the areas of sequential and parallel algorithms, graph theory, linear algebra, and scientific computing.

Background

My research has focused on connections between graph properties and eigenvalues and eigenvectors of the corresponding Laplacian matrices, and more generally real symmetric and Hermitian matrices. Such connections are useful in the analysis of algorithms for scientific computation. My most recent work seeks to exploit the properties of graphs associated with sparse matrices in designing practical algorithms for solving numerical problems, and in creating techniques for better analyzing such algorithms. This is an outgrowth of research I have done on generalizing graph embedding techniques for bounding eigenvalues of Laplacian matrices.

In embedding techniques for bounding eigenvalues, matrices are considered in terms of their underlying graphs. Embeddings involving the underlying graphs are constructed, and properties of these embeddings (e.g., congestion and dilation) can be used to compute bounds on eigenvalues of the original matrices. It can be proved that the standard embedding techniques do not provide tight eigenvalue bounds for certain graphs.

Working with Gary Miller, I showed the reason why standard embedding techniques did not give tight bounds for all Laplacians is that edge directions were not considered in the representation of embeddings [GM98, GM00]: the underlying graph of a Laplacian is determined from a factorization of the matrix into a product U^TWU, where U is a directed edge-vertex incidence of matrix and W is a diagonal matrix of edge weights. In the standard approach, embeddings were specified without considering the directions specified in U. By including the directions, it is possible to use the embedding to construct a matrix whose eigenvalues can be put into an exact correspondence with the eigenvalues of the original Laplacian. The map between corresponding pairs of eigenvalues is a simple function involving reciprocals. In the case of a generalization of the Laplacian that has non-zero row sums (this corresponds to the existence of a Dirichlet boundary condition), our generalized embedding matrix can be used to construct the inverse of the original matrix. I generalized this result to all real symmetric matrices [Gua98].

Embedding techniques have also been used to bound the eigenvalues of preconditioned systems, where the eigenvalue bounds can be used in bounding the rate of convergence of iterative solution methods applied to these systems.

It turns out that the ideas used in bounding the eigenvalues of a matrix and those used in bounding eigenvalues of preconditioned systems can be expressed in a general way that encompasses both applications. I have recently completed a paper [BGH] in a collaboration with Bruce Hendrickson and Erik Boman at Sandia National Laboratories. The paper proves structural results that generalize the ideas underlying a preconditioning technique referred to as support preconditioning. It ties together results about support theory with results about the structure of matrix inverses and pseudoinverses. In particular, the paper explains a way to generate factors of the inverse (or pseudoinverse if the matrix is singular) of any Hermitian matrix. The result suggests possible ways in which factors of an *approximate inverse* of a matrix might be efficiently computed.

Goals of a Sandia Visit

I believe the Computer Science Institute at Sandia National Laboratories would be an excellent place to spend part of my sabbatical year. My interests mesh well with those of the CSRI, in particular my interest in problems related to large-scale scientific computations. I have a number of goals for the proposed appointment:

- I am interested in applying and in extending my past work. A particular area of interest is approximate inverse preconditioning. My work with Bruce Hendrickson and Erik Boman formulates the inverse of a matrix in terms of a generalized embedding. It suggests some possibilities for computing factors of approximate inverses. For example, computing approximate solutions to one of the equations in our result may provide approximate factors to a matrix inverse that can be used in preconditioning. It is also worth investigating if there are other ways to apply these ideas in practice as well (e.g., there may be ways to apply generalized embedding ideas to construct preconditioners in a graph-theoretical way).

  More generally, I am interested in learning about the range of approaches to generating approximate inverses. There are a number of potential benefits:

  First, my methods may be applicable to better analyzing or understanding these approaches. Second, a broader understanding of current approaches may help me develop ideas for applying our generalized embedding results. Third, I want to understand if there are any general principles that are common to all such methods.

  Work on these topics would also allow me to continue my collaboration with Bruce Hendrickson and Erik Boman.

- I want to learn more about the practical issues involved in preconditioning linear systems. I have focused mainly on the theoretical issues up to this point. Theoretical ideas can suggest directions, but in scientific computation the details of implementations can have a large effect on the efficiency and stability of the computation, and the accuracy of the results. I would like to develop a deeper understanding of the issues involved in applying these techniques in practice so I can move my ideas more directly and quickly into applications.

- I would like to expand my collaborations with Sandia researchers. I work in a small department in a fairly small university that focuses on undergraduate education. There are not other faculty members working in my area, nor are there graduate students who can work with me. I want to spend the sabbatical developing working relationships. Sandia offers excellent possibilities for collaboration. It has a contingent of people working on the sorts of problems that interest me, both from the theoretical and practical directions. This will allow me to extend what I already do, and to focus on new directions, as discussed in the previous point.

  I should mention that my connection with Sandia dates to 1998, when I was a finalist for the von Neumann Fellowship. I have worked with Bruce Hendrickson and Erik Boman (off-and-on as permitted by my teaching schedule) since summer 2001. I am interested in working with other people in related areas, particularly those that support large-scale scientific computations.

Other Details

As I note in my CV, I am a U.S. citizen. As noted in the introductory paragraph, I would like to spend five months working at Sandia, starting early-to-mid January and continuing through early-to-mid June. Exact start and finish dates can be specified later to fit our schedules, as can any arrangements to meet administrative or other constraints imposed by Sandia or Bucknell.

Title: Parallel Graph algorithms for Numerical Optimization

PI: Fredrik Manne, University of Bergen, Norway

Dates: August 2004 – June 2005

CSRI POC: Bruce Hendrickson, (505) 845-7599

Project Summary: With this I apply to spend my sabbatical term for the academic year 2004-2005 at CSRI, Sandia National Laboratories. The following document gives a closer description of my background and the research I intend to perform while at CSRI.

Background
I am employed as a professor at the Department of Informatics, University of Bergen, Norway. This is a position I have held for the last 6 years, the last 3 years as vice chairman of the department. My primary research interests include Combinatorial Scientific Computing, graph algorithms, parallel algorithms, and distributed computing.

Current Research
My research is focused on discrete algorithms in large scale applications, a field known as "Combinatorial Scientific Computing". Within this field I have worked with parallelizing various linear algebra codes, load balancing on parallel computers, and developing parallel algorithms for graph coloring. Lately, I have also worked on distributed algorithms.

My research spans from practical implementations on state of the art super computers to more theoretical computer science, particularly algorithm development and analysis.

Within the field of graph coloring I have together with my former student A. Gebremedhin developed the first greedy truly scalable parallel graph coloring algorithm.

The Graph coloring problem (GCP) consists of assigning as few colors as possible to nodes in a graph such that two adjacent nodes are assigned different colors. The GCP arises in a number of scientific computing and engineering applications such as parallel numerical computations, optimization, time tabling and scheduling and frequency assignment.

In a parallel application a graph coloring is usually performed in order to partition the work associated with the nodes into independent subtasks such that the subtasks can be performed concurrently. The underlying assumption being that the nodes represent tasks and that the edges are dependencies among these. The tasks with nodes that have the same color will then make up an independent set that can be performed concurrently. One example of this is in parallel adaptive mesh refinement where a graph coloring can be used to determine independent sets of triangles that can be refined simultaneously without interfering with each other.

In numerical optimization graph coloring is used to minimize the number of function evaluations needed to compute Jacobian and Hessian matrices via finite differences or though automatic differentiation.

There exists several linear, or close to linear, time greedy coloring heuristics that have been shown to produce colorings of high quality. Even though these algorithms are fast they can on very large or distributed data sets still be costly to evaluate. For this purpose it is desirable to develop fast parallel and/or distributed algorithms to solve the GCP. Previous efforts have lead to parallel algorithms that allows the graph to remain distributed while the coloring is computed. Unfortunately, these algorithms did not give a reduced execution time as the number of processors were increased.

Our *shared memory* algorithm was the first algorithm for the GCP that achieved speed-up as the number of processors were increased. The algorithm is based on letting each processor color its portion of the graph

independently of what the other processors are currently doing. As this might lead to an inconsistent coloring a second stage is used to detect any conflicts which are then later corrected.

Our initial motivation for working on this problem was to speed up a parallel eigenvalue computation. Later collaboration with Alex Pothen from The Old Dominion University has led to applications of this work in the area of numerical optimization methods to compute Jacobian and Hessian matrices. In this setting one uses function evaluations to determine rows or columns (or rows) using one function evaluation if the desired columns are structurally orthogonal. This is important to exploit as the function evaluations can be costly. Partitioning the columns into the minimum number of groups such that each group can be evaluated with one function evaluation reduces to a graph coloring problem. In this setting the notion of "neighbor" is somewhat more general than for the traditional problem and depending on the type of matrix one wish to compute one might have to take into account nodes further away than just the closest neighbors. Theoretical and experimental results have shown that our algorithm can be modified to handle these conditions well.

As a continuation of my work on parallel algorithms I have lately also started to work on distributed algorithms. My main interest concerns "self -stabilizing" algorithms. These are distributed algorithms that run on a set of mobile units that try to solve some common problem without the use of a central command. Each unit can only "see" and communicate with its nearest neighbors. The combined effort of all the units should together be able to come up with a global solution to a given problem without any particular unit having complete knowledge of the structure of the entire system. In addition, if the configuration changes due to movement or malfunctioning among the participants the solution should also change to reflect this. Applications of these types of algorithms can be found in various mobile ad-hoc networks. This could for instance be for maintaining a communication network among autonomous robots in a hostile environment or as a general purpose network without base stations. My research in this area has been performed in close cooperation with Jean Blair from The United States Military Academy at West Point. As of 2004 I have received a grant from the Norwegian Research Council for a project to study frequency allocation problems in these types of networks. This is again closely related to my work on parallel graph coloring algorithms.

I have also been working on a project with the goal of computing exact or lower bounds on the minimum amount of fill in grid graphs when performing sparse factorization. The purpose of this work is to establish bounds for a number of test cases such that one can compare the quality of existing software for reordering sparse matrices. So far this has turned out to be a very difficult problem and we have only been able to come up with exact lower bounds for relatively small instances.

Research Plan

It is my wish to come to CSRI to continue my work on parallel graph coloring in close collaboration with the researchers at CSRI. CSRI has expertise on the types of applications such as the Zoltan project and ASCI simulation codes, where this work can have an impact. In particular, the current version of the parallel graph coloring algorithm has as previously mentioned so far only been implemented on shared memory parallel computers. It is my intention to develop and test a *distributed memory* version of this algorithm. This is an important step in making the algorithm more useful to others as nearly all of the Sandia codes that could make use of a parallel graph coloring algorithm are themselves implemented on distributed memory platforms.

There are several issues involved in developing such a code. One of these is that on a shared memory computer each processor has direct access to the colors that the other processors have assigned so far, but on a traditional distributed memory computer using the MPI-1 message passing library for communication this information would have to be requested and sent back explicitly thus requiring the processors to synchronize which would slow down the computation. There is also a possible danger of flooding the recipient with requests that it is not yet ready to process. A more desirable solution would be to use explicit non-intrusive "get" commands to fetch information from the other processors without them having to directly participate in the exchange of data. These types of commands are now available through the message passing library MPI-2. However, this needs to be tested and MPI-2 has still not been implemented on all platforms that support MPI-1.

I also intend to continue the work on computing lower bounds for the amount of fill in grids graphs making use of the extensive knowledge about algorithms for sparse reordering at CSRI.

Although I am not aware of any research at CSRI (or Sandia Labs in general) related to self-stabilizing algorithms I plan to continue this work partly in cooperation with Jared Saia from the University of New Mexico.

Contacts within CSRI and UNM

I have had regular contacts with Erik Boman and Bruce Hendrickson the last couple of years. During "SIAM 50th Anniversary and 2002 annual meeting" in Philadelphia Hendrickson and Alex Pothen invited me to give a talk on "Parallel algorithms for distance-2 graph coloring". During the spring of 2003 I visited CSRI for three days to prepare for spending my sabbatical there. I had discussions with several researchers about possible future research projects and also gave a talk about my own research. The most concrete outcome of this were discussions with Karen Devine about developing a parallel graph coloring algorithm for distributed memory and merging this with Zoltan. One possible application might then to be parallelize the hypergraph partitioner currently under development in Zoltan.

While I was in Albuquerque I also visited The Department of Computer Science at The University of New Mexico where I had talks with Professor Jared Saia bout common interest within the field of distributed algorithms. I also talked with the department chairman Professor Deepak Kapur about spending part of my sabbatical at the University of New Mexico. As a consequence I have since received a letter of invitation from Dr. Kapur inviting me to do so.

Another connection with Sandia concerns work that I did some years ago on efficient load balancing algorithms that has since been continued by Ali Pinar who was a student of Bruce Hendrickson and by Johan Steensland who is now a post doc at Sandia Labs in Livermore.

Chapter 4. Workshops/Sponsorships

Title: Workshop on Large-Scale Robust Optimization

PI: Bart van Bloemen Waanders

Dates: August 31 – September 2, 2005

CSRI POC: Bart van Bloemen Waanders, (505) 284-6746

Project Summary: The purpose of this workshop is to bring together leading researchers in the area of robust optimization and provide Sandia critical technological information for future development in this area. Uncertainty in simulation models and data measurements is a wide spread issue for many important engineering and other scientific problems. Robust optimization offers a capability to account for variations and uncertainties by typical posing the problem as a "min max" problem. The solution methods of such a formulation is at the core of this workshop. Solving large scale problems is eventually our goal, but thus only small scale problems can be solved with existing solution methods. We will invite participants to specifically focus on this particular issue.

Title: Frontiers of Extreme Computing

PI: Erik DeBenedictis

Dates: October 23-27, 2005

CSRI POC: Erik DeBenedictis, (505) 284-4017

Project Summary:

Transitioning Moore's Law to the Next Generation

Supercomputers have revolutionized science and defense in the last several decades, but additional effort will be required to maintain the trend. Moore's Law has driven computers to be ever faster and progressively more capable of simulating larger problems with more sophisticated physics, thereby solving problems of progressively more importance to society. There is growing alarm that the current trend of "Moore's Law" is reaching its end. Applications specialists in several fields foresee requirements up to 1 Zettaflops (10 million times the speed of the currently-fastest supercomputer) to complete just their missions. These requirements not only exceed the growth rate of Moore's Law, but also exceed the physical limits of computers based on the physics currently underlying their operation. This opens a "mission gap" between the peak performances of supercomputers based on current trends and the mission needs of applications. There are options for filling this gap by developing new supercomputers based on disruptive technologies, yet the community must commit to one or a few to have sufficient resources to develop a solution. Furthermore, new and important problems based on optimization, inverses, and data analysis will have fundamentally larger resource requirements than simulations, with these new problems driving even higher computing requirements.

An important new workshop is being organized to match the continuum of important supercomputing applications with over-the-horizon computing methods fostered by the approaching nanoscale devices and to determine the limits of practical computing imposed by the constraints of basic physics and technology. Although not asserting a particular target performance value, a roadmap for staging advances coordinated with likely technology progress will be developed that will traverse the end of the reign of transistorized microprocessors and cross in to the domain of post-transistor nanotech devices and reversible logic at the end of the next decade. But even beyond this, participants will consider the factors determining the ultimate capabilities and what technologies may enable them and the problems these supercomputers will solve.

Organization and Key Topics:

Problems to Solve. Identify important problems that can be solved with supercomputing and their resource requirements. In some cases, scaling up today's applications will enable solution to new and important problems. In other cases, new problems will require new applications. Characterize the resource needs of the applications in decadal levels (1, 10, 100 Petaflops, 1, 10, 100 Exaflops, etc…), being mindful of the basic capabilities offered by Moore's Law and disruptive technologies in terms of FLOPS, memory, I/O, interconnect, and other relevant features.

Baseline Technologies. Identify the physical limits of supercomputer classes in use today, such as clusters, MPPs, and other approaches, based on the principles of physics and available roadmaps.

Disruptive Technologies. Identify other classes of computation that may succeed the baseline technologies based on:
New architectures, such as Field Programmable Gate Arrays (FPGA), Processor in Memory (PIM), the Vector architecture (reborn), and others.

New devices capable of computing, such as RSFQ, CNFET, RTD, SET, Y-junctions, Moltronics, Quantum Dots, spintronics, and other devices of which the participants may be aware.
New ways of using devices, such as adiabatic logic design or reversible logic.

Programming Methods. Billions have been invested in code for supercomputers, which is predominately composed of Fortran or C/C++ code with MPI or OpenMP communications. Determine which disruptive technologies are incompatible with existing code or programming methods, thereby classifying application missions into those that can leverage the existing software investment and which cannot.
Reconciliation and Planning. Determine the suitability of running each application at particular scale levels with particular computing technologies, providing a matrix matching the ability of each relevant technology to run an application.
Nanoscale Technology directions and alternatives
Advanced computer architecture and parallel execution models
Low power technology and Reversible logic
Peta/Exascale applications and algorithms
Scalable operating and runtime system software
System engineering
Programming models and languages
Reliability and fault tolerance
Measuring success: metrics and methods

Charter to Groups
To establish the first long-term roadmap for the future of high performance computing from the near-term range of Petaflops-scale through to the asymptotic realm of extreme computing beyond an Exaflops as determined by the capabilities and limitations of future enabling technologies in order to identify critical research directions for continued growth in sustained performance for future applications and to determine the ultimate bounds on real world problems.
Organizers propose four groups of about 20 participants each in two general groups as follows:

Application Pull	**Technology Push**
Real world problems that become solvable with each new generation of computer technology	End of Moore's Law according to ITRS (SIA) roadmap Nano-scale technology implementations of digital logic functions Extreme computing technologies

Working Group 1: Applications Pull
This group will attempt to model society's process for committing to develop and use new computer technology. The groups will consider progressively more powerful and esoteric technology options. To understand society's motivation for supporting technology research, the group will first characterize new problems that become soluble with an uptick in technology. Then, the challenges of developing the technology will be assessed with a specific emphasis on the challenges (hardware, software, science, experimentation) needed to practically solve the problems identified. Reconciling the technology development costs with benefit to society becomes a roadmap for selling the process.

Working Group 2: End of ITRS
The ITRS extends semiconductor CMOS and DRAM technologies to a few tens of nanometers with increasing uncertainty due to cost, fabrication process, and power concerns. Using these projections at the end of the predicted technology path, this working group will summarize the technology operational properties, determine likely performance characteristics of both conventional and custom architectures implemented at this technology design point, and consider one or more challenging applications of importance that will both require and benefit from this level of capability. Key technical issues and challenges will be identified to achieve this level of capability and required research directions will be specified.

Working Group 3: Nano-scale and other Leap-frog Technologies
Nano-scale technology of the future will permit the definition, design, and fabrication of logic devices at near atomic scale that will yield component densities unlikely to be exceeded due to quantum limitations. Alternative technologies like RSFQ logic exploit superconductivity to deliver unprecedented switching rates (100s of GHz) at low power. Such innovations in asymptotic technologies will achieve ultimate capabilities from physical devices still operating within the domain of Boolean logic. This group will determine the level of performance that may be achieved within this technology realm, the computational

structures and architectures that may be required and employed to exploit it, and the applications that may be enabled by it.

Working Group 4: Extreme Technologies and Methods
This group will explore computers unconstrained by anything except the laws of physics. It is known that reversible logic and quantum computing create another performance tier while still being recognizable as computers. Of course, participants are free to speculate on more far-fetched technologies like neural nets and DNA computers. Since the groups will have already considered simpler technologies in the Monday and Tuesday sessions, this session could find important problems that can only be solved with these esoteric technologies.

Title: Software Institutes Planning Workshop

PI: David Womble

Date: August 4-5, 2005

CSRI POC: David Womble, (505) 845-7471

Project Summary: The Software Institutes Planning Workshop is one of several workshops being held in preparation for the SciDAC II program.

One of the new elements being proposed for the SciDAC II program is "Institutes." An institute in envisioned to be a forum or vehicle for collaborations between laboratory, academic and industrial research. Within an institute, researchers will be expected to tackle major unsolved problems in the computer sciences.

The goal of this workshop will be to produce a white paper of approximately ten pages that "lays out the case" for institutes. To accomplish this, it must clearly motivate the institutes in terms of the benefits to the Department of Energy and to the nation; it must refine the concept of an institute, including proposed scope, activities and governance model; it must develop a list of topics that might be amenable to attack by an institute; it must estimate the cost of an institute; and it must consider interactions with other agencies.

The workshop will be structured to produce a white paper. The tentative agenda and organization of the meeting will be as follows.

There are many examples of successful (and unsuccessful) institutes that participants are encouraged to consider before arriving at the workshop, e.g., Aspen Institute, Institute for Theoretical Physics. In addition, many of the workshop participants have had some experiences with institutes of varying types and should come prepared to share these experiences, both good and bad. Finally, participants should give some thought before the meeting both to questions that might be successfully addressed by an institute and to the structure of their "ideal" institute.

The confirmed attendees for this workshop are

Al Geist (ORNL) gst@ornl.gov
Bart Miller (Wisconsin) bart@cs.wisc.edu
Bill Gropp (ANL) gropp@mcs.anl.gov
Bill Kimmerly (PNNL) william.kimmerly@pnl.gov
Bill Saphir (LBNL) WCSaphir@lbl.gov
Chris Oehmen (PNNL) christopher.oehmen@pnl.gov
Craig Steffen, (Illinois/NCSA) csteffen@ncsa.uiuc.edu
David Bernholdt (ORNL) bernholdtde@ornl.gov
David Jefferson (LLNL) jefferson6@llnl.gov

David Womble (SNL) dewombl@sandia.gov
Dennis Gannon (Indiana) gannon@indiana.edu
Fred Johnson (DOE) fjohnson@er.doe.gov (by telephone)
Jack Dongarra (Tennessee) dongarra@utk.edu
John Mellor-Crummy (Rice) johnmc@rice.edu
John P. Ziebarth (LANL) ziebarth@lanl.gov
Karsten Schwan (Georgia Tech) karsten.schwan@cc.gatech.edu
Kim Yates (LLNL) yates2@llnl.gov
P. Sadayappan (Ohio State) saday@cis.ohio-state.edu
Paul Hargrove (LBNL) PHHargrove@lbl.gov
Pete Beckman (ANL) beckman@mcs.anl.gov
Rao Garimella (LANL) rao@lanl.gov
Ron Brightwell (SNL) rbbrigh@sandia.gov
Steven Parker (Utah) sparker@cs.utah.edu
Wu Feng (LANL) feng@lanl.gov

The workshop organizers are David Womble, Sandia National Laboratories and Al Geist, Oak Ridge National Laboratories.

Title: Workshop on High Performance Interconnects for Distributed Computing (HPI-DC)

PI: Ada Gavrilovska
Georgia Tech.

Dates: July 24, 2005

CSRI POC: Rolf Riesen (505) 845-7363)
Keith Underwood (505) 284-9283

Project Summary: The purpose of this workshop is to explore the confluence of WAN technologies with high performance interconnects, as applicable or applied to realistic high end applications. The intent is to create a venue that will act as a bridge between researchers developing tools and platforms for high-performance distributed computing, end user applications seeking high performance solutions, and technology providers aiming to improve interconnect and networking technologies for future systems.

Topics of interest include:

The emergence of 10.0 GigE, InfiniBand, programmable NICs, network processors, and protocols like DDP and RDMA over IP, make it possible to create tightly linked systems across physical distances that exceed those of traditional single cluster or server systems. Further, these technologies can deliver communication capabilities that achieve the performance levels needed by high end applications in enterprise systems and like those produced by the high performance computing community.

Topics of interest include:

- Hardware/software architectures for communication infrastructures for HPC
- Data and control protocols for interactive and large data volume applications
- Novel devices and technologies to enhance interconnect properties
- Interconnect-level issues when extending high performance beyond single machines, including architecture, protocols, services, QoS, and security
- Remote storage (like iSCSI), remote databases, and datacenters, etc.

The expected attendance is 30. Requested amount $5,025.

Title: 16th International Conference on Domain Decomposition Methods (DD-16)

PI: Ray Tuminaro

Dates: January 11-15, 2005

CSRI POC: Ray Tuminaro, (925) 294-2564

Project Summary: This letter follows up on our telephone conversation on CSRI co-sponsorship of the 16th International Conference on Domain Decomposition Methods (henceforth "DD-16") of 21 May 2004, which in turn follows up on my letter with enclosures describing the conference series and the context of the 16th meeting of 30 March 2004.

Sandia was a major sponsor of the 10th International Conference on Domain Decomposition Methods in 1997, when it was held in Boulder, Colorado. I am hoping that Sandia will want to maintain close identification with the series through co-sponsorship of its first return to the United States since that very successful meeting, which was the second-largest in the history of the series. Sandia's perspective on the local organizing committee is represented by Ray Tuminaro. It is also likely that there will be a plenary speaker from Sandia in 2005, due to Sandia's technical prowess in the field of scalable PDE simulations. (The ultimate decision on plenary speakers is rendered by the international scientific committee that governs this series, independent of co-sponsorship relationships.)

In my opinion, the benefits of Sandia co-sponsorship of this meeting will be several, beginning with association. The domain decomposition meeting series is respected internationally and draws top ranked applied mathematicians with interests in large-scale computational science. Its proceedings are covered in MathSciNet. It has an informative organizational website, which contains technical reference material. The 16th meeting will be held in co-operation with the SIAM Activity Group on Supercomputing. I believe that Sandia will be happy to have its logo displayed (assuming you so allow) on conference materials, since many scientists with whom Sandia has collaborative technical interests will participate in this meeting.

I have before me a one-pager on the technical focus of the CSRI. Well-written, it leads with scalable solvers. That is the lead subject of DD-16. CSRI interests also include the following topics that are welcome at and likely to make an appearance at DD-16, as they have at past DD meetings: optimization, PDEs, graph-based algorithms, load balancing, scalable and heterogeneous platforms. Developments in these areas reported at the conference that are oriented towards massively parallel distributed memory machines, such as Red Storm, can carry over into practice at Sandia.

To give the meeting something unique that would not be possible without Sandia's technical participation or co-sponsorship, I would like to propose a minisymposium on domain decomposition methods in PDE-constrained optimization. Having just come from a day of talks on this subject at a Sandia-sponsored workshop in Santa Fe this past week, I am aware both of its relevance to DD-16 and of a group of potential presenters, who would be excellent for opening up this rapidly developing area for a larger national and international audience.

To support a plenary speaker to anchor this minisymposium and fellowships that will allow U.S. citizen students to attend the meeting and to participate in it, I request a donation to the conference budget of $5,000.00. I hope the CSRI advisory board will look favorably upon this meeting, as beneficial to Sandia and to the community in which Sandia operates. If you agree, I will follow up with detailed instructions on where to send a check, which we can deposit during either the remainder of FY'04 or early in FY'05, according to your convenience, for expenditures that will be incurred in January 2005.

Chapter 5. Seminar Abstracts

The CSRI hosts an active short-term visitor program, which is closely aligned with a seminar schedule. Short term visit typically last between two days and two weeks and include a seminar as well as a broad range of meetings and collaborations with Sandia staff. The longer visits are encouraged to increase the likelihood that a significant collaboration will develop. The abstracts for the CSRI seminar are listed below in alphabetical order by speaker.

Title: Parameter Estimation for a Nonlinear HIV Dynamics Model

Speaker: Brian M. Adams, LTE Candidate, North Carolina State University

Date/Time: Wednesday, March 9, 2005, 10:00-11:00 am (PDT)

Location: Building 980 Room 95 (Sandia NM)
Building 940 Room 1182 (Sandia CA)

Brief Abstract: Human Immunodeficiency Virus (HIV) currently affects 38 million people worldwide. Since discovering HIV in the early eighties, researchers have made great strides in understanding this immune system-decimating retrovirus. They have consequently implemented successful treatment strategies that delay progression to AIDS, thus prolonging a patient's life. Mathematical models of in-host infection have played key roles in these developments by helping scientists gain insight into viral dynamic mechanisms and rates.

This talk begins with an introduction to HIV infection and clinical data from a Boston-based acute HIV infection study, which motivates desirable model features. A nonlinear ODE model for HIV infection dynamics is described and a statistical/mathematical inverse problem methodology for estimating its dynamic parameters from data is presented. This methodology is motivated by varied patient outcomes over the disease time course, and therefore assumes that dynamic parameters vary between individuals. The estimates resulting from this process are therefore general probability density functions, describing the distribution of parameters across the population. In the formulation considered, the optimization problem reduces to a quadratic programming problem, albeit one that often requires regularization to achieve good results.

The proposed methods are validated on simulated data and compared to traditional nonlinear least squares methods of data fitting for obtaining individual parameter estimates. In addition to the interindividual variability captured by the parameter distribution, quantification of uncertainty in the inverse problem process is discussed, with implications on reliability of estimates from the process. Finally, parameter estimation results using viral load and T-cell count data from patients in the acute HIV infection study are discussed. It is hoped that models validated with such data can predict the efficacy of vaccines or novel treatment strategies.

CSRI POC: Scott Mitchell, (505) 845-7594

Title: Quantum Computation for Quantum Chemistry

Speaker: Alan Aspuru-Guzik, University of California, Berkeley

Date/Time: Monday, December 12, 2005, 10:00-11:00 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: The calculation time for the energy of atoms and molecules scales exponentially with system size on a classical computer, but polynomially using quantum algorithms. We demonstrate that such algorithms can be applied to problems of chemical interest using modest numbers of quantum bits. Calculations of the H2O and LiH molecular ground-state energies have been carried out on a quantum computer simulator using a recursive phase estimation algorithm. The recursive algorithm reduces the number of quantum bits required for the read-out register from approximately twenty to four. Mappings of the molecular wave function to the quantum bits are described. An adiabatic method for the preparation of a good approximate ground-state wave function is described and demonstrated for stretched H2. The number of quantum bits required scales linearly with the number of basis functions used and the number of gates required grows polynomially with the number of quantum bits.

CSRI POC: Alex Slepoy, (505) 284-3650

Title: Architectural Requirements for Efficient Execution of Graph Algorithms

Speaker: David A. Bader, University of New Mexico

Date/Time: Thursday, March 24, 2005, 2:00-3:00 (MT)

Location: Building 980 Room 95 (Sandia NM)
Building 915 Room S45 (Sandia CA)

Brief Abstract: Combinatorial problems such as those from graph theory pose serious challenges for parallel machines due to non-contiguous, concurrent accesses to global data structures with low degrees of locality. The hierarchical memory systems of SMP clusters optimize for local, contiguous memory accesses, and so are inefficient platforms for such algorithms. Few parallel graph algorithms outperform their best sequential implementation on SMP clusters due to long memory latencies and high synchronization costs. In this talk, we consider the performance and scalability of two graph algorithms, list ranking and connected components, on two classes of shared-memory computers: symmetric multiprocessors (SMP) such as the Sun Enterprise servers and multithreaded architectures (MTA) such as the Cray MTA-2. While previous studies have shown that parallel graph algorithms can speedup on SMPs, the systems' reliance on cache microprocessors limits performance. The MTA's latency tolerant processors and hardware support for fine-grain synchronization makes performance a function of parallelism. Since parallel graph algorithms have an abundance of parallelism, they perform and scale significantly better on the MTA. We describe and give a performance model for each architecture. We analyze the performance of the two algorithms and discuss how the features of each architecture affects algorithm development, ease of programming, performance, and scalability.

This is joint work with Guojing Cong and John Feo.

CSRI POC: Bruce Hendrickson, (505) 845-7599

Title: Duality-based error estimates and their application to inverse problems

Speaker: Wolfgang Bangerth, Texas A&M

Date/Time: Tuesday, April 12, 2005, 1:00-2:00 pm (MT)

Location: Building 980 Room 95 (Sandia NM)
Building 915 Room S101 (Sandia CA)

Brief Abstract: A posteriori error estimates for finite element discretizations have been investigated for some 25 years now, and there has been remarkable progress in estimating the discretization error. However, almost all of the approaches used in the field use various analytical tools that essentially measure the stability of solutions with regard to perturbations. While such stability estimates are available for many model equations, they are lacking for many practical cases of nonlinear equations, even if experimental evidence indicates that solutions of these equations indeed are stable.

One approach to avoid this problem is to numerically approximate the stability properties along with the computation of the solution. This leads to the formulation of "duality based error estimates" that are not only applicable to energy-norm estimates, but readily generalize to any output functional requested.

In this talk, we will explain the derivation of such estimates, and in particular apply it to the inverse problem of identifying parameters in PDEs from measurements of the state variable. This is a typical case of a problem for which stability-measures are notoriously hard to derive or may be even non-existent in general, and one that because of its numerical complexity greatly benefits from the adaptive meshes that can be derived from error estimators.

CSRI POC: Scott Collis, (505) 284-1123

Title: Rehabilitating POD

Speaker: Chris Beattie, Virginia Tech

Date/Time: Wednesday, July 27, 2005, 10:00-11:00 am

Location: Building 980 Room 95 (Sandia, NM)

Brief Abstract: This talk describes some refinements of the proper orthogonal decomposition (POD) methods for model reduction placing it on firmer mathematical foundations: large-scale or infinite-order dynamical systems arising from PDE models are replaced with systems of relatively low dimension having similar response characteristics. Subspaces of the state space are classified according to characteristic energies of solution trajectories, yielding a variational characterization of optimal subspaces that POD methods implicitly approximate. The effect of weighting/quadrature strategies and a dual formulation are illustrated as well.

CSRI POC: Richard Lehoucq, (505) 845-8929

Title: Studies of Algorithms for Gyrokinetics & The Effects of Shaping on Plasma Turbulence

Speaker: E. A. Belli, Ph.D Candidate, Princeton University

Date/Time: Monday, November 7, 2005, 9:30-10:30 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: The gyrokinetic equation is the fundamental equation describing drift-like microinstability-driven turbulence in magnetized plasmas. Presently, gyrokinetic simulations of plasma turbulence and transport are one of the primary tools used to design the next generation of experimental fusion devices with optimal neoclassical confinement and stability properties. However, solving the full 5D, nonlinear, integro-differential gyrokinetic equation in realistic geometry can be computationally intensive, requiring many hours of computing time on massively parallel machines for a single set of parameters. In this talk, I will describe various improved numerical algorithms and order-reduction models for Eulerian gyrokinetic codes, including a recently developed iterative implicit scheme based on numerical or analytic approximations to the plasma response. This method eliminates the long time to set up implicit arrays, yet still has the same larger time-step advantages of an implicit method. Various model preconditioners and iteration schemes, including Krylov-based solvers, will be compared. Applications to studies of the effects of flux surface shape on the gyrokinetic stability and transport of tokamak plasmas using the GS2 code and a new local linear trial function-based gyrokinetic code will also be discussed.

CSRI POC: John Shadid, (505) 845-7876

Title: Smoothed Aggregation (SA) multigrid

Speaker: Marian Brezina, University of Colorado at Boulder

Date/Time: Tuesday, March 29, 2005, 3:30-4:30 pm

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Algebraic multilevel methods have been developed to tackle whole classes of problems based on little or no knowledge about the geometrical configuration of the problem solved. This makes them attractive for solving problems for which standard geometric multigrid implementation proves impractical or impossible, such as problems resulting from discretizations over unstructured meshes, problems featuring varying coefficients or problems associated with no meshes at all. For difficult problems, however, these methods can often take advantage, or even require, additional information about the problem solved. Sometimes, this information takes the form of implicit assumptions. We will discuss ways of relaxing these requirements and enhancing robustness for a particular method, the Smoothed Aggregation (SA) multigrid.

The talk will attempt to describe SA in a self-contained form and is targeted at the audience interested in learning the principles on which SA methods are based. In particular, no prior knowledge of SA will be assumed. We will start by describing SA multigrid and proceed to discuss some of the recent developments and future research and development directions.

CSRI POC: Scott Collis, (505) 284-1123.

Title: Network Tomography in the Real World

Speaker: Hal Burch, Interview Candidate, Carnegie Mellon University

Date/Time: Tuesday, March 8, 2005, 1:00-2:00 pm

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Network operators, particularly Internet Service Providers, routinely monitor end-to-end and link delays in their network to ensure their network is operating optimally. Deploying large number of monitors around the network is time-consuming and expensive to both install and maintain. Network tomography allows an operator to leverage the data collected from monitors to infer delays not directly measured.

Traditional network tomography techniques make a variety of assumptions that may not be true in the real world. These include tree topologies, symmetry, and the ability to address internal routers (which may not be true on tunneled networks). Moreover, some require more than a thousand probes per path, which results in a high measurement cost.
Some require a condition that implies a monitor must be at every router, making network tomography techniques unnecessary.

Hal will present a network tomography technique that works on arbitrary topologies, does not require addressing internal routers, can model asymmetric networks, and requires few probes per paths. Hal will present results from running the presented network tomography technique on one of AT&T's networks, including comparison of the output to known behavior and the ability of the technique to identify the link(s) responsible for changes in measured delays. In addition, results from simulated networks will be compared against actual link delays.

CSRI POC: Suzanne Rountree, (505) 844-4379

Title: An Active Set Method for Large-Scale Constrained Optimization

Speaker: Professor Richard H. Byrd, University of Colorado

Date/Time: Thursday, February 3, 2005, 10:00-11:00 a.m. (PT)

Location: Building 980 Room 95 (Sandia NM)
Building 921 Room 137 (Sandia CA)

Brief Abstract: The two dominant approaches for solving nonlinear inequality constrained optimization problems are currently interior point methods and active set methods. However, the principle active set approach, successive quadratic programming, is limited in the number of free variables it can handle. Interior point methods can currently handle much larger problems, but have certain weaknesses, especially when warm starts are available.

This issue has motivated us to develop SLIQUE, an active set algorithm that uses a linear programming model rather than quadratic programming to deal with inequalities. After solution of a linear programming subproblem has selected a working set, SLIQUE minimizes a quadratic approximation of the Lagrangian subject to the selected equality constraints.

To force convergence from infeasible starting points SLIQUE uses an exact nondifferentiable penalty function. To determine the penalty parameter, we use an LP subproblem. We have developed a more powerful global convergence theory for exact penalty methods based on this procedure.

CSRI POC: Paul Boggs, (925) 294-4630

Title: Multi-scale Saliency Search in Image Analysis

Speaker: Anthony Campisi, University of New Mexico
Alex Slepoy, Sandia National Laboratories

Date/Time: Tuesday, September 13, 2005, 2:00-3:00 pm

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Our program implements a visual saliency search mechanism, based upon the model described by Itti and Koch (*), that is purely bottom-up, that is, without any information about a specific target to find. When presented with an image, it identifies areas of interest by their conspicuity, and generates a saliency map which guides the "attention" of the model. Note that the model is purely bottom-up, it can only point out areas of interest, it cannot provide any more information about those areas.

By suppressing the most salient point, attention can be shifted to the subsequent salient points, allowing the program to identify multiple areas of the image that may be of interest. This results in a kind of pre-attentive selection, that could for instance be used as a precursor to object recognition. Furthermore, it could be used to study what kinds of objects, in what kind of environment, stand out, and what kind blend into their surroundings.

The algorithm relies on interesting multi-scale decomposition of data that is strongly resonant with issues inherent in multi-scale simulation. A variant of this algorithm is employed in the JPEG2000 standard for image compression as a much faster alternative to FFT.

* Itti, L. & Koch, C. (1999) A saliency-based search mechanism for overt and covert shifts of visual attention Vision Research 40, 1489 – 1506.

CSRI POC: Richard Lehoucq, (505) 845-8929

Title: Molecular dynamics at constant temperature - Some mathematical and numerical results

Speaker: Eric Cances, CERMICS, Ecole Nationale des Ponts et Chaussees

Date/Time: Wednesday, October 12, 2005, 2:45-3:45pm (MT)

Location: Building 980 Room 95 (Sandia NM)
Building 915 Room S101 (Sandia CA)

Brief Abstract: Phase-space integrals are widely used in Statistical Physics to relate the macroscopic properties of a system to elementary phenomena at the molecular scale. In this talk, I will present and compare, from both theoretical and numerical viewpoints, different NVT (constant temperature) phase-space sampling methods. I will focus in particular on the Langevin dynamics, the Hybrid Monte Carlo method and deterministic thermostatting methods (Nose'-Hoover chains, Nose'-Poincare' and Recursive Multiple Thermostats).

CSRI POC: Richard Lehoucq, (505) 845-8929

Title: Averaging Based Estimators for Elliptic Equations

Speaker: Varis Carey, Cornell University

Date/Time: Monday, March 28, 2005, 9:00-10:00 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Averaging operators provide inexpensive and robust error control for finite element discretization of elliptic PDEs. We present a framework, first introduced by Schatz and Wahlbin, that modifies existing popular recovery operators (such as the Zienkiewicz-Zhu Estimator) to produce estimators that are *locally* bounded in maximum-norm, and can produce asymptotically exact estimators on unstructured meshes. We illustrate the excellent performance of these operators in the preasymptotic range with a variety of numerical examples. Certain features of the averaging operators that make them very convenient for parallel computation are also detailed. In addition, we give new numerical results illustrating the effective use of these operators on problems with "pollution", and develop a similar framework of local error control for function values, employing some recent local negative-norm *a priori* estimates of Schatz.

CSRI POC: Kevin Copps, (505) 844-4521

Title: A comprehensive spatio-temporal numerical scheme for fully implicit 3D extended MHD

Speaker: Luis Chacon (Speaker) and Dana Knoll
Los Alamos National Laboratory

Date/Time: Monday, January 31, 2005, 1:00-2:00 pm

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: An accurate and efficient numerical solution of the magnetohydrodynamics (MHD) in general geometries is of the essence for the adequate simulation and proper design of the next generation MFE fusion reactors. Temporally, a fully implicit treatment is desirable (vs. an explicit or semi-implicit treatment) to effectively deal with the numerical stiffness of MHD fast waves without compromising accuracy. Spatially, the desired scheme should be conservative, solenoidal in the magnetic field to machine precision, stable in the absence of numerical and physical dissipation, and suitable for mesh adaptivity in curvilinear geometries. In this talk we will discuss the details of both the spatial and temporal aspects of a comprehensive numerical scheme that satisfies the requirements outlined above. In particular, we will provide details of a novel cell-centered finite-volume 3D MHD spatial scheme that is solenoidal, conservative, non-dissipative, stable, and second-order accurate even in curvilinear grids. We will also discuss the current status in the development of a fully implicit Newton-Krylov algorithm for the compressible 3D extended MHD model, focusing on the details of the physics-based preconditioning concept. We will present results that demonstrate optimal solver scalability under grid refinement and time step size for the resistive MHD system.

CSRI POC: John N. Shadid, (505) 845-7876

Title: An introduction to Chapel: Cray Cascade's High Productivity Language

Speaker: Brad Chamberlain, University of Washington/Cray Inc.

Date/Time: Thursday, March 17, 2005, 10:00-11:00 am (PST)

Location: Building 980 Room 24 (Sandia NM)
Building 915 Room W133 (Sandia CA)

Brief Abstract: DARPA's HPCS program has challenged the High-End Computing Industry with improving user productivity by the year 2010. One of the ways that Cray Cascade's team is responding to this challenge is by investigating the development of a new parallel language named Chapel. In this talk, I will describe the motivations and foundations for Chapel. I will also introduce several Chapel concepts in detail, illustrating their use in some sample kernels written in the language.

CSRI POC: Matt Leininger, (925) 294-4842

Title: Fast Simulation-based Optimization for Engineering Design Under Uncertainty

Speaker: Chun-Hung Chen, George Mason University

Date/Time: Monday, June 6, 2005, 11:00 - 12:00 noon (MT)

Location: Building 980 Room 95 (Sandia NM)
Building 915 Room S101 (Sandia CA)

Brief Abstract: Simulation is a popular tool for designing large, complex, stochastic systems, since closed-form analytical solutions generally do not exist for such problems. While the advance of new technology has dramatically increased computational power, efficiency is still a big concern when using simulation for engineering design with uncertainty, in which case many alternative designs must be simulated. This presentation gives our new development to address such an efficiency issue. A key component of our methodologies is a new control-theoretic simulation technique called Optimal Computing Budget Allocation (OCBA) invented by the speaker. OCBA is developed in the context of Ordinal Optimization and is employed to provide the most efficient number of simulation runs for a desired accuracy. With integration with Pattern Coordinate Search method, it can be effectively applied to continuous stochastic optimization problems.

(Joint work with Dr. Vicente Romero at Sandia-New Mexico)

CSRI POC: Monica Martinez-Canales, (925) 294-3157

Title: Methods for Engineering Design under Uncertainty

Speaker: Dr. Wei Chen, Northwestern University

Date/Time: Wednesday, July 20, 2005, 2:00-3:30 pm (MST)

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: With today's increasing and global competition, probabilistic design methods with consideration of uncertainty in design have been gaining much attention. The consideration of uncertainty enables engineers to make reliable decisions, to lessen quality loss, to manage risk, and to avoid over-conservative designs. However, decision making under uncertainty poses a challenging and complex optimization problem. A rigorous design framework and efficient computational techniques are therefore needed. In this talk, recent research developments in efficient robustness and reliability (R&R) assessments and affordable probabilistic optimization, metamodeling techniques, probabilistic sensitivity analysis, multidisciplinary design optimization under uncertainty, and model validation will be presented.

CSRI POC: Tim Trucano, (505) 844-8812

Title: Evaluating I/O Characteristics and Methods for Storing Structured Scientific Data

Speaker: Avery Ching, Northwestern University

Date/Time: Thursday, October 13, 2005, 2:00-3:00 pm

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Many large-scale scientific simulations generate large, structured multi-dimensional datasets. Data is stored at various intervals on high-performance I/O storage systems for checkpointing, post-processing, and/or visualization. This time for data storage is very I/O intensive and can dominate the overall running time of the application depending on the characteristics of the I/O access pattern. When and how to describe these I/O access patterns greatly affects performance. In this paper, we examine the impact of various I/O parameters and methods to store scientific data using the MPI-IO interface in an optimized parallel file system.

CSRI POC: Lee Ward, (505) 844-9545

Title: RDMA Locks and their Usefulness in High Performance Computing

Speaker: Kenin Coloma, Northwestern University

Date/Time: Tuesday, September 14, 2005, 11:00-12:00 noon

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: A number of modern high performance interconnects available on today's supercomputers support low latency Remote Direct Memory Access (RDMA) or one-sided communication. Given the similarity to shared memory programming, one can imagine the importance of having some means of globally protecting parts of memory. Not surprisingly, some of the distributed lock algorithms developed for shared memory architectures can be used in the RDMA environment. The MPI-2 RMA specification presents some unique challenges which make a performance comparison between distributed locking algorithms using Sandia's Portals communication library directly and an MPI-2 based locking algorithm developed at Argonne National Laboratories intriguing.

CSRI POC: Lee Ward, (505) 263-3322

Title: Ocean's Slow Burn Process and the Ocean Modeling for Climate Research

Speaker: David Dietrich, President, AcuSea, Inc. & C. Aaron Lai, LANL

Date/Time: Thursday, January 13, 2005, 1:00-2:00 pm

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Dr. Dietrich will present a detailed numerical model of the North Atlantic Ocean that once required supercomputing power and that can now be done using a personal computer. This is because of advances both in hardware and ocean modeling. This permits reliable and efficient studies addressing a variety of ocean-related climate change theories, including abrupt climate change. Low numerical dispersion and dissipation are needed to maintain the narrow, thin deep current water properties (temperature, salinity and kinetic energy) with sufficient intensity during their 5-10 year flow from their source regions (northern seas and Arctic Ocean) to realistically affect the Gulf Stream separation and path, which is crucial in climate change studies. Results will be shown from a robust 4th-order-accurate ocean model based on collocated (non-staggered) control volumes, which achieves this numerically demanding task using a 2 GHz P4 PC.

Dr. Lai will present the finding of "ocean's slow burn process" and its impacts on climate research. The importance of the deep currents is amplified by their effects on the dissociation of big methane hydrate deposits on the continental shelf/slopes. Methane hydrates can dissociate into methane bubbles and water. Bacteria metabolize the methane, starting a food chain that results in CO2, water and heat as final products. This kind of heat remains trapped by the overlying warm water as it spreads laterally. Eventually, it enters the Arctic Ocean, where rapid deep warming has been detected recently. A detailed analysis of the ocean biogeochemical data collected during the World Ocean Circulation Experiment (WOCE) 1987 - 1998 had revealed the evidences of ocean's slow process. The "Slow burn process", as described above, depletes dissolved oxygen in seawater. The observation data match the predicted amount of oxygen depletion very well. Therefore, we can estimate the amount of heat generated in the slow burn process. This finding shows that there is an internal heat generator in ocean that can significantly alter the physical/dynamical process of ocean circulation. It also offers an explanation for the observed warming of intermediate depth seawater since 1955. The impacts of this ocean's slow burn process on climate change will be discussed in the context of the interactions among the atmosphere, sea ice and the oceans.

CSRI POC: Mark Boslough, (505) 845-8851

Title: Monte-Carlo Algorithms for Matrices and Applications

Speaker: Petros Drineas, Rensselaer Polytechnic Institute

Date/Time: Wednesday, August 24, 2005, 9:00-10:00 (MT)

Location: Building 980 Room 95 (Sandia NM)
Building 915 Room S107 (Sandia CA)

Brief Abstract: We are interested in developing and analyzing Monte Carlo algorithms for performing useful computations on large matrices, and applying them on (massive) datasets. Examples of such computations include matrix multiplication, the computation of the Singular Value Decomposition of a matrix, the computation of CUR type decompositions of matrices, testing the feasibility or infeasibility of linear programs, and solving least squares problems.

We present a Pass-Efficient model of data streaming computation in which our algorithms may naturally be formulated and present algorithms that are efficient within this model. We also discuss various applications of these methods to the analysis of data sets.

CSRI POC: Bruce Hendrickson, (505) 845-7599

Title: Global Optimization: For Some Problems, There's HOPE

Speaker: Daniel M. Dunlavy, Graduate Research Assistant
University of Maryland
John Von Neumann Research Fellowship Candidate

Date/Time: Tuesday, March 1, 2005, 11:00–12:15 am (MST)

Location: Building 980 Room 95 (Sandia NM)
Building 915 Room S101 (Sandia CA)

Brief Abstract: A new method for solving unconstrained nonlinear optimization problems will be presented. The method, homotopy optimization using perturbations and ensembles (HOPE), is an extension of previously-developed homotopy optimization methods and has been used successfully to solve the protein structure prediction problem---a challenging optimization problem in biochemistry---for several protein models.

Given a template function for which the global minimizer is known or easy to compute and a target function to be minimized, HOPE generates a sequence of local minimizers of a continuous homotopy function between the two functions, starting at the global minimizer of the template function. To increase the likelihood of finding the global minimizer of the target function, points in the sequence are perturbed and used as starting points for generating new sequences, creating an ensemble of sequences in the end. Each convergent subsequence leads to a local minimizer of the target function, and HOPE chooses the best of these local minimizers as its approximation to the global minimizer.

A brief discussion of the convergence properties of HOPE will be given, followed by a comparison of HOPE with simulated annealing and Newton's method (with a trust region) applied to solving the protein structure prediction problem.

CSRI POC: Patricia D. Hough, (925) 294-1518

Title: Development of a 2-D convection model to simulate phase transition accurately over long time integration*

Speaker: Katherine J. Evans, Los Alamos National Laboratory

Date/Time: Monday, December 19, 2005, 10:00-11:00 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: We present a two-dimensional convection phase change model using the incompressible Navier-Stokes equation set and enthalpy as the energy conservation variable. Major algorithmic challenges such as stiff nonlinearities at the phase front are posed by adding self-consistent convection. The equation set is solved with the Jacobian-Free Newton-Krylov (JFNK) nonlinear inexact Newton's method. The generalized-minimum residual (GMRES) algorithm is the linear Krylov solution approximation for the outer Newton loop. SIMPLE, a pressure-correction algorithm, is used as a physics-based preconditioner for GMRES. This algorithm is compared to solutions using SIMPLE as the main solver. Algorithm performance is assessed for a benchmark problem, phase change convection within a square cavity of a solid pure material cooled below the melting temperature. A time step convergence analysis demonstrates that the JFNK model is second order accurate in time. A Gallium melting simulation is also performed and evaluated; in this configuration multiple roll cells develop in the melted region at early times when the aspect ratio is high. The use of spatial discretization greater than first order is necessary to capture the multiple cellular structure and subsequent phase front variations. The JFNK-SIMPLE solution algorithm converges with greater efficiency than SIMPLE as a stand alone solver, and the effect becomes more pronounced for problems with increased size and complexity. JFNK-SIMPLE can provide second order accurate converged solutions to the nonlinear equation set with larger time steps and finer grids than SIMPLE, which performs sequential linear solves on each equation. There are advantages to using SIMPLE solvers, for example their ease of convergence to a linear tolerance. When SIMPLE is incorporated as a preconditioner to JFNK, these benefits are retained, plus the ability to model more complex and realistic problems with minimal and quantified error.

*Katherine J. Evans and Dana A. Knoll, Fluid Dynamics Group, T-3 Michael Pernice, CCS-3, Los Alamos National Laboratory, LA-UR-05-5843

CSRI POC: John Shadid, (505) 845-7876

Title: Rule-based Modeling of Biochemical Networks

Speaker: James R. Faeder, Los Alamos National Laboratory

Date/Time: Monday, January 19, 2005, 9:00-10:00 am

Location: Building 980, Room 95 (Sandia NM)

Brief Abstract: We present a method for generating a biochemical reaction network from a description of the interactions of components of biomolecules. The interactions are specified in the form of reaction rules, each of which defines a class of reaction associated with a type of interaction. Reactants within a class have shared properties, which are specified in the rule defining the class. A rule also provides a rate law, which governs each reaction in a class, and a template for transforming reactants into products. A set of reaction rules can be applied to a seed set of chemical species and, subsequently, any new species that are found as products of reactions to generate a list of reactions and a list of the chemical species that participate in these reactions, i.e., a reaction network, which can be translated into a mathematical model.

CSRI POC: Alex Slepoy, (505) 284-3650

Title: Adjoint based optimization of free-shear-flow noise control

Speaker: Jonathan Freund, Theoretical and Applied Mechanics,
University of Illinois Urbana-Champaign

Date/Time: Thursday, April 21, 2005, 9:00-10:00 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Direct numerical solutions of the adjoint of the perturbed and linearized compressible flow equations are used to circumvent the complexity of an unsteady free shear flow and reduce its noise directly. This affords a unique opportunity to contrast a baseline noisy flow and its slightly perturbed but much quieter counterpart. We focus on a two-dimensional mixing layer as a model for the near-nozzle region of a turbulent jet. Before and after application of the control, the flow is superficially unchanged. The mean flow spreading, turbulence kinetic energy, and apparent evolution and pairing of the unsteady flow structures appear unaffected despite a factor of 10 reduction in the acoustic intensity. However, a POD decomposition of the data illuminates a key change in the flow field: the normal basis modes come more closely in $\sin$/$\cos$-like pairs whose time coefficients trace more circular orbits in phase planes. This is consistent with a smoother downstream advection of the flow's fluctuation energy which is not obvious in standard visualizations of the flow. Smoother advection is less efficient at putting energy into radiation capable (supersonic) modes. This is consistent with our observation that the randomly excited flow can be quieted to the same level as a harmonically excited highly regular flow.

CSRI POC: Scott Collis, (505) 284-1123

Title: Numerical Simulations Based on Adaptive Delaunay Meshing

Speaker: Pascal Frey, Lab. Jacques-Louis Lions
University Paris VI France

Date/Time: Thursday, September 15, 2005, 10:00-11:00 (MT)

Location: Building 980 Room 95 (Sandia NM)
Building 915 Room S101 (SNL CA)

Brief Abstract: Numerical simulations based on adaptive Delaunay meshing P. Frey, C. Dobrzynski and O. Pironneau.

CAD-based Navier-Stokes solvers are too expensive in some engineering fields; architecture and medical application to for blood flows are two such examples. In such fields, mesh generation and adaptation is also a bottleneck because the users are not expected to have the know-how.

In this seminar we report on a Navier-Stokes solver for incompressible temperature and time dependent flows dedicated to architectural applications. The building blocks are not new: a finite element method with time implicit pressure projection steps and mesh adaptativity; but putting them together in an easy to use and efficient 3D code is the challenge which motivates this work.

For non-engineering applications the user interface is a big problem. In an earlier attempt we designed Freefem3D based on a fictitious domain discretization, thus avoiding boundary fitted mesh. However it turned out that the display of the solutions requires a boundary fitted mesh; it is possible to generate a feasible surface mesh for graphics but it is much more difficult to generate a feasible surface mesh for FEM.

In this project the user interface is taken from Freefem3d; then, with a marching cube algorithm we produce a graphic-only feasible mesh; finally a surface mesh, adapted to a FEM computation is constructed with an adaptation module and the result is used as input to a Delaunay volumic mesh generator (the solver is optimized and parallelized, all modules are the authors' work).

CSRI POC: Philippe P. Pebay (925) 294-2024

Title: Overview of AMPL

Speaker: David M. Gay, Sandia National Laboratories

Date/Time: Tuesday, October 18, 2005, 10:00-11:00 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: AMPL is a comprehensive and powerful algebraic modeling language for stating, solving, and analyzing linear and nonlinear optimization problems, in discrete or continuous variables. AMPL lets you use common notation and familiar concepts to formulate optimization models and examine solutions, while the computer manages communication with an appropriate solver. AMPL's flexibility and convenience render it ideal for rapid prototyping and model development, while its speed and control options make it an especially efficient choice for repeated production runs. This talk gives an overview of AMPL by one of its creators, David Gay, who joined Sandia in 2003. Sandia has free access to AMPL for internal use, and the talk will include some illustrations from Sandia applications.

CSRI POC: Scott Mitchell, (505) 845-7594

Title: On the TCAT Approach for Obtaining Multiphase Subsurface Conservation Equations

Speaker: William G. Gray
Environmental Sciences and Engineering
University of North Carolina at Chapel Hill

Date/Time: Tuesday, June 28, 2005, 10:00-11:00 am (PT)

Location: Building 980 Room 24 (Sandia NM)
Building 915 Room S101 (Sandia CA)

Brief Abstract: Modeling of flow and transport processes in natural porous media requires that conservation equations be formulated at a length macroscale much larger than the pore diameter but much smaller than the system length scale. The Thermodynamically Constrained Averaging Theory (TCAT) approach provides a means to upscale the needed equations. This method extends the standard method of averaging pore scale conservation equations in at least two important ways: i) conservation equations may be employed for the properties of interfaces between phases and common curves where three phases come together; and ii) pore scale thermodynamics are averaged to the macroscale so that a relation between microscale and macroscale properties, such as temperature and pressure, are explicit. The importance and effect of using the TCAT approach in comparison to standard approaches is demonstrated for porous media systems involving single phase flow, two phase flow, and the expression for the solid phase stress tensor.

CSRI POC: Genetha Gray, (925) 294-4957

Title: First-Principles Modeling of Chemical Reactions on Heterogeneous Catalysts: Fundamental Understanding and Catalyst Design

Speaker: Jeff Greeley, Technical University of Denmark

Date/Time: Wednesday, May 18, 2005, 1:30-2:30 pm

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: First-principles, quantum-based simulations of elementary chemical processes are emerging as a powerful tool for the analysis and design of heterogeneous, transition metal catalysts. These simulations can provide detailed, molecular-level information about the energies, geometries, and electronic structures of chemical species on metal surfaces. This information, which is often inaccessible through experiment alone, permits the determination of catalytically-relevant quantities such as adsorption energies, adsorbate entropies, and activation barriers of elementary reaction steps. A thorough understanding of these important catalytic parameters can, in turn, be used as a starting point for the first-principles design of new catalysts.

In this talk, I present two illustrations of the use of first-principles calculations to analyze chemical reactions on heterogeneous catalysts. The first example, of relevance to the study of anodes for direct methanol fuel cells, focuses on the development of *fundamental understanding* of the methanol decomposition reaction mechanism on Pt(111). The second example, of interest for a broad variety of catalytic processes involving oxygen, demonstrates how a first-principles approach can be used as a preliminary *screening technique for new catalysts*. For the latter example, two screening methodologies are illustrated. First, the development of *databases* of catalytic parameters for a significant number of binary metal alloys is described, and second, the use of *evolutionary algorithms* to screen a large search space of quaternary metal alloys is discussed. The application of both methodologies to important problems in catalysis and electrocatalysis is examined.

CSRI POC: Aidun, John, (505) 844-1209

Title: Discrete exterior calculus with convergence to the smooth continuum

Speaker: Jenny Harrison, University of California, Berkeley

Date/Time: Thursday, August 11, 2005, 10:00-11:00 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: By articulating the concept of an infinitesimal within the framework of classical analysis, we may define integral and derivative at a single point in space. Theorems of Stokes and Gauss can be stated and proved at a single point. The theory extends immediately to a discrete theory by taking sums over finitely many points in space. Convergence to the smooth continuum is satisfied, leading readily to a full theory of calculus beyond the classical setting, including calculus on fractals and soap films.

The discrete theory has well behaved operators and products that avoid common problems of other approaches. For example,

1. The product on discrete forms is associative and graded commutative.
2. Operators on domains, such as boundary and Hodge star, are geometrically defined as limits in the strong topology, not as duals to differential forms, or through meshes.
3. Forms and cochains are isomorphic
4. Convergence to the smooth continuum is guaranteed.

CSRI POC: Pavel Bochev, (505) 844-1990

Title: A comparison of some mesh adaptation schemes with FreeFem++

Speaker: Frederic Hecht, Lab. Jacques-Louis Lions
University Paris VI France

Date/Time: September 15, 2005, 11:00-12:00 noon (MT)

Location: Building 980 Room 95 (Sandia NM)
Building 915 Room S101 (SNL CA)

Brief Abstract: FreeFem++ is a free software for the solution of 2D PDEs with the finite element methods, that has been developed at the J-L Lions Laboratory. The main characteristics of FreeFem++ are:

- A large variety of available finite element: linear and quadratic Lagrangian elements, discontinuous P1 and Raviart-Thomas elements, elements of a non-scalar type, mini-element, etc....
- Automatic interpolation of data on different meshes to an overmesh, storage of the interpolation matrix.
- Linear problems description (real or complex) thanks to a formal variational form, with access to the vectors and the matrix if needed.
- Includes tools to define discontinuous Galerkin formulations. Analytic description of boundaries. When two boundaries intersect, the user must specify the intersection points.
- Automatic mesh generator, based on the Delaunay-Voronoi method.
- Inner points density is proportional to the density of points on the boundary [cf. P-L George, "Meshing", Hermes 2000].
- Metric-based anisotropic mesh adaptation. The metric can be computed automatically from the Hessien of a solution.
- Every variable is typed.
- Available solvers: LU, Cholesky, Crout, CG, GMRES, UMFPACK linear solver, eigenvalue and eigenvector computation.
- Online graphics, C++-like syntax.
- Many examples: Navier-Stokes, elasticity, Fluid structure, Schwarz's domain decomposition method, Eigen value problem, residual error indicator, etc.
- Parallel version using MPI

This presentation will use FreeFEM++ to compare the following scheme combined with mesh adaptation:

- Hessian mesh adaptation,
- residual error indicator,
- hierarchical error indicator.

CSRI POC: Philippe P. Pebay, (925) 294-0224

Title: AM PERCS Interconnect and IO Research

Speaker: Jay Herring, IBM

Date/Time: Thursday, September 22, 2005, 10:00-11:00 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: In order to improve the percent usable performance of HPC clusters, improvements in the interconnect structure are required. As part of the PERCS project, IBM has been looking at different interconnect architectures to understand their advantages and disadvantages. Currently there are four main categories of interconnects that IBM is researching, this presentation will present all four at a high level. Along with the base communication requirements for the interconnect, there are requirements on the interconnect to support storage and IO traffic. Therefore, these interconnect options are being evaluated based on message passing performance (latency and bandwidth), storage and IO capabilities, and cost estimates.

CSRI POC: Erik DeBenedictis, (505) 284-4017

Title: Large Files, Small Writes and pNFS

Speaker: Dean Hildebrand, Ph.D. Candidate, University of Michigan

Date/Time: Wednesday, August 24, 2005, 3:00-4:00 pm

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Numerous workload characterization studies have highlighted the prevalence of small and sequential data requests in scientific applications. Parallel file systems excel at large data transfers, often at the expense of small I/O performance. pNFS is an emerging NFSv4 extension that provides operating system and hardware independent high performance data access to parallel file systems. In this talk, I will discuss how pNFS can improve the overall write performance of parallel file systems by utilizing both the parallel file system's I/O protocol for large write requests and the NFSv4 protocol for small write requests.

CSRI POC: Lee Ward, (505) 263-3322

Title: Phase Field Methods for Flows with Elastic Membranes

Speaker: Judith Hill, Carnegie Mellon University

Date/Time: Wednesday, January 26, 2005, 9:00 to 10:00 a.m.

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: An open area of research in computational fluid dynamics is the study of multi-fluid flows, phenomena which arise in many real-world fluid mechanics applications. In this talk, a phase field method for implicitly capturing the location of the interfaces in binary fluids will be discussed. This approach avoids the computational geometry complexities associated with fully Lagrangian or front-tracking Eulerian methods. These complexities are particularly pronounced on parallel computers.

Additionally, for flows with elastic membranes separating the fluids, a phase field formulation for the membrane stresses will be presented. The coupled membrane-fluid flow formulation is discretized by a combined continuous/discontinuous Galerkin method, which lends itself to an inherently parallel implementation. A block Schur-complement preconditioner neutralizes the ill-conditioning of the coupled system due to disparate material properties. Parallel performance results demonstrate the efficacy of this preconditioner.

CSRI POC: Bart van Bloemen Waanders, (505) 284-6746

Title: An Inverse Problem in Nondestructive Evaluation of Spotwelds

Speaker: Thomas Hoft, Interview Candidate, University of Minnesota

Date/Time: Tuesday, April 26, 2005, 9:00-10:15 am (PT)

Location: Building 980 Room 95 (Sandia NM)
Building 915 Room S101 (Sandia CA)

Brief Abstract: A thermal imaging method for nondestructive evaluation of spotwelds has been proposed. In this method, a transducer is employed to generate heat near the weld while the surface temperature is measured. The inverse problem is to access the quality of the weld from the temperature.

In this presentation we develop a simple model for the thermal diffusion problem. The inverse problem we seek to solve, amounts to finding a heat source in a 2-D domain, given temperature as a function of space and time. We solve this classically ill-posed problem by devising a time-stepping algorithm which solves a regularized problem at each time step. We illustrate the main ideas with numerical examples.

CSRI POC: Patty Hough (925) 294-1518
Monica Martinez-Canales (925) 294-3157

Title: Sequential Kriging Optimization Methods for Stochastic and Multi-Fidelity Evaluations

Speaker: Deng (Dan) Huang, Ph.D., Post Doctorial Candidate
Department of Industrial & Systems Engineering, Ohio State University

Date/Time: Thursday, May 12, 2005, 10:00-11:00 am (MDT)

Location: Building 980 Room 24 (Sandia NM)
Building 915 Room S145 (Sandia CA)

Brief Abstract: Sequential Kriging Optimization (SKO) is a method developed in recent years for solving non-linear expensive black-box problems in areas such as large-scale circuit board design and manufacturing process improvement. The method is based on a stochastic process meta-model (kriging) that provides a global prediction and a measure of prediction uncertainty. The subsequent evaluations are selected by maximizing the so-called expected improvement function.

We extended SKO to address stochastic systems by modifying the expected improvement function to achieve the desired balance between the need for global and local searches. In studies using test functions, SKO compared favorably with alternative approaches in terms of consistency in finding global optima and efficiency as measured by number of evaluations.

We proposed another extension of the SKO method, named Multiple Fidelity Sequential Kriging Optimization (MFSKO), where surrogate experimental systems are exploited to reduce the total evaluation cost. In this method, data on all experimental systems are integrated to build a kriging meta-model. The location and fidelity level of the evaluations are selected by maximizing an expected improvement function that is related to the evaluation costs. The proposed method was applied to 1) a metal-forming process design problem via Finite Element simulations, and 2) discrete event simulations of hospital emergency departments.

CSRI POC: Laura Painton Swiler, (505) 844-8093

Title: IMEX-RK methods

Speaker: Alex Kanevsky, Brown University
von Neumann Fellowship candidate

Date/Time: Tuesday, March 8, 2005, 9:00 - 10:00 a.m.

Location: Building 980 Room 95 (Sandia NM)
Building 915 Room 133W (Sandia CA)

Brief Abstract: We are interested in removing one of the main limitations of high-order fully explicit time integration schemes for problems which suffer from high levels of geometry-induced stiffness: the severe stability-based time-step restriction. Geometry-induced stiffness, or scale-separation stiffness, is a result of attempting to simultaneously simulate a system that has geometric features of drastically varying scales. We implement high-order Implicit-Explicit Runge-Kutta (IMEX-RK) methods to overcome geometry-induced stiffness in various linear and nonlinear problems on unstructured grids, by solving the non-stiff portions of the domain using explicit methods, and solving the more expensive stiff portions using implicit methods. We follow the method of lines approach, and discretize space using a nodal Discontinuous Galerkin Spectral Element method.

Surprisingly, IMEX-RK methods are not only more efficient than ERK methods for sufficient levels of stiffness, but are also more accurate when nonconsistent filters are used. We identify the fundamental mechanism by which filtering can become an additive process. We demonstrate analytically and numerically that filtering a numerical approximation which is at or near steady-state, for which the time step is small, or a combination thereof will result in a growing filtering-in-time error which will erase modes from the Fourier or polynomial approximation when nonconsistent filters are used, and suggest the development and implementation of time-dependent, time-consistent filters to control this error.

This is joint work with David Gottlieb and Jan S. Hesthaven (Brown University) and Mark H. Carpenter (NASA Langley).

CSRI POC: Scott Collis, (505) 284-1123

Title: Binding of phenylalanine analogs to phenylalanyl-tRNA synthetase

Speaker: Peter Kekenes-Huskey, California Institute of Technology

Date/Time: Friday, January 28, 2005, 10:00-11:00 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: A computational strategy has been developed that captures the binding characteristics of phenylalanine analogs in its respective aminoacyl-tRNA synthetase (aaRS). Binding to the aaRS is a requisite for incorporation into a translated protein, therefore this approach can provide an indirect measure of incorporation for a candidate analog. The protocol couples a ligand torsion sampling method, moleculeGL, with a side chain replacement program, SCREAM, to generate conformations that maximize binding. This protocol has been applied to binding of the twenty natural amino acids and several phenylalanine analogs in select hosts and compared to measured kcat/km data. The reported binding energies correlate well with measured kcat/km data for AMP activation, thus establishing the efficacy of the approach. Furthermore, this strategy has been applied to several amino acid analogs of interest, yielding binding data for wild-type proteins, as well as proposed mutations to improve binding.

CSRI POC: Paul Crozier, (505) 845-9714

Title: Recent advances in nonlinear output feedback and applications to unmanned vehicles

Speaker: Farshad Khorrami, Six Metrotech Center, Polytechnic University

Date/Time: Wednesday, May 25, 2005, 11:00-12:00 noon

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: In this talk, an overview of recent advances in nonlinear control based on a new dynamic high-gain scaling based design will be presented. This design technique has been shown to be a highly flexible design methodology providing a unified framework for control of nonlinear systems of both lower triangular (feedback) and upper triangular (feedforward) structures and also to some systems without any triangularity properties. Furthermore, the design paradigm provides both state feedback and output feedback results and can be extended to adaptive and robust frameworks and also to decentralized control of large-scale systems. The power of the design approach has been demonstrated through output feedback designs for nonlinear systems with unknown parameters coupled with unmeasured states and with appended Input-to-State Stable (ISS) dynamics driven by all states. A central ingredient in the dynamic high-gain scaling based design is the solution of particular pairs of coupled Lyapunov equations.

CSRI POC: David Schoenwald, (505) 284-6285

Title: Numerical Methods for the Wigner-Poisson Equations

Speaker: Matthew Lasater, North Carolina State University

Date/Time: Monday, May 17, 2005, 9:00-10:00 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Resonant tunneling diodes are prototypical nanoscale semiconductor devices that theory and numerical simulation predict can develop terahertz frequency current oscillation. Since these devices are developed at such a small size, quantum physics (instead of classical physics) dictate their operation. The Wigner-Poisson equations describe quantum mechanical electron transport and are used in modeling the resonant tunneling diode. In this talk, time-dependent and time-independent simulation of the device are discussed and compared. While the time-dependent simulation can immediately detect current oscillation, it is computationally intensive to perform. In contrast, while the time-independent simulation can more quickly generate the steady-state current-voltage relationship of the device, determining if current oscillation will occur requires solving an eigenvalue problem which is also computationally expensive. To speed-up solving the eigenvalue problem, a spectral transformation, the Cayley transform, is implemented. Numerical results will be presented.

CSRI POC: Eric Keiter (505) 284-6088

Title: A parallel adaptive mesh refinement algorithm of an unsplit staggered mesh scheme for magnetohydrodynamics

Speaker: Dongwook Lee, Ph.D. Candidate
University of Maryland at College Park

Date/Time: Monday, December 5, 2005, 9:00-10:00 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Many numerical solutions of multidimensional MHD problems have been solved using operator splitting, which is simply an extension of the one-dimensional formulation in Godunov type finite-volume schemes. Although the formal accuracy of such schemes has been shown to be second order, there are generally unphysical errors due to the inherent dimensional splitting in numerical time integration.

To avoid there errors we have developed a new operator unsplit staggered mesh algorithm (USM-MHD) that has been implemented on massively parallel computers. The USM-MHD scheme has shown to efficiently ensure the divergence-free magnetic fields constraint up to computer round-off error. The method is a second order finite-volume Godunov algorithm which has a TVD version of the MUSCL-Hancock scheme, and also incorporates a full consideration of multidimensional structures in MHD. The development of a parallel adaptive mesh refinement (AMR) version of this scheme is ongoing and is being implemented in the FLASH code computational framework using the PARAMESH library.

We will discuss the formulation of the USM-MHD scheme, and as well as results of several numerical test problems.

CSRI POC: John Shadid, (505) 845-7876

Title: Adaptive Parameter Space Exploration with Gaussian Process Trees

Speaker: Herbert K. H. Lee, University of California, Santa Cruz

Date/Time: Wednesday, April 6, 2005, 11:00-12:00 (MT)

Location: Building 980 Room 24 (Sandia NM)
Building 921 Room 137 (Sandia CA)

Brief Abstract: Traditionally, to obtain even a qualitative understanding of the output of a computer simulation, runs have been made over a complete grid of input parameter configurations. For large scale simulations, such sweeps can be prohibitively expensive, and fixed designs such as Latin hypercubes can be still difficult yet inefficient. Thus, there is a need for computationally inexpensive surrogate models that can be used in place of simulation to adaptively select new settings of input parameters and map the response with far fewer simulation runs. We provide a general methodology for modeling and adaptive sampling to greatly speed up parameter sweeps. Binary trees are used to recursively partition the input space, and Gaussian process models are fit within each partition. Trees facilitate non-stationarity and a Bayesian interpretation provides a measure of uncertainty in the sample space which can be used to guide future sampling. Our methods are illustrated on several examples, including the motivating example involving computational fluid dynamics simulation of a NASA reentry vehicle.

CSRI POC: Genetha Gray, (925) 294-4957

Title: Materials Simulation of Carbon Materials

Speaker: Kwang-Ryeol Lee, Korea Institute of Science and Technology

Date/Time: Friday, April 29, 2005, 10:30-11:30 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: In this presentation, I will talk on recent results of molecular dynamic simulation of ta-C films which guide us to atomic scale understanding of residual compressive stress of the highly distorted covalent bond materials. Effect of third element addition such as Si or W to amorphous carbon film was investigated by both empirical MD simulation and an initio calculation. Small amount of Si incorporation significantly reduce the residual stress by relaxing the atomic bond angles by removing carbon atoms in metastable site. W atoms dissolved in a-C matrix without forming WC clusters played a role of pivot site where the atomic bond distortion can occur without inducing a significant in elastic energy. Finally, I will introduce a research project recently launched at KIST to combine massive MD and electron transport calculation to simulate the nanoscale CMOS device.

CSRI POC: Steven J. Plimpton, (505) 845-7873

Title: Analysis of a prototypical multiscale method coupling atomistic and continuum mechanics

Speaker: Frederic Legoll, University of Minnesota

Date/Time: Thursday, July 14, 2005, 10:00-11:00 am (MT)

Location: Building 980 Room 95 (Sandia NM)
Building 915 Room S145 (Sandia CA)

Brief Abstract: In order to describe a solid which deforms smoothly in some region, but non smoothly in some other region, many multiscale methods have been recently proposed, that aim at coupling an atomistic model (discrete mechanics) with a macroscopic model (continuum mechanics). We present here a theoretical analysis for such a coupling in a one-dimensional setting. We study both the general case of a convex energy and a specific example of a nonconvex energy, the Lennard-Jones case.

In the latter situation, we prove that the discretization needs to account in an adequate way for the coexistence of a discrete model and a continuous one. Otherwise, spurious discretization effects may appear. We also consider the effect of the finite element discretization of the continuum model on the behaviour of the coupled model.

This work is joint with Xavier Blanc (Paris 6) and Claude Le Bris (CERMICS, ENPC).

CSRI POC: Richard Lehoucq, (505) 845-8929

Title: Mathematical contributions to electronic structure calculations: one overview and two examples

Speaker: Claude Le Bris, CERMICS, Ecole Nationale des Ponts et Chaussees

Date/Time: Wednesday, October 12, 2005, 1:30-2:30pm (MT)

Location: Building 980 Room 95 (Sandia NM)
Building 915 Room S101 (Sandia CA)

Brief Abstract: We will overview three decades of mathematical contributions to computational chemistry. The emphasis is laid on electronic structure calculations. Theoretical contributions will be mentioned, but the talk will be biased toward numerical issues. Two examples of recent contributions will be detailed: convergence of SCF algorithms, and domain decomposition methods.

CSRI POC: Richard Lehoucq, (505) 845-8929

Title: Numerical methods for interface and contact problems

Speaker: Patrick Le Tallec, Ecole Polytechnique

Date/Time: Thursday, July 21, 2005, 10:00-11:00am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: The numerical simulation of fluid structure interaction problems in large deformation or of rolling tyres in rotation require - the derivation of energy conserving time integration schemes in presence of nonlinear kinematic constraints such as incompressibility and contact, - proper handling of the weak continuity constraint at the structure interface both with an internal fluid, and with a rigid obstacle. This informal talk will address several aspect related to these two issues.

CSRI POC: Richard Lehoucq, (505) 845-8929

Title: Development of Comprehensive Grid Generation Technology

Speaker: Vladimir D. Liseikin, Institute of Computational Technologies of Siberian Branch of the Russian Academy of Sciences, and Professor at Novosibirsk State University

Date/Time: Friday, April 1, 2005, 11:00-12:00 noon

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Development of grid generation techniques based on the numerical solution of Beltrami and modified Beltrami equations will be discussed. Control of the grid behavior is realized with the specification of coefficients in the equations. Formulas for the coefficients providing various types of grid adaptation will be presented. Examples of numerical experiments to some plasma-related problems are demonstrated, in which the principal grid property is alignment with a given magnetic field.

CSRI POC: Patrick Knupp, (505) 284-4565

Title: Wide-range, Multi-phase Equation of State for Metals

Speaker: Igor V. Lomonosov, Institute for Problems in Chemical Physics of the Russian Academy of Sciences, Chernogolovka, RUSSIA

Date/Time: Wednesday, October 19, 2005, 9:00-10:00 am

Location: Building 960 Room 1001B (Sandia NM)

Brief Abstract: I will discuss the high pressure and high temperature phase diagram of metals, and general problems of equation of state (EOS) construction. It is pointed out that the available high pressures, high temperatures information covers a broad range of the phase diagram, but has a heterogeneous character and, as a rule, is not thermodynamically complete; its generalization can be done only in the form of a thermodynamically complete EOS.

Principles of EOS construction are described using as examples the quasi-harmonic model of solids and the Mie-Grueneisen EOS. The extension of Mie-Grueneisen EOS to regions of high temperatures and low densities, and methods of taking into account effects of melting and evaporating are discussed.

I present a multi-phase EOS model accounting for solid, liquid, gas and plasma states, as well as two-phase regions of melting and evaporating. The thermodynamic properties of metals and their phase diagrams are calculated with the use of this model; typical examples for nickel and uranium are presented. Theoretical calculations of thermodynamic properties of the solid, liquid, and plasma phases, and of the critical point are compared with results of static and dynamic experiments.

I discuss our determination of the high pressure melting and vaporization curves, which were developed using direct temperature measurements and indirect information on phase transitions resulting from shock-wave experiments. These include measurements of sound velocity in shocked metals, principal Hugoniots of cooled and heated metals, porous metal Hugoniots and release isentropes. Conclusions are made for high-pressure melting and vaporization, as well as calculated parameters of the critical point parameters.

General regularities obtained from the analysis of 30 metals are presented. The correspondence between Isobaric Expansion Experiment (IEX) and shock-wave data in the liquid is demonstrated for the metals investigated. Birch's law for liquid metals and its applicability are discussed. Accurate EOS calculations are presented for possible applications of heavy-ion beam for high-energy-density physics. Application of multi-phase EOS concerns the problem of shock-wave stability in metals is alluded to.

CSRI POC: John Aidun, (505) 844-1209

Title: Interior Methods for Mathematical Programs with Complementarity Constraints (MPCCs)

Speaker: Gabriel Lopez-Calva, Interview Candidate, Northwestern University

Date/Time: Monday, April 18, 2005, 11:00-12:15 am (MT)

Location: Building 980 Room 24 (Sandia NM)
Building 915 Room S101 (Sandia CA)

Brief Abstract: I will start with a brief introduction to the notion of 'complementarity' and to Mathematical Programs with Complementarity Constraints (MPCCs), as a motivation. I will present then a framework for the numerical solutions of MPCCs using interior methods, based on an exact penalty reformulation. I will go over the main contributions of our work, namely, global and local convergence results, and strategies to handle the penalty parameter for effective practical implementations. Numerical results will also be presented. This is joint work with Sven Leyffer (Argonne National Laboratory) and Jorge Nocedal (Northwestern University). [Time permitting, I will briefly talk about my current research projects, at the end.]

CSRI POC: Patty Hough (925) 294-1518
Monica Martinez-Canales, (925) 294-3157

Title: Superconducting Quantum Computation - Hardware and Applications

Speaker: Dr. Peter J. Love, D-Wave Systems Inc.

Date/Time: Monday, December 12, 2005, 9:00-10:00 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: The implementation of one- and two-qubit gates in superconducting qubit prototypes has confirmed their utility for quantum computation. Such qubits are readily fabricated, and highly scalable. Superconducting architectures are suitable for implementation of the adiabatic model quantum computing, avoiding the difficulties of pulsed control. I will discuss D-Wave's approach to the realization of superconducting quantum computers, and the commercial applications of quantum computing.

CSRI POC: Alex Slepoy, (505) 284-3650

Title: Geometric Algorithms for Intensity-Modulated Radiation Therapy Delivery Planning

Speaker: Shuang (Sean) Luan, Department of Computer Science and Department of Radiology University of New Mexico

Date/Time: Wednesday, April 13, 2005, 10:00-11:00 am

Location: Building 980 Room 95 (Sandia NM)
Building 921 Room 137 (Sandia CA)

Brief Abstract: Intensity-modulated radiation therapy (IMRT) is a modern cancer treatment technique and aims to deliver a highly conformal dose to a target tumor while sparing the surrounding sensitive structures. A key to performing IMRT is the accurate and efficient delivery of discrete dose distributions using the linear accelerator and the multileaf collimator. The IMRT delivery planning problems arise in such scenarios, whose goal is to determine a control sequence of the linear accelerator and the multileaf collimator that creates a prescribed dose distribution both accurately and efficiently. In this talk, I will present some new delivery planning algorithms and software. Our new algorithms, which are based on geometric optimization techniques, are very efficient and guarantee the optimal quality of the output treatment plans. Comparisons between our software with commercial planning systems and the current most well known leaf sequencing algorithm in the medical literature showed substantial improvement.

Our software has already been used for treating cancer patients at a couple of medical centers.

CSRI POC: Bruce Hendrickson, (505) 845-7599

Title: The Parallel Boost Graph Library

Speaker: Andrew Lumsdaine, Indiana University

Date/Time: Wednesday, June 29, 2005, 10:00-11:00 am

Location: Building 980 Room 24 (Sandia NM)

Brief Abstract: We present a library of generic software components for parallel and distributed computations on graphs, based on the Boost Graph Library (BGL). The BGL consists of a rich set of generic graph algorithms and supporting data structures, but it was not originally designed with parallelism in mind. In this talk, we revisit the abstractions comprising the BGL in the context of distributed-memory parallelism, lifting away the implicit requirements of sequential execution and a single shared address space. We illustrate our approach by describing the process as applied to one of the core algorithms in the BGL, breadth-first search. The result is a generic algorithm that is unchanged from the sequential algorithm, requiring only the introduction of external (distributed) data structures for parallel execution. More importantly, the generic implementation retains its interface and semantics, such that other distributed algorithms can be built upon it, just as algorithms are layered in the sequential case.

By characterizing these extensions as well as the extension process, we develop general principles and patterns for using (and reusing) generic parallel software libraries. We demonstrate that the resulting algorithm implementations are both efficient and scalable with performance results for several algorithms.

CSRI POC: Jonathan Berry, (505) 284-4021

Title: Mobility as a Predictive Measure of Evolutionary Algorithm Effectiveness

Speaker: Monte Lunacek, CSRI Ph.D Student

Date/Time: Wednesday, June 22, 2005, 1:00-2:00 pm

Location: Building 980 Room 24 (Sandia NM)

Brief Abstract: The global search properties of heuristic search algorithms are not well understood. In this paper, we introduce a new metric, mobility, that quantifies the dispersion of local optima visited during a search. This allows us to explore two questions: How disperse are the local optima visited during a search? How does mobility relate to algorithm performance? We compare local search with two evolutionary algorithms, CHC and CMA-ES, on a set of non-separable, non-symmetric, multi-modal test functions. Given our mobility metric, we show that algorithms visiting more disperse local optima tend to be better optimizers.

CSRI POC: Jean-Paul Watson, (505) 845-8887

Title: Optimization of large-scale water systems: Performance, Adaptation, and Security

Speaker: Professor Jay Lund, University of California, Davis

Date/Time: Thursday, November 10, 2005 10:00-11:00 am (PT)

Location: Building 980 Room 95 (Sandia NM)
Building 915 Room S101 (Sandia CA)

Brief Abstract: Advances in computing power and software now make it possible to apply optimization techniques and formulation to very large problems. This talk presents an application of generalized network flow model to California's intertied water system, covering about 90% of the state's urban and agricultural water uses. Several applications are reviewed include large-scale coordinated water management, climate change, and restoration of Hetch Hetchy Valley. Potential applications to water system security from natural and human threats are discussed. The limitations and benefits of such analysis are discussed. Some areas where optimization and system analysis development would be useful are discussed.

CSRI POC: Vicki Howle, (925) 294-2204

Title: Traveling Threads: A New Multi-Threaded Execution Model

Speaker: Richard C. Murphy, Notre Dame University

Date/Time: Monday, April 11, 2005, 2:00-3:00 pm

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: The primacy of single processor performance, and the ease of the corresponding programming model, has forced the architects of parallel machines to accept limited scalability of processor and memory hierarchies. While architectures at any scale are plagued by the von Neumann bottleneck, the upcoming generation of trans-petaflop machines represent the hardest scalability problem. While interconnection networks provide more bandwidth, they have done little to address the problem of relative latency. In a typical machine that includes both near (on-node) and far (off-node) memory, both types of memory are relatively more distant from the processing resources. Furthermore, finding the concurrency in existing instruction streams to mask that latency is an increasingly daunting task. Advances in architecture are required to address this problem.

This work presents a new multi-threaded execution model based on architecture-level thread migration that combats both the problem of providing additional concurrency (to mask latency) and reducing interconnection network pressure (by converting request-reply operations to one-way operations). The architectural parameters of this model are analyzed in the context of a set of Sandia benchmarks (CTH, LAMMPS, sPPM, and a driver for Trilinos). Furthermore, the fundamental properties of the Sandia codes are assessed in relation to the SPEC Floating Point Benchmark Suite (SPEC-FP), the gold standard for computer architecture research. It will be shown that the Sandia codes exhibit significantly more memory activity than their SPEC-FP counterparts.

CSRI POC: Keith Underwood, (505) 284-9283

Title: Some Linear Algebra of Quantum Computing

Speaker: Dianne P. O'Leary, University of Maryland

Date/Time: Wednesday, September 28, 2005, 11:00-12:00 noon (PT)

Location: Building 980 Room 95 (Sandia NM)
Building 915 Room S101 (Sandia CA)

Brief Abstract: Conventional computer circuits perform logic operations on bits of data through a sequence of gates. Quantum computers transform data by multiplication by a unitary matrix, using a sequence of gates determined by a factorization of that matrix into a product of allowable elementary factors. This talk gives a brief introduction to quantum computing and describes how to use matrix factorization to design quantum circuits with optimal-order gate counts for computing with qubits (0-1 logic) and qudits (multilevel logic). This is joint work with Stephen S. Bullock and Gavin K. Brennen.

CSRI POC: Tamara Kolda, (925) 294-4769

Title: Informatics for Higher Order Finite Element Simulations

Speaker: Philippe P. Pebay, Sandia National Laboratories

Date/Time: Monday, August 8, 2005

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Advanced finite element applications (FEAs) are required by Sandia in order to accomplish its role as stockpile steward without weapons testing. New FEAs use a combination of hierarchical (h) and polynomial (p) adaptivity to simulate a test with fewer compute resources than traditional simulations. However, there is currently a paucity of tools for dealing with the many additional complexities these codes present. In particular, choosing whether a region of the simulation domain should be h-, p-, or hp-adapted is currently accomplished with heuristics. When these heuristics fail, diagnosing their cause is difficult. Also, the way mesh adaptations are currently performed consumes a great deal of time, as estimates of the error can only be made after an adaptation has been performed. In practice, this results in an iterative process where multiple mesh adaptations must be made before the desired error bounds are reached. Both these problems reduce the gains that higher order FEAs would otherwise have over traditional FEAs.

In this talk, we will discuss how Bayesian techniques can help analysts and developers devise an "a priori" error estimator based on both data obtained from previous runs, and expert knowledge. In particular, we will examine why optimal quadrature is critical for such approaches.

CSRI POC: Richard Lehoucq, (505) 845-8929

Title: New Computational Methods in Quantum Dynamics, and Molecular Applications

Speaker: Bill Poirer, Texas Tech

Date/Time: Wednesday, July 27, 2005, 2:00-3:00 pm

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Accurate quantum dynamics calculations of molecular systems, for which both nuclear and electronic degrees of freedom are treated quantum mechanically, are notoriously challenging to perform, even on modern-day computers. However, such calculations are essential for answering fundamental dynamical questions that cannot be resolved any other way (e.g. via classical simulation). This talk addresses new methodologies and algorithmic developments that provide orders-of-magnitude improvements in computational efficiency, thus allowing accurate quantum dynamics calculations to be performed for larger and more complex systems than previously considered.

CSRI POC: Richard Lehoucq, (505) 845-8929

Title: Generalizations of an Inverse Free Preconditioned Algorithm for Symmetric Generalized Eigenvalue Problems

Speaker: Pat Quillen, University of Kentucky

Date/Time: Wednesday, April 20, 2005, 9:00-10:00 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: In their 2002 paper, Golub and Ye present an inverse free preconditioned Krylov subspace method for computing the extreme eigenvalues of a symmetric definite pencil (A,B). In this talk we discuss application of this algorithm to a symmetric definite pencil (C,D) whose extreme eigenvalues correspond to eigenvalues of the pencil (A,B) nearest to some target μ. In general, convergence of the algorithm applied to this transformed pencil is prohibitively slow and a high quality preconditioner is required. Our focus is largely on preconditioning techniques for this problem, and this will be the basis for much of the discussion. We also discuss a block generalization of the inverse free algorithm which exhibits convergence behavior superior to that of the original algorithm, especially when combined with a preconditioning strategy. Examples will be provided demonstrating the viability of this method.

CSRI POC: Jonathan Hu, (925) 294-2931

Title: Why Polyhedra Matter in Nonlinear Equation Solving

Speaker: J. Maurice Rojas, Texas A&M

Date/Time: Monday, October 24, 2005, 9:30-10:30 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Polyhedra are fundamental in linear optimization and their importance in nonlinear optimization is now being recognized in many ways.

In this talk, we introduce some of the most useful connections between polyhedra and polynomials. We start by showing how the number of complex solutions of any system of polynomial equations can be expressed in terms of volumes of polytopes. Integer linear equations also enter in an important way.

Time permitting, we will discuss some recent algorithmic speed-up techniques for solving nonlinear equations, including polyhedral homotopy and tropical geometry. We will also touch upon the subtleties behind counting just the real solutions of a nonlinear system.

CSRI POC: Mark (Danny) Rintoul, (505) 844-9592

Title: Automated Nonlinear Macromodelling and Applications

Speaker: Jaijeet Roychowdhury, University of Minnesota

Date/Time: Monday, August 8, 2005, 2:00-3:00 pm

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Over the last few years, techniques for automatically abstracting small, computationally efficient macromodels of nonlinear systems have started making an impact on electronic circuit design and CAD. In this talk, we will describe three nonlinear macromodelling research areas that have shown particular promise. The first, trajectory-based piecewise polynomial (PWP) macromodelling, has been applied successfully to a variety of stable non-oscillatory nonlinear circuits (including op-amps, comparators, and I/O buffers) to generate useful general-purpose macromodels faster by about an order of magnitude.

The second topic pertains to algorithms for reducing time-varying systems using multitime differential equations combined with Krylov-subspace methods. Applications include RF circuits such as mixers and sampling circuits such as switched-capacitor filters. Another promising and timely application - addressing a critical missing need for digital switching noise prediction - is abstracting computationally-efficient interference macromodels of switching digital gates.

The third area to be described concerns macromodelling oscillatory systems. We will provide a brief description of numerical techniques to abstract nonlinear phase macromodels of oscillators. We will describe the application of these macromodels to predict phase noise and injection locking in oscillators, as well as uses for fast phase-locked loop (PLL) simulation. If time permits, we will also describe recent biological and nanoscale applications to predict spontaneous pattern formation and image-processing behaviors in very large systems of coupled oscillators. Open problems and challenges will be mentioned in the course of the talk.

CSRI POC: Scott Hutchinson, (505) 845-7996

Title: Making Chord Robust to Byzantine Attacks

Speaker: Jared Saia, University of New Mexico

Date/Time: Monday, August 29, 2005, 2:00-3:00 pm

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Chord is a distributed hash table (DHT) that requires only O(log n) links per node and performs searches with latency and message cost O(logn), where n is the number of peers in the network. Chord assumes all nodes behave according to protocol. We give a variant of Chord which, for any fixed epsilon>0, is resilient to (1/4-epsilon)z bad nodes joining the network over a time period during which 1) there are always at least z total nodes in the network and 2) the number of peer joins and leaves is no more than z^k for some tunable parameter k. We assume there is an omniscient and computationally unbounded adversary controlling the bad peers and that the IP-addresses of all the bad peers and the locations where they join the network are carefully selected by this adversary.

Our notion of resilience is rather strong in that we not only guarantee that searches can be performed but also that we can enforce any set of ``proper behavior'' such as contributing new material, etc. In comparison to Chord, the resources required by this new variant are only a polylogarithmic factor greater in communication, messaging, and linking costs.

This is joint work with Amos Fiat (U. Tel Aviv), Valerie King (U. Victoria), Maxwell Young (UNM) and Vishal Sanwalani (UNM) and represents work previously published in PODC and ESA.

CSRI POC: Cindy Phillips, (505) 845-7296

Title: On the Feasibility of Optical Circuit Switching for High Performance Computing Systems

Speaker: Eugen Schenfeld, IBM

Date/Time: Thursday, September 22, 2005, 11:00-12:00 noon

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: The interconnect plays a key role in both the cost and performance of large-scale HPC systems. The cost of future high-bandwidth electronic interconnects is expected to increase due to expensive optical transceivers needed between switches. We describe a potentially cheaper and more power-efficient approach to building high-performance interconnects.

Through empirical analysis of HPC applications, we find that the bulk of inter-processor communication (barring collectives) is bounded in degree and changes very slowly or never. Thus we propose a two-network interconnect: An Optical Circuit Switching (OCS) network handling long-lived bulk data transfers, using optical switches; and a secondary lower-bandwidth Electronic Packet Switching (EPS) network. An OCS could be significantly cheaper than an all electronic network as it uses fewer optical transceivers. Collectives and transient communication packets traverse the electronic network.

We present compiler techniques and dynamic run-time policies, using this two-network interconnect. Simulation results show that our approach provides high performance at low cost.

CSRI POC: Erik DeBenedictis, (505) 284-4017

Title: Towards Rational Mutation of an Enzyme: Understanding Rubisco's Specificity

Speaker: Dr. Jurgen Schlitter, Ruhr-Universität Bochum

Date/Time: Wednesday, May 18, 2005, 10:30-11:30 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: The plant enzyme Rubisco plays a prominent role in photosynthesis as it fixes gaseous carbon dioxide to Ribulose molecules, thus being the main material entry of the biosphere. During evolution it developed increasing specificity for carbon dioxide as compared to oxygen binding which opens an energetically unfavorable pathway. Our today's understanding of this interesting progress started with comparative studies of sequences which indicated the importance of the C-terminal tail and led us to successful mutational experiments. The later developed time-window hypothesis connects the kinetics of the conformational change that opens the binding niche with the specificity of the binding process. Recent MD simulations support the hypothetical correlation between structure, conformational kinetics and specificity. It seems possible now to predict beforehand semi-quantitatively the specificity of new mutants.

CSRI POC: Paul Crozier, (505) 845-9714

Title: A Parallel Dense-Matrix Eigensolver

Speaker: Mark Sears, Sandia National Laboratories

Date/Time: Thursday, October 20, 2005, 10:00-11:00 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Good eigensolver technology has been around for some time, for example reference implementations of several algorithms in LAPACK are readily available. However, there has been a need for a high performance parallel dense-matrix eigensolver for the symmetric/Hermitian case especially in the electronic structure community. High performance in the modern computing environment means good parallel scaling in addition to using the best possible BLAS implementation. This has been somewhat difficult to achieve with currently available algorithms, primarily because of the reduction to tridiagonal form. Many people have suggested that reducing to a band form would be more efficient, but then the problem of dealing with a banded matrix is difficult to solve.

This talk will cover a new algorithm which I have developed for the dense matrix Hermitian eigenproblem. This algorithm is based on reduction to a band form rather than tridiagonal form, and this is coupled to a direct band eigensolver, which uses a combination of bisection and a new form of inverse iteration based on QR, both highly parallel. However, this is just the beginning. For a practical eigensolver a high performance parallel implementation using the best available BLAS is essential. In addition for the general case $Ax = \lambda Bx$ a number of other high performance parallel algorithms are needed, including matrix multiply, Cholesky factorization, and triangular solves. My implementation uses a block-cyclic decomposition for all of these algorithms. I will also present some impressive results for scaling and absolute performance.

CSRI POC: Richard Lehoucq, (505) 845-8929

Title: Multimaterial Interface Reconstruction*

Speaker: M. J. Shashkov, Los Alamos National Laboratory

Date/Time: Tuesday, December 13, 2005, 10:00-11:00 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Lagrangian methods for flow simulations automatically keep track of material interfaces whereas Eulerian and Arbitrary-Lagrangian-Eulerian (ALE) methods must incorporate special procedures to track or infer interfaces contained in mixed cells.

In the first part of this talk we focus on interface reconstruction methods in which an approximate interface is recreated in each cell using volume fraction information in the cell and its neighbors. It addresses two important but often ignored issues in such procedures.

The first issue is the accurate recovery of the interface while avoiding the creation of holes and fragments. Creation of such artifacts by the interface reconstruction procedure independent of the physical processes is obviously undesirable leading to a loss of accuracy and a divergence of the simulation from reality.

The second important subject is the treatment of interfaces in mixed cells with three or more materials. Many existing procedures recover only some interface topologies correctly and that too, only if the materials are processed in a specific order. Therefore, it is very important to devise interface reconstruction algorithms that are order-independent. Results of the interface reconstruction procedure developed with these issues in mind will be presented.

In the second part of this talk we describe a new Moments of Fluid Method (MoF) for interface reconstruction. In the MoF method, in addition to the volume fractions we use information about position of centroids of each material. Results of interface reconstruction using MoF are presented.

*R.V. Garimella, V. Dyadechko, M.J. Shashkov, B.K. Swartz
T-7, Mathematical Modeling and Analysis
Los Alamos National Laboratory
Los Alamos, NM 87544

CSRI POC: Allen C. Robinson, (505) 844-6614

Title: Preconditioners for Generalized Saddle-Point Problems

Speaker: Christopher Siefert, University of Illinois

Date/Time: Monday, August 29, 2005, 9:00-10:00 am (MT)

Location: Building 980 Room 95 (Sandia NM)
Building 915 Room S101 (Sandia CA)

Brief Abstract: Generalized saddle point problems arise in a number of applications, ranging from optimization and metal deformation to fluid flow and PDE-governed optimal control. We examine two types of preconditioners for these problems, one block-diagonal and one indefinite, and present analyses of the eigenvalue distributions of the preconditioned matrices.

We also investigate the use of approximations for the Schur complement matrix in these preconditioners and develop eigenvalue analysis accordingly. We examine new developments in probing methods (derived from graph coloring methods for sparse Jacobians) for building approximations to the Schur complement and consider their effect in the context of our preconditioners. To illustrate our results, we consider a model Navier-Stokes problem as well as a real-world application involving the deformation of aluminum strips.

CSRI POC: Ray Tuminaro, (925) 294-2564

Title: Developing multi-agent systems as a new approach for engineering design and optimization

Speaker: John Siirola, Interview Candidate, Carnegie Mellon University

Date/Time: Tuesday, June 21, 2005, 10:00-11:00 am

Location: Building 980 Room 95

Brief Abstract: The explosive growth of computing capacity raises a new question for the process systems engineering community: How can we take what is effectively an unlimited computing resource and apply it to solving new, larger, harder engineering problems?

In this seminar, I present the framework and implementation of an agent-based system for general single- and multi-objective optimization. This system creates and manages a collaborative environment wherein a diverse library of many independently-running algorithms communicate through the sharing of intermediate and final solutions to an optimization problem. These algorithms span the entire range of available approaches, including rigorous gradient-based or branch-and-bound optimization methods, stochastic search methods, and heuristic methods. The agent system also introduces the new concept of "polymorphic optimization." Analogous to diversity in the algorithmic library, polymorphic optimization provides diversity in the formulation of the original problem. This formalism allows algorithms working on fundamentally different formulations of the original optimization problem to collaborate and seamlessly share solution information.

My approach represents the confluence of two primary areas of optimization research: development of optimization algorithms and development of formulation methodologies. The former aims to create new, more robust algorithms that can solve a larger variety of problem classes and/or solve specific problem classes faster. The latter aims to develop new approaches for representing problems in a form that existing optimization algorithms will more likely solve successfully. Current research in both areas focuses primarily on "traditional" serial optimization: optimizing a single objective using one algorithm running on a stand-alone personal computer.

In contrast, the emerging standard for high-performance computing is the distributed cluster of inexpensive personal computers. Historically, researchers have used clusters for large-scale parallel algorithms. Engineering optimization is a member of a sizable body of problems where, although there are several approaches to solving the problem, these approaches are primarily serial algorithms that do not readily lend themselves to classical parallelism. Agent-based systems form a promising approach to this next generation of parallel systems: parallelism at the method and formulation level.

In this talk, I illustrate the benefits, future potential, and remaining challenges of an agent-based optimization approach through several examples, including the single-objective [1] and multi-objective [2] optimization of non-convex functions, and the design and multi-objective optimization of microfluidic separation systems [3,4].

References:

[1] J.D. Siirola, S. Hauan, and A.W. Westerberg. "Toward Agents-Based Process Systems Engineering: Proposed Framework and Application to Non-convex Optimization." Comp.Chem.Engng. 27(12), pp. 1801-1811 (2003).
[2] J.D. Siirola, S. Hauan, and A.W. Westerberg. "Computing Pareto Fronts using Distributed Agents". Comp.Chem.Engng. 29(1), pp. 113-126 (2004).
[3] A.J. Pfeiffer, J.D. Siirola, and S. Hauan. "Optimal design of microscale separation systems using distributed agents." In the Proceedings of FOCAPD 2004, pp. 381-384 (2004).
[4] J.D. Siirola and S. Hauan. "Polymorphic Optimization." Submitted to Comp.Chem.Engng. May, 2005.

CSRI POC: David Schoenwald, (505) 284-6285

Title: Fast Electrostatics and Polarizable Forces

Speaker: Robert D. Skeel, Purdue University

Date/Time: Tuesday, November 1, 2005, 9:30-10:30 am

Location: Building 980 Room 24 (Sandia NM)

Brief Abstract: The calculation of nonbonded, and especially electrostatic, interactions is a major bottleneck in molecular simulations. To make matters worse, a consensus is emerging among researchers concerning the general inadequacy of fixed point-charge models and the desirability of including electronic polarizability in the models. Presented here is a report of recent work on two projects, each yielding substantial performance gains. The first is the use of hierarchical interpolation of interaction potentials on nested grids to calculate energies and forces in linear time for both periodic or nonperiodic boundary conditions. The second is the self-consistent solution of large dense linear systems for the induced dipole model. This is joint work with David J. Hardy and Wei Wang.

CSRI POC: Richard Lehoucq, (505) 845-8929

Title: Artificially Intelligent Alternative to Equation-less modeling: Dare to Delegate

Speaker: Alex Slepoy, Sandia National Laboratories

Date/Time: Wednesday, May 4, 2005, 3:00-4:00 pm

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: An important methodology, emerging in multi-scale modeling, uses an equation-less method to overcome lack of descriptive equations at the higher scale. Unfortunately, such methods rely on great degree of smoothness in the meso-scale derivatives for their efficiency.

I propose an alternative strategy to equation-less modeling using approaches from the automatic reasoning formalism. I apply the approach to classical force field fitting using ab initio electronic structure information. I motivate the approach in the context of the "missing scales".

CSRI POC: Richard B. Lehoucq, (505) 845-8929

Title: Symmetry Preserving Principle Component Analysis for Dynamical Systems

Speaker: Danny Sorensen, Rice University

Date/Time: Tuesday, June 21, 2005, 10:00-11:00 am

Location: Building 980 Room 24 (Sandia NM)

Brief Abstract: Model reduction seeks to replace a large-scale system of differential or difference equations by a system of substantially lower dimension, that ideally, has the same response characteristics as the original system, yet requires far less computational resources for realization. Such large-scale systems arise in structural analysis, circuit simulation, protein dynamics, spatial discretization of certain time dependent PDE control systems and in many other applications.

Principle Component Analysis based upon the singular value decomposition of a discrete trajectory is central to a number of important model reduction methods in both linear and nonlinear settings. This talk will introduce the ideas of model reduction for dynamical systems in this context and give a brief introduction to methods for large scale problems.

In the molecular dynamics of proteins, there are often symmetry conditions on the molecular conformations that should be respected. We shall introduce a relatively new technique for a symmetry preserving SVD suitable for a reduced basis analysis in molecular dynamics that respects known symmetries.

CSRI POC: Heidi Thornquist, (505) 284-8426

Title: Near-Wall Treatments for Coarse Grid Large-Eddy Simulation

Speaker: Jeremy Templeton, Interview Candidate
Stanford University, Stanford, California

Date/Time: Wednesday, April 20, 2005, 11:00-12:15 am (MT)

Location: Building 980 Room 95 (Sandia NM)
Building 915 Room S101 (Sandia CA)

Brief Abstract: The major bottleneck preventing large-eddy simulations (LES) of turbulent fluid flow from being used in the complex geometries and high Reynolds numbers typically encountered in engineering practice is the need to resolve the near-wall layer. This is because current sub-grid scale (SGS) models cannot accurately account for the highly energetic and anisotropic small scales found in this region. To avoid the cost of resolving these scales, wall treatments have been used in LES to allow it to accurately capture the important physics of the flow without resolving the near-wall turbulence. In this work, optimal control theory has been used to derive a technique that provides boundary conditions for LES so that it can be used on coarse grids that do not resolve the wall region. The controller is able to compensate not only for the unresolved physics but also for the numerical and SGS modeling errors that occur when coarse grids are used. Results from high Reynolds number plane channel flow have demonstrated that this approach can accurately reproduce the theoretical mean velocity profile for this configuration. Higher order statistics have shown that the rms profiles are also well predicted in the outer flow, but near the wall are elevated in order to compensate for SGS model deficiencies. To improve the prediction of these higher order quantities, a new technique for augmenting LES SGS models by correcting them based on the resolved turbulence is currently being pursued. Preliminary results have indicated that this method has the potential to accurately predict both mean and rms profiles, and has the additional advantages of being computationally inexpensive and easy to implement.

CSRI POC: Patty Hough (925) 294-1518
Monica Martinez-Canales, (925) 294-3157

Title: Compatible discretizations for PDEs

Speaker: Kate Trapp, University of Richmond

Date/Time: Monday, August 8, 2005, 10:00-11:00 am (MT)

Location: Building 980 Room 95 (Sandia NM)
Building 915 Room S101 (Sandia CA)

Brief Abstract: Compatible discretization methods for solving partial differential equations are often used in electromagnetics to preserve the underlying physical properties expressed in the equations. The support operator (or mimetic) method and the covolume method are two compatible discretization methods that rely on a discrete vector calculus structure. These two methods can be placed in a common framework through the use of discrete differential complexes. In this talk we describe this framework, how it captures fundamental relationships among first-order differential operators, its connection to the finite element methods of Raviart-Thomas and Nedelec, and demonstrate how convergence results for div-curl systems can be extended to a larger class of algorithms.

CSRI POC: Pavel Bochev, (505) 844-1990

Title: Finite Element Method on the Sphere

Speaker: Necibe Tuncer, Auburn University

Date/Time: Thursday, April 21, 2005, 10:00-11:00 (PT)

Location: Building 915 Room S101 (Sandia CA)

Brief Abstract: The numerical approximation of partial differential equations on the sphere has been an attractive subject of late. We study the model problem −_u = f(x, y) where u(x, y) = g(x, y) on the boundary of the domain. Here the domain is the unit disc. In our research, we generate the FEM on the unit disc, by mapping the unit square to the domain with the radial projection. We introduce these mapped finite elements on the unit disc as “radially projected finite elements” We discuss the approximation properties of radially projected finite elements on the unit disc and the sphere.

CSRI POC: Monica Martinez-Canales, (925) 294-3157

Title: Correlations in planar quantum dots: A quantum Monte Carlo study

Speaker: Professor Cyrus Umrigar, Cornell University

Date/Time: Monday, March 28, 2005, 10:00-11:00 am

Location: Building 897 Room 1033 (Sandia NM)

Brief Abstract: Quantum dots, also known as "artificial atoms" are not only of considerable technological interest but also of theoretical interest because it is possible to go from a weak correlation to a strong correlation regime either by increasing the relative strength of electron-electron interaction to the external potential or by increasing the magnetic field.

We employ diffusion Monte Carlo to study the ground and excited states of dots in various regimes and compare the results to those from the Hartree Fock (HF) method and from density functional theory within the local spin density approximation (LSDA). In the absence of a magnetic field we find, in contrast to the situation for real atoms, that the total energies and addition energies obtained from LSDA are much more accurate than those from HF. This is because the relative magnitude of the correlation energy to the exchange energy is much larger in dots than in atoms and the density is less inhomogeneous in dots. LSDA predicts reasonably accurate excitation energies for many states, but in those cases where the LSDA states are spin contaminated it predicts excitation energies that are too low, whereas, in those cases where there is considerable multideterminantal character in the excited state it predicts excitation energies that are too high. Hund's first rule is satisfied for all electron numbers studied, but for N=10 there is a near degeneracy.

In the large magnetic field limit the determinants can be limited to those arising from the lowest Landau level. For finite magnetic fields Landau level mixing is important and can be taken into account very effectively by multiplying the infinite-field determinants by a Jastrow factor which is optimized by variance minimization. We apply these wave functions to study the transition from the maximum density droplet state (integer quantum Hall state, L=N(N-1)/2) to lower density droplet states (L>N(N-1)/2). Composite-fermion wave functions, projected onto the lowest Landau level and multiplied by an optimized Jastrow factor, provide an accurate and efficient alternative form of the wave functions.

CSRI POC: Kevin Leung, (505) 844-1588

Title: Dynamic Differential Data Protection in High-Performance Middleware

Speaker: Patrick Widener, University of New Mexico

Date/Time: Thursday, July 21, 2005, 1:00-2:00 pm

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Modern distributed applications are long-lived, are expected to provide flexible and adaptive data services, and must meet the functionality and scalability challenges posed by dynamically changing user communities in heterogeneous execution environments. The practical implications of these requirements are that reconfiguration and upgrades are increasingly necessary, but opportunities to perform such tasks offline are greatly reduced. Developers are responding to this situation by dynamically extending or adjusting application functionality and by tuning application performance, a typical method being the incorporation of client- or context-specific code into applications' execution loops.

Prior work has highlighted the performance advantages provided by dynamic code extension or specialization. Our work addresses a basic roadblock in deploying such solutions, which is the protection of key application components and sensitive data in distributed applications.

Our approach, termed Dynamic Differential Data Protection (D3P), provides fine-grain methods for providing component-based protection in distributed applications. Context-sensitive, application-specific security methods are deployed at runtime to enforce restrictions in data access and manipulation. D3P is suitable for use in low- or zero-downtime environments, since such deployments are performed while applications run, D3P is appropriate for high performance environments and for highly scalable applications like publish/subscribe, because it creates native codes via dynamic binary code generation. Finally, due to its integration into middleware, D3P can run across a wide variety of operating system and machine platforms.

This talk introduces the need for D3P, using sample applications from the high performance and pervasive computing domains to illustrate the problems addressed by our D3P solution. It also describes how D3P can be integrated into modern middleware. Experimental evaluations demonstrate the fine-grain nature of D3P, that is, its ability to capture individual end users' or components' needs for data protection, and they also describe the performance implications of using D3P in data-intensive applications.

CSRI POC: Rolf Riesen, (505) 845-7363

Title: Efficient High Order Schemes for Long-Time Wave Propagation of Complex Multiscale Flows

Speaker: Helen Yee, Exploration Technologies Directorate
NASA Ames Research Center

Date/Time: Tuesday, May 24, 2005, 9:00-10:15 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: A general framework for the design of efficient adaptive low-dissipative high order schemes is presented. It encompasses a rather complete treatment of the numerical approach based on four integrated design criteria:

- For stability considerations, condition the governing equations before the application of the appropriate numerical scheme whenever it is possible.
- For consistency, compatible schemes that possess stability properties, including physical and numerical boundary condition treatments, similar to those of the discrete analogue of the continuum are preferred.
- For the minimization of numerical dissipation contamination, efficient and adaptive numerical dissipation control to further improve nonlinear stability and accuracy should be used.
- For practical considerations, the numerical approach should be efficient and applicable to general geometries, and an efficient and reliable dynamic grid adaptation should be used if necessary.

These design criteria are, in general, very useful to a wide spectrum of flow simulations. However, the demand on the overall numerical approach for nonlinear stability and accuracy is much more stringent for long-time integration of complex multiscale viscous shock/shear/turbulence/acoustics interactions and numerical combustion. Robust classical numerical methods for less complex flow physics are not suitable or practical for such applications.

The present approach is designed expressly to address such flow problems, especially unsteady flows. The minimization of employing very fine grids to overcome the production of spurious numerical solutions and/or instability due to under-resolved grids is also sought. The incremental studies to illustrate the performance of the approach are summarized.

CSRI POC: Randall M. Summers, (505) 844-6296

Title: Filtering and Limiting in High Order Methods for Non-ideal MHD

Speaker: Helen Yee, Exploration Technologies Directorate
NASA Ames Research Center

Date/Time: Tuesday, May 24, 2005, 10:30-11:45 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: The adaptive nonlinear filtering and limiting in spatially high order schemes [Yee et al. (1999), Yee & Sjogreen (2002) and Sjogreen & Yee (2004)] for the compressible Euler and Navier-Stokes equations have been recently extended to the ideal and non-ideal magnetohydrodynamics (MHD) equations.

The numerical dissipation control in these adaptive filter schemes consists of automatic detection of different flow features as distinct sensors to signal the appropriate type and amount of numerical dissipation/filter where needed and leave the rest of the region free from numerical dissipation contamination. The numerical dissipation considered consists of high order linear dissipation for the suppression of high frequency oscillation and the nonlinear dissipative portion of high-resolution shock-capturing methods for discontinuity capturing. The applicable nonlinear dissipative portion of high-resolution shock-capturing methods is very general.

The objective of this talk is to present the performance of three commonly used types of discontinuity capturing nonlinear numerical dissipation as nonlinear filters for the viscous, resistive and hall MHD.

CSRI POC: Randall M. Summers, (505) 844-6296

Title: Building Blocks for Reliable Complex Nonlinear Numerical Simulations

Speaker: Helen Yee, Exploration Technologies Directorate
NASA Ames Research Center

Date/Time: Wednesday, May 25, 2005, 9:00-10:30 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: This talk describes some of the building blocks to ensure a higher level of confidence in the predictability and reliability (PAR) of numerical simulation of multiscale complex nonlinear problems. The focus is on relating PAR of numerical simulations with complex nonlinear phenomena of numerics.

To isolate sources of numerical uncertainties, the possible discrepancy between the chosen partial differential equation (PDE) model and the real physics and/or experimental data is set aside. The discussion is restricted to how well numerical schemes can mimic the solution behavior of the underlying PDE model for finite time steps and grid spacings. The situation is complicated by the fact that the available theory for the understanding of nonlinear behavior of numerics is not at a stage to fully analyze the nonlinear Euler, Navier-Stokes and MHD equations. The discussion is based on the knowledge gained for nonlinear model problems with known analytical solutions to identify and explain the possible sources and remedies of numerical uncertainties in practical computations. Examples relevant to numerical prediction of flow transition/chaos by direct simulation are included.

CSRI POC: Randall M. Summers, (505) 844-6296

Chapter 6. Fellowships

The Computer Science Research Institute supported two students during CY2005 through the DOE High Performance Computer Science (HPCS) Fellowship administered by the Krell Institute. The objective of the DOE High-Performance Computer Science Fellowship program is to encourage the training of computer scientists by providing financial support to talented students to enter a period of study and research in computer science with an emphasis on high-performance computing, accompanied by practical work experience with reseachers at Los Alamos National Laboratory, Lawrence Livermore National Laboratory and Sandia National Laboratories.

The fellowship program requires a program of study that will provide a solid background in high-performance computing. Examples of research specializations of interest to the program include:

- Parallel and novel architectures, including clusters of SMPs
- Three-dimensional scientific visualization
- High-speed network interconnects
- Parallelizing compiler technology
- Parallel systems software (OS kernel technology, file systems, etc.)
- Performance evaluation and modeling
- Scalable computer security
- Object-oriented scientific programming frameworks
- High-performance computing software component frameworks
- Parallel scalable algorithms

After the start of the fellowship, each participant will participate in a research assignment (practicum) at Los Alamos national laboratory (LANL), Lawrence Livermore National Laboratory (LLNL), or Sandia National Laboratories (SNL) working with high-performance computing researchers.

Students must be United States citizens planning full-time, uninterrupted study toward a Ph.D. degree at a U.S. university. Students who are senior undergraduates or in their first or second year of graduate study in computer science are eligible to apply for the DOE high-Performance Computer Science Graduate Fellowship (DOE HPCSF).

For additional information regarding the Krell Institute and the HPCS fellowship, please see the web site http://www.krellinst.org.

Chapter 7. For More Information

For more information about the CSRI, please contact one of the following:

David E. Womble
Computer Science Research Institute, MS1318
PO Box 5800
Sandia National Laboratories
dewombl@sandia.gov
(505) 845-7471 (voice)
(505) 845-7442 (fax)

S. Scott Collis
Computer Science Research Institute, MS1320
PO Box 5800
Sandia National Laboratories
sscoll@sandia.gov
(505) 284-1123 (voice)
(505) 845-7442 (fax)

Deanna R. Ceballos
Computer Science Research Institute, MS1319
PO Box 5800
Sandia National Laboratories
drcebal@sandia.gov
(505) 845-7753 (voice)
(505) 284-4520 (fax)

Bernadette Watts
Computer Science Research Institute, MS1319
PO Box 5800
Sandia National Laboratories
bmwatts@sandia.gov
(505) 844-3936 (voice)
(505) 284-4520 (fax)

Appendix A. Distribution

Internal:

1 MS1318 David E. Womble, 01410
1 MS1320 S. Scott Collis, 01416

1 MS0899 Technical Library, 9536 (electronic copy)


Sandia National Laboratories