SAND REPORT
SAND2006-0001
Unlimited Release
Printed March 2006

Computer Science Research Institute 2003 Annual Report of Activities

David E. Womble, Barbara J. DeLap, 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.osti.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

SAND2006-0001
Unlimited Release
Printed March 2006

Computer Science Research Institute 2003 Annual Report of Activities

David E. Womble, Barbara J. DeLap, and Deanna R. Ceballos
Computer Science Research Institute
Sandia National Laboratories
P. O. Box 5800
Albuquerque, NM 87185-1110

Abstract

This report summarizes the activities of the Computer Science Research Institute (CSRI) at Sandia National Laboratories during the period January 1, 2003 to December 31, 2003. During this period the CSRI hosted 164 visitors representing 78 universities, companies and laboratories. Of these 78 were summer students or faculty members. The CSRI partially sponsored 5 workshops and also organized and was the primary host for 3 workshops. These 3 CSRI sponsored workshops had 178 participants--137 from universities, companies and laboratories, and 41 from Sandia. Finally, the CSRI sponsored 18 long-term collaborative research projects and 5 Sabbaticals.

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, 2003 to December 31, 2003.

1.2. Technical Focus of the CSRI

A number of potential long-term focus areas for the Sandia program in 2003 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

This chapter summarizes the major projects funded by the Computer Science Research Institute. Project proposals can be submitted to the CSRI at any time and are considered for funding by the Executive Board. Criteria for funding projects include technical excellence and quality of the work, impact on the ASCI program, and the strength and breadth of collaborations with Sandia technical staff. Generally these projects are intended to develop new collaborations with university researchers. Also, the work described in these proposals is performed on-site at Sandia National Laboratories, which improves both the strength and breadth of the collaborations. This distinguishes these projects (and the CSRI) from the normal contracting process at Sandia National Laboratories. The projects are listed in alphabetical order by the PI.

1) Scalable Fault Tolerance through Compiler-Driven Communication Induced Checkpointing
   Lorenzo Alvisi, The University of Texas at Austin

2) EMU code for shape optimization
   Dr. Florin Bobaru, University of Nebraska-Lincoln

3) 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

4) Hardware-Accelerated Nonlinear Finite Element Analysis
   Richard H. Crawford, University of Texas at Austin

5) Surrogate-Based Optimization
   John E. Dennis, Jr., Rice University

6) User Friendly, Cache Aware, Parallel PDE Solvers
   Craig C. Douglas, University of Kentucky and Yale University

7) Global-Basis method for fluid problems
   Professor Jacob Fish, Rensselaer Polytechnic Institute
   Haim Waisman, Ph.D. Student, Rensselaer Polytechnic Institute

8) Large-Scale PDE-Constrained Optimization
   Omar Ghattas, Carnegie Mellon University
   Lorenz Biegler, Carnegie Mellon University

9) Harvey Greenberg, University of Colorado at Denver

10) Research in Finite Element Simulations
    Max D. Gunzburger, Florida State University

11) 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

12) System Performance Benchmark Testing, Evaluation and Presentation
    Chu J. Jong , Illinois State University

13) CSRI Executive Board Member
    Deepak Kapur, University of New Mexico

14) Understanding Performance of New PIM-Based Execution Models
    Peter M. Kogge, University of Notre Dame

15) Supporting MPI Collective Communication Operations with Application Bypass
Dr. Panda, Ohio State University
Dr. Wyckoff, Ohio State University
Dr. Sadayappan, Ohio State University

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

17) The Red Summit Project: Petaflops Scale Computing through Processor in Memory Architecture
Dr. Thomas Sterling, California Institute of Technology (Caltech)

18) Developing Pattern Search Methods for Constrained Optimization
Virginia Torczon, College of William and Mary
Robert Michael Lewis, College of William and Mary

Title: Scalable Fault Tolerance through Compiler-Driven Communication-Induced Checkpointing

PI: Lorenzo Alvisi, The University of Texas at Austin

Investigators: Calvin Lin, The University of Texas as Austin
Jeff Napper, The University of Texas at Austin
Thomas Bressoud, Bell Laboratories

Dates: April 25, 2002 – April 24, 2004

CSRI POC: Patty Hough, (925) 294-1518

Project Summary:

We propose to investigate new algorithms for supporting low-overhead fault tolerance for long running, distributed scientific applications. In particular, we propose to derive algorithms that exploit a combination of compiler analysis and program annotations to determine an efficient schedule for check pointing the state of distributed applications.

Motivation: The goal of this research is to develop effective techniques for providing fault tolerance to very large distributed computing systems such as those employed by the ASCI program. The need for fault tolerance in these systems is becoming increasingly clear. For instance, measurements taken for the ASCI Blue Mountain cluster at Los Alamos National Labs indicate that the 6000-node system has a mean-time-to-failure (MTTF) of about 1.6 hours. This figure, which is consistent with the prediction that the MTTF for an off-the-shelf single processor system is about 10,000 hours, is already sufficiently low to jeopardize the usability of the cluster. Barring an unforeseen improvement in the reliability of off-the-shelf processors, it will become practically impossible to exploit the full computing potential of the larger ASCI systems, such as Sandia's Red Storm, expected in the near future. Still, the cpu-hungry users of ASCI machines would find unacceptable any solution that requires dedicating a considerable chunk of processors to achieving fault tolerance (for instance, by using active replication). These users are also unwilling to adopt fault tolerance techniques that require them to rewrite their highly tuned code.

Background: Rollback recovery protocols are ideally suited to provide transparent fault tolerance in a distributed system that needs to tolerate common failures with little overhead in terms of dedicated resources and execution time. Briefly, these protocols record information during process execution that is then used during recovery to restore a crashed process to a state consistent with that of the other application processes. One of the objectives of our research agenda over the past five years has been to increase our understanding of these techniques, using a combination of algorithmic work, system building, and experimental analysis.

In this proposal, we focus on a specific rollback recovery technique, called *communication-induced checkpointing* (CIC). To put the proposed research in context, we first discuss how communication-induced checkpointing compares with other approaches used to save the global state of distributed computations. We then describe the results of some of our previous work, which has provided the first experimental evaluation of the performance of CIC protocols.

1 Checkpoint-Based Rollback-Recovery The simplest form of rollback recovery is based on checkpointing the state of the distributed application. Checkpoint-based rollback-recovery techniques fall into three categories.

Uncoordinated checkpointing allows each process to decide independently when it is most convenient to take a checkpoint. For example, a process may reduce overhead by taking checkpoints when the amount of state information to be saved is small [9,12]. However, there are several disadvantages to uncoordinated checkpointing. First, during recovery, processes must search for a consistent global checkpoint out of the set of individual checkpoints. If this search is unsuccessful, it can ignite a series of cascading rollbacks,

known as the *domino effect* [14], which may cause a substantial loss of useful work. Second, a process may take a *useless* checkpoint, i.e. one that will never be part of a global consistent state, thus incurring overhead without advancing the global state from which recovery can be initiated. Third, each process must maintain multiple checkpoints and periodically invoke a garbage collection algorithm.

Coordinated checkpointing [5,10] requires processes to orchestrate their checkpoints in order to form a consistent global state. Every process recovers from its most recent checkpoint, thereby simplifying recovery and avoiding the domino effect. Also, coordinated checkpointing requires only one permanent checkpoint per process on stable storage, reducing storage overhead and eliminating garbage collection. However, processes must incur the overhead of achieving coordination, and all processes send their checkpoints to storage at approximately the same time. The latter characteristics raise serious doubts about the scalability to thousands of processors.

Communication-induced checkpointing (CIC) protocols [2,3,6] avoid the domino effect without requiring all checkpoints to be coordinated. Processes take two kinds of checkpoints: *local* checkpoints can be taken independently, while *forced* checkpoints guarantee the progress of the recovery line by preventing useless checkpoints. Unlike coordinated checkpointing, CIC protocols do not exchange coordination messages but instead piggyback information on each application message. This information is used to determine if communication and checkpoint patterns can lead to the creation of useless checkpoints, and a forced checkpoint is taken to break these patterns. This intuition has been formalized in an elegant theory based on the notion of *Z-cycles* [11]. A key result is that a local checkpoint is useless if it is involved in a Z-cycle [7,11].

CIC protocols are believed to have several advantages. First, they allow processes notable autonomy in deciding when to take checkpoints. Second, they are believed to scale up well with the number of processes since they do not require global checkpoints. Finally, CIC protocols never generate useless checkpoints. Unfortunately, there is a price to pay for these advantages. In order to prevent useless checkpoints, the information piggybacked on application messages occasionally induces the processes to take forced checkpoints before they can process the messages. Second, processes have to pay the overhead of piggybacking information on top of application messages, and they also need to keep several checkpoints on stable storage.

2 An Analysis of Communication-Induced Checkpointing The above discussion suggests that CIC protocols promise to scale to systems with thousands of nodes. Unfortunately, the literature presents only a qualitative analysis of CIC protocols, leaving open the question of how these protocols truly perform in practice. To shed some light on this question, we have performed the first experimental analysis of CIC protocols [1].

Our study compared the performance of three CIC protocols [2,3,6] for a set of applications in the NAS NPB2.3 benchmark suite [4]. Our analysis has shown that the theoretical advantages of CIC protocols do not materialize in practice. In particular, our results show the following:

1. CIC protocols that use an eager approach to preventing Z-cycles by taking forced checkpoints whenever they suspect the formation of a Z-cycle perform worse than lazy protocols that use a time-stamping function to prevent a Z-cycle at the last second.
2. CIC protocols do not scale well with a larger number of processes. We have found that the number of forced checkpoints increases almost linearly with the number of processes.
3. A process takes at least twice as many forced checkpoints as local ones. Therefore, the touted benefit of allowing processes to take independent checkpoints does not materialize.
4. There is a considerable unpredictability in the way CIC protocols behave. The amount of stable storage required, performance overhead, and number of forced checkpoints depend greatly on the number of processes, the application, and the communication pattern. This unpredictability makes of CIC protocols more cumbersome than other alternatives.
5. A successful placement policy of local checkpoints must be dynamic, must account for forced checkpoints, and must adapt to changes in the application behavior.
6. CIC protocols seem to perform best when the communication load is low and the pattern is random. Regular, heavy load communication patterns seem to fare worse.

Proposed Work: We propose to explore a new approach to checkpoint-based rollback recovery that we believe can yield in practice the theoretical advantages that traditional CIC protocols fail to deliver.

The principal reason for this failure is that CIC protocols are prone to generating a large number of forced checkpoints. In these protocols, a process determines whether to take a forced checkpoint on the basis of information piggybacked on the application messages that it receives. This information however is insufficient to provide the recipient with the knowledge necessary to make an optimal decision. In particular, the piggybacked information does not allow the recipient to rule out communication patterns that would lead to Z-cycles. As a result, recipients conservatively take forced checkpoints that would not be otherwise necessary, negatively impacting performance. A second negative consequence of frequent forced checkpoints is that they nullify CIC's promise of maintaining the advantages of uncoordinated checkpointing.

To address these problems, we propose to develop algorithms that allow a source-to-source compiler to determine an efficient checkpointing schedule. This approach can potentially result in significant advantages. For example, the compiler can build an internal representation of the communication patterns exhibited by an application and use it to place checkpoints appropriately in order to build a consistent recovery line and prevent the domino effect. This can yield the same advantages of coordinated checkpointing (easy garbage collection, minimal storage requirements) but without the coordination overhead and reduced scalability. In particular, this approach would allow skewing the time at which different processes checkpoint their state, reducing the risk of overwhelming the I/O to stable storage. Also, the compiler can try to build a "smart" recovery line by scheduling local checkpoints at the most appropriate times in the execution of each the application's processes. To help the compiler, programmers could also annotate their program with hints that indicate times in the execution of a process where checkpointing would be most convenient.

The steps towards realizing this vision are the following:

- develop an internal representation of the *space-time diagrams* [8] used to express the events that occur in the execution of a distributed application,
- use this representation to develop algorithms that build recovery lines by scheduling independent checkpoints in the execution of the application processes,
- define criteria to determine the most preferable recovery line among a set of alternative ones and develop algorithms based on these criteria,
- instantiate these algorithms in a source-to-source compiler that will automatically insert calls to *Egida* [1, 13], a library for low-overhead fault-tolerance developed at UT Austin, into an application and demonstrate on a Sandia application code.

SNL Collaboration: This research will be conducted in collaboration with a number of technical staff members at Sandia. Curtis Janssen and Matt Leininger will participate in coupling software with the Massively Parallel Quantum Chemistry (MPQC) code and developing MPI-based tools. Mike Goldsby and Edward Walsh will participate in characterizing Sandia applications as well as in developing and implementing the software interfaces. Lee Ward will provide hardware and OS expertise for Sandia's ASCI machines. Patty Hough and Vicki Howle will be developing numerical algorithms that take advantage of the tools developed.

Most of the visits will take place at the California site; however, at least one visit will be to the New Mexico site in order to establish new collaborations with others interested in fault tolerance. In particular, we intend to develop working relationships with researchers in 9200 and with the developers of ASCI finite-element applications such as ALEGRA and SIERRA. The work conducted under this proposal will be incorporated into software tools that will be made available to Sandia and others in the ASCI community.

References

[1] L. Alvisi, E. N. Elnozahy, S. Rao, S. A. Husain, and A. De Mel. An analysis of communication-induced checkpointing. In *Proceedings of the IEEE Fault-Tolerant Computing Symposium (FTCS-29),* pp. 242-249, Madison, WI, June 1999.

[2] R. Baldoni, F. Quaglia, and B. Ciciani. A vp-accordant checkpointing protocol preventing useless checkpoints. In *Proceedings of International Symposium on Reliable Distributed Systems,* pp. 61-67, West Lafayette, IN, October 1998.

[3] D. Briatico, A. Ciuffoletti, and L. Simoncini. A distributed domino-effect free recovery algorithm. In *Proceedings of the IEEE International Symposium on Reliability Distributed Software and Database,* pp. 207-215, December 1984.

[4] NASA Ames Research Center. NAS parallel benchmarks. http://science.nas.nasa.gov/Software/NPB/1997.

[5] E. N. Elnozahy, D. B. Johnson, and W. Zwaenepoel. The performance of consistent checkpointing. In *Proceedings of the Eleventh Symposium on Reliable Distributed Systems,* pp. 39-47, October 1992.

[6] J. M. Hélary, A. Mostefaoui, R. H. B. Netzer, and M. Raynal. Preventing useless checkpoints in distributed computations. In *Proceedings of IEEE International Symposium on Reliable Distributed Systems,* pp. 183-190, 1997.

[7] J. M. Hélary, A. Mostefaoui, and M. Raynal. Virtual precedence in asynchronous systems: concepts and applications. In *Proceedings of the 11th Workshop on Distributed Algorithms.* LNCS press, 1997.

[8] L. Lamport. Time, clocks, and the ordering of events in a distributed system. *Communications of the ACM,* 21(7):558-565, July 1978.

[9] C. C. Li and W. K. Fuchs. Catch: Compiler-assisted techniques for checkpointing. In *Proceedings of the 20th International Symposium on Fault-Tolerant Computing,* pp. 74-81, 1990.

[10] G. Muller, M. Hue, and N. Peyrouz. Performance of consistent checkpointing in a modular operating system: Results of the ftm experiment. In *Proceedings of the First European Dependable Computing Conference (EDDC-1),* pp. 491-508, October 1994.

[11] R. Netzer and J. Xu. Necessary and sufficient conditions for consistent global snapshots. Technical Report 93-32, Department of Computer Sciences, Brown University, July 1993.

[12] J. S. Plank, M. Beck, and G. Kingsley. Compiler-assisted memory exclusion for fast checkpointing. *IEEE Technical Committee on Operating Systems Newsletter, Special Issue on Fault Tolerance,* pp. 62-67, December 1995.

[13] S. Rao, L. Alvisi, and H. Vin. Egida: An Extensible Toolkit For Low-overhead Fault-Tolerance. In *Proceedings of the 29th Fault-tolerant Computing Symposium (FTCS-29)*, Madison, Wisconsin, June 1999, pp. 48-55.

[14] R. B. Strom and S. Yemeni. Optimistic recovery in distributed systems. *ACM Transactions on Computer Systems,* 3(3):204-226, April 1985.

Title: EMU code for shape optimization

PI: Dr. Florin Bobaru, Department of Engineering Mechanics, University of Nebraska-Lincoln

Dates: July 6 – August 23, 2003

CSRI POC: Stewart Silling, (505) 844-3973

Project Summary:

My research interests are in the general area of computational mechanics. More specifically, I am interested in developing numerical algorithms for optimization of materials and structures. I have recently shown the benefits one can obtain in shape optimization of elastic and thermoelastic bodies if, instead of the classical Finite Element Method, a meshfree approach is used. I am interested in new algorithms that would push the current limits in optimal design of materials and structures. Combining shape and material optimization, or shape and topology optimization are issues I am currently considering. With powerful algorithms for optimization, one would expect breakthroughs in designing individual components of complex systems ranging from new paradigms for MEMS design to improved designs of large scale systems such as better thermal protective shields for space shuttle or damage tolerant armors. Nature has been perfecting and evolving systems of higher and higher complexity to adjust to challenging environments. My goals are to find algorithms that would use the fundamental minimum principles of mechanics and physics to produce designs of comparable efficiency and versatility as nature's answers to complex environments.

Research Proposal: This project plans to advance the methods of shape design sensitivity analysis to problems involving regular boundaries subject to singular transformations, and singular boundaries under regular or singular transformations (see Figure 1). The peridynamic theory developed by Dr. S.A. Silling of Sandia will be the analysis tool instead of the classical theory of elasticity.

The boundaries to be considered are either interior boundaries (that may represent cracks, material interfaces, phase changes, etc.) or outer boundaries. Due to this generality of the formulation, the same framework can be used to treat shape optimization problems as well as material optimization of two or multi-species composites (searching for the best shape of inclusions in an elementary volume element by varying the interface).

The problem of calculating sensitivities of a response functional to singular transformations of the boundary (inner or outer, regular or singular) of a body subject to mechanical loading has not been studied so far due to intrinsic difficulties of treating singular boundaries and discontinuous fields in formulations of boundary value problems. In particular, the classical formulation of linear fracture mechanics gives rise to the well-known square-root singularity of the stress field. Clearly, sensitivities of singular boundary (the crack for example) shape variations cannot be addressed in the classical setting.

The problem of calculating sensitivities of a response functional to singular transformations of the boundary (inner or outer, regular or singular) of a body subject to mechanical loading has not been studied so far due to intrinsic difficulties of treating singular boundaries and discontinuous fields in formulations of boundary value problems. In particular, the classical formulation of linear fracture mechanics gives rise to the well-known square-root singularity of the stress field. Clearly, sensitivities of singular boundary (the crack for example) shape variations cannot be addressed in the classical setting.

The peridynamic formulation in elasticity eliminates the difficulties of the classical continuum theories when dealing with spatial derivatives of discontinuous fields. In this formulation, long-range forces are represented by a pairwise force function with compact support acting between two nodes of the discretization. Cusp-like profiles are obtained for the opening of a crack and infinite stresses at the crack tip are eliminated. The ad-hoc assumptions that were proposed in the past to correct the unphysical infinite stresses from the classical linear fracture mechanics theory are now no longer needed. The same model is used in the peridynamic formulation on and off the discontinuity present in the fractured elastic body. The meshfree character of the peridynamic theory makes it extremely suitable for numerical treatment of shape optimization problems based on sensitivities.

We propose to develop analytical and numerical sensitivities for singular transformations in elasticity using the peridynamic formulation. Based on these sensitivities it will be possible to develop shape optimization of elastic bodies with singular boundaries including bodies with cracks, notches, cusp-like shapes, etc. At the same time, material optimization of composite (searching for optimal shapes of inclusions) or Functionally Graded Materials (finding the optimal material gradation) will be possible since the boundaries in our formulation are both external and internal. Another possible application would be the sensitivity and uncertainty estimation for elastic bodies with large numbers of random cracks.

Collaborations: The current proposal is seeking collaboration with the author of the peridynamic theory, Sandia's Dr. Stewart A. Silling. Dr. Silling is also the primary developer of the EMU code that implements the peridynamic theory. Dr. Silling is strongly supporting the current proposal and has provided a supporting letter for a proposal submitted by the author of this proposal to the DOE Early Career PI Program. EMU is a meshfree method designed to model cracks, discontinuities, and other singularities that may exit in a body initially or emerge as a result of deformation. In addition, the EMU code is able to represent interfaces between materials by modifying the bond properties between the discretization points. Modeling complex heterogeneous materials is one of the areas in which the peridynamic theory seems to have a great potential. We propose to couple the sensitivity calculations for singular transformation to the EMU code and build an optimal design code based on the peridynamic formulation. Ideally, when completed, the peridynamic code and the new sensitivity formulations will interface and enhance the capabilities of the DAKOTA system developed at SANDIA.

Length of Participation and Location of Visit: The proposed length of participation is for one-month summer faculty for three years at the Sandia National Laboratories, Albuquerque, NM. The starting date for our collaboration is 1st of August 2002. The visit during the 2002 summer will be focused on familiarization with the numerical implementation of the peridynamic theory, the EMU code. The basics ingredients of the sensitivity analysis with singular transformations for domains with singularities will be developed. For the remaining part of 2002 and before the summer of 2003, our collaboration will continue in various forms: Dr. Silling will be invited to present a seminar at the Department of Engineering Mechanics at University of Nebraska-Lincoln in the Fall of 2002. The author of the current proposal has a Master student who will be working to develop the computer programs for a proper coupling of the shape sensitivity analysis into Sandia's EMU code. Dr. Silling will be invited to participate in the graduate committee of the Master student.

The second visit (June 2003) will focus on the development of a series of optimization algorithms for shape and material optimization based on the developed sensitivity programs. We will test peridynamic formulation against some classical problems in shape optimization and, moreover, we will attempt solutions to singular problems never considered before. Parallel versions for the sensitivity calculations will be advanced. The third visit (August 2004) will be dedicated to large-scale simulations involving hundreds and possibly thousands of design variables and the assessment of the benefits of the peridynamic formulation in optimal design of complex systems (complex geometries and complex material behavior). Problems of different scales will be considered. Comparisons with results provided by the DAKOTA system will be made.

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

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

Investigator: Kenin Coloma, Northwestern University

Dates: August 16, 2003 – August 15, 2004

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

Project Summary:

Background: 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, 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.

FY04 Objectives:

Objective 1. Optimizations for Persistent File Domain (PFD)
Investigate enhancements to Persistent File Domains (PFD). MPI_File_set_view() contains useful information that potentially can be used to further optimize the I/O performance. Since MPI_File_set_view() is collective, the file visible domain to one process can also be known by others in which the opportunities of I/O optimizations are expected. Contractor will investigate the use of this information together with the data access extent to determine an optimal PFD in each. Contractor will investigate dynamical re-assignment of PFD to reflect the change of the access patterns issued from the current I/O calls. Contractor will investigate I/O performance by means of MPI-IO file access.

Objective 2: Extend PFD Techniques for MPI Atomicity
Contractor will investigate incorporating the implementation for MPI atomicity into the PFD framework. Many file systems perform client-side caching together with read-ahead and write-behind strategy in order to speed up the sequential access to the file data. The size of cache buffer is usually in the unit of file blocks. A concurrent overlapping I/O request can easily result in the same file blocks cached in multiple clients where the cache coherence problem emerges. A traditional solution to this is through using the file locking mechanism to bypass the write-behind, so the file modification is visible to all cached clients. However, using file locking can seriously degrade the I/O parallelism as demonstrated in our previous work. We will extend the PFD work to cover this problem. Basically, our idea is to have an I/O request performed through the clients that cached the requesting data. Since file locking is not required in this approach, the degree of I/O parallelism can be maintained.

Objective 3: Design Cache Coherence Solution for MPI Non-collective I/O
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. Contractor will investigate extending the PFD approach to MPI non-collective I/O calls. Contractor will investigate PFD assignment at file open and set view since these two MPI calls are synchronous. For the implementation of non-collective I/O requests, contractor will investigate two approaches: using MPI one-side communication and using a multi-thread implementation. Contractor will also evaluate the impact of using different file block sizes to the concurrent overlapping I/O performance.

FY04 Tasks:

First quarter:
Develop a bookkeeping approach for tracing the change on the MPI fileview and data request size (data type for I/O buffers.)
Implement the PFD re-assignment.
Implement the trigger condition for PFD re-assignment.
Collect potential file hints that can be used to fine-tune the persistent file domain partitioning.
Determine which ROMIO codes should incorporate this optimization.
Evaluate I/O benchmarks to examine the effect of PFD implementation.

Second quarter:
Incorporate the MPI atomicity implementation to the PFD framework.
Study MPI one-side communication.
Design techniques that use the PFD concept with MPI one-side communication.

Identify the ROMIO code fragments for inserting our implementation.

Third quarter:
Implement the PFD for MPI non-collective.
Study the data access patterns of a few parallel scientific applications that can potentially benefit.
Study potential file hints, either provided from users or generated from the data types used in the program, for determining proper file domain partitioning method.

Fourth quarter:
Examine the connection of modified interfaces between ROMIO and MPI-IO to ensure all MPI-IO applications perform without changing their MPI-IO calls.
Benchmark for regular and common access patterns for PFD on non-collective I/O operations.

FYO4 Deliverables:
Quarterly Status Report and a Final Report

FY05 Objectives:

Objective 1: User-Space Cache Coherence Solution for MPI Non-collective I/O
As mentioned in the FY04 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:
Design MPI-level caching management policy for non-collective MPI-IO optimizations
Set up Lustre environment and study its native interface
Design the group pipelined collective I/O algorithms for MPI-IO

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

Third quarter:
Implement first MPI-IO version on top of Luster file system
Evaluate MPI-level caching for non-collective I/O
Evaluate group pipelined collective I/O using different access patterns

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

FYO5 Deliverables:
Quarterly Status Report and a Final Report

Title: Hardware-Accelerated Nonlinear Finite Element Analysis

PI: Richard H. Crawford, University of Texas at Austin

Dates: May – September 2003

CSRI POC: David Thompson, (925) 294-6588

Project Summary:

1 Background and Motivation: Sandia has developed finite element analysis code that models fields with nonlinear basis functions (i.e. quadratic instead of linear). Although these codes exist, they are not in wide use because analysts do not have a wide array of tools for dealing with them. Particularly lacking in this respect are visualization aids.

Our goal is to add support into VTK for nonlinear finite elements in general and quadratic hexahedra and tetrahedra in particular. For each element type, we will provide boundary surface rendering (where both surface deflection and color are defined by field values) and isosurface rendering of these elements. Volume rendering is a topic for future work but not included here due to time constraints.

Although we will be focusing on quadratic elements, the interpolant is easily extended to higher orders and other elements. The interpolant is of the form

$\phi(\mathrm{r}) = \Sigma\ \phi_i N_i\ (\mathrm{r})$

where ϕ is the scalar field, ϕ_i is the value of the scalar field at the i-th node – a result of the finite element simulation, N_i is the shape function associated with the i-th node, and $\mathrm{r} =< r, s, t >$ is the set of parameterspace coordinates. Here, ϕ may be any scalar field including x, y, and z, the model-space coordinates of all points in the element. ϕ is also used to represent deflections δ_x, δ_y, δ_z and any arbitrary field used to color the element boundary or form a level set.

The shape functions use a tensor product Lagrange interpolant of the form

$N_i(r) = T_i(r)T_i(s)T_i(t)$
With

$$\mathrm{Ti}(x) = \begin{cases} (1 - 2x)(1 - x) & \text{node at } x\text{-axis origin} \\ 4x(1 - x) & \text{middle node on } x\text{-axis} \\ x(2x - 1) & \text{node furthest from } x\text{-axis origin} \end{cases}$$

2 Progress To The Present: Last summer, Rahul Khardekar, a graduate student at the University of Texas at Austin working with Dr. Rich Crawford, visited Sandia as a prelude to Dr. Crawford's intended visit this summer. During that time, we developed a technique to render element boundary surfaces using recent graphics hardware. The advantage is not only fast processing, but per-pixel coloring based on the field value. That research has lead to a paper (Khardekar & Thompson 2003) to be presented in February 2003.

We have continued to fund Rahul during the Fall semester and he has made considerable progress into the much larger problem of locating and parameterizing isosurfaces in nonlinear elements. The problem of isosurfacing nonlinear elements is two-fold:

• *Element range*. Because we know only nodal values for an element and because Lagrange interpolants are not convex combinations of the input values, there is no simple way to determine the range of scalar values inside an element.
• *Isosurface tracking*. Once the range of the interpolant is known, a starting point on each component of the level set must be located.

To find the range of the interpolant over the domain of an element, we must find where the Jacobian of ϕ vanishes; these points are the maxima and minima of the scalar field. This requires the solution of 3 nonlinear equations in 3 unknowns for each element. There may be up to 6 solutions for a quadratic element. There are several ways to accomplish this. Mann & Rockwood (2002) have used the divergence theorem to numerically solve for both isolated and non-isolated critical values. Another technique that is more numerically robust but slower is homotopy. Since this may be considered a preprocessing step to visualization and storing critical points with each element does not require much space, we plan to use homotopy continuation techniques. This is an embarrassingly parallel process; the critical points of each element are independent of all others.

Because we are searching for minima and maxima over a bounded domain, we must also search the boundaries of each element for extrema. These searches take the same form as above, but with one or two parameter-space coordinates held constant. By storing these extrema along with each cell we can easily determine whether there are any isolated components of a level set inside an element.

Once the extrema are identified, we are prepared to find level sets inside elements. Seed sets or contour trees will be used to quickly identify elements that intersect a level set but we must still generate starting points on each component of the level set.

After starting points have been identified, the level set can be parameterized for rendering or further postprocessing.

We hope to explore both adaptive triangulation and higher-order parameterizations. However, exact parameterizations are not guaranteed to exist and even if they did, they would be of exceedingly high degree.

3 Proposed Work: We hope to continue Rahul's work this summer by comparing several numerical and analytical techniques we have identified for finding critical points of a scalar field and parameterizing segments of isosurfaces. We also hope to extend our hardware-accelerated per-pixel rendering technique to arbitrary interior surfaces of elements using the next generation of video cards coming out in Q1 of CY2003. The deliverable will be a set of VTK classes embodying the algorithms we have developed and a paper ready for submission to an archival quality journal. Dr. Crawford is an outstanding professor in mechanical engineering at the University of Texas at Austin who served a stint at the Purdue CAD Lab as a graduate student alongside the likes of Warren Waggenspack (who was a visited Sandia for a summer recently). His curriculum vitae is attached along with a copy of the paper on the completed work.

3.1 Deliverables

- A VTK module that supports isosurface computation and accelerated rendering of 27-node hexahedra and 10-node tetrahedra.
- A paper ready for submission to an archival-quality journal. Applicable journals might be CAD, Computer Aided Geometric Design (CAGD), or the Journal of Graphics Tools.

3.2 Schedule: Our timetable for the completion of this work is given below:

Date	Event
Jan 03	nVidia GeForceFX released
Jan-May 03	Continuing remote collaboration with UT Austin
Jan-Apr 03	Implementation of critical-point preprocessing
Jun-Aug 03	Design architecture of VTK implementation
May-Sep 03	Dr. Crawford and student visit Sandia
May-Jun 03	Implementation of accelerated isosurface extraction
Jun-Aug 03	Benchmark and optimize
Jun-Aug 03	Finalize VTK architecture
Aug-Sep 03	Write and submit paper
Aug-Sep 03	Full VTK Implementation

The nVidia GeForceFX will be the line of video cards we will target for hardware acceleration – Khardekar & Thompson 2003) describes why the current generation of video cards will not work for per-pixel-accurate rendering.

The preprocessing work should be completed during the spring semester and will become Rahul's master's thesis. Once critical points of the finite element function have been identified, the isosurfacing code may be implemented. There are several existing approaches to this work and we hope to have time to compare several, although time may not permit this. Our first approach will be the continued use of adaptive triangulation. However, if an algorithm that generates nonlinear surface patches proves fast enough, it would be preferable since more accurate normal and curvature information is retained. This information is critical to feature identification and tracking techniques that are in development at Sandia as part of the Scientific Data Management (SDM) project.

Once the software prototype is complete, we will benchmark and optimize the result. At the same time, the design of the full implementation into the VTK pipeline will be finalized.

Title: Surrogate-Based Optimization

PI: J.E. Dennis, Jr., Rice University

Dates: July 1, 2003 – June 30, 2004

CSRI POC: Paul Boggs, (926) 294-4630

Project Summary:

The purpose of this proposal is to fund the PI's interaction with Sandia National Laboratories at NM and CA. These interactions should be of two sorts. First, algorithmic advances and interesting practical experience of the PI and his collaborators in universities and industry would be transferred to the labs, and conversely, lab problems and experience would inform the research of the PI and his collaborators. Bidirectional technology transfer is our goal and firm belief. We seek to be both brokers and developers; to provide support and receive feedback. The point of the visits for which support is requested is to increase the total pool of knowledge about engineering design problems of the particularly nasty sort discussed below.

The class of optimization problems that we target is largely from engineering design and widely regarded by practicing designers as intractable. We have focused for several years not on solving problems better, but on solving problems that can not be solved at all well by existing methods. We have resisted the temptation succumbed to by some university engineering groups of designing elaborate (soft) design support systems based on naive expectations of the spontaneous generation of supporting algorithms. Instead, we have focused on providing the rigorous mathematical infrastructure to underpin design support systems such as Boeing's Design Explorer and Sandia's DAKOTA, though to date, we have had measurable impact only on Design Explorer, where our influence has been considerable. Dr. Greg Shubin, Head of Mathematics and Engineering Analysis, or Dr Evin Cramer, Technical Fellow of Boeing Phantom Works can verify our contributions. <shubin or cramer>@redwood.rt.cs.boeing.com.

Another tenet of our research is that engineers generally understand their problems better than we can, and so while we do provide default choices, all our work has been directed by the rule that engineering intuition should be able to be seamlessly incorporated into our algorithmic framework. We take over only when the user's approach has reached the limit of its resolution, and we proceed to "clean up" the final design or to get the designer out of a rut fallen into by conventional approaches.

Our algorithms are carefully structured so that if a user has a favorite method for solving the surrogate problem, then it can be provided to our FOCUS software as a SEARCH procedure in a direct search method. Thus, we provide direct search methods to act as meta-algorithms supporting user favored heuristics. So far, we have the algorithms and software in place to deal with general nonlinearly constrained optimization for problems of the type we see often in engineering design:

- f(x), C(x) are expensive and have few correct digits. This happens because there is an underlying expensive state equations that needs to be solved for a given x to generate the state variables that are used to evaluate the optimization problem functionals.
- Evaluation may fail (f(x)=infinity) expensively and unexpectedly for feasible x and f and C may be discontinuous even when they are defined. This happens because the state equation solver mentioned above is often a kludgey coupling of table lookups and state equation solvers from different physical disciplines. For example, one might use successive replacements to try to couple a structures solver with a dynamics solver. This procedure may work for some x, but it may not for one nearby. In a Boeing helicopter rotor design problem, the function failed to return a value about 67% of the time. Someone unfamiliar with commercial reality might suggest that the solvers should be tightly coupled to smooth out this behavior. Indeed, a part of the economic justification for making such an effort can derive from a proof of concept we might provide by the techniques suggested here. Before bandoning legacy solvers to reimplement a solver more amenable to SAND type approaches, try this approach to estimate the economic gains.

- Evaluation will usually fail unless some selected simple constraints hold. This happens when there are simple bounds, like nonnegativity, on some of physical parameters. It means that we must be able to be always feasible with respect to some constraints, though for efficiency, we would like to allow infeasibilities during the course of the interaction in any constraints for which that is reasonable.

In addition, since we are usually dealing with interpolatory surrogates like polynomials, (kriging) splines, or neural nets, there is a fairly low limit on the number of decision variables we can treat. However, this is a problem with constructing interpolants more than with the underlying approach. Thus, if the surrogates come from a less detailed description of the underlying physics, then we can treat more decision variables. We are aiming for an ExxonMobil problem with a couple hundred each discrete and continuous variables, and for which the surrogate is based on a simplified model of flow-in-porous media. Still, the largest problem we actually have solved was 256 continuous decision variables.

To meet these challenges of our problem class, we have been able, for the very nice Lewis and Torczon barrier algorithm (f(x) is infinite if x is not feasible), to use a crucially different proof to strengthen their convergence results for nonlinear optimization under simple linear constraints by dropping their assumption of continuous differentiability on an open set containing the feasible region. Indeed, if the initial iterate has finite problem function values, then we show the existence of a limit point even when the functions are discontinuous and extended real valued. If the problem is locally smooth at such a point, then we show appropriate optimality conditions from the Bourbaki/Clarke nonsmooth analysis. In addition, the proofs based on our weaker hypotheses are much shorter and simpler.

Our general constrained algorithm FPS adapts Fletcher's new penalty function-free filter step acceptance idea to direct search methods. Thanks to our improved convergence analysis mentioned above, we can allow discontinuities and extended values in the function to which the algorithm is applied, Thus, we can apply our NLP filter method directly to the Lewis and Torczon barrier function and thus enforce every iterate feasibility for simple linear constraints with no need for extra complication in the theory. This FPS algorithm is being used at Boeing in the current release of the Boeing Design Explorer package. It had a great success in wing planform design.

The focus of our efforts supported by research funds will be to:

- **Implement our mixed discrete/continuous variable algorithm MVP in FOCUS during the next 3 years.** This is the longest-term goal of the proposal. The difficulty lies in the fact that our main focus is on categorical discrete variables and nonlinear nonconvex problems. These are very common in engineering design, and they are usually handled by heuristic parameter studies. For our purposes, a categorical variable is a variable, which is an input to the simulation driving the optimization, but it has the property that if its input value is not from a prescribed finite set, then the simulation will not run and the function or the constraints cannot be evaluated. This property precludes the use of continuous relaxations for discrete variables, and hence branch and bound. Our MATLAB MVP has been quite successful on some artificial test problems as well as on a problem from cryogenic engineering design from a mechanical engineer at the Automotive Research Center at Ann Arbor. In that problem, MVP increased or decreased the number of heat intercepts and changed the neighboring insulation material of as the iteration proceeded. This means that number of heat intercepts, which determines the number of optimization variables was itself an optimization variable. We were able to obtain a 60% reduction in the require power by our approach over the best published result that chose the categorical variables by a parametric study. We regard this as highly promising behavior. We continue work on the algorithm. The software difficulty is in designing a user interface to capture the user's domain specific knowledge concerning the relationships between variables like different insulation materials.

- **Extend our MVP work to generally constrained mixed variable problems**. We have long advocated using the Ferris-Mangasarian variable distribution techniques to extend our algorithms to higher dimensions. Those techniques can be thought of as an adaptive approach to the linking variables long used by design engineers and explained in the book by Vanderplaats.

- **Continue development of multiobjective filter based tools to support engineering decisions on trading off various competing objectives**. This is the holy grail. Our preliminary MATLAB experiments indicate that the surrogate/filter approach may extend current capabilities, but there is much to do here. We would like to provide a surrogate based representation of the trade-off surfaces, which we would refine as the user homes in on the interesting region.

The visits supported by this funding would aim to:

- **Work with Bill Hart on incorporation of evolutionary approaches to SEARCH for all our algorithms**. It is important to provide effective default SEARCH procedures for casual and evaluative use or for naive users. Bill Hart's work seems very well suited in robustness and general applicability to this application. The way our framework is designed, we give the user the opportunity to embed their own optimization procedures to be applied to the surrogate to identify promising candidates for improved designs. This is an important feature of our approach. Many designers have ad hoc approaches.

- **Work with Mike Eldred on algorithmic enhancements to DAKOTA**.

- **Work with Juan Meza, Paul Boggs, Patty Hough, and Tammy Kolda on enhancements to the PDS/trust-region algorithm**. Specifically, we advocate the use of the local quasi-Newton quadratic model as a surrogate, and we suggest the restriction of the SEARCH based on this surrogate to a lower dimensional subspace as in the work of Byrd, Schnabel and Schultz or in the work of Boggs, Kearsley, and Tolle.

Title: User Friendly, Cache Aware, Parallel PDE Solvers

PI: Craig C. Douglas, University of Kentucky and Yale University

Dates: October 1, 2003 – September 30, 2004

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

Project Summary:

This project is joint work with Jonathan Hu, Ray Tuminaro, and Jaideep Ray of Sandia/California. Danny Thorne, a graduate student at the University of Kentucky, is also involved. Danny is supported for 10 months at Kentucky through ASCI funds and also spends two months a year as a Sandia summer student. The codes we are producing are being tested on combustion problems that Jaideep Ray is providing. We are also interested in collaborations with other Sandia projects.

Many cache aware algorithms have several shortcomings. Implementing them is intrusive to already existing codes and usually leads to ugly, hard to maintain spaghetti style programs. In addition, the modified codes are usually not portable, even to upgraded CPUs in the same processor line. Even new codes are hard to construct and maintain. All of these bad features are normally independent of the type of mesh structure that is used.

A great deal of progress has been made in eliminating these shortcomings for typical academic research codes. Part of the work leads to codes that are two to three times faster than traditional codes for a variety of architectures, but are not optimal for any particular processor/cache combination. The approach is to minimize the number of parameters to one (usable cache size) and to live with a code that is much faster, i.e., one that is "good enough, but portable" instead of perfect. In short, a small amount of inefficiency goes far in making a cache aware code useful and easy to tune.

Producing cache aware codes for adaptively refined meshes is hard work. Producing one that is user friendly is much harder. Having it work on p processors, p=1,2,…, is harder still. There are many issues that have to be resolved: data structures, recognizing and taking advantage of local structure in a grid, domain substructuring, dynamic load balancing, grid management, error estimation, parallel I/O, and (lastly) interacting with a real simulation code.

Luckily Sandia already has a number of the pieces that are needed (e.g., the Zoltan package for dynamic load balancing). What we are producing is not necessarily a standalone code, but one that can be used by many different software packages and applications inside of Sandia.

At the suggestion of Jaideep Ray, we have used the software package Grace to do the grid generation and load balancing. The code is very effective when the number of elements in a grid patch is a power of two. Unfortunately, the code runs poorly when this assumption is not met. Our primary test examples do not fit this assumption. Portions of Grace may have to be replaced, either with code from a Sandia package or replacement code that we write. Using an existing Sandia code is the first choice.

During the four weeks that Craig Douglas spends at Sandia, we expect to start with running codes for both 2D and 3D parallel PDE solvers for adaptively refined grids. We hope to integrate the codes into Jaideep Ray's combustion code and gain experience from how effective it is. If possible, we would like to integrate the code into Ray Tuminaro's ML package.

One of the possible outcomes of the four weeks will be a new version of Jonathan Hu's academic code from his graduate school career. This is an unstructured and quasi-unstructured grid cache aware multilevel solver for two and three dimensions. The original code works with either scalar or coupled elliptic PDEs. The academic code requires Hu in order to set it up, which is not optimal. We will look into producing a new code in C and C++ that takes advantage of the work that is ongoing.

Title: Global-Basis method for fluid problems

PI: Professor Jacob Fish, Rensselaer Polytechnic Institute
Haim Waisman, Ph.D. Student, Rensselaer Polytechnic Institute

Dates: May 2 – August 20, 2003

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

Project Summary:

Research efforts will focus on (i) implementation of the Global-Basis method in SALSA and ML-pack and (ii) validation of the Global-Basis method on fluid problems to be provided by Sandia. The graduate student, Haim Waisman (US Permanent Resident), will be closely working with Ray Tuminaro at Livermore on ML pack issues and John Shadid on issues related to SALSA, fluid flow problems and code validation.

Haim will be faced with the following three tasks:

1. Implementation of a complex prolongation operator
2. Construction of hybrid prolongation operator consisting of combination of local and global basis vectors
3. Implementation of the Global-Basis method as a filter to the existing multilevel and single level methods within SALSA and ML-pack

Task 1: The smoothing iteration matrix is nonsymmetric and for indefinite system the spectral radius is greater than one. Therefore, the optimal prolongation operator spanning the space orthogonal to the smother may be complex. Currently, ML-pack only works with real numbers. Thus one of Haim's talks would be to work with Ray Tuminaro on generalizing the framework of ML-pack to allow for complex algebra.

Task 2: One of the multilevel versions of the proposed Global-Basis methods consists of expanding the sparse prolongation matrix based on the local basis (for example obtained from the Smoothed Aggregation method) with global basis vectors obtained from the Arnoldi method (AR-pack). This will require development of coarse level matrices from a mixture of sparse (based on local basis) and dense (based on global basis) prolongation and restriction operators.

Task 3: This task calls for the implementation of a filter within or on top of the ML-pack. Most of the local basis multilevel methods will fail for highly indefinite systems. In this implementation Global-Basis method will detect the modes, which cause divergence of the iterative method of choice and will construct an additional level that will filter out the "bad" modes.

This is obviously a very ambitious program and we do not expect that within the three months of the summer program that Haim will be able to successfully complete all the goals stated in the three tasks. We hope however, that during his stay he will be able to make significant progress on the three fronts by working together with John and Ray, so that when he returns to RPI later in the Fall semester in 2003 he would be well positioned to successfully complete all the tasks.

Title: Large-Scale PDE-Constrained Optimization

PI: Omar Ghattas, Carnegie Mellon University
Lorenz Biegler, Carnegie Mellon University

Dates: October 1, 2003 – July 31, 2004

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

Project Summary:

1 Introduction: This document proposes research for FY2003–FY2004 within the PDE-Constrained Optimization Project under the auspices of Sandia National Laboratories' Computer Science Research Institute. This project is collaboration between faculty, postdocs, and students at Carnegie Mellon, Courant, and Rice Universities, and Sandia Albuquerque and Livermore researchers. The project's principal aims are to develop, implement, and demonstrate algorithms and software tools for optimization of systems governed by partial differential equations (PDEs). PDE-constrained optimization is a critical enabling technology as the engineering community moves beyond simulation to simulation-based design, control, and parameter estimation. While the field of computational nonlinear optimization has made steady progress over the past 30 years, the class of PDE-constrained optimization problems we target presents unique and pervasive challenges.

Particularly when the PDE simulations are large and complex—as embodied by Sandia's suite of large-scale production engineering simulation codes—and the optimization variables are finely parameterized, entirely new classes of scalable, efficient, and robust optimization algorithms are required. This imperative is made all the more critical by the maturing state of PDE simulation technology. For a number of problem classes, it is only now—that the mathematical models are sufficiently physically descriptive, the numerical discretizations are sufficiently robust, and the "forward" PDE solvers are sufficiently scalable to high-resolution problems—that PDE-optimization becomes not only warranted but necessary for adding the most value to simulation.

2 Research Tasks: This section describes research tasks we feel are essential for advancing the technology of PDE optimization to the point where it can be readily applied to large, complex, parallel applications typical of Sandia's suite of simulation codes. The primary mechanism for conducting this research will be through short and long term visits by university faculty, postdocs, and students to Sandia at both Albuquerque and Livermore locations.

At the same time, this research program will leverage ongoing algorithmic research at Carnegie Mellon, Courant, and Rice that is funded by other sources. The visits will help motivate and direct this work to the complexities presented by Sandia applications.

Task 1: Investigate domain decomposition approaches for time-dependent and stationary optimization problems with relevance to Sandia applications

The solution of quadratic programming sub problems is the most computationally time-consuming component of a full space SQP method. When applied to PDE constrained optimization, Krylov subspace methods combined with domain decomposition (DD) based preconditioners have proven to be successful for these tasks. One such approach we have developed, the Lagrange-Newton-Krylov-Schur (LNKS) method, as 1 implemented in the Veltisto package, preconditions by decomposing the full space of unknowns into state, adjoint, and decision subspaces, with approximate solves within each subspace in turn effected by domain decomposition preconditioners. The approach can be very effective for steady PDE optimization; for example, solution of 3D Navier Stokes optimal flow control problems with O(106) states and O(104) controls has been achieved in as little as five times the cost of the flow solution, and with high parallel efficiency. There are several remaining challenges to ensure a scalable general-purpose method for PDE optimization. First, we need better-reduced Hessian preconditioners, which at the moment are on based on limited memory quasi-Newton ideas. We will be pursuing preconditioners based on the character of the underlying infinite dimensional operator, for general classes of problems. Second, scalable preconditioners are required for incorporating inequalities on state and decision variables. We have recently proposed such a preconditioner that for certain problems seems insensitive to barrier terms introduced by

interior-point methods. We will extend and explore this DD preconditioner for a broader class of applications arising at Sandia. This complements the recent work of the project on interior-point SQP. In particular, the resulting preconditioners can be directly applied to solving tangential subproblems in MOOCHO. Third, time-dependent PDE optimization remains a formidable task, which often requires different solution techniques, than those applicable for steady state problems. We plan to continue our work that addresses these issues, both on time-domain decomposition methods, as well as on improved preconditioners for time-dependent reduced Hessians. The Sundance–MOOCHO interface will serve as a prototyping tool to explore different time-domain decomposition strategies before applying the most promising to Sandia simulation tools.

Task 2: Investigate discretization accuracy for optimization for select Sandia applications, using Sandia simulation codes and frameworks

In PDE-constrained optimization, it is often assumed that accurate PDE solvers combined with robust optimization tools are sufficient to compute accurate optimal solutions. This is often true when applied to simple test problems, but may not hold when more sophisticated discretizations are applied to complex problems. In this case, a straightforward interface of PDE solver and optimizer may compute an optimal control/design/parameter at a computational cost that is much higher than required for the quality of the computed solution or, worse, may compute a control/design/parameter that has little to do with the optimal control/design/parameter. In such cases, discretizations can sometimes be modified to remedy this inferior behavior. We have analyzed this in the context of some model problems discretized using stabilized finite elements. We will explore this issue on selected Sandia applications, such as source detection. We plan to use the Sundance–MOOCHO interface as a prototyping tool.

Task 3: Develop Eulerian methods for shape optimization

The development of efficient general-purpose algorithms for shape optimization will have a significant impact on simulation-based optimal design. However, shape optimization presents several additional difficulties beyond those encountered in non-shape PDE-constrained optimization, including the computation of shape derivatives and representation and management of large shape motion and the resulting mesh movement (which has yet to be solved cleanly in a parallel environment). As an alternative to conventional moving-mesh Lagrangian methods, we will investigate level-set like techniques for representing shape in the context of PDE-based shape optimization. Level set methods overcome the difficulties of regenerating meshes and permit topological changes that classical methods are incapable of allowing. However, most level set methods are based on controlled evolution of a pseudo-time Hamilton-Jacobi system, which 2 amounts to steepest descent solution of the reduced space optimality conditions. Instead, we are pursuing a full-space, Newton-based approach that builds on our work in fast algorithms for PDE optimization to directly solve for the optimal shape. This promises to be more robust and up to several orders of magnitude faster. Since we are dispensing with controlled evolution (which has regularizing properties), we must find alternate regularizations to render the shape problem well posed. The shape optimization work is currently being prototyped in Sundance, which gives access to MOOCHO solvers. Driving Sandia applications include aerodynamics, magneto hydrodynamics, and structural dynamics applications.

Task 4: Study the influence of inexact problem information on PDE optimization methods

PDE simulation tools used at Sandia and elsewhere apply iterative solvers, mesh adaptation, etc., to efficiently compute the PDE solution for a given set of parameters. As a consequence, if such PDE solvers are interfaced with a PDE optimization tool such as MOOCHO, the latter is not supplied with exact function values or derivative information. In some cases the error in function and derivative information can be controlled, e.g., through the setting of stopping tolerances in iterative PDE solvers. In other cases at best only asymptotic information for the error is available. We have developed a theory that allows us to integrate PDE simulations that have controllable errors with SAND optimization tools. Our approach tightens the error tolerances as needed, based on the progress of the optimization, allowing inexpensive solves whenever feasible. We will implement and study our approach in the context of the Sandia PDE optimization environment. The effects of inaccuracies for which only asymptotic error bounds are available on the performance of optimization algorithms and on the quality of the computed solution are much less understood. We plan to explore these effects on Sandia applications using Sundance–MOOCHO, MPSalsa–MOOCHO, and other existing simulation–optimization interfaces.

Task 5: Improve barrier/filter strategy for rSQP++/MOOCHO
The superior convergence properties of the interior point code IPOPT and the initial benchmark results of the barrier and filter implementations in rSQP++ (recently renamed MOOCHO) have led to a number of proposed refinements for this strategy. Working with Andreas Waechter (IBM), a new restoration strategy has been proposed for the filter line search and we will implement multi-algorithm versions of this approach into MOOCHO over the next few months. In addition, we will investigate full space SQP algorithms with barrier and filter features to complement the reduced space features that we have already implemented. The full space option allows us to assess domain independent comparisons more directly and also leverages the work that has been done on LNKS as well as implementations and applications in Sundance. Finally, we intend to explore novel preconditioners for solving tangential subproblems in MOOCHO that can also exploit the structure of barrier terms in this subproblem, as mentioned above.

Task 6: Address optimization under uncertainty
Uncertainty quantification is a strategic element of CSRI. Through developments in MOOCHO we hope to complement these efforts by providing capabilities for robust (or multiscenario) optimization strategies. These approach lead to optimal designs and decision-making that is insensitive to changes in uncertainties in the problem definition (including model parameters, external data and inputs, or performance requirements).

Optimization formulations for multiscenario problems are currently much more expensive to solve than design optimization problems. However, in the context of SAND optimization, very efficient decomposition 3 strategies can be applied that take full advantage of parallelism as well as the features of the barrier/filter approach that we will extend in MOOCHO. With this capability in MOOCHO, we will be able to develop optimal designs quickly that are tolerant to a wide variety of uncertainty descriptions for challenging large-scale PDE-based models.

Task 7: Improve SAND optimization with existing engineering packages
Over the past three years at CSRI we have developed a hierarchy of optimization implementations to PDEbased modeling tools and solvers. This hierarchy allows us to gauge how the structure of existing and developing tools can be exploited for SAND optimization. For example, the rSQP++ link to MPSalsa has led to efficient SAND optimization applications for problems with few decision variables. On the other hand, through new developments such as Sundance, much larger and more challenging applications can be considered. With the addition of barrier and filter line search strategies in MOOCHO, these implementation levels can be refined further and far less information may be required from the solver (e.g., multipliers are no longer required) than with previous methods. Therefore, we intend to extend SAND optimization to additional engineering modeling tools at Sandia.

Task 8: Target Sandia applications
The work from these tasks will be applied to a number of important Sandia applications. For these, we would like to be involved in the problem formulation and solution stages. Working with the Optimization and Uncertainty Quantification team, we plan to collaborate on (and are already becoming involved in) applications related to homeland security, including source detection in water-borne and airborne contaminant transport, and the optimization of reactive and fluid flow systems. These applications will validate the algorithmic improvements described in Tasks 1–7 and lead to extensions that apply to more challenging problem characteristics.

Title: Combinatorial Optimization

PI: Harvey Greenberg, University of Colorado at Denver

Dates: October 1, 2003 – December 31, 2004

CSRI POC: Cynthia Phillips, (505) 845-7296
William Hart, (505) 844-2217

Project Summary:

Harvey Greenberg will work with Cynthia Phillips and William Hart to refine research ideas that will be prominent in various FY04 funding proposals. In particular, he will educate and/or brainstorm on one or more of the following topics: linear-programming/integer-programming validation/certification, computation of multiple "sufficiently structurally different" solutions to (discrete) optimization problems, multiobjective branch and bound, nonlinear branch and bound, handling uncertainty in optimization, Lagrangian methods for derivative-free optimization, sensitivity analysis, post-solution analysis, and general techniques for finding robust solutions to discrete optimization problems.

Title: Research in Finite Element Simulations

PI: Max D. Gunzburger, Florida State University

Dates: May 1, 2001 – December 31, 2004

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

Project Summary:

There are a variety of potential research projects and collaborations that I would very much to establish with CSRI and other Sandia Laboratory personnel. The list that follows is not an inclusive one. For example, I would be interested in getting involved, if CSRI and Sandia personnel feel that it would be beneficial, in ongoing projects such as finite element methods for MHD flows and finite element methods for compressible flows.

Stochastic PDE's and design under uncertainty – I believe it is fortuitous that john Red-Horse and I have a common interest in numerical methods for stochastic partial differential equations and their use for design under uncertainty, and that both of us are also in contact with engineers at the Air Force lab in Dayton about such problems. This is fortuitous because these areas of research are emerging as very important to numerous applications and at the same time remain very difficult to treat. I believe interacting with John will enable us to make significant progress in the theoretical and practical treatment of these problems. There is also the possibility of interaction with the Thermal Sciences Department on their response surface techniques for design under uncertainty; here, my work on centroidal Voronoi tessellations may have a beneficial role to play.

Least-squares finite element methods – Pavel Bochev (a CSRI visitor) and I are developing a theory of least-squares finite element methods, which unifies all known variants of this class of methods under a single abstract setting. This effort could have considerable practical consequences in addition to the perhaps its more obvious theoretical importance. In particular, the unified theory is very likely to facilitate in the selection of the best method to apply in specific situations. Pavel and I are also in the midst of writing a book on least squares finite element methods, which will collect known theoretical, algorithmic, and implementation results about this emerging methodology.

Shape and value control of complex systems – CSRI already has some excellent collaborators, e.g., Matthias Heinkenschloss and Omar Ghattas that certainly can provide much of what might be needed in the area of control for complex systems. However, I see a number of opportunities, perhaps a little down the road, for my interaction with CSRI and other Sandia Laboratory personnel on flow control problems and

other control problems involving complex systems. For example, my expertise on shape control problems may come in handy when such efforts are needed in the future. Furthermore, it is my understanding that Martin Berggren will possibly be making an extended visit to CSRI. Martin and I have, on a number of occasions, discussed some work that we would like to do together; these projects, I believe, would be of benefit to Sandia and I would certainly welcome the opportunity of working with Martin.

Centroidal Voronoi tessellations – I have already committed myself to a project which will be funded separately by Rebecca Brasson; this is a very focused project in which I will produce, using centroidal Voronoi tessellations, an initial set of points for use in her group's particle-in-cell codes. I believe there are a number of other opportunities for interacting with CSRI and other Sandia Laboratory personnel on research projects based on my previous work on centroidal Voronoi tessellations. This includes unstructured, adaptive mesh generation and meshless computing methods. I would very much like to continue the development of PDE computing for solving Sandia problems based on using centroidal Voronoi tessellations in concert with Sandia geometry and discretization software.

I will also participate in a number of other activities as part of my association with CSRI. A partial list of these include the following.

Students – I would very much like for some of my Ph.D. students to spend time at CSRI or other units of the Lab. Ideally, these students would find problems for their dissertations that are of interest to them, to Lab personnel, and to me. In this case, lab personnel would be actively involved in advising the students. The process of matching students to Lab problems and Lab personnel could begin with summer visits. Once firm connections are established, students would spend additional time at the lab. I have already encouraged two of my current students to apply for visits to CSRI for next summer.

Short courses or lecture series – I would be willing to deliver short courses or lecture series on any topic within my areas of expertise; I have already participated in such activities a number of times both in the US and abroad. Notably, I have given short courses on subjects such as flow control, mathematical and computational aspects of superconductivity, finite element methods for fluid problems, and basic aspects of finite element methods.

Workshop organization – I would be willing to co-organize, with CSRI or other Sandia personnel, workshops on emerging topics in computational and applied mathematics and scientific computing. I have lots of expertise in the organization of both small and large workshops and conferences.

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: January 1, 2003 - September 30, 2003

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

Project Summary:

Introduction: The numerical solution of partial differential equation systems (PDEs) arising in engineering and the applied sciences involves implicitly 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." These scales essentially correspond to unresolved sub-grid scale physics that is 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 add-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 developed 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. The work to be performed under the direction of Dr. Thomas Hughes at the University of Texas at Austin will be done in collaboration with Sandia National Laboratories personnel, namely Drs. John Shadid, Pavel Bochev, Mark Christon, and David Gartling, among others. In addition, Dr. Hughes will spend time in residence at Sandia National Laboratories and will teach a short course there on the multiscale methods. The collaborative research, which is viewed as broad-based and long-term, will address the following areas.

Proposed Topics

1) Shock Hydrodynamics (Sandia technical lead: Dr Mark Christon)

   We will investigate: 1) the physical and numerical aspects of the von Neumann - Richtmeyer viscosity in a multiscale 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. Finally, extensions to an Eulerian formulation will be

developed that are suitable for treating multi-material problems. The proposed work will be demonstrated initially on suitable Lagrangian strong-shock problems and will be documented in an archival publication.

2) Compressible and Incompressible Navier-Stokes Equations and Turbulence (Sandia technical leads: Drs. John Shadid, Pavel Bochev and David Gartling)

   We intend to pursue the development, implementation and analysis of a sequence of multiscale methods for transport/reaction systems. The eventual target application is the Navier-Stokes equation with thermal energy and mass transfer with chemical reactions. The major component systems will be: 1) Incompressible Navier-Stokes, 2) Navier-Stokes, thermal energy with Boussinesq interaction model, 3) Navier-Stokes, thermal energy and mass transfer with chemical reactions, 4) Compressible Navier-Stokes with a low Mach number assumption, 5) Navier-Stokes compressible flow. This effort will attempt to derive the appropriate multiscale interaction models for these systems and provide consistent turbulence models to include relevant physical mechanisms.

3) Viscoelastic Fluid Systems (Sandia technical lead: Dr. David Gartling)

   We have a further interest in pursuing multiscale models for viscoelastic fluid systems. This effort will consider velocity, pressure and stress formulations. Additional efforts will attempt to couple thermal energy transport. Formulations for specific constitutive models will be pursued.

4) Large Deformation Solid Dynamics (Sandia technical lead: Dr. Mark Christon)

   Multiscale and stabilized method analysis will be applied to large deformation solid mechanics problems. These methods will allow a broad choice of basis functions in contrast to classical mixed finite element methods. It will also provide an initial basis for a consistent formulation of unresolved sub-grid scale physics in a few selected applications. Possible applications include microstructure issues with PZT, and macro-scale deformation of bodies with unresolved cellular microstructure.

5) Additional Topics (Sandia technical lead: Dr. Pavel Bochev)

   Possible extensions of the multiscale variational formulation to control-volume schemes, discontinuous Galerkin methods, semi-conductor device modeling with drift diffusion and hydrodynamic models, and plasma simulations will be addressed.

Tasks for FY03

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

Deliverables for FY03

1) Short Course on multiscale analysis and stabilized methods at Sandia National Laboratories, Albuquerque (10 hours).
2) Course notes.
3) Reference bibliography.
4) Two weeks on site at Sandia.
5) Consultation/collaborations with Sandia technical personnel on specific research topics.
6) Year end report and compilation of any technical reports and articles produced by supported work.

The research tasks and deliverables will be the responsibility of Dr. Hughes who will be assisted by one full-time graduate student.

Title: System Performance Benchmark Testing, Evaluation and Presentation

PI: Chu J. Jong , Illinois State University

Dates: May 19, 2003 – August 15, 2003

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

Project Summary:

I. INTRODUCTION

The purpose of benchmark testing is to measure the performance of a computer system under different configurations. Results from benchmark testing are normally used by the system developers, both hardware and software engineers, to identify the weak areas of their system.

From the testing results, developers can prioritize improvement items and work from the top one on the list down to the bottom to optimize their efforts making a better future machine or system software. Many benchmark-testing suites are built (or used) by a set of "de facto" testing applications; some of them are composed by a large set of general testing strategies. Their results are either too specific to certain applications or too abundant to become useful. When developing a computer system, both system (hardware and software) and application developers are working together to set specifications about how their new system should be built. Requirements from the application developers should be integrated into the entire design/build process to make the system valuable. However, most application requirements are only used by the initial design stage and missing from the subsequent revision stages. This phenomenon is mainly due to the lack of adequate association between the benchmark testing results and the actual runtime application behaviors. Application developers usually have a hard time to comprehend the performance matrixes generated by the prototype systems. Thus they miss the opportunities to make revision for their requirements. In other words, the benchmark testing results do not contribute, in large, to help application developers substantiate their needs to the design of a new system. Therefore, it is essential to formalize a mechanism that bridges the gap between the benchmark testing results and application runtime requirements.

To bridge the gap, we laid out three phases: First, understanding phase, second, realizing phase, and third, integrating phase. This document emphasize on how to construct the first phase of bridging the gap. In the following sections we outline the implementation strategies, the execution sequence, and time-line activities and deliverable for the first phase, the understanding phase. 2

II. OBJECTIVE

The driving force of the understanding phase is that the system performance benchmark testing results should be easily understood by the users. The understanding of the relation between the system performance matrixes and the application's behavior will not only help system developers build a better system but also help application developers develop applications that manage the system resources effectively. The goal of this project is to construct the understanding phase during the summer of 2003 to help both system and application developers reduce their coding efforts on making their system more valuable. Our strategy is to break down this phase into three components: a benchmark testing suite, a processing and evaluation software, and a presentation interface layer (both system and application), and make each component a sub-layer of this phase.

III. METHODOLOGY

There are different views on the performance benchmark testing results. From the system developers, they need to understand the system performance under different execution environments. From the application developers, they need to understand the meaning of performance matrixes under their application execution. When developing the above mentioned three components, we need provide views that allow both system and application developers grasp concrete ideas about what their system will do on behalf of their needs. Each of the following paragraphs describes the basic approach that we plan to apply to these components.

Benchmark testing suite: The main purpose of the benchmark-testing suite is to provide the ability to gather as much performance data as possible under variety system configurations. We plan to enhance the existing MPI Benchmark Testing Suite, written by the system developers at Sandia National Laboratories a couple of years ago, by adding portable feature and probably more performance data collection mechanisms. We expect to use all the existing code but do not exclude the possibility of re-engineering part of the testing strategies if there will be a need.

Processing and evaluation software: Raw data are hard to digest. The purpose of this component is to pre-process the raw data, such as put them in a uniformed format, and then evaluate the preprocessed data under user's (mainly the system developers) choice, such as filtering, selection, and cross referencing. We plan to write a user program, possibly a performance matrix configurator, to help user organize their choices.

Presentation interface: While a configurable performance matrix program may be sufficient to the system developers, it does not give enough higher-level abstraction to the application users. A comprehensive graphical user interface layer will be needed to provide application users a better interpretation on the performance benchmark testing results. Developing this interface requires substantial involvement from the application developers and should maintain a close working relation with them. The key issue to make this interface layer successful relies on the understanding of user requirements. A 3-D graphical user interface is always good to have; a successful 2-D user interface with sufficient abstraction to satisfy most users' needs can be an interim solutions. The conversion to a 3-D one in the future should not incur much extra work.

IV. EXECUTION TIME-LINE AND DELIVERABLES

All three components either require frequent accessing target systems or interacting with application developers, it is better to build them on site. We plan to have them completed at the end of summer of 2003. Our proposed time-line activities and deliverable day of the first stage of "bridging the gap" are shown in the following table. The completion of this stage alone should construct sufficient connection between system and application developers. In the next section, Section V. Future Work, a brief description and a coarse time-line will be given on the second stage of "bridging the gap."

WW / Tasks	21 *5/19*	22 *5/26*	23 *6/02*	24 *6/09*	25 *6/16*	26 *6/23*	27 *6/30*	28 *7/07*	29 *7/14*	30 *7/21*	31 *7/28*	32 *8/04*	33 *8/11*
Benchmark Testing Suite: Enhancement and modification if needed (starts on 5/19)	X	X	X										
Processing and Evaluation: Set Uniform Data Format, Filtering, Referencing Interface			X	X	X	X	X	X				X	X
Representation Interface: Configurator, Graphical User Interface						X	X	X	X	X	X	X	X
Progress Report and Presentation (tentative 8/13)													X

SUMMER, 2003 – 13 WEEKS, FROM 5/19/2003 TO 8/15/2003

BIOGRAPHICAL SKETCH

Dr. Chu J. Jong is an assistant professor of Computer Science at Applied Computer Science Department at Illinois State University. He holds a Ph.D. from University of New Mexico and a Master degree of Computer Science from University of Minnesota. His research interests include: Parallel systems and parallel performance tools.

V. FUTURE WORK

The second stage of "bridging the gap" has two phases, realizing and recursive integrating phases. This is a possible extension of the first stage. In the following, we give a brief description on each and provide a coarse implementation time-line at the end.

In realizing phase, users should realize the behavior of their applications and then be able to set criteria for their system resource usage. In recursive integrating phase, revised requirements should feed back to the system at each revision step. Two layers, analyzing and implementation layers, one for each phase will be implemented.

Analyzing layer – The major function of this layer is to analyze applications and generate system resource utilization matrixes. These matrixes will be used to determine where and how the instrumentation should be used; they should also be used by system developers for their initial system design. Tasks include a parser to extract system calls, a system resource utilization matrix generator, and a system resource usage presenter.

Implementation layer – The major functions of this layer are adding instrumentations to the suggested areas from analyzing layer, gathering runtime resource utilization data, and revising instrumentation strategies. The gathering and feeding back to the system will be performed recursively until an accurate result will be generated. Tasks include a runtime instrumentation library, a problem and system configuration generator, and a performance result analyzer.

The possible time-line activities as below:

Year/Month / Phases/Tasks	2003 8	9	10	11	12	2004 1	2	3	4	5	6	7	8	9	10	11	12
Realizing: Preparation (identify user applications)	X																
Realizing: Request Parser; Utilization Matrix Generator		X	X	X	X	X	X	X	X								
Realizing: Progress Report and Presentation									X								
Recursive Integrating: Instrumentation, Feed Back								X	X	X	X	X	X	X			
Recursive Integrating: Generator, and analyzer.												X	X	X	X	X	X
Recursive Integrating: Wrap Up															X	X	X
All phases: Progress Report and Final Presentation																	X

TIME-LINE ACTIVITIES FOR FUTURE WORK

Title: CSRI Executive Board Member

PI: Deepak Kapur, University of New Mexico

Dates: October 1, 2003 – September 30, 2004

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: Understanding Performance of New PIM-Based Execution Models

PI: Peter M. Kogge, University of Notre Dame

Dates: February 1, 2003 – December 31, 2003

CSRI POC: Neil Pundit, (505) 845-7601

Project Summary:

Background: In recent years, the so-called "memory wall" (the ratio of latency to main memory to the cycle time of a modern high performance CPU) has had a dramatic effect on the design of computers at all levels. At the desktop level, up to three levels of caches and faster memory busses have been unable to keep up with the growing disparity. At the supercomputer level, attempts to not only allow access to very remote memories, and still support some sort of memory coherence has overwhelmed interconnection bandwidth. The situation is so bad that efficiencies of only a few percent are "typical." New programming models and languages, such as MPI, co-array Fortran, etc, have by themselves added only incremental improvement.
One alternative, termed "Processing-In-Memory," has the potential to significantly affect the wall directly. With this technology, significant processing logic can be placed literally next to main memory arrays (on the memory chips themselves). Tremendous reductions in latencies are coupled with tremendous bandwidth increases to these local memories.

If used in conventional computing models, these features should allow some improvement in the situation. However, if used with new models of execution, these PIM technologies hold the potential of near revolutionary improvements in high end computing systems, especially as we move into massive parallelism. At the very simplest, the ability to expand memory access requests to include atomic "operate at the memory" radically simplifies access and updates to shared data structures. At an intermediate level, the ability to initiate a method invocation directly in the memory holding an object, and have it executed there (where latencies are short, and access to all components of the object can be done without network traffic), has the potential to eliminate many latency causing events totally, and reduce others from expensive two-way latencies to simpler one-way trips. At an even more significant level, the ability to devise architectures that support extremely light-weight threads, that can be executing remotely at a memory node and then literally "pick up and leave" when the next object referenced is not local, have the potential to radically change the way we think about memory versus processing, particularly when dealing with exactly those sparse data structures that so bedevil current supercomputers.
Over the last decade, Notre Dame has become a leader in such PIM architectures, especially when targeted to the high end of computation, as in the DIVA, HTMT, and HPCS Cascade projects. This proposal endeavors to leverage that experience in directions of direct interest to Sandia.

Objectives: The objectives of this project are two-fold. First is to help explore and understand the real value these PIM-based architectures and execution models have on problems of real interest to the very high end computing community, especially as found at Sandia National Labs. Second is to build an understanding of the nature and potential of programming support tools that might allow future high-end applications to leverage such systems.

Work Description: This project will continue the development of a variety of simulation and application analysis tools that will aid in understanding how these new PIM-based architectures might behave under performance stressing applications. The simulators will include both trace based and virtual prototypes. The former extract information about these new opportunities from execution traces of real applications and kernels, and would allow projections of potential execution times, and the frequencies of key events in a PIM-based system. The latter emulates the enhanced "memory" accesses possible in a large PIM-based memory by a library of network based routines, and then implements the code that would run at individual memory nodes by tailored code sequences. Such "runnable" codes allow demonstrations of correct operation, scaling of execution to problem sizes above what trace-based simulations might do, and provide a basis for what to ask for from compilation and program generation technologies.

Such tools would be developed at Notre Dame during the Spring of 2003, and then taken to Sandia during the summer, when Sandia-relevant applications can be run against them in a controlled environment. Post analysis of the data would occur during the Fall of 2003 back at Notre Dame. It is expected that no Sandia-developed applications will leave Sandia.

It is expected that early discussions with Sandia would in particular help guide selection of applications and kernels, and definition of relative measures of effectiveness that can be used to evaluate the results.

Expected Outcomes: The expected outcomes from this project are primarily in the areas of increased understanding of new models of execution as supported by PIM-based architectures, what are the key measures to look for in evaluating such new models, and how applications of real interest might measure up when run on such systems.

Key deliverables will be copies of the analysis programs and a final report summarizing the results.

Budget Justification: The budget for this project assumes a one-year time frame, from approximately Jan. 1, 2003 to Dec. 31, 2003. The labor component of this funding would cover a staff programmer full time from Jan. 1 through the end of the summer, ¾ of a semester and the summer for each of two Ph.D. level graduate students. The two graduate students would be in residence at Sandia for the bulk of the summer, with a brief period in the middle for an extended debrief back at Notre Dame. Travel includes nine person-trips to Sandia at $500 each, not counting the summer travel by the graduate students. Six of these person trips would nominally be in the Spring of 2003, to verify that the analysis tools being developed will capture Sandia application characteristics, and the final three would be in the Fall, to provide a final presentation and discussion.

The project will be managed by Dr. Peter M. Kogge of the Univ. of Notre Dame
Dr. Kogge received his Ph.D. in EE from Stanford in 1973. From 1968 until 1994 he was with IBM's Federal Systems Division, and was appointed an IBM Fellow in 1993. In August, 1994 he joined the University of Notre Dame as first holder of the endowed McCourtney Chair in Computer Science and Engineering. He is currently the Associate Dean for Research, College of Engineering. He is also an IEEE Fellow and a Distinguished Visiting Scientist at the Center for Integrated Space Microsystems at JPL. He holds over 20 patents and is author of two books, including the first text on the now ubiquitous technique of hardware pipelining.

His current research areas include massively parallel processing architectures, advanced VLSI and nano technologies, non von Neumann models of programming and execution, parallel algorithms and applications, and their impact on computer architecture. Major current projects currently include PIM architectures (especially for massively parallel systems), inherently low power computer architectures, and the use of nano level technologies such as Quantum dot Cellular Automata. Prior projects include a multi-threaded parallel processor which has flown on every Space Shuttle, the IBM 3838 Array processor which was for a time the fastest floating point machine marketed by IBM, EXECUBE - the world's first parallel processor on a DRAM chip, and a major role in the HTMT petaflops project.
A more complete vitae can be found at www.cse.nd.edu/~kogge.

Year/Month / Phases/Tasks	2003					2004										
	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
Realizing: Preparation (identify user applications)	X																
Realizing: Request Parser; Utilization Matrix Generator		X	X	X	X	X	X	X	X								
Realizing: Progress Report and Presentation									X								
Recursive Integrating: Instrumentation, Feed Back								X	X	X	X	X	X	X			
Recursive Integrating: Generator, and analyzer.												X	X	X	X	X	X
Recursive Integrating: Wrap Up															X	X	X
All phases: Progress Report and Final Presentation																	X

Title: Supporting MPI Collective Communication Operations with Application Bypass

PI: Dr. Panda, Ohio State University

Investigators: Dr. Wyckoff, Ohio State University
Dr. Sadayappan, Ohio State University

Dates: July 1, 2003 – June 30, 2004

CSRI POC: Ron Brightwell, (505) 844-2099

Project Summary:

Overview: For large-scale parallel systems supporting MPI, it is desirable that the MPI implementation ensures "progress" in order to achieve good performance and scalability [1]. Currently, collective communication operations in MPI are implemented by explicit send/recv calls by the processes. However, if a single node gets delayed (say an intermediate node of a broadcast, reduction, or barrier operation), the whole operation gets delayed. This leads to increased execution time for applications and limited scalability.

Modern interconnects are supporting new communication mechanisms such as remote memory operations (RDMA Read and RDMA write). Similarly, modern NICs are providing programmable interface and memory to support collective communication operations with minimal interaction from processors [2,3,4]. These advances allow communication operations to be implemented without explicit use of send/recv calls by the processes. This leads to the following two open challenges:

1) Can we implement MPI collective operations with application bypass by taking advantage of RDMA operations and NIC-level support?

2) How much performance benefit can be delivered to applications with this bypass property?

In this research, we propose to investigate both these challenges.

Proposed Work: The proposed work is divided into four broad directions: developing suitable mechanisms and semantics to implement collective communication with application bypass with host-level support, taking advantage of RDMA operations, taking advantage of NIC-level support, and evaluating the performance benefits at the application level. These four directions are briefly discussed below.

1) Mechanisms/Semantics to support application bypass with host-level support: Currently collective operations are implemented with explicit send/recv calls by the participating processes. In order to implement these operations with application bypass, new mechanisms and semantics are needed to be developed and evaluated. For example, for supporting broadcast at an intermediate node without the support of the application processes, do we need to have a dedicated thread or interrupt handler to take care of sending the messages to other nodes? How should this dedicated thread or interrupt handler be designed with minimal overhead? How the collective communication library needs to be interfaced with the dedicated thread or interrupt handler? What additional information a collective communication message should be carrying to invoke the appropriate thread or interrupt handler?

For supporting collective operations where an intermediate node has to supply data (such as barrier, reduction, or complete exchange), there is a need to define a clear semantic to support such operations at the application level. In current MPI implementations, all collective operations are blocking. In order to support collective operations with application bypass, we need to design and develop a new semantic. One possibility is to design a split-phase collective operation semantic which can support non-blocking/fuzzy collective operation. Under this semantic, a process can supply data to the dedicated thread/handler while entering the collective operation, then leave for performing computation, and come back later to check for the completion of the collective operation.

We plan to study and analyze the above mechanisms and semantic to determine the right combination to support collective operations with application bypass and with minimal overhead. We will implement and evaluate these alternatives under MPICH/GM on Myrinet, MVICH/VIA over GigaNet cLAN, and MPICH/EMP over Gigabit Ethernet.

2) Taking advantage of RDMA operations: Modern interconnects support Remote DMA operations. Two kinds of RDMA operations are possible: RDMA Read and RDMA write. Myrinet and GigaNet cLAN implement RDMA write only. Compared to send/recv, RDMA write allows data to be written from a sender to a receiver without the exclusive involvement of the receiver. This one-sided operation is currently being exploited in many MPI implementations to support efficient point-to-point communication. We plan to study the benefit of such one-sided operation for supporting collective communication operations with application bypass.

The main issues here are as follows: How to define a region of pinned memory in a system to support collective operations with RDMA writes? How to organize this pinned area into different buffers and allocate these buffers for different collective operations in a unique manner across processes? How to integrate such pinning and buffer organization with the dedicated host-level thread/handler for collective operations so that the send operations at intermediate nodes can be replaced with appropriate RDMA write operations?

In our preliminary work along this direction with MVICH/VIA on cLAN, we have observed that barrier operation on an 8-node can be implemented with a 29% improvement with RDMA write operations compared to the traditional send/recv operations (31 microsec vs. 44 microsec). For larger systems, we hope to see much more benefit. In this evaluation, the benefits associated with RDMA write operation is only explored. The intermediate nodes still perform send/recv operations. We plan to extend this framework with appropriate pinning and buffering schemes for all collective operations and integrate it with appropriate mechanisms from step (1) to support collective operations with application bypass.

3) Taking advantage of NIC-level support: Modern NICs are becoming smarter with programmable processors, DMAs, and larger amount of memory. Potential benefits of such smarter NICs are being exploited to support fast point-to-point communication operations [6,7], collective communication operations [3,4], and remote atomic operations [2]. For collective operations, we have explored schemes where the NIC processors at intermediate nodes try to take care of most of the send/recv operations instead of the host nodes. For example, for a barrier operation, the host processors inform their respective NIC processors about reaching their barrier points. Then the NIC processors independently perform the required pair-wise exchange operations without the help of host processors. At the end of the exchange operations, the NIC processors inform the host processors to exit the barrier.

Such NIC-level support provides a natural mechanism to support a collective operation in an application bypass manner. For the time being, all our implementations still support blocking collective operations. We plan to extend this framework to support non-blocking collective operations. Also, on current generation NICs, NIC processors do not have sophisticated floating-point units. Thus, it is very difficult to design fast implementation of reduction operation with complete NIC-level support. We plan to use a hybrid scheme where the NIC processors can perform the required communication operations and gather the data to some pinned memory and then the dedicated thread/handler at the host can perform the required computation steps. Such a hybrid scheme will deliver good performance and scalability to support collective communication operations in an application bypass manner.

4) Evaluating performance benefits at the application level: The designs proposed above have multiple trade-offs, such as RDMA support vs. NIC-level support, dedicated thread vs. interrupt handler support at the host, different buffering and pinning schemes, and different algorithms for collective operations. While designing and developing these schemes we first plan to use performance evaluations with micro-benchmarks to select the stronger and effective ones. As the entire application bypass collective operation framework is developed, we also plan to use in-depth application-level evaluations to study the overall benefits of this research.

At OSU/CIS, we have access to the standard NAS benchmarks as well as some parallel visualization applications. At Ohio Supercomputer Center (OSC), we have access to many production-level application codes. As the schemes for application bypass collective operations get developed, we will compare their benefits with the existing (conventional) schemes for a range of applications. Such evaluations will demonstrate the benefits of application bypass collective operations and provide guidelines for designing next generation parallel systems and applications with high performance and scalability.

Milestones: The term of the proposed work will be one year starting July 1, 2002. The following list of milestones represents the approximate completion date of the research directions as described above.

Sept. 2002 - Design and Development of separate thread/interrupt handler to support application bypass and their impact on performance

Design, development, and evaluation of non-blocking semantics for collective operations to support application bypass

Dec. 2002 - Design of suitable pinning and buffering schemes to support collective operations with RDMA operations

Developing algorithms for frequently used collective communication operations (barrier, broadcast, and reduction) with RDMA support

Integrating RDMA support with separate thread/interrupt handler at the host to support application bypass and studying the performance benefits

Mar. 2003 - Design of NIC-based support for collective operations

Integrating the schemes with separate thread/interrupt handler at the host to support application bypass and studying the performance benefits

June 2003 - Performance evaluation of both schemes for a range of applications and systems

Deriving guidelines about which schemes work best to support collective operations with application bypass

Deliverables: The following items will be delivered during the course of the contract:

Different mechanisms (thread/interrupt handler, RDMA support, NIC-based support), semantics (non-blocking collectives), and their implementations on contemporary communication subsystems (Myrinet/GM, GigaNet/VIA, and Gigabit Ethernet/EMP) to support collective operations with application bypass.

Performance and scalability analysis of the new scheme for a range of applications.

References

[1] Portals 3.0: Protocol Building Blocks for Low Overhead Communication, Ron Brightwell, Bill Lawry, Arthur B. Maccabe, Rolf Riesen, CAC '02 Workshop, April 2002.

[2] D. Buntinas, D.K. Panda, and W. Gropp, NIC-Based Atomic Operations on Myrinet/GM, SAN-1 Workshop, held in conjunction with High Performance Computer Architecture (HPCA) Conference, Feb 2002.

[3] D. Buntinas, D. K. Panda and P. Sadayappan, Fast NIC-Based Barrier over Myrinet/GM, Int'l Parallel and Distributed Processing Symposium (IPDPS), April 2001.

[4] D. Buntinas, D. K. Panda, J. Duato, and P. Sadayappan, Broadcast/Multicast over Myrinet using NIC-Assisted Multidestination Messages, Fourth Int'l Workshop on Communication, Architecture, and Applications for Network-Based Parallel Computing (CANPC '00), Jan 2000, pp. 115-129.

[5] R. Gupta, The Fuzzy Barrier: A Mechanism for the High Speed Synchronization of Processors, Proceedings of the International Conference on Architectural Support for Programming Languages and Operating Systems, 1989, pp. 54-63.

[6] P. Shivam, P. Wyckoff, and D. K. Panda, EMP: Zero-copy OS-bypass NIC-driven Gigabit Ethernet Message Passing, Supercomputing (SC '01), November 2001.

[7] P. Shivam, P. Wyckoff and D.K. Panda, Can User Level Protocols Take Advantage of Multi-CPU NICs?, Int'l Parallel and Distributed Processing Symposium (IPDPS '02), April 2002.

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

PI: Dan C. Sorensen, Rice University

Dates: August 1, 2002 – September 30, 2003

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: The Red Summit Project:
Petaflops Scale Computing through Processor in Memory Architecture

PI: Dr. Thomas Sterling, California Institute of Technology (Caltech)

Dates: October 1, 2003 – September 30, 2004

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

Project Summary:

Introduction
The California Institute of Technology seeks to renew and continue its successful research collaboration with the Sandia National Laboratory in the area of advanced high end computing architecture. Caltech is currently completing a two-year research project sponsored by the CSRI to explore the opportunity of achieving performance gain for MPPs through the augmentation of such systems with advanced processor in memory or PIM components substituted for at least part of the system main memory. Significant progress towards this goal has been accomplished through the Caltech project although additional work is required to fully satisfy all of the original program objectives. As a result of this effort, significant potential has been demonstrated, not just for PIM-enhanced MPPs but also for large arrays of PIM components alone. Caltech proposes to actively collaborate with Sandia National Laboratory to aggressively pursue this important opportunity with CSRI towards the goal of accelerating the path to practical Petaflops-scale computing, for possible delivery well within this decade. This proposal constitutes a request for one-year renewal and extension of the existing project to complete its goals and to establish a three-way working relationship combining the interests and talents of CSRI, Caltech, and the University of Notre Dame towards the potential opportunity of PIM technology for real world applications. This proposal is being submitted in conjunction with a second proposal to CSRF to conduct a parallel effort in detailed system studies related to implementation, integration, and application of PIM.

Towards a PIM Petaflops Computer Architecture
PIM combines computing logic and memory stacks on the same integrated semiconductor die. Structures enabled by this merger permit low latency access to data stored in memory, simultaneous access to data contents of an entire memory row (typically 256 bytes), and the opportunity to partition the total memory capacity of a single chip in to many memory/logic nodes (potentially 64 by 2005) to greatly expose and exploit on-chip memory bandwidth and provide high degree of concurrent execution on a single die. Estimates of potential peak performance vary significantly but 64 Gflops peak performance per chip is a reasonable estimate (+/- 50%) with half a Gbyte of memory in the second half of this decade. Computation performed by such a system is very different from conventional clustered solutions. Instead of a process residing on a single node, many threads of computation will be performed concurrently, migrating through memory by means of parcels between nodes and threads on each node. The node processors themselves perform as fine grain transactional processing elements in response to incident parcels and sharing a virtual global address space with all other PIM chips and nodes. This demand driven paradigm is highly adaptive at runtime, automatically overlaps communication with computation, and exposes a fine grain application parallelism intrinsic to sparse irregular data structures. It also permits flexibility in architecture by allowing the integration of shared functional units with a PIM chip to be dynamically scheduled on a fine grain basis. It is the relationship of the new computing model enabled by this advanced PIM concept to important applications and the opportunity to exploit shared functional units that are the focus of this one-year proposal to SNL CSRI.

Proposed Project Goal and Objectives
The goal of the proposed Red Summit Project is to enable and accelerate the realization of practical effective general-purpose Petaflops-scale computation in the earliest time frame for DOE applications critical to national security and industrial commerce. Implicit in this goal is the need to achieve breakthrough innovation in scalable computer architecture and its application to real world problems. To achieve this goal, the Red Summit Project is being proposed to make distinct and complementary advances within a broader context of research that is exploring a set of interrelated concepts. Within this research

framework, the Red Summit Project will investigate two key aspects of the challenge: exploiting separate functional units and applying the architecture to key applications. Therefore, the objectives of the Red Summit project are:

1. To achieve a detailed understanding of the effect of the MIND architecture on key applications, and
2. To understand how to incorporate semantic hooks in to the evolving MIND instruction set architecture in order to exploit various "drop-in" functional units that may be shared among multiple MIND nodes on a given chip.

An important aspect of this proposal is the nurturing of close collaborations and frequent interactions with research scientists at Sandia National Laboratory. Experience with the first project shows that even in spite of best intentions, without mutual interest and involvement of both institutions, effective communications will be limited and visits fewer than desired, even as their value is less than wished. To address this challenge and build a strong collaboration with close interactions, the partnership described above will provide specific shared tasks, identified points of contact, and visits by SNL scientists to Caltech as well as visits by Caltech scientists to Sandia. In addition, Caltech will organize a small workshop on the emerging technology under separate Sandia sponsorship. Finally, as has been the case in the past, the PI and Caltech colleagues will actively contribute to Sandia hosted workshops in topics related to high end computing throughout the duration of the project award period.

Project Tasks
The tasks specified below are those to be conducted explicitly under sponsorship of Sandia National Laboratory in coordination with CSRI and do not represent the entire body of work being performed related to the goals of this project under separate contract as discussed above. The tasks are focused on achieving coordinated and cooperative research of strong interest to both institutions and engaging key researchers in a strong collaboration. In particular, it is intended to work directly with SNL computational scientists in the development of application codes that can run on the MIND architecture and to port such codes on to the MIND FPGA prototype when it becomes available. This work was begun during the first CSRI research project on one such application, LJS. It is also intended to work with SNL scientists in the extension of the MIND instruction set to provide extensible logical interfaces to shared functional units.

Year 1
1. Collaborate with SNL scientists to develop instruction set architecture extensions for support of external functional units and test on the MIND prototype.
2. Develop two application codes of significance to SNL, and port them to run on the MIND prototype.

Title: Developing Pattern Search Methods for Constrained Optimization

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

Dates: June 1–December 31, 2003

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

Project Summary:

Goals:
- Finish and install a C++ implementation of our Matlab prototype for robust pattern search to handle optimization problems with linear constraints.
- Test this implementation, preferably on some Sandia problems.
- Experiment with various enhancements based on ideas we have had while writing a review paper on direct search methods with Tammy Kolda for *SIAM Review*.
- Initiate a serious investigation with Tammy of efficient ways to handle degenerate linear constraints.

Context: Based on prior visits to the CSRI and an ongoing collaboration with Tammy we delivered a Matlab prototype of a pattern search algorithm for the linearly constrained optimization problem

$$\begin{array}{ll} \text{Minimize} & f(x) \\ \text{subject to} & Ax \le 0 \\ & \ell \le x \le u, \end{array} \tag{1}$$

where $f: \mathrm{R}^n \rightarrow \mathrm{R}$ and $A \in \mathrm{R}^m \times \mathrm{R}^n$.

The prototype implementation does not handle the case of degenerate constraints (which can occur in practice). While our analysis covers the case of degeneracy [10], it is well known from linear programming that degeneracy can be an inherently difficult to handle. However, we do know how to deal with degeneracy algorithmically, and this is one of the issues we will explore with Tammy during the proposed visit.

During our visit to CSRI last summer, we began writing with Tammy an invited review paper on direct search methods for *SIAM Review*. As part of this project, we undertook a thorough review of other recent work on direct search methods (e.g., [1, 2, 4, 12, 13, 14]). From our review we see ways to combine a lot of these ideas with our own (e.g., [3, 5, 6, 7, 8, 9, 10, 11, 15]) to make pattern search even more flexible.

Algorithmic development - One intriguing idea from the work of Coope and Price [1] is to incorporate into pattern search conjugacy of search directions. We see this as a promising way to help deal with scaling issues, which can bedevil direct search methods in general.

Another of the recent developments in direct search methods has been the introduction of derivative-free sufficient decrease conditions for direct search methods [4, 12, 13, 14]. This allows us to relax or eliminate the traditional lattice structure that underlies pattern search algorithms. Tammy has already had some success incorporating ideas from [4] into APPS [5]. We believe we can use similar ideas from [12, 13] in connection with conjugate directions.

Software development - We are now well positioned to make considerable progress on software development. Anne Shepherd, our William & Mary graduate student and a DOE High-Performance Computer Science Graduate Fellow, is now actively involved in the project, having finished all her other Ph.D. candidacy requirements this past fall. She will be working on the C++ implementation we hope to finish this spring. This will be the first step in her long-term plan to develop effective parallel variants.

Applications - There are new applications at Sandia (i.e., transmembrane protein structure prediction) for which pattern search may prove to be effective. Preliminary results recently obtained by Genetha Gray and Tammy, using the current version of APPS, are very promising. This was joint work with members of the Biosystems Research group at Sandia National Laboratories, Livermore. Another application includes optimization for circuit simulation using ChiliSpice and Xyce. In both applications, a question that remains is how to deal effectively with constraints, which arise in the definition of these problems.

Participation - The proposal is that we spend a month during the summer of 2003 at Sandia National Laboratories, Livermore. Our primary contact will be Tammy Kolda.

Deliverables - The proposed deliverable is an alpha version of a C++ implementation of pattern search capable of handling general nondegenerate linearly constrained optimization problems.

Impact - The success of pattern search has led to its incorporation in several software projects at Sandia National Laboratories, including OPT++, the DAKOTA software toolkit, and APPSPACK. Throughout the development of our new software for handling linear constraints, we will consult with all the parties who have already made use of our pattern search technology to ensure that our 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] I. D. Coope and C. J. Price. A direct search conjugate directions algorithm for unconstrained minimization. *Australian & New Zealand Industrial and Applied Mathematics (ANZIAM) Journal*, 42, electronic Part C:C478–498, November 2000.

[2] I. D. Coope and C. J. Price. On the convergence of grid-based methods for unconstrained optimization. *SIAM Journal on Optimization*, 11(4):859–869, 2001.

[3] Elizabeth D. Dolan, Robert Michael Lewis, and Virginia J. Torczon. On the local convergence properties of pattern search. Technical Report 2000-36, Institute for Computer Applications in Science and Engineering, September 2000. Revised December 2002 for *SIAM Journal on Optimization*.

[4] U. M. García-Palomares and J. F. Rodríguez. New sequential and parallel derivative-free algorithms for unconstrained minimization. *SIAM Journal on Optimization*, 13(1):79–96, 2002.

[5] Patricia D. Hough, Tamara G. Kolda, and Virginia J. Torczon. Asynchronous parallel pattern search for nonlinear optimization. *SIAM Journal on Scientific Computing*, 23(1):134–156, 2001.

[6] Tamara G. Kolda and Virginia J. Torczon. On the convergence of asynchronous parallel pattern search. Technical Report SAND2001–8696, Sandia National Laboratories, Livermore, California, February 2002. Revised July 2002 for *SIAM Journal on Optimization*.

[7] Tamara G. Kolda and Virginia J. Torczon. Understanding asynchronous parallel pattern search. In G. DiPillo and A. Murli, editors, *High Performance Algorithms and Software for Nonlinear Optimization*, pages 316–335. Kluwer Academic Publishers B.V., 2003.

[8] Robert Michael Lewis and Virginia Torczon. Rank ordering and positive bases in pattern search algorithms. Technical Report 96–71, Institute for Computer Applications in Science and Engineering, Mail Stop 132C, NASA Langley Research Center, Hampton, Virginia 23681–2199, 1996.

[9] Robert Michael Lewis and Virginia Torczon. Pattern search algorithms for bound constrained minimization. *SIAM Journal on Optimization*, 9(4):1082–1099, 1999.

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

[11] 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.

[12] Stefano Lucidi and Marco Sciandrone. A derivative-free algorithm for bound constrained optimization. *Computational Optimization and Applications*, 21(2):119–142, 2002.

[13] Stefano Lucidi and Marco Sciandrone. On the global convergence of derivative free methods for unconstrained optimization. *SIAM Journal on Optimization*, 13(1):97–116, 2002.

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

[15] Virginia Torczon. On the convergence of pattern search algorithms. *SIAM Journal on Optimization*, 7(1):1–25, February 1997.

Chapter 3. Sabbaticals

The following university faculty did all or part of a sabbatical stay at Sandia National Laboratories during calendar year 2003.

Dr. Martin Berggren,
Aeronautical Research Institute of Sweden, and
Uppsala University

Dr. Mathias Heinkenschloss
Rice University

Dr. John E. Renaud,
University of Notre Dame

Dr. James D. Teresco,
Williams College

Homer F. Walker,
Worcester Polytechnic Institute

Title: Numerical Methods For Partial Differential Equations with Emphasis on the Equations of Fluid Mechanics

PI: Martin Berggren, Aeronautical Research Institute of Sweden, and Uppsala University

Dates: February 1, 2002 – June 30, 2003

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

Project Summary:

My research interests are centered around numerical methods for partial differential equations with emphasis on the equations of fluid mechanics. This includes modeling and discretization issues as well as algorithms for solving the resulting algebraic problem and software aspects.

Numerical methods for fluid-mechanics problem have reached a certain maturity (even if there is still a long way to go to achieve good prediction of, say, massively separated flow at high Reynolds numbers). Therefore, I believe that many of the scientific breakthroughs in the future will appear when the numerical solution is part of a larger interdisciplinary effort. This is a reason why my specific research activities have been directed to such "compound" problems; fluid-structure interaction, analysis of flow instabilities, flow control, and aerodynamic shape optimization. The complexity of interdisciplinary problems makes it essential to interact closely with experts in different disciplines. A visit to the CSRI would allow me a unique opportunity to interact with scientists associated with Sandia. Collaboration involving some of the following Sandia staff members would in particular be of interest: Mike Eldred, Richard Lehoucq, Walter Rutledge, Kambiz Salari, Andy Salinger, and Bart van Bloemen Waanders. I have also been informed that Professor Max Gunzburger is making arrangements for regular visits. My research interests intersects those of Gunzburger, so it would certainly be of interest to cooperate. I also hope and believe that my specific expertise will be of value for Sandia.

Shape Optimization

Computational Fluid Dynamics (CFD) is increasingly used for analysis purposes, for instance in the design stage of components in which fluid-mechanical properties are important, such as for vehicles, turbines, or pumps. As the turnaround time for an analysis cycle decreases with increasing hard- and software performance, it is natural to aim for using CFD in an automatic search of the best designs. This is the purpose of shape optimization. Structural optimization methods, such as minimizing weight or maximizing stiffness under prescribed loads, appear to be more developed than shape optimization in a fluid-dynamics context; there are even several commercial packages, from Altair engineering for instance, performing shape or topology optimization of structures. Although shape optimization in a fluid-dynamics context is a less mature field, some substantial results already exists, perhaps best exemplified by the work of Jameson and coworkers [11]. The industrial interest for the technique can be exemplified by the EU-sponsored project *Aeroshape*, a three-year project devoted to aerodynamic shape optimization that started January 2000, in which most of the leading European aeronautical industries and research institutes participate (Parts of my work at FFA is within this project.)

One important, fundamental issue is well-posedness. Oscillatory shapes often appear in shape optimization computations, as discussed by Pironneau [12], for instance. This can have several reasons, physical as well as numerical. For instance, it is known that so-called riblets, small grooves in the streamwise direction on the surface of embedded solids, may decrease the viscous drag [2]. Also, the sensitivity of the objective function to oscillatory changes in the shape may be low and this may cause numerical instabilities for the discrete problems, similarly to what happens in mixed-methods for incompressible flow (checker-boarding, inf-sup, LBB conditions, etc.) it is thus important to find a sound formulation of the problem, covering what is wanted out of the optimization (for instance, do we want those riblets or not?). The formulation should yield a well-posed mathematical problem, not prone to numerical instabilities, but one that still covers a large span of designs; the easy way of achieving well-posedness is otherwise to shrink the design space sufficiently. Some results in this direction can be found in the literature; two recent contributions are by Bedivan [4] and Gunzburger & Kim [9], but many issues still have to be addressed.

Many aspects of these more fundamental issues can be studied on a model-problem level, say finite-element approximations of the impressible Navier-Stokes equations in 2D. However, it is also important to consider some type of "real-world" analysis codes.

A numerical technique that is gaining in popularity in industrial-type codes for solving the Navier-Stokes equations is cell-vertex finite-volume discretizations in an edge-based formulation combined with unstructured meshes [3]. The use of unstructured meshes simplifies automatic mesh generation and adaptation for complex geometries. The edge-based formulation has the advantage that the implementation is basically unchanged when going from 2D to 3D, and it is also independent of the shape of the control volumes. This facilitates the use of hybrid meshes, where different control-volume shapes are used in the same mesh, typically so that regular hexahedrals with high aspect ratios are used to resolve the boundary layer close to solid boundaries, whereas tetrahedrals fill the area outside the boundary layer.

The applications in mind are of the final-design type. That is, at least a rough preliminary design of the component is assumed to be known. It is the purpose of the optimization to find the "best" design through moderate modifications of the preliminary design. It thus seems appropriate to use a local, gradient-based optimization technique combined with adjoint-based methods for the calculations of gradients, to allow for efficient gradient evaluation when the number of design variables is large and the objective-function evaluations are expensive. My own experience from similar applications strongly indicate that a fast convergence of the optimization requires in general extremely accurate gradient evaluations [5, 8, 10]. It would be interesting to investigate this approach in the context of edge-based unstructured finite-volume solvers, which particular emphasis on very accurate, "discrete" implementation of the adjoint equations. Practical implementation and testing in an industrial-type code should be a part of this.

To summarize, two possible projects could be the following:

Investigations regarding problem formulations, well-posedness, approximation properties, and numerical algorithms. Studies on model problems not more complicated that incompressible Navier-Stokes equations at low Reynolds numbers in 2D.

The use of gradient-based optimization for shape optimization using an industrial-style, edge-based finite-volume code.

Alternate Approaches to Linear Stability Analysis

Stability analysis for fluid flows is concerned with deciding if a particular flow is stable for infinitesimal or finite perturbations. Linearized for perturbations u around a laminar flow state U and performing eigenvalue analysis of an operator $A(U)$ appearing in an equation of evolution, such as $u_t + A(U)u = 0$, for the disturbance u shows whether it will be linearly *asymptotically* stable. For a long time it has been recognized that this analysis does not explain the transition to turbulence for certain flows, particularly shear flows such as pipe Poiseuille flow. When the laminar state U is strongly sheared, as in a jet, in the flow over a surface, or in a pipe, the operator $A(U)$ will be highly nonnormal and thus susceptible to significant transient effects, regardless whether the eigenvalues of $A(U)$ indicate stability or not. The importance of this mathematically well-known fact for so-called subcritical (or bypass) transition to turbulence has only surprisingly recently been recognized [6, 7, 13, 14]. In these articles, transient effects have been analyzed using methods that is only applicable to paralled flow cases. Recently, methods very similar to the ones used for optimal-control problems and shape optimization problems have been applied to the study of transient effects [1]. This approach is quite general and applicable to transient studies of almost any flow. However, it has only been applied to the growth of steady disturbances in boundary layers. It would be interesting to study transients in more general situations using this approach.

Title: Application for a Four-Month Visit to the Sandia Computer Science Research Institute

PI: Matthias Heinkenschloss
Rice University

Dates: February 1, 2003 – October 24, 2003

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

Project Summary:

PDE (partial differential equation) simulation-constrained optimization in the context of design, control, parameter identification or uncertainty analyses arises in many engineering applications and in particular in many Sandia applications. Solution of such optimization problems poses many mathematical and computational challenges. These challenges even increase when one wants to design general frameworks for the solution of such optimization problems. There are many ongoing exciting efforts at Sandia to develop optimization tools for complex engineering applications, to assist in the interface between optimization and simulation tools, and to implement "optimization aware" simulation frameworks.

During my visit at Sandia, I will assist in the transfer of optimization methods developed in my research into Sandia frameworks such as DAKOTA, rSQP++, investigate difficulties arising in the interface between simulation codes and optimization (such as discretization accuracy for optimization, errors in objective and constraint function and derivative evaluations arising from iterative linearized PDE solves, effect of adaptive gridding on the optimization, optimization of time dependent problems) for selected Sandia applications, using Sandia simulation codes and frameworks (such as PREMO, Xyce, SIERRA), and to analyze approaches to overcome the above mentioned difficulties, collaborate in the development and application of preconditioners for optimization subproblems, document the research in technical reports.

Title: Optimization Under Uncertainty

PI: John E. Renaud
Professor of Aerospace and Mechanical Engineering
University of Notre Dame

Investigator: Victor M. Perez
Harish Agarwal

Dates: August 15, 2003 – May 31, 2004

CSRI POC: Mike Eldred, (505) 844-6479

Project Summary:

Overview: This proposal involves three participants currently working at the University of Notre Dame. Professor John E. Renaud will be on a leave of absence from Notre Dame and will visit Sandia National Laboratories during the fall semester (August 22, 2003 – January 6, 2004). Victor M. Perez, who will defend his doctoral dissertation at Notre Dame this spring, and will begin working at Sandia as a post-doctoral researcher beginning June 1, 2003 and continuing through May 31, 2004. Mr. Harish Agarwal will be participating as a summer graduate student intern beginning on June 1, 2003 and working through December 31, 2003. The focus of this research participation is in the general areas of optimization algorithm development, surrogate models and uncertainty estimation.

The integration of uncertainty estimation within a framework of numerical optimization, under uncertainty offers designers new tools for validating and certifying improved designs in a simulation based design environment. Sandia is moving toward an engineering process in which decisions are increasingly based on computational simulation. In the absence of physical testing, the validation of these computational models, and the certification of designs generated by these simulation tools, remains an imposing challenge. During their stay at Sandia, participants will address the modeling of uncertainty in systems whose complexity results from the interaction of multiple disciplines. These uncertainties will be integrated in a variety of approaches for optimization under uncertainty, including the interior point trust region model management approach developed by the participants at Notre Dame (Perez, et al., 2002, Rodriguez, et al., 2001). The goal is to develop an approach for optimization under uncertainty that accounts for both epistemic uncertainties (i.e., model form uncertainty) and aleatory uncertainties (variational uncertainty in materials and manufacturing). This work builds on the strengths of each of the participants. Multilevel parallel optimization methods for simultaneous analysis of disciplines will used to drive reliability-based design optimization approaches.

Background: Dr. Renaud's experience includes five years as a manufacturing systems design engineer with the Eastman Kodak Company. He is a National Science Foundation National Young Investigator Award winner. He is currently the ex-officio chair of the AIAA Multidisciplinary Design Optimization Technical Committee. He is also a member of the executive committee of the ASME Design Automation Committee and served as the Conference Chair for the 27th ASME Design Automation Conference. He is an associate editor of the *ASME Journal of Mechanical Design* and serves on the editorial boards of *Engineering Optimization* and the *AIAA Journal of Aircraft*. His research activities include design optimization, simulation based design, uncertainty modeling, multidisciplinary design and rapid prototyping. Funding in support of his research efforts has come from the National Science Foundation, NASA, Wright Laboratories, General Electric, General Motors Corporation, Ford Motor Company, Parametric Technology Inc., Stratasys Inc., and Andersen Consulting.

Victor Perez is a Ph. D. candidate at the University of Notre Dame. He came to Notre Dame as a Fulbright fellow. His research in the areas of reduced order approximations, simulation based design, and multidisciplinary optimization has been developed under the direction of Dr. John E. Renaud. He has published eight conference papers and three journal papers in the area. His experience includes four years as product development engineer at Robert Bosch, Mexico, and 3 years as instructor at the Iberoamericana University in Mexico City.

Harish Agarwal completed his undergraduate education in Mechanical Engineering in June 2000 from the Indian Institute of Technology (IIT), Kharagpur, India. Since then he has been pursuing graduate studies (Ph.D.) in Mechanical Engineering at the University of Notre Dame. His research interests include design optimization, simulation based design, robust design, multidisciplinary design, uncertainty modeling, reliability based design optimization and optimization under uncertainty. Currently he is working on projects involving reliability-based optimization and uncertainty modeling using evidence theory.

Potential Collaborations: The participants plan to work with the following Sandia National Laboratory scientists and engineers as part of this visit: Michael S. Eldred, Anthony A. Giunta, Scott A. Mitchell, Steven F. Wojtkiewicz, Timothy G. Trucano, Laura P. Swiler, William L. Oberkampf, Vicente J. Romero.

Dr. Michael Eldred will serve as the primary technical contact during the participant's stay. As part of his visit Dr. Renaud will present two seminars related to his research efforts in 1.) trust region model management for sequential approximate optimization and 2.) uncertainty quantification using evidence theory in simulation based design optimization. In addition Dr. Renaud will develop and offer a short course on multidisciplinary design optimization methods. Participants will have an opportunity to work with and support the development of Sandia's Dakota Framework for optimization and uncertainty analysis.

Introduction: Optimization Under Uncertainty – Reliability-Based Design Optimization
In deterministic design optimization, designs are often driven to the limit of the design constraints (active constraints at the optimum), leaving little or no latitude for uncertainty in the mathematical modeling and simulations. In addition, deterministic design optimization does not account for the variational uncertainty associated with the randomness of physical quantities that make up the system. Optimized designs determined without due consideration of variability can be unreliable leading to early life cycle failures. To account for physical uncertainties and simulation model uncertainties the use of reliability analysis must be incorporated within the design optimization process.

In the past ten years increased emphasis has been focused on the development of procedures to combine design optimization techniques with probabilistic analysis/design methods. Many new methods have been suggested by researchers for reliability-based design optimization (RBDO) such as the performance measure approach (PMA) and the reliability index approach (RIA). Variants of PMA and RIA include the advanced mean value (AMV) method, conjugate mean value (CMV) method, moving least square (MLS) method and the hybrid mean value (HMV) method (see Choi and Youn, 2001 and Choi et al., 2001). Each of these variants provides a numerical tool for probabilistic constraint evaluation. An innovative framework for reliability based multidisciplinary design optimization is developed in Sues and Cesare, 2000 and modified in Sues et. al 2001. Pettit and Grandhi, 2000 have investigated the reliability-based design optimization of aerospace structures. Qu et. al., 2000 have investigated the use of response surface approximation for reliability-based optimization of composite laminates.

Advances in Reliability Based Design Optimization
In Agarwal and Renaud, 2002, a reliability based design optimization (RBDO) framework is developed for structural design applications. The framework accounts for variational uncertainties in the structural materials used in design. Performing non-deterministic design optimization under uncertainty requires approximately n^2 additional CPU time, where n is the number of design variables, as compare to deterministic design optimization. The RBDO framework of Renaud and Agarwal makes use of response surface techniques to significantly reduce the computational cost of performing reliability analysis. The resulting designs perform robustly with respect to the variability in the material properties of the system.

Variational uncertainty, which is associated with the randomness of physical quantities, is easily modeled by statistical means using probability and cumulative density functions. *Model and simulation uncertainty are much more difficult to characterize and have to be modeled using other means such as possibility theory, fuzzy sets, evidence theory, etc Oberkampf et al., 2001.* In Agarwal, et al., 2003, evidence theory is used to account for model and simulation uncertainty in multidisciplinary design optimization. Performance constraints are formulated using *belief and plausibility measures* developed using evidence theory. Evidence theory has received considerable attention in recent years from scientists at Sandia National Laboratories, where all designs must be certified through the exclusive use of modeling and simulation. This work builds on the earlier efforts of Gu et al., 2000 in which a robust design optimization strategy is

developed. In that study a methodology for estimating worst case propagated uncertainties in coupled multi-physics systems is developed. The worst case propagated uncertainties account for both the variability in physical quantities as well as the epistemic uncertainties (i.e., bias errors) in the numerical simulations used to drive the design process.

In Gu and Renaud, 2001, 2002 an implicit approach to modeling uncertainties within a multilevel optimization framework is developed. The multilevel optimization framework makes use of decomposition and therefore does not enforce consistency of the multidisciplinary analyses during the optimization. Implicit methods are required in order to estimate uncertainties.

Evidence Theory for Estimating Simulation Uncertainties

The reliability index approach discussed earlier does not account for uncertainties in the simulation models used to drive the optimization process. In general, a distinction can be made between aleatory uncertainty (also referred to as, variability, stochastic uncertainty, irreducible uncertainty, inherent uncertainty), and epistemic uncertainty (also referred to as model form uncertainty, reducible uncertainty, subjective uncertainty, or simply *uncertainty*) (see Oberkampf et al., 1998, 1999 2001). Material and manufacturing variations (aleatory uncertainties) can be modeled using probabilistic methods. While the uncertainty in simulation models (epistemic uncertainties) must be modeled using other mathematical theories, distinct from probability theory.

Agarwal et al., 2003 focuses on accounting for uncertainties in mathematical models (i.e., simulation tools) used for non-deterministic engineering systems design and optimization. The non-deterministic nature of the mathematical model of the system exists from the fact that: a) the system responses of the model can be non-unique due of the existence of uncertainties in the input parameters of the model, or b) there are multiple alternative mathematical models for the system and the environment. The simulation tool, however is deterministic in the sense that given the same input data, the simulation tool gives unique values of response quantities.

Agarwal et al., 2003 model epistemic uncertainty using the theory of belief measures, popularly known as *evidence theory* or *dempster-shafer theory* (see Klir and Wierman, 1998, Parsons, 2001). The advantage of using evidence theory lies in the fact that it can be successfully used to quantify the degree of uncertainty when the amount of information available is sparse. Like most modern uncertainty theories, evidence theory provides two uncertain measures. They are known as belief and plausibility. In Agarwal et al., 2003 the use of these uncertain measures are used to model non-deterministic performance constraints within a design optimization framework.

As part of this activity, the evidence theory approach developed in Agarwal et al., 2003 for optimization under uncertainty will be combined with traditional reliability approaches which account for aleatory uncertainties, in optimizing multidisciplinary systems. This approach will account for both the variation in material properties and manufacturing (aleatory uncertainty) and the imprecision in the simulation tools (epistemic uncertainty) developed for multidisciplinary design.

Single Level Method Development for Optimization Under Uncertainty

Many reliability based design optimization methods are typically nested (i.e., multilevel) optimization frameworks. The significant computational expense in evaluating reliability constraints derives from the fact that the constraint *reliability analysis* methods are themselves formulated as an optimization problems (Rackwitz, 2000). To overcome this difficulty, single level approaches have been proposed in Kuschel and Rackwitz, 2000, Wang and Kodiyalam, 2002, and Chen et al., 1997. Single level reliability based optimization problems can be formulated in two different ways. In one of the approaches, the reliability constraints (lower-level optimization) are replaced by Karush-Kuhn-Tucker (KKT) optimality conditions for the first order reliability problem (Kuschel and Rackwitz, 2000). This formulation suffers from some serious drawbacks. First, it requires Hessian information of the limit-state functions, which leads to the failure of the traditional optimization algorithms. Second, an explicit transformation from the standard normal space to the original space of random variables is necessary. Third, it is also limited to FORM and SORM approximations. The other single level approach appears to offer improved efficiencies. This formulation completely eliminates the lower level optimization of the constraints and the most probable

point (MPP) of failure is instead determined implicitly as the algorithm proceed towards the optima in a single level iteartive fashion Wang and Kodiyalam, 2002. *The development of computational improvements in single level reliability based design optimization will be investigated during the participant's visits.*

The participants are currently investigating the use of multilevel parallel optimization methods for the simultaneous analysis of disciplines within the Wang and Kodiyalam method. Initial results indicate significant computational savings when using the multilevel parallel optimizer as a driver. The method can be implemented in parallel and the opportunity to work with the parallel computation capabilities of Sandia will benefit this research.

Test Problems

Dr. Renaud's laboratory at the University of Notre Dame has access to a suite of smaller test problems for validating new numerical optimization methods. The opportunity to work with larger test problems available at Sandia such as the ICF capsule design problem is very appealing to the investigators. Dr. Renaud's laboratory has been developing a morphing aircraft optimization test problem involving an unmanned arial vehicle (UAV). The fully unstructured Navier –Stokes code FUN2D developed at NASA is used for the CFD analysis of the aircraft's performance.

Summary

The focus of this research participation is in the general area of optimization under uncertainty (OUU). This collaborative effort will allow the participants to investigate and develop a variety of methods including new parallel methods for simultaneous analysis within the OUU approach of Wang and Kodiyalam. The use of reduced order surrogate models within an interior point approach for optimization under uncertainty will be investigated. The use of evidence theory for estimating model form uncertainty combined with probabilistic methods to account for aleatory uncertainties within an OUU framework will also be investigated. This research will exploit the parallel computational capabilities available at Sandia. Personnel at Sandia will have an opportunity to attend seminars and a short course offered by the participants. The integration of uncertainty estimation within a framework of numerical optimization, under uncertainty offers designers new tools for validating and certifying improved designs in a simulation based design environment.

Title: System-sensitive Balancing and Octree/SFC Procedure

PI: James D. Teresco
Department of Computer Science, Williams College

Investigator: Luis Gervasio, Williams College
Jamal Faik, Williams College

Dates: September 29, 2003 – May 21, 2004

CSRI POC: Karen Devine, (505) 845-7585

Project Summary:

1 Research Background

My research interests are in dynamic load balancing for parallel adaptive scientific computation. In particular, I am interested in dynamic load balancing when the target computer systems are heterogeneous and/or hierarchical, and for highly adaptive problems.

Target parallel environments for modern, adaptive scientific computation range from the largest tightly-coupled supercomputers to heterogeneous clusters of workstations, with a variety of processor, memory, and communication hierarchies. Hierarchical and heterogeneous systems are increasingly common, a trend that is likely to continue with the increasing popularity of clusters of multiprocessors and the emergence of grid technologies, or metacomputing. This presents challenges for the development of efficient software, particularly influencing dynamic load balancing procedures. Whereas partitioning for a flat network with uniform processing nodes involves the determination of partitions of equal size and minimal boundary, optimal partitioning in other environments must consider the heterogeneity and hierarchy of processors, networks, and memories.

I have been investigating procedures for optimal task scheduling and dynamic load balancing in these nonuniform computational environments. I began this work by developing the Rensselaer partition model (RPM) [11] as part of my dissertation. Original RPM development was concerned with supporting the Meshing Environment for Geometry-based Analysis (MEGA) [10] and the Trellis framework [1]. RPM has also been used as the foundation of the parallel version of the Algorithm-Oriented Mesh Database (AOMD) [9]. The software interfaces with Sandia's Zoltan library [4] for dynamic load balancing.

My recent work on system-sensitive dynamic load balancing involves the design and implementation of a persistent and dynamic machine model for use with Zoltan, which incorporates relative capabilities of the composing blocks of the execution environment. The model has a tree structure that can easily be traversed by balancing procedures through a general interface that allows a topology-driven, yet transparent, execution of hierarchical partitioning. Each node in the tree contains its relative computing power and percentage of load. These capabilities are assessed initially by running a benchmark program and are updated dynamically by agents, threads that independently monitor the performance of each node while the user application is running. Processing nodes at the leaves of the tree have data representing the relative computing power of their processors. Network nodes have a bandwidth property and an aggregate computing power calculated as a function of the powers of their children and the network characteristics (e.g., bandwidth and latency). For load balancing purposes, we interpret a node's power, once normalized, as the percent of the load of its immediate parent that it should be assigned, based on its communication and computational capabilities. These relative “power" values can be used with any procedure that can produce variable-size partitions.

Preliminary results have shown significant benefit to this approach. We solve a two-dimensional Rayleigh-Taylor instability problem on a rectangular domain using eight processors of the Sun cluster at Williams College: five 450MHz Sparc UltraII processors, and three 300MHz Sparc UltraII processors, connected by fast (100 Mbit) Ethernet. Given an equal distribution of work, the fast nodes will be idle 1/3 of the time waiting for the slow nodes to complete their work. By giving 50% more work (as determined by the benchmarks) to each of the five fast nodes, an overall speedup of 24% is possible. We achieve a total wall clock time (including setup and all computation and communication) reduction from 3396 seconds to 2912

seconds, an improvement of 14.25%. We expect that the improvement in the actual computation time is much closer to the 24% theoretical maximum. Similar tests using the dynamic monitoring capabilities are underway. I have also been heavily involved in work on Zoltan's Octree/SFC partitioning procedure. Since an octree is easily constructed from an arbitrary mesh, the procedure is independent of any octree procedures that are used for mesh generation or adaptive refinement. Partitioning may be done in parallel [5, 6] by distributing the tree across the participating processors. Gervasio [7] developed the Octree/SFC partitioning procedure within Zoltan and studied traversals of the octree by space-filling curves (SFCs) corresponding to Morton, Hilbert, and Gray code orderings. The Hilbert ordering generally achieves the best partitions of the three orderings, which likely results from the greater locality of the Hilbert curve [3]. Campbell [2, 3] redesigned the SFC implementation to provide better efficiency, scalability, and extensibility. Gervasio has modified the Octree/SFC procedure to compute non-uniform partition sizes, to allow its use for system-sensitive load balancing experiments. Our continued interest in the Octree/SFC and other SFC-based partitioning and load-balancing procedures is motivated by their incremental nature. For highly-adaptive problems where dynamic load balancing is needed frequently, incremental procedures provide the necessary speed and limited data migration to allow for efficient rebalancing.

2 Proposed Research

During a visit to Sandia, I would continue work on both system-sensitive load balancing and on the Octree/SFC procedure. Working closely with the Zoltan group will help us to perform the proposed research more quickly, and will allow us to find new and interesting projects of mutual interest.

2.1 System-sensitive Balancing

- We will conduct a study to test our system-sensitive balancing procedures on a variety of heterogeneous and hierarchical systems of interest. We have worked primarily on a cluster of Sun servers and workstations at Williams College. We will test the current benchmarking and dynamic monitoring approach on other systems available at Sandia, Williams, Rensselaer, and elsewhere.
- Our current node power estimates take into account relative processor speeds and network capabilities. We would like to extend this model, including the dynamic monitoring agents, to include memory availability. A process that begins to use virtual memory often incurs unacceptable time penalties. We will place a threshold on the amount of work that can be assigned to one process based on its available main memory.
- We will use the machine model to guide a hierarchical load balancing procedure, to achieve partitions that use different algorithms or parameters, as appropriate, at different levels of the machine model hierarchy. Automating this requires that each Zoltan balancing procedure be able to produce numbers of partitions not equal to the number of processes. Such hierarchical balancing has the potential for benefit on any hierarchical system, but will become essential if distributed cluster architectures are to be utilized. Consider an adaptive computation on two clusters, connected by a campus-area network or even a wide-area network. These computations are enabled by grid tools such as MPICH-G2 [8]. Here, it is essential to minimize communication across the slow network interface. We intend to investigate the use of hierarchical balancing in this extreme situation.
- We plan to conduct a thorough analysis of the performance of Zoltan's load balancers. The studies will examine running time, load balancing accuracy, partition quality statistics such as surface indices (surface-to-volume ratios) and interprocess adjacencies, as well as incrementality, for real adaptive solution processes.

A preliminary study was conducted in 2002 [12]. Such information will be generally useful to Zoltan users, who could make better informed decisions when choosing from among Zoltan's procedures. It is of particular interest to the development of system-sensitive balancers that choose automatically from among the available procedures, based, at least in part, on their performance on the target architecture. We intend to include (i) a more thorough set of available procedures and parameters (HSFC, BSFC, Octree/SFC, more options of Jostle and Parmetis), (ii) a variety of computer systems (Sun cluster, PC clusters with different interconnects, networks of workstations, IBM SP and other ASCI-class supercomputers, metacomputers), (iii) a variety of applications (two- and three-dimensional simulations, highly adaptive, applications of interest to Sandia), and (iv) a variety of mesh structures (different element types, both simple and complex geometries and adaptive behavior).

2.2 Octree/SFC Procedure

For the Octree/SFC procedure, efficiency improvements and a detailed incrementality study are planned.

- We will consider a more adaptive approach to tree construction for the Octree/SFC procedure. The tree structure is already adaptive to the extent that a leaf octant is only refined if an object is inserted that overfills that octant, resulting in a deeper (more refined) tree in areas where objects are concentrated. However, we propose to take an approach similar to that taken by Zoltan's HSFC and BSFC algorithms and recursively refine only those octants that lie at partition boundaries until a satisfactory load balance is achieved. The startup cost of Octree/SFC would be reduced, yet the same partitions would be produced.
- As part of the performance comparisons described in the previous section, we will study the incrementality of Octree/SFC and other Zoltan procedures. A fast, incremental procedure is essential when solving highly adaptive problems. We will measure running times and the amount of data moved during load balancing in response to small changes in the mesh.
- We believe that Octree/SFC and other procedures in Zoltan can benefit from knowledge retained across successive load balancing invocations. While startup costs are often small relative to repartitioning and migration costs, we expect that startup costs will be more significant when highly-adaptive, terascale problems are being solved. Reconstruction of the octree (or analogous internal structures used by other Zoltan procedures) and reinsertion of all objects when only a fraction of these objects have changed requires work on the order of the total number of objects rather than the total number of objects that have changed. For Octree/SFC, this would require the fully-maintained tree structure, previously implemented by Campbell [2]. It would also require that objects' global Zoltan identifiers be consistent across calls. If an application can guarantee this, and it calls Zoltan functions to allow the update of the tree or other intermediate data structures when the collection of objects is modified, a truly dynamic and scalable repartitioning could be achieved. Similar approaches could benefit other Zoltan procedures that use intermediate structures to compute the partitions.

Title: Robust nonlinear solution algorithms for large-scale complex coupled nonlinear PDE systems

PI: Homer F. Walker, Professor
Worcester Polytechnic Institute

Dates: September 2003 – May 2004

CSRI POC: John Shadid, (505) 845-7876 and Roger Pawlowski, (505) 284-3740

Project Summary:

I am requesting support to visit Sandia National Laboratories in Albuquerque during my sabbatical from Worcester Polytechnic Institute in the 2003-2004 academic year. The specific proposal is to visit for nine two-week periods (one per month) from September 2003 through May 2004. During these visits and also the interim periods, I will pursue joint research with collaborators at SNL.

A major research activity during the sabbatical will be to advance and extend current work with John Shadid, Roger Pawlowski and others on robust nonlinear solution algorithms for large-scale complex coupled nonlinear PDE systems. This research will continue our current investigation of globalizations of inexact Newton methods applied to steady-state problems, with further development and study of linesearch and trust-region methods, and will also include consideration of new possibilities, such as nonlinear preconditioning (domain-based and multigrid), that may enhance robustness. We will also investigate extensions of Newton's method that use higher-order information, such as tensor methods. Broad application categories are likely to be (1) transport reactions systems with low heat release, such as fluid flow, CVD, and biological cell modeling, and (2) transport reactions systems with high heat release, such as combustion and catalytic reactors. The overall goal will be to develop methods and implementations that are robust and efficient when applied to these problems on massively parallel computers.

I will also explore opportunities for research in new areas with other investigators. Possibilities at present include working with (1) Scott Hutchinson and others on the Xyce team to explore nonlinear solvers and homotopy/continuation methods that exploit special structure in coupled device-circuit systems, (2) Andrew Salinger and others involved in Loca development to investigate linear and nonlinear solution methods in continuation and bifurcation algorithms, and (3) Rich Lehoucq, David Day, Teri Barth, and others to collaborate on eigensolvers and iterative linear algebra methods.

Chapter 4. Workshops/Sponsorships

The following workshops were either fully hosted by the CSRI or partially supported during 2003.

---

1) SOS7 – March 4-6 2003 – Pundit

2) Eleventh Copper Mountain Conference on Multigrid Methods, 2003 – March 30-April 3, 2003 – Shadid

3) Fast OS Workshop – July 8-10, 2003 – Brightwell

4) CSRI Student/Young-Investigator Support for the Seventh U.S. National Congress on Computational Mechanics – July 27-31, 2003 – Christon

5) Harvey Mudd College Mathematics Clinic – September 2003 – May 2004 – Romero, L.

6) 2nd Annual International Conference on IEEE Engineering in Medicine and Biology Society September 17-21, 2003 – May

7) The 2003 International Conference on Preconditioning Techniques for Large Sparse Matrix Problems in Scientific and Industrial Applications – October 27-29, 2003 – Womble

8) Solution Methods for Saddle Point Systems in Computational Mechanics Workshop December 3-6, 2003 – Lehoucq

Title: SOS7 Workshop

PI: Neil Pundit, Sandia National Laboratories,
Bill Camp, Sandia National Laboratories

Dates: March 4-6, 2003

CSRI POC: Neil Pundit, (505) 845-7601

Project Summary:

This is a formal request for CSRI Sponsorship of the SOS7 Workshop on Distributed Supercomputing to be held in Durango, CO in late Feb or early March of 2003.

This will be the 7th year of the Workshop started jointly by Sandia, Oakridge, and the Swiss Inst of Technology. The 2002 Workshop was held in Switzerland, and the 2003 (host SNL), and 2004 (host ORNL) will be held in the USA.

In 2000 the Workshop was opened to all DOE Labs, collaborating universities, and selected vendors. The expected attendance (by invitation) is about 70.

The 2003 program plans to have the following:

- a session on new machines (Red Storm, ASCI Purple, PNNL's cluster, ..)
- a session on reports from recent workshops (Architecture, Fast-OS, Cluster Perf Enhancement, and Scalable IO, etc)
- a session on selected state-of the-art papers
- key note speeches
- a few panel discussions on controversial topics.

While the details of the Program are being worked out jointly, one of the traditions of this Workshop is ample discussions on topics of mutual interest in distributed supercomputing. I am the Technical Chair, and Bill Camp is the host.

Title: Eleventh Copper Mountain Conference on Multigrid Methods, 2003

PI: John Shadid, Sandia National Laboratories

Dates: March 30-April 3, 2003

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

Project Summary:

We are requesting funding support in the amount of $5,000 from the Computer Science Research Institute for the Copper Mountain Conference on Multigrid Methods, 2003 in Copper Mountain, Colorado.

The Eleventh Copper Mountain Conference on Multigrid Methods will take place March 30-April 3, 2003. The web site is http://amath.colorado.edu/faculty/copper/2003/. The conference purpose is to present cutting edge information and technology in the Multigrid mathematical area. We have academic, governmental, and commercial attendance. We also have a student paper competition from around the world, so it also works as a job fair for students entering the marketplace post PHD.

Title: Fast-OS Workshop

PI: Ron Brightwell, Sandia National Laboratories,
Barney Mccabe, University of New Mexico

Dates: July 8-10, 2003

CSRI POC: Neil Pundit, (505) 845-7601

Project Summary:

We wish to hold another FAST-OS Workshop in Washington DC area (Doubletree in Rockville, MD), on July 8-10. The goal of the workshop is as follows:

- share in the developments since the last workshop, particularly LWK work
- be cognizant of the accepted ideas in runtime system in the HECRTF workshop (June 16-18, in Wash DC)
- identify areas of concentration and researchers and organizations to carry out an expanded program of $3M to be funded by MICS in FY04 in which Sandia plays a leading role - identify gaps remaining which put the realization of a petaflops in 2010 at risk

The expected attendance is 35; the organizers are Ron Brightwell and Barney Maccabe (UNM).

I estimate the total cash budget to be $4.9K at the most. Your kind approval is highly appreciated.

Title: CSRI Student/Young-Investigator Support for the Seventh U.S. National Congress on Computational Mechanics

PI: Mark Christon and Thomas Bickel

Dates: July 27-31, 2003

CSRI POC: Mark Christon

Project Summary:

The 7th US National Congress on Computational Mechanics (USNCCM7) will be the first where a national laboratory is taking the lead organizational role. In this capacity, Sandia will have a unique opportunity to showcase its research capabilities at both the national and international level. As a part of the USNCCM7, we are seeking support to provide Congress Fellowships for student and young-investigators that wish to participate in the technical program. Our goal is to recruit approximately 75 outstanding young-investigators and students for the 2003 meeting. A Congress Fellowship will consist of a fixed award of approximately $500 for students and $1000 for young-investigators. All Fellowship awardees are expected to present a paper at the Conference. The funds provided by the CSRI will be used with the Fellowship support funds from the Lawrence Livermore and Los Alamos National Laboratories, the National Science Foundation, and industry. CSRI funds will only be used to provide Fellowship support for US citizens.

Conference Background and Scope

The US National Congress on Computational Mechanics is the official congress of the U.S. Association for Computational Mechanics which is an affiliate of the International Association for Computational Mechanics. The biannual Congresses of the U.S. Association for Computational Mechanics are major scientific events that attract an international audience. The seventh U.S. Congress, hosted by Sandia National Laboratories, will highlight the latest developments in all aspects of computational mechanics --- from new applications in nanotechnology and bioengineering to recent advances in numerical methods and high-performance computing. The technical program features invited plenary lectures by distinguished experts as well as minisymposia that focus on topics of current scientific interest. Participants will have a special opportunity to experience the unique scientific environment of Sandia National Laboratories through interactions with research staff and to envision its future in initiatives such as the Center for Integrated Nanotechnologies (CINT) and the Microsystems and Engineering Sciences Applications Complex (MESA). A number of short courses and vendor exhibits are planned, and the social program will take advantage of the numerous recreational and cultural offerings of the scenic Albuquerque region. Additional details on the conference, such as the scientific program committee, may be found at http://www.esc.sandia.gov/usnccm.html.

Title: Harvey Mudd College Mathematics Clinic

PI: Louis Romero

Dates: September 2003 – May 2004

CSRI POC: Louis Romero, (505) 845-7512

Project Summary:

Topic in Geo-location
This project aims at addressing two different problems involving a geo-location system that is very similar to GPS. It differs from GPS in that it uses frequency of arrival as well as time of arrival information. This system is already used on two SPS satellites, and will soon be added to a third. There is an excellent change that the results from this project will be implemented on the next generation of Sandia's geo-location systems.

Mathematically, our geo-location problem can be formulated as a system of five quadratic equations in five unknown. The first problem involves finding robust techniques for numerically solving these equations. Techniques for solving these sorts of equations (using resultants) were developed in the 19^{th} century and can be found in older books involving the theory of equations. However, these classical references pay no attention to issues concerning the numerical stability of these algorithms.

Recently Steve Vavasis (Cornell University) has developed a technique (for a different problem) that is a modern day implementation of these classical techniques. We would like to implement this technique for the geo-location problem. This would involve

- Understanding the formulation of the geo-location problem.
- Understanding the mathematics of simultaneous systems of quadratic equations.
- Writing a simple MATLAB code to implement the algorithm for solving these equations.

The main goals for this project should be easily obtainable in a year's time. However, we could push forwards along these lines to learn more about the solutions of these sorts of equations. These sorts of equations occur in many different classes of problems where distance constraints are involved. In particular, in robotics and protein folding.

Depending on the time and interest of the team members, there is a second problem involving correcting for the wave propagation through the ionosphere. At the current time errors from ignoring this effect are the main obstacle to getting better resolution in geo-location systems. Analyzing these effects will require numerical analysis signal analysis, and the theory of dispersive wave propagation.

Title: 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society

PI: Elebeoba (Chi-Chi) May

Dates: September 17-21, 2003

CSRI POC: Elebeoba (Chi-Chi) May, (505) 844-9933

Project Summary:

I am requesting funding in the amount of $4000.00 for the Communication Theory, Coding Theory and Molecular Biology special session which is part of the 25th ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY (September 17-21, 2003 in Cancun, Mexico).

The proposed special session, Communication Theory, Coding Theory and Molecular Biology, will focus on advances in the application of information technology concepts to the study of molecular systems and processes as well as the use of information technology in the design and development of molecular computing (DNA computing). This session is an excellent example of the intersection of traditional engineering with biology and exemplifies truly cutting-edge work in the extension of communication technology methods and algorithms beyond traditional challenges. This area is highly correlated with Sandia's research in computational sciences, algorithms, and biotechnology. Support of this special session will showcase Sandia's leadership initiative in the application of computational sciences and algorithms to new research frontiers that can potentially impact our nation.

Title: The 2003 International Conference on Preconditioning Techniques for Large Sparse Matrix Problems in Scientific and Industrial Applications

PI: Esmond G. Ng, Lawrence Berkeley National Laboratory

Dates: October 27-29, 2003

CSRI POC: David Womble (505) 845-7442

Project Summary:

Conference support requested for: The 2003 International Conference on Preconditioning Techniques for Large Sparse Matrix Problems in Scientific and Industrial Applications. Location of the conference: The Embassy Suites Napa Valley, Napa, California. Dates of the conference: October 27-29, 2003

Several invited speakers at the conference will present their recent work on preconditioning techniques that will have significant impact on the large-scale simulation efforts in the ASCI Program. For example, Jorge Nocedal of the Northwestern University, who is well known in the numerical optimization community, will speak on preconditioning techniques for large-scale optimization; Henk van der Vorst of the University of Utrecht in the Netherlands, who is an expert in iterative methods, will discuss his recent work on preconditioning for indefinite linear systems; Bruce Hendrickson of the Sandia National Laboratories will speak on tools for analyzing preconditiones; and Panayot Vassilevski of the Lawrence Livermore National Laboratory will present a framework for algebraic multigrid. Advances in these areas are relevant to many areas of the ASCI Program.

Title: Solution Methods for Saddle Point Systems in Computational Mechanics Workshop

PI: Richard Lehoucq, Sandia National Laboratories,
John Shadid, Sandia National Laboratories
Ray Tuminaro, Sandia National Laboratories
Gene Golub, Stanford University

Dates: December 3-6, 2003

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

Project Summary:

Motivation
The numerical approximation of several important scientific and engineering applications described by PDE systems lead to block structured indefinite linear systems. Applications include: Stokes and Navier-Stokes fluid flow; incompressible elasticity formulations; mixed formulations for Poisson type problems; and limiting cases of electromagnetics problems. In these applications, the successful design of robust, scalable, and efficient preconditioners for Krylov methods is intimately connected with an understanding of the structure of the resulting block matrix system. Effective preconditioners are often based on an approximate block decomposition of the system that carefully considers the spectral properties of the component block operators and the Schur complement operators. Through this purely linear algebraic view of preconditioning, a simplified system of block component equations is developed that encodes a specific "physics based" decomposition. Recent progress using these ideas has demonstrated the construction of a number of effective preconditioners with nearly optimal convergence in various applications. Our workshop is interested in a general view of these applications and methods to enable further research efforts.

Purpose of the Workshop
The purpose of this workshop is to bring together leading researchers in the design of solution methods for saddle point systems. Our intent is to host a small workshop (~25 people), of international researchers from a number of fields. A series of invited talks in a workshop format (45 min talks, 15 min for questions, with no parallel sessions) is planned. These talks will be selected to highlight recent progress in the field, facilitate interactions, and to provide an over-view of current research. Formal workshop and informal interactions will be strongly encouraged and the schedule will be designed to promote these interactions. We believe this activity will serve as a catalyst to accelerate further developments in this area. On a Sandia specific (ASCI) view of the proposed workshop, our goal is to invite leading Sandia ASCII algorithm research PIs to interact with this group. Our intent is to use this forum to provide a concentrated, intensive view of the sate-of-the-art in this field that generalizes understanding of these methods to a larger set of computational mechanics applications. Through the workshop environment, we not only expect significant transfer of technical information and techniques but also the possible development of collaborative efforts that will support our ASCI Algorithms research in critical areas.

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 seminars are listed below in alphabetical order by speaker.

Title: A Unilevel Method for Reliability-Based Design Optimization

Speaker: Harish Agarwal, University of Notre Dame
Design Automation Laboratory

Date/Time: Wednesday September 24, 2003, 3:00-3:30 pm

Location: Building 980 Room 95

Brief Abstract: Reliability based design optimization (RBDO) deals with finding optimum designs that are characterized by a low probability of failure. Traditional RBDO involves the solution of a nested-optimization problem which is computationally intensive. Present research efforts are focused towards the development of an efficient and robust unilevel method for performing RBDO. The proposed method is mathematically equivalent to the original nested-optimization formulation as it satisfies the constraint qualification of the first-order KKT conditions. Numerical examples will be used to illustrate the methods potential.

CSRI POC: Mike Eldred, (505) 844-6479

Title: Models of protein interactions at the cellular and molecular length scales

Speaker: Susan R. Atlas, University of New Mexico

Date/Time: Monday, September 22, 2003, 9:00-10:00 am

Location: Building 980 Room 95

Brief Abstract: New experimental technologies such as gene expression microarray analysis, high-throughput protein screening, and techniques for the precise manipulation of atoms and molecules at the nanoscale are transforming our ability to study protein interactions within the cellular environment. This flood of new data is providing essential clues for deducing mechanisms of signal transduction in gene regulatory networks, with application to DNA damage response cascades in diseases such as cancer. At the molecular level, it is enabling us to begin unlocking the mechanisms by which motor proteins such as myosin and kinesin naturally convert chemical energy from the reaction of ATP into mechanical motion. A deeper understanding of the atomic interactions governing these processes would potentially enable the tailored design of novel biomimetic materials for nanotechnology applications, as well as providing insight into fundamental biophysical processes such as ion transport through cell membranes, interfacial phenomena, protein folding in solution, and catalysis.

In this talk, I will describe recent work on these two distinct aspects of protein interactions, performed in collaboration with colleagues at UNM and Los Alamos. In the first part of the talk, I will describe how the formal reasoning framework of Bayesian networks is being used to represent and merge the complex and context-dependent biological factors that can influence protein interactions within a regulatory subnetwork. In the latter part of the talk, I will present a hierarchical theory of atomic interactions that provides a unified, multiscale framework for incorporating quantum-mechanical charge-distortion and charge-transfer effects into atomistic potentials for the dynamical simulation of biophysical systems and materials.

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

Title: Metrics for Information Extraction and Recognition

Speaker: Alex Backer, California Institute of Technology

Date/Time: Monday, March 3, 2003, 11:00-12:00 am (MST)

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

Brief Abstract: The fundamental problem of machine and biological learning for recognition consists of storing patterns and then retrieving them upon occurrence of a similar or related pattern. How 'similar or related' is defined is thus critical. Good metrics are crucial to the robustness of the recall operation in the face of noise and variation between different instances of the patterns of interest.

I will introduce a cost-based metric for sparse binary time series, related to the Sellers algorithm for evolutionary distances between genetic sequences, and an algorithm based on dynamic programming that allows the efficient calculation of minimal cost transformations. These allow the 'recognition' of the sensory stimulus presented to an animal in any given trial. Despite previous claims that the signal-to-noise ratio of single olfactory bulb neurons is of 10^-4, I will show use of an appropriate metric allows the recovery of enough information about odor identity in noisy spiking patterns to classify correctly among several stimuli on up to 97% of all trials.

I will compare the performance of different read-out codes. An extension of the algorithm to multiple neurons will be presented and used to show that spikes from different neurons have different "meaning" in terms of the information they convey about the stimulus. This allows the computation of the non-redundant information per neuron, allowing a prediction of the total bandwidth of an insect olfactory system. I will discuss several improvements to the classification algorithm that greatly reduce computation time. An application of the metric to the detection of significant changes in spiking patterns will be presented and used to demonstrate learning in neuronal responses. I will briefly discuss a proposed extension of the cost-based metric approach to the recognition of images and multidimensional objects in other domains.

Of particular interest among the interactions between multiple neurons is synchronization, an ubiquitous characteristic of neuronal responses whose function has remained speculatory for decades. I will use the algorithms described to show synchronization is critical for the readout of neuronal assemblies by downstream neurons, and briefly discuss work in progress to define measures of synchronization and higher-order interactions for large numbers of non-periodic time series.

If time allows, I will touch briefly on a novel class of nonlinear learning systems termed Winnerless Competition (WLC) introduced to model the olfactory system, and show that a local learning rule can create the global connectivity pattern required to create the dynamic attractors characteristic of WLC.

CSRI POC: Bill Hart, (505) 844-2217

Title: A Tensor-Krylov Method for Solving Systems of Nonlinear Equation: An Overview with Future Extensions

Speaker: Brett Bader, Interview Candidate, University of Colorado

Date/Time: Tuesday, March 4, 2003, 10:00-11:00 am (MST)

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

Brief Abstract: This talk will present an overview of an iterative, Krylov-subspace based approach for solving large-scale systems of nonlinear equations using tensor methods. Standard tensor methods have performed especially well on small, dense problems where the Jacobian matrix at the solution is singular or ill conditioned, something that happens on many classes of large-scale problems, such as bifurcation tracking.

This research extends tensor methods to large-scale problems by developing a tensor-Krylov method that bases each iteration on a linear model augmented with a limited second-order term and employs a line search globalization scheme that imitates a trust region method. Two advantages over existing methods are its ability to solve the local tensor model to a specified accuracy and its reliance on block methods, which can be more memory efficient than standard techniques. Numerical comparisons with Newton-GMRES on several benchmark PDE problems are presented and provide evidence that tensor-Krylov methods are more robust and efficient than Newton-GMRES on some important and difficult problems. In addition, the tensor-Krylov method still exhibits the superlinear convergence behavior of direct tensor methods on singular and ill-conditioned problems. Future extensions to tensor method research will also be discussed in some detail.

CSRI POC: Roger Pawlowski, (505) 284-3740

Title: Uncertainty models in structural engineering

Speaker: Michael Beer, Rice University

Date/Time: Tuesday, November 18, 2003, 11:00-12:00 noon (MST)

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

Brief Abstract: Structural behavior and structural safety can be realistically assessed only if the uncertainty in the structural parameters is appropriately taken into consideration and realistic computational models are applied. Uncertainty of data and models must be accounted for in its natural form. Stochastic models are not always capable of fulfilling this task without restrictions, as uncertainty may also be characterized by fuzzy randomness or fuzziness.

On the basis of fuzzy set theory a general method for fuzzy structural analysis is presented. This method is formulated in terms of a-level optimization combined with a modified evolution strategy. Every analysis algorithm for the realistic simulation of load-bearing behavior may be applied in this fuzzy structural analysis in the sense of a deterministic fundamental solution.

A fuzzy probabilistic safety concept is introduced on the basis of the theory of fuzzy random variables. This concept permits fuzziness, randomness and fuzzy randomness to be accounted for simultaneously. All uncertainty of structural parameters is appropriately mapped onto fuzzy structural responses and the fuzzy failure probability.

The fuzzy structural responses and the fuzzy safety level are compared to permissible values and assessed using an analogon of Shannon's entropy and defuzzification algorithms. Referring to permissible structural responses and a required safety level uncertain structural design parameters are derived by applying a fuzzy cluster analysis algorithm to the computed fuzzy results.

The comprehensive application of the developed algorithms is demonstrated by way of example. Hereby, a geometrically and physically nonlinear algorithm for the analysis of reinforced concrete structures is adopted.

CSRI POC: Vicente Romero, (505) 844-5890

Title: Discrete Sensor Placement Problems in Distribution Networks

Speaker: Tanya Berger-Wolf, University of New Mexico

Date/Time: Thursday, September 4, 2003, 10:00-11:00 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: We consider the problem of placing sensors in a building or a utility network to monitor the air or water supply. Our goal is to quickly detect and identify the source of any supply contamination. Various related versions of this problem have been studied and modeled numerically. We propose discrete models of the sensor placement problems that aim to either minimize the number of sensors given a time limit on the contamination detection time or minimize contamination detection time given a fixed budget of sensors. We give the necessary and sufficient condition for source identification and analyze the detection model. We believe that the combinatorial optimization approach to this problem is useful both as a stand alone solution method and in combination with the traditional continuous approach, providing an initial solution that can be improved numerically.

CSRI POC: Bill Hart, (505) 844-2217

Title: Using Stochastic Programming Problem Structure to Gain Computational Efficiency

Speaker: John R. Birge, Northwestern University

Date/Time: Thursday, January 23, 2003, 10:00-11:00 am (PST)

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

Brief Abstract: The inclusion of random parameters into optimization problems leads to the formulation of stochastic programs that have specific problem structures. Standard optimization procedures may not take advantage of these structures, leading to long solution times or inaccurate results. In this talk, we will describe the basic form of stochastic programs and common structures. We will describe techniques for taking advantage of these structures in algorithms based on interior point, decomposition, and Lagrangian methods.

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

Title: New Approach for Characterization of Binding-Site Search Space

Speaker: Mike Brown, Interview Candidate, University of New Mexico

Date/Time: Tuesday, April 1, 2003, 10:30-11:30 am

Location: Building 980 Room 95 (Sandia – NM)
Building 916 Room 101 (Sandia – CA)

Brief Abstract: A new approach for defining the Cartesian spatial boundaries of binding pockets is presented. The method involves calculation of a macromolecule-encapsulating surface (MES) that separates binding pocket volume from outside space. The surface provides means for identification of binding sites and calculation of their volume. Additionally, the MES can be used to limit the search space for ligand docking and de novo design algorithms via identification of accessible atoms within the binding pocket or limitation of translation ranges to binding pocket space. The approach has been shown to be efficacious based on testing with 50 enzyme-ligand complexes for which the binding pockets are known. Additionally, we have modified the flexible docking program AutoDock 3.0 to incorporate MES boundaries using an energetic term. The results show increased efficiency of the genetic algorithm for ligand docking characterized by a larger percentage of successful runs and a decrease in required run times. MES incorporation also facilitates search of an entire enzyme for ligand docking, without the requirement of a predetermined binding pocket location.

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

Title: Numerical Solutions of Hyperbolic Conservation Laws: Incorporating Multi-Resolution Viscosity Methods into the Finite Element Framework

Speaker: Marcus Calhoun-Lopez, Iowa State University

Date/Time: Wednesday, June 25, 2003, 10:00-11:00 am

Location: Building 980 Room 95

Brief Abstract: It is well known that the classic Galerkin finite-element method is unstable when applied to hyperbolic conservation laws, such as the Euler equations for compressible flow. Adding a diffusion term to the equations stabilizes the method but sacrifices accuracy in regions of sharp gradients. An elegant solution devised for spectral methods is to add diffusion only to the high frequency modes of the solution, which stabilizes the method without overly smoothing the sharp gradients.

This idea has been incorporated into the finite-element framework by using hierarchical functions as a multi-frequency basis. For this method, we are able to prove convergence for a one-dimensional scalar conservation law. Numerical results will be presented for one- and two-dimensional hyperbolic conservation laws.

Marcus Calhoun-Lopez is a student intern from Iowa State University. He has been at Sandia for the last year working on his dissertation, and will be going to the University of Maryland to begin post-doc research with Eitan Tadmor. He plans to continue development of the hierarchical basis functions under the SIERRA/Premo framework.

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

Title: A Brief Introduction to UPC

Speaker: William W. Carlson, IDA Center for Computing Sciences

Date/Time: Tuesday, September 23, 2003, 2:00-3:00 pm

Location: Building 980 Room 95

Brief Abstract: UPC is a parallel extension of the C programming language which provides programmers with a shared global address space. A descendant of Split-C, AC, and PCP, UPC has two primary objectives: 1) to provide efficient access to the underlying machine, and 2) to establish a common syntax and semantics for explicitly parallel programming in C. The quest for high performance means in particular that UPC tries to minimize the overhead involved in communication among cooperating threads. When the underlying hardware enables a processor to read and write remote memory without intervention by the remote processor (as in the Cray T3E), UPC provides the programmer with a direct and easy mapping from the language to low-level machine instructions. At the same time, UPC's parallel features can be mapped onto existing message-passing software or onto physically shared memory to make its programs portable from one parallel architecture to another. As a consequence, UPC has seen implementations on a variety of hardware platforms and is becoming a standard language.

CSRI POC: Zhaofang Wen, (505) 284-0206

Title: A Distributed Passive File Caching Scheme

Speaker: Kenin Coloma, SNL Student Intern, Northwestern University

Date/Time: Thursday, September 4, 2003, 1:00-2:00 pm

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: We're designing a client-side file-caching scheme that uses passive MPI Remote Memory Access (RMA) for both data transfer and coherency. So far we've concentrated our efforts on a standalone system, which allows only one process to cache a section of file at any given time. By maintaining a distributed table of where each file block is cached, we allow for access to remote caches, and can enforce a lax or strict consistency model. After initialization of the system, there need not be any barriers or any cooperative communication between processes until tear down of the system. We look at some further design considerations and some other near term goals.

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

Title: Exact Analytical Formulation for coordinated Motion in Polypeptide Chains

Speaker: Evangelos A. Coutsias, University of New Mexico

Date/Time: Wednesday, April 2, 2003, 10:00-11:00 am (MST)

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

Brief Abstract: We present a new exact formulation for the motion of a segment of a polypeptide chain under coordinated changes in the "soft" degrees of freedom ϕ_i, ψ_i. Our method differs from previous analyses in its use of a reduced set of intrinsic angle coordinates τ_i describing the orientation of rigid subunits of the chain and it allows non-canonical values for the C^{α} bond angles and for the ω dihedrals. Backbone motions are performed by means of the concerted movement of pairs of adjacent rotatable bond vectors at a fixed angle to each other and constrained to lie on two conjoined cones. The resulting systems of "2-cone" equations in the τ_i can be analyzed exhaustively and all possible solutions are found as the real eigenvalues of a sparse generalized eigenvalue problem of size 24x24. Applications to date include a new fast algorithm for homology loop modeling and a new set of local moves for loop refinement through Monte Carlo Energy Minimization.

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

Title: A column pre-ordering strategy for the unsymmetric-pattern multifrontal method

Speaker: Tim Davis, University of Florida, Stanford University,
and Lawrence Berkeley National Laboratory

Date/Time: Monday August 11, 2003, 3:00– 4:00 pm MDST

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

Brief Abstract: A new method for sparse LU factorization is presented that combines a left-looking column pre-ordering strategy with a right-looking unsymmetric-pattern multifrontal numerical factorization. The column ordering is selected to give a good a priori upper bound on fill-in and then refined during numerical factorization (while preserving the bound). Pivot rows are selected to maintain numerical stability and to preserve sparsity. Existing left-looking methods do not select pivot rows to preserve sparsity. As a result, the new method nearly always obtains better orderings than existing left-looking methods (such as that used in MATLAB 6.1), and the prior unsymmetric-pattern multifrontal method on which it is based (MA38).

Recent results will be present comparing UMFPACK v4.2 with a recent implementation of Gilbert & Peierl's sparse partial pivoting algorithm, with Eisenstat & Liu's symmetric pruning, and a symmetric preordering (AMD). The results show that this simpler code (500+ lines) is significantly faster than UMFPACK v4.2 (30,000+ lines of code) for circuit simulation matrices. Exactly the opposite occurs for matrices from most other disciplines, and even for some circuit simulation domains (frequency-domain analysis, for example), where UMFPACK is much faster.

CSRI POC: Mike Heroux**, (**320) 845-7695

Title: Talk on Russian Supercomputing Tech. (Russian ASCI program)

Speaker: Erik DeBenedictis, SNL, Org. 09223

Date/Time: Friday, October 17, 2003, 2:00-3:00 pm

Location: Building 980 Room 95

Brief Abstract: I will give an overview of Supercomputer technology in Russia's NW establishment and give an overview of doing business with Russians in Russia.

This is a report from my attending a Supercomputing conference last week at the Russian NW facility known as VNIIEF in Sarov (and formerly known as Arzamas-16). This is the Russian equivalent of Los Alamos.

Topics:

* History of Russia's acquisition of Supercomputer technology and export control

* Current state of Supercomputer technology. Specifically, Russia has developed a domestic supercomputer capability similar in some ways to Cplant. This capability includes a broad range of systems and systems software efforts by a highly organized group at VNIIEF.

* The overall supercomputing effort in Russia is similar to ASCI in many ways. This includes attempts to "push" Moore's Law, a powerful effort in algorithms and applications, and a V&V effort.

* I will have copies of the English language abstracts to the conference. This is a 100-page document.

* A discussion of the issues involved in dealing with the Russians, including travel and security barriers and the generous hospitality the VNIIEF lab offers to DOE personnel in an effort to get business.

* The VNIIEF organization is interested in business with counterpart organizations at Sandia, LLNL, and LANL. There are many agreements in place that impact this sort of thing in both directions.

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

Title: Fast Eigenvalue/eigenvector Computation for Dense Symmetric Matrices

Speaker: Inderjit S. Dhillon, University of Texas at Austin

Date/Time: Monday, August 4, 2003, 1:30-2:30 pm

Location: Building 980 Room 95

Brief Abstract: Computing the eigenvalues/eigenvectors of a dense symmetric matrix is a classical problem and usually proceeds in 3 phases: (i) reduce the matrix to tridiagonal form, (ii) solve the tridiagonal eigenproblem and (iii) back-transform to obtain eigenvectors of the original matrix. In this talk, we present the first O(n^2), numerically stable, embarrassingly parallel algorithm for phase (ii). All previous algorithms for this purpose were O(n^3) in complexity, and in practice, often accounted for 80% of the total time in solving the dense eigenproblem. The new algorithm achieves O(n^2) complexity by avoiding all Gram-Schmidt orthogonalization.

The crucial ingredients of the new algorithm are:
(1) using factored forms to achieve provably high relative accuracy,
(2) using "twisted factorizations" to compute approximate eigenvectors with provably small relative residual norms,
(3) using multiple factored forms to "separate" clustered eigenvalues, thereby automatically obtaining numerically orthogonal approximations to eigenvectors (no Gram-Schmidt).

An interesting facet of our work is that high accuracy in intermediate computations leads to a much faster overall algorithm.

After describing the algorithm, I will give detailed numerical results. Impressive speedups are obtained on matrices from the real-life applications of computational quantum chemistry (PNNL) and finite-element modeling (from Jeff Bennighof). The largest dense problem solved "in-core" on a 256-node parallel computer is of size 128,000 x 128,000 which takes about 8 hours of CPU time. A preliminary version of the sequential software is now included in the LAPACK public-domain library.

The work on the sequential algorithm is joint with Beresford Parlett of UC Berkeley, while the parallel algorithm has been developed jointly with Paolo Bientinesi and Robert van de Geijn of UT Austin.

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

Title: When Energy Isn't Enough: Topics in Nonequilibrium and Mesoscale Phenomena

Speaker: Brian Dodson, SNL, Interview Candidate

Date/Time: Thursday, January 16, 2003, 9:00-10:00 am

Location: Building 980 Room 95

Brief Abstract: Properties of materials at elevated pressure and temperature have long been studied using a combination of quasi-static and shock-wave experimental techniques and continuum and atomic-scale theoretical methods. A widespread feature of this body of work is that much of the physics and chemistry underlying the actual behavior of a given material is encoded in empirical or semiemperical effective parameters. The use of such parameters often allows experimental results to be approximately fit to theoretical models, albeit with ill-defined and often narrow limits on the regime of acceptable accuracy. Similar limitations commonly appear in the study of nanomechanical effects, for example friction and adhesion, and particularly in the case of incommensurate surfaces. This talk will discuss a variety of nonequilibrium and/or mesoscopic phenomena, which serve to illustrate hidden assumptions in current models and potential approaches toward achieving improved predictability.

CSRI POC: John Aidun (505) 844-1209

Title: Efficient Iterative Algorithms for the Stochastic Finite Element Method with Application to Acoustic Scattering

Speaker: Howard Elman, University of Maryland

Date/Time: Thursday, June 12, 2003, 11:00 am -12:00 noon (MST)

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

Brief Abstract: In this study, we describe the algebraic computations required to implement the stochastic finite element method for solving problems in which uncertainty is restricted to right hand side data coming from forcing functions or boundary conditions. We show that the solution can be represented in a compact outer product form, which leads to efficiencies in both work and storage, and we demonstrate that block iterative methods for algebraic systems with multiple right hand sides can be used to advantage to compute this solution. We also show how to generate a variety of statistical quantities from the computed solution. Finally, we examine the behavior of these statistical quantities in one setting derived from a model of acoustic scattering.

Joint work with Oliver Ernst, Dianne O'Leary and Michael Stewart.

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

Title: PERCS: IBM Effort in HPCS

Speaker: Mootaz Elnozahy, IBM

Date/Time: Tuesday, October 28, 2003, 2:00-3:00 pm

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: The High Productivity Computing Systems (HPCS) is a DARPA-sponsored initiative that signals a fundamental shift in the way high-end computing systems are to be built and evaluated. Instead of the traditional myopic focus on performance as the most important system property, users in the technical computing community now have a broader definition of productivity of a system. This definition includes issues of usability, robustness, system management, and ease of programming. A productive system is one that delivers a high level of performance while scoring equally well on the other aspects of the system. Recognizing the difficulty of this task, the HPCS program aims at reinvigorating the research community by sponsoring groundbreaking ideas that could yield a commercially viable system for the 2010 timeframe. This talk will cover the general vision behind our effort and how we envision adaptable systems that could well both on commercial and technical workloads.

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

Title: How does Non-Symmetry affect Performance of Iterative Eigensolvers?

Speaker: Mark Embree, Rice University

Date/Time: Tuesday, June 17, 2003 10:00-11:00 am

Location: Building 980 Room 95

Brief Abstract: Large symmetric eigenvalue problems are routinely solved via the Lanczos algorithm, whose popularity benefits from a mature convergence theory. Nonsymmetric eigenvalue problems pose a greater computational challenge for two reasons: they require more intricate algorithms (Arnoldi, bi-orthogonal Lanczos), and the eigenvalues themselves can be ill conditioned.

Such ill conditioning can have important consequences for both the underlying application, and the performance of algorithms designed to compute such eigenvalues. In this talk, we will illustrate how ill conditioned eigenvalues can complicate the performance of Krylov subspace algorithms. In this context, one seeks only some subset of the spectrum relevant to the application. Using recent results with Christopher Beattie and Dan Sorensen, we will demonstrate how ill-conditioning of these desired eigenvalues has little impact on convergence behavior, while ill-conditioning of the part of the spectrum we do not wish to compute has a central impact, manifested in both convergence plateaus and the asymptotic convergence rate. In particular, we will demonstrate a perplexing implication of our convergence bounds: restarting an eigenvalue algorithm using information derived from exact knowledge of the unwanted spectrum can perform far worse than similar methods based only on approximate knowledge of those eigenvalues.

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

Title: MOLECULES HAVE SIZE
(Electronic Hopping Between Large Molecules)

Speaker: David Emin, University of New Mexico

Date/Time: Thursday, January 30, 2003, 10:30-11:30 am

Location: Building 980 Room 95

Brief Abstract: Electronic carriers can move between large polyatomic molecules by phonon-assisted hopping. Such inter-molecular hopping is often described with a theory developed to describe electronic hopping between monatomic ions. In particular, the electronic states were assumed to be small enough to be approximated simply as points.

Here the theory of multi-phonon hopping is extended to address motion between finite-size molecules. Increasing the size of the electronic states between which carriers hop increases the range of the predominant electron-phonon interactions. Then the hopping activation energy increases strongly with the jump distance.

The point-state model necessarily ignores intra-molecular electronic redistribution in response to atomic motion. These effects alter the magnitude and temperature dependence of the intermolecular jump rate.

All told, this work details qualitative differences between the physics of electronic hops between large states and hops between severely localized states. These size-related effects can explain qualitative differences between observations of inter-ionic and inter-molecular hopping transport.

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

Title: Anisotropic Mesh Adaptation: Application to CFD Simulations

Speaker: Pascal J. Frey, INRIA, Rocquencourt

Date/Time: Thursday, September 11, 2003, 10:00-11:00 am (PST)

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

Brief Abstract: In the context of numerical simulations, mesh adaptation has revealed very efficient as it allows to compute an accurate solution while minimizing the number of nodes. Anisotropy is expected to increase the potentiality of mesh adaptation while reducing even further the number of nodes.

During this talk, we will show how to construct anisotropic metrics suitable to govern the mesh generation algorithms. In addition, we propose a new adaptation scheme, based on the resolution of a transient fixed-point problem for the couple mesh-solution as well as a metric intersection in time.

CSRI POC: Monica-Martinez-Canales, (925) 294-3157
Philippe Pebay, (925) 294-2024

Title: Quantum Computing

Speaker: Anand Ganti, SNL

Date/Time: Thursday, June 5, 2003, 1:00-2:00 pm

Location: Building 980 Room 1

Brief Abstract: I will go over the principles of quantum computing

1) Definition of a Q-bit string
2) Permissible Operations on Q-bits
3) Measurements of Q bits
4) Composite Quantum states

I will also go over Grover's searching algorithm and if time permits Shor's factoring algorithm.

The talk will be from a computer science perspective and will not include how quantum computing is physically realized.

Maybe some other time I can go over quantum error correction once I learn it myself :)

CSRI POC: Neil Pundit, (505) 845-7601

Title: Multilevel solutions, least square extrapolation and a posteriori error estimate

Speaker: Marc Garbey, University of Houston

Date/Time: Wednesday, September 10, 2003, 9:00-10:00 am

Location: Building 980 Room 95

Brief Abstract: A posteriori error estimators are fundamental tools used to provide confidence in the numerical computation of PDEs. Unfortunately, the main theories of a posteriori estimators have been developed largely in the finite element framework, first for linear elliptic operators and second for non-linear PDEs in the absence of disparate length scales such as boundary layers.

On the other hand, there is a strong tradition in using grid refinement combined with Richardson extrapolation to produce CFD solutions with improved accuracy and, therefore, a posteriori error estimates on coarse grid solutions. But in practice, the effective order of a numerical method often depends on space location and is not accurately satisfied on different levels of grids used in the extrapolation formula. The Richardson extrapolation method then becomes unreliable. We propose a more robust and numerically efficient method that automatically finds the order of a method as the solution of a least square minimization problem on the residual. We introduce a multi-level least square extrapolation method that post-processes several grid solutions to recover accuracy and provide a posteriori error estimate. This method is not restricted to uniform refined grid solutions, nor strictly embedded grid levels. Our least square extrapolation method is a post-processing of data produced by existing PDE codes, that is easy to implement and does not require detailed knowledge of the PDE code itself.

CSRI POC: Matthew Hopkins, (505) 284-6376

Title: Random Parameters – an AMPL Extension for Stochastic Programming

Speaker: David Gay, Interview Candidate

Date/Time: Thursday, February 20, 2003, 11:00 am - 12:00 noon

Location: Building 980 Room 95

Brief Abstract: AMPL is an algebraic modeling language and system for expressing and manipulating linear and nonlinear optimization problems in discrete or continuous variables. AMPL has so far only dealt with deterministic problems, but as interest in stochastic optimization grows, we are adding facilities that will make working with some kinds of stochastic optimization problems easier. In particular, "random parameters" will permit expressing some multi-stage problems while letting suitable stochastic solvers do their own sampling. This talk gives an overview of AMPL and our current work to extend it.

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

Title: Two Examples of Simulation-Based Optimization at Sandia

Speaker: Genetha Gray, SNL

Date/Time: Wednesday, November 12, 2003, 10:00-1:00 am (PST)

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

Brief Abstract: Simulation is an important tool that is commonly used for the study of complex systems in science and engineering. In recent years, simulation has been paired with optimization in order to design and control such systems. In this presentation, we will discuss two simulation-based optimization applications.

We will begin with a problem that emerges in biology: determining the structure of transmembrane proteins. Difficulties with traditional protein structure determination methods necessitate new computational approaches to the problem. In particular, one innovative approach includes the optimization of an empirical scoring function used in Sandia's Bundler code. We will review the background for this problem, describe the Bundler scoring function and resulting optimization problem, and compare and contrast results obtained using two different derivative-free optimization algorithms.

The second problem we will discuss arises in the modeling of semiconductor bridges (SCBs) in circuit simulation. Using laboratory data for comparison, we want to test the validity of SCB models and identify appropriate parameters for Chilespice (or Xyce) computer simulations. We will explain our approach to this problem as well as the numerous challenges we face. We will present results and describe some resulting modifications to the SCB model.

Time permitting, we will briefly digress to discuss the speaker's thesis research on inverse problems in tomography.

Finally, we will conclude with a summary of future work on these and related applications.

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

Title: Numerical challenges in granular flows modeling

Speaker: Pierre Gremaud, North Carolina State University

Date/Time: Wednesday, November 5, 2003, 10:00-11:00 am

Location: Building 980 Room 95

Brief Abstract: In spite of its industrial importance, the study of the flow of granular materials lags far behind studies of corresponding problems in fluid dynamics. In this talk, I will review some basic continuum models for granular materials. Two problems will be considered in detail: first, the formation of granular heaps around obstacles, second, the computation of granular flows in silos. The first problem admits an Eikonal formulation while the second one can lead to hyperbolic or elliptic systems. In all cases, numerical methods will be introduced and discussed. Numerical results will be presented and analyzed. In particular, the presence of secondary circulation in some flows will be discussed.

CSRI POC: Todd Coffey, (505) 845-0520
David Day, (505) 844-1868

Title: Probabilistic Models for Microstructures

Speaker: Mircea Grigoriu, Cornell University

Date/Time: Tuesday, August 5, 2003, 10:00-11:00 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Probabilistic models available for two-phase microstructures and polycrystals are reviewed, and some of their limitations are discussed. It is shown that the level-cut Poisson field is a more versatile model for two-phase materials than the traditional level-cut Gaussian field, and can produce inclusions with a broad range of geometrical features. Also, a new model is presented for polycrystals based on a representation of the atomic lattice orientation by a non-Gaussian random field with piece-wise constant samples. The field can be viewed as an extension of the Poisson square wave stochastic process. The proposed models are used to generate samples of both two-phase microstructures and polycrystals.

CSRI POC: Steve Wojtkiewicz, (505) 284-5482

Title: Optimal Sizing and Shape Optimization in Structural Mechanics

Speaker: Gundolf Haase, Institute of Computational Mathematics
Johannes Kepler University of Linz

Date/Time: Monday, April 7, 2003, 11:00 am -12:00 noon (MST)

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

Brief Abstract: We consider an industrial application consisting in the mass minimization of a frame in an injection-moulding machine. This frame has to compensate the forces acting on the mould inside the machine and has to fulfill certain critical constraints. The deformation of that frame with constant thickness is described by the plain stress state equations for linear elasticity. If the thickness varies then we use a generalized plain stress state with constant thickness in the coarse grid elements. These direct problems are solved by an adaptive multigrid solver.

The mass minimization problem leads to a constrained minimization problem for a non-linear functional, which will be solved by some standard optimization algorithm, which requires the gradients with respect to design parameters.

For the shape optimization problem, we assume that the machine components consist of simple geometrical primitives determined by a few design parameters. Therefore, we calculate the gradient in the shape optimization by means of numerical differentiation, which requires the solution of approximately 4 direct problems per design parameter. The adaptive solver guarantees the detection of critical regions automatically, and ensures a good approximation to the exact solution of the direct problem.

A different strategy has been implemented for optimal sizing problems. Here, we have to handle hundreds of varying thickness parameters in the optimization problem, which makes numerical differentiation non-competitive. The first approach to calculate the gradient quite fast consists in using Automatic Differentiation (AD) of our direct problem solver. This approach works fine for direct solvers for the direct problem but requires huge memory and disk capabilities to handle iterative, and especially adaptive solvers. The other approach consists in writing the total derivative of the functional and get many partial derivatives by solving only one adjoint problem by means of our adaptive solver. Here, a special handling of the geometrical design parameters accelerates the optimization significantly.

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

Title: Source Level Debugging of FPGA Circuits Synthesized from General Purpose Programming Languages

Speaker: K. Scott Hemmert (Interview Candidate), Brigham Young University

Date/Time: Thursday, June 26, 2003, 1:45 - 2:30 pm

Location: Building 980 Room 95

Brief Abstract: With the growing popularity of using high-level synthesis tools to map programs written in general-purpose programming languages to FPGA hardware, it has become necessary to provide comprehensive, intuitive debugging tools in order to verify the correctness of the synthesized hardware. This research is studying the issues with providing debugging information to the programmer in the context of the original source code from which the circuit was derived (source level debugging). The difficulty in creating a source level debugger lies in the fact that typical synthesizing compilers provide no information about how the source code is mapped to hardware. To enable source level debugging, a synthesizing compiler has been modified to provide a two-way mapping of source code to circuit elements. This mapping is called the debug database and must account for optimizations, which may insert, remove, combine and/or modify instructions and variable assignments.

The source level debugger uses the information in the debug database to control and observe the state of the application running in an FPGA in the context of the original source code. It does this by providing the ability to single step, run to breakpoint, and watch and set variable values. The debugger runs in one of two modes: clock step mode or source step mode. Clock step mode provides truthful behavior and defines a single step as a single cycle of the hardware clock. Source step mode shows expected behavior and defines a single step as executing a single line in the source code.

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

Title: Efficient numerical methods for solving a class of Helmholtz problems

Speaker: Ulrich Hetmaniuk, SNL, Org. 9214

Date/Time: Thursday, March 6, 2003, 10:00-11:00 am

Location: Building 980 Room 95

Brief Abstract: The Helmholtz equation models wave phenomena. It arises, for instance, in the scattering of acoustic waves and in electrodynamics. Unfortunately, the numerical solution of the Helmholtz equation is theoretically and computationally challenging. For example, at a wavelength equal to the length of a submarine divided by 100, the finite element discretization of such a scattering problem typically requires hundreds of millions of grid points. Therefore, efficient numerical procedures for solving the Helmholtz problem are important and would facilitate the simulation of many physical phenomena.

My talk will consider the scattering of acoustic waves by partially axisymmetric obstacles. These obstacles are obtained by assembling a major axisymmetric component and a few features. Examples include rockets, guided missiles, and submarines. For such scatterers, an axisymmetric acoustic scattering analysis is not applicable. On the other hand, a straightforward three-dimensional (3D) analysis is inefficient because the geometrical properties of the axisymmetric components are not exploited.

My proposed methodology exploits such local axisymmetry with a fictitious domain approach. This fictitious domain decomposition method transforms the original large-scale 3D problem into a set of two-dimensional Fourier problems connected by Lagrange multipliers to another set of small-scale auxiliary problems associated with the features of the scatterer. The result is an algorithm that reduces the solution time and memory requirements by more than an order of magnitude.

CSRI POC: Steve Plimpton, (505) 845-7873

Title: CITI, NFSv4, and ASCI Review

Speaker: Peter Honeyman and Andy Adamson, University of Michigan, CITI

Date/Time: Wednesday, August 13, 2003, 10:00-11:00 am

Location: Building 980 Room 95

Brief Abstract: Since April 2002, faculty and students at the University of Michigan's Center for Information Technology Integration have been busy adapting NFSv4, the emerging IETF standard for distributed filing, to high-performance, clustered computers, such as those common in the ASCI infrastructure. CITI is also helping the Tri-Labs investigate NFSv4 as the direction for evolving the desktop and workstation environment in moving away from DCE services.

In this talk, Peter Honeyman and Andy Adamson will give an overview of the goals, progress, and challenges of CITI's efforts to exploit innovative features of the NFSv4 protocol and to adapt the protocol to meet the needs of clustered and integrated computing environments. There will be time for questions and suggestions.

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

Title: Generalizing Discriminant Analysis Via the Generalized Singular Value Decomposition

Speaker: Peg Howland, University of Minnesota

Date/Time: Wednesday, March 5, 2003, 10:00-11:00 am

Location: Building 980 Room 95

Brief Abstract: Discriminant analysis has been used for decades to extract features that preserve class separability. It is defined in terms of covariance matrices that represent the scatter within and between clusters. By optimizing the trace of a combination of scatter matrices, the dimension of data vectors can be reduced while preserving their cluster structure. The requirement that one of these matrices be nonsingular limits the application to data whose full dimension does not exceed the number of samples. When applied to information retrieval, for example, the number of terms is restricted by the number of documents.

Using the generalized singular value decomposition in this new context, we circumvent this restriction and extend the applicability of a number of optimization criteria. In addition, we avoid the numerical problems inherent in explicitly forming the scatter matrices. Our experiments confirm the effectiveness of this approach when compared to other current methods. They also lead to further questions about how to refine the algorithm and how to formulate better optimization criteria. These questions, as well as the potential application to facial recognition, provide our current challenges.

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

Title: Variational Multiscale Methods: Physical Models and Numerical Methods

Speaker: Tom Hughes, The University of Texas at Austin

Date/Time: Monday-Friday, July 7-11, 2003, 9:00-11:00 am

Location: Building 980 Room 95

Brief Abstract: In this short course we present an introduction to the variational multiscale method, which represents a unified approach to physical modeling of multiscale phenomena governed by partial differential equations and the development of effective numerical methods for their resolution. We emphasize finite element applications but the ideas are general and apply to other numerical methods as well. (We have used them so far in the development of finite difference, finite volume, and spectral methods, in addition to finite element methods.) We start by describing a few examples of multiscale phenomena including, aeroacoustics, turbulence, and advective-diffusive processes. The major application area so far has been fluid dynamics but there have been some applications to solid mechanics as well. The analytic ideas are first illustrated for time-harmonic wave-propagation problems in unbounded domains governed by the Helmholtz equation. This leads to the well-known Dirichlet-to-Neumann formulation. We then present a general treatment of the variational multiscale method in the context of an abstract Dirichlet problem, which is applicable to advective-diffusive processes and other processes of physical interest. We show how the exact theory represents a paradigm for subgrid-scale models and a posteriori error estimation. We examine hierarchical p-methods and bubble function methods in order to understand and, ultimately, approximate the "fine-scale Green's function" which appears in the theory. We illustrate relationships between so-called residual-free bubbles, element Green's functions, and stabilized methods. We then generalize these ideas to a class of non-symmetric, linear evolution operators formulated in space-time. The variational multiscale method also provides guidelines and inspiration for the development of stabilized methods (e.g., SUPG, GLS, etc.), which have attracted considerable interest and have been extensively utilized in engineering and the physical science. We present an overview of stabilized methods for advective-diffusive equations and illustrate their use on problems of compressible flow. We also develop a variational multiscale treatment of incompressible viscous flows including turbulence. This represents a new and alternative formulation of Large Eddy Simulation, which provides a simplified theoretical framework of LES, and a construct accommodating improved modeling. Numerical results exhibit the accuracy of the procedures on homogeneous isotropic flows, equilibrium and non-equilibrium turbulent channel flows, and transitional boundary layer flows. We then describe current research trends in which the variational multiscale method is used to derive nonlinear stabilized methods. This may provide a framework for understanding nonlinear numerical mechanisms in current use (e.g., shock-capturing viscosities) and a more rational basis for deriving new and improved variants.

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

Title: A Stabilized Mixed Discontinuous Galerkin Method for Darcy Flow Thomas J.R. Hughes(a) and Arif Masud(b)

Speaker: Tom Hughes, The University of Texas at Austin

Date/Time: Tuesday September 16, 2003, 10:00-11:00 am (MDST)

Location: Building 980 Room 95

Brief Abstract: Previously we developed a stabilized mixed finite element method for Darcy flow [1]. Velocities were assumed continuous or in H(div,Omega) and pressures could be continuous or discontinuous. Almost all combinations were shown to be convergent. An interesting feature of the formulation is that there are no mesh-dependent parameters in the stabilization terms. In this work we generalize to a discontinuous Galerkin formulation in which velocities may also be discontinuous. This permits complete freedom in the choice of interpolations. We present numerical results and error estimates in the "stability norm."

References:
[1] A. Masud and T.J.R. Hughes, "A stabilized mixed finite element method for Darcy flow," Computer Methods in Applied Mechanics and Engineering, v. 191, p. 4341-4370, 2002.

CSRI POC: John N. Shadid, (505) 845-7876
Pavel B. Bochev, (505) 844-1990

Title: Optimization under Uncertainty: From Trust-Region Algorithms to Spatial Statistics

Speaker: Takeru Igusa, Johns Hopkins University

Date/Time: Thursday March 13, 2003 11:00-12:00 noon (MST)

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

Brief Abstract: Trust region optimization algorithms are provably convergent provided that exact values for the objective function and its gradient are available. When the objective function is in terms of uncertain variables, however, only variable fidelity estimates for the objective function can be computed. While stochastic approximation algorithms (based on Robbins-Monro or Kiefer-Wolfowitz) are provably convergent for this OUU problem, they share some limitations, which often preclude their use in large-scale problems.

In this talk, we examine simulation-based methods of checking the conditions under which trust region algorithms, based on variable fidelity estimates of the objective function, are provably convergent. It will be shown that a spectrum of approaches can be used to satisfy the convergence conditions, beginning with naive Monte Carlo and its variants. Further, it will be shown that some of the more efficient approaches, which are based on spatial statistics and asymptotic local averaging, have a theoretical basis that is consistent with surrogate-based trust-region optimization algorithms being developed at SNL. It is believed that the most efficient techniques will entail an adaptive sampling process, which is an optimization problem within the spatial statistics context.

CSRI POC: Steve Wojtkiewicz, (505) 284-5482

Title: Inferring Genomic Regulatory Systems: The Davidson Model

Speaker: Sorin Istrail, Celera/Applied Biosystems

Date/Time: Tuesday, August 19, 2003, 1:00-2:00 pm

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Understanding the functional meaning of the genomic DNA sequence might well be considered the most important problem in bioscience. In the post-assembly phase of genomics, aiming towards the functional-maps of the cell, a most exciting phase has been the construction of the cis-regulatory maps. The literature of the area in the last few years witnessed an explosion of new technologies and experimental data. This accelerated progress in understanding the genomic control apparatus of gene expression led to the inference of developmental regulatory networks of extraordinary complexity. By all accounts, it appears that the unveiling of the cis-regulatory code is near.

In 2001 Eric Davidson published the book "Genomic Regulatory Systems," where he reports on 30 years of work at Caltech, assisted by 300 docs and postdocs, on sea urchin. The Davidson Lab work provided a general experimental framework for the study of a gene's cis-regulatory region and for regulatory network inference. The work proceeded hierarchically to uncover the "hardwired information processing logic" of gene regulation. Their technology, intertwined with analytical inference, resulted in the most completely understood transcriptional systems to date. For the area, their major accomplishments might as well be called: the "First Gene:" endo16, and the "First Network:" A Provisional Regulatory Gene Network for Specification of Endomesoderm in the Sea Urchin Embryo.

We will give an introduction and a mathematical analysis of the basic modeling theory developed by the Davidson Lab, organized as a "Davidson for Computer Scientists" presentation. We will also present a glance into our recent work with Eric Davidson towards the identification of the regulatory circuitry building blocks involved in regulatory information processing. As a context for the presentation, I will talk about two Applied Biosystems projects. The first focuses on providing industrial strength acceleration of the discovery of the cis-regulatory code. The second is devoted to integrative road-maps, part of our effort to build a programming language for genomics ("G").

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

Title: Micromechanics of materials with pores/microcracks

Speaker: Mark Kachanov, Tufts University

Date/Time: Wednesday, April 16, 2003, 10:00 – 11:00 am

Location: Building 980 Room 95

Brief Abstract: This seminar will focus on basic issues of micromechanics, such as microstructure-property relations, cross-property correlations and proper quantitative characterization of microstructures. After an overview of the theoretical framework, we will discuss computational issues to be addressed.

Some of the key computational issues, that constitute long-term goals, are as follows:

- Computational ("virtual") experiments on deterministic crack/pore arrays with the focus on the effective (anisotropic) elastic properties
- Computational testing of the elastic-conductive cross-property connections on deterministic crack/pore arrays
- Realistic microstructures contain mixtures of cracks and pores of diverse, "irregular" shapes. It is important to computationally generate a "database" that would quantify effects of various "irregularity factors", like irregular defect shapes, non-planar cracks, etc.
- Alternating numerical scheme for multiple interacting cracks.
- Investigation of curvilinear trajectories of crack growth under complex loading conditions, by incremental techniques.
- Computational ("virtual") experiments on clustering of defects: to verify the impact of clustering on the effective properties and on the fracture-related properties (like stress intensity factors).
- To verify the hypothesis that the precise information on the orientation distribution of cracks is not needed, as far as the effective properties are concerned (in the sense that the effective properties have low sensitivity to this information).

CSRI POC: Rebecca Brannon, (505) 844-5095

Title: Massive Parallelism in the HTMT Project: A Retrospective

Speaker: Peter Kogge, University of Notre Dame

Date/Time: Monday, March 10, 2003, 3:00-4:00 pm

Location: Building 980 Room 95

Brief Abstract: This talk will address one of the most ambitious massively parallel computer efforts of the last decade - a multi-institutional attempt to define a machine with a sustainable performance in excess of a petaflop. The talk wills overview the technologies, organization, and execution models behind the HTMT, with emphasis on the memory hierarchy, where coordination across literally millions of "memory nodes" (using Processing-In-Memory technology) becomes a necessity.

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

Title: Towards a Cost Model for Processing-In-Memory

Speaker: Peter M. Kogge, University of Notre Dame

Date/Time: Monday, July 28, 2003, 3:30-4:30 pm

Location: Building 980 Room 95

Brief Abstract: This talk will address a very important issue in making PIM technology economically viable, that is "what does it cost?" The tack chosen here is to do so by estimating a "per bit" premium for PIM over commodity DRAM. Factors such as die size, yield, and intrinsic opportunities for internal redundancy, along with pin outs and test considerations will be discussed. Existing models for classical DRAM will be used as a basis.

CSRI POC: Keith Underwood, (505) 09223

Title: Using MPPS for on-line visualization of large-scale scientific simulations

Speaker: Michael Lampe, University of Heidelberg

Date/Time: Monday, August 4, 2003, 11:15 am – 12:00 noon

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

Brief Abstract: Traditionally, the results of scientific simulations have been visualized in a post-processing step. Current data set sizes make this procedure questionable if not even unfeasible. Processing the data on the spot, i.e. right on the machine where it is computed, is one alternative. Visualization is becoming worthy of massively parallel processor power itself.

Another valuable aspect of using MPPs for visualization is the possibility for on-line visualization, i.e. one can get immediate feedback about the state of a computation. This is useful for end users of simulation tools and developers alike. End users can quickly explore complex parameter sets or abort jobs gone awry. Developers have an additional debugging tool at hand.

The software toolbox UG [1] that has been used as the basis for many different simulation tools, also from the field of computational mechanics, includes a visualization subsystem that realizes some of the above-mentioned ideas.

This subsystem implements two methods for parallel visualization of 3D grids and fields. The first one is based on a list-priority algorithm that solves the hidden-surface problem by drawing from back to front. The second one follows a sort-last approach (z-buffer algorithm) that renders partial images from per processor local data in parallel and produces the final image in a composition step.

Both usability and scalability aspects of UG's built-in graphical capabilities will be discussed.

References
[1] P. Bastian, K. Birken, S. Lang, K. Johannsen, N. Neuß, H. Rentz-Reichert, and C. Wieners, "UG: A flexible software toolbox for solving partial differential equations," Computing and Visualization in Science, v. 1, p. 27–40, 1997.

CSRI POC: Carter Edwards, (505) 284-4640

Title: A Parallel Infrastructure and Programming Model for Adaptive, Irregular Applications

Speaker: Dr. Stefan. Lang. University of Heidelberg

Date/Time: Monday, August 4, 2003, 10:00-10:45 am

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

Brief Abstract: Parallel applications based on the message-passing paradigm are difficult to design and implement. The development process sharpens when furthermore solution adaptive techniques are needed and computations on complicated 3D-geometries are focused.

Many the arising difficulties are addressed inside the presented infrastructure 'UG' (Unstructured Grids), for the computation of pde's, e.g. dynamic load migration and load balancing, parallel local grid adaption, and parallel IO.

We present proper abstractions for each of the different functionality parts of a parallel, adaptive and unstructured software system. This architecture assures a maximal degree of code reuse. Therefore the treatment of various partial differential equations is possible without superfluous coding. The parallel grid manager handles a hierarchical multigrid data structure, describing a two- or three-dimensional geometry. On top of this manager works the parallel and local adaption process, which controls manipulation of the distributed meshes. It allows refinement and/or coarsening of grids consisting of mixed element types. Each adaption pattern results in a grid closure such that no hanging nodes occur and the (multi) grid is always kept consistent. In our implementation grid adaption shows a truly local behavior, thus it is well suited for parallelization. In a parallel context grid adaption involves the need to rebalance the computational load. This stage involves both determining a new load balancing and dynamically redistributing the objects of the grid parts. We discuss load balancing for multigrids in general and present methods based on multiple constraints, which are able to dynamically rebalance the multigrid.

A key feature of 'UG' is the capability to dynamically migrate grid objects between the processors during run time. Thus the computation need not be interrupted, but continues after load transfer with a balanced work load on each processor. This difficult task is supported by 'DDD' (dynamic distributed data), a new parallel programming model. The migration process with its various stages is discussed: packing the data objects into buffers, sending and receiving the message buffers and unpacking the data objects. The DDD implementation contains several components with different functionality,

- Interfaces, to supports communication operations efficiently,
- Transfer, providing procedures for creating and deleting distributed object copies,
- Identification/Join, to create one global logical object out of several, and processor local objects
- Prio for changing priorities in a global consistent manner.

Finally we present parallel adaptive calculations of real-world problems. Timings, efficiency and scalability of parallel grid adaption, load balancing and load migration are compared with the numerical phases, discretization and solution. The results clearly show the enormous possibilities when the above acceleration strategies - multigrid, adaptivity and parallelism - are combined.

References
[1] UG, http://cox.iwr.uni-heidelberg.de/_ug

CSRI POC: Carter Edwards, (505) 284-4640

Title: Optimal Control of an Integro-difference Population Model

Speaker: Suzanne Lenhart, University of Tennessee & Oak Ridge National Lab

Date/Time: Thursday, May 1, 2003, 11:00 am - 12:00 noon (MST)

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

Brief Abstract: Integro-difference equations are models that are discrete in time and continuous in space. These equations model populations with discrete non-overlapping generations with separate growth and dispersal stages. The dispersal is modeled by an integral of the population density (after the growth) against a kernel. Optimal control of such a hybrid equation is a new area and involves a combination of the techniques from the discrete version of Pontryagin's Maximum Principle and from control of partial differential equations. Analysis and characterization of an optimal harvesting control will be given using an adjoint equation. Numerical algorithms and illustrations will be included for a variety of dispersal kernels and growth functions. About the Speaker Lenhart is a full professor in the Mathematics Department at the University of Tennessee and a part-time research staff member at Oak Ridge National Laboratory. She received her PhD at the University of Kentucky in partial differential equations.

Her research involves partial differential equations, ordinary differential equations and optimal control. She works on a variety of applications, including population models, disease models, bioremediation, lasers and resource management.

She served six years on the SIAM Council from 1995-2000 and currently serves on the SIAM Education Committee. She just finished a two-year term as the President of the Association for Women in Mathematics. She is an elected member of the Board of the Society for Mathematical Biology. Lenhart has been the director of the Research Experiences for Undergraduates summer program at UT since 1990.

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

Title: Request Algorithms for Freenet-Style Peer-to Peer Systems

Speaker: Jens Mache, Lewis & Clark College

Date/Time: Wednesday, May 21, 2003, 10:00-11:00 am

Location: Building 980 Room 95

Brief Abstract: As the Internet continues to experience rapid growth and ever increasing numbers of people requesting particular pieces of information, it gets exceedingly difficult to fulfill these requests (cf. Slashdot effect). Yet, to fulfill all the requests for one piece of information, there is an alternative to one powerful server (or server cluster): to rely on copies of the piece of information distributed across the network, so that many computers can fulfill these requests. This scalable and decentralized approach is an example of peer-to-peer networking.

In most peer-to-peer systems, edge resources self-organize into overlay networks. At the core of Freenet-style peer-to-peer systems are insert and request algorithms that dynamically change the overlay network and replicate files on demand.

We ran simulations to test how effective these algorithms are at improving the performance of subsequent queries. Our results show that for the original Freenet algorithms, performance improved less rapidly with a ratio of 99 requests to 1 insert than with an equal number of requests and inserts. This motivated us to design and test the performance of several new request algorithms. By changing the overlay network after failed requests and by further rewarding the fulfillers of successful requests, our new algorithms improved median pathlength by up to a factor of 9.25.

CSRI POC: Vitus Leung, (505) 844-1896

Title: Graph Coloring in Optimization Revisited

Speaker: Fredrik Manne**,** University of Bergen, Norway

Date/Time: Thursday, May 15, 2003, 10:00-11:00 am

Location: Building 980 Room 95

Brief Abstract: We revisit the role of graph coloring in modeling a variety of matrix partitioning problems that arise in numerical determination of large sparse Jacobian and Hessian matrices. The problems considered correspond to the various scenarios under which a matrix computation, or estimation, may be carried out, i.e., the particular problem depends on whether the matrix to be computed is symmetric or non-symmetric, whether a one-dimensional or a two-dimensional partition is to be used, whether a direct or a substitution based evaluation scheme is to be employed, and whether all nonzero entries of the matrix or only a subset need to be computed.

The resulting complex partitioning problems are studied within a unified graph theoretic framework where each problem is formulated as a variant of a coloring problem. Our study integrates existing coloring formulations with new ones. The insight gained from the unified graph theoretic treatment is used to develop and analyze several new heuristic algorithms.

CSRI POC: Erik Boman, (505) 844-2003

Title: Analysis and Implementation of Multi-rate PDE Methods for the Simulation of Circuits with Multiple Time Scales

Speaker: Ting Mei, University of Minnesota

Date/Time: Friday, August 22, 2003, 10:00-11:00 am

Location: Building 980 Room 95

Brief Abstract: In systems containing signals with widely separated time scales (so-called multi-rate problems), many periods of the fast scale must be resolved in order to capture the more interesting slow scale response. This computational cost can result in extremely long run times for simulations of interest. Recently, multiple-rate partial differential equation (MPDE) methods have shown dramatic reduction in the computation times for certain multi-rate problems. In these MPDE methods, the time domain is factored via periodicity into a product space of different time domains. This transformation moves the governing DAEs to PDEs in time that are then solved using methods similar to those applied to spatial PDE problems. Early research into these approaches has demonstrated speedups in excess of two orders of magnitude.

A prime motivation for Sandia's interest in these methods comes from the need to model oscillator circuits as part of a larger charging circuit. These autonomous oscillator circuits generate waveforms with modulated frequencies (frequency drift). This requires the addition of phase variables into the MPDE framework and leads to warped-time MPDEs (WaMPDE). In this talk, we present an analysis of MPDE methods in both the time and frequency domains with an emphasis on the stiffly-stable numerical methods for the MPDE and WaMPDE solutions. We further examine the questions of stability and accuracy for the MPDE solution of DAEs.

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

Title: Toward Computational Materials Design

Speaker: Richard P. Muller, Caltech Materials and Process Simulation Center

Date/Time: Monday, February 17, 2003, 9:00-10:00 am

Location: Building 980 Room 95

Brief Abstract: We have developed first-principles-based multiscale techniques for studying a wide variety of materials, catalysts, and molecules. These techniques provide the ability to investigate the structures and properties of a material prior to synthesis and experimental characterization, and have the potential to dramatically decrease the cost and the time required for the development of new materials. We are now at a point where computational techniques can begin to play a leading role in the design of new materials.

My talk will summarize the overall development of software tools and methods for computational materials design in our research group. I will describe projects underway in our group in polymerization, alkane activation, and fuel cell electrode catalysis, as well as in the design of energetic materials.

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

Title: New Algorithms for Protein Folding in the HP Model

Speaker: Alantha Newman, Massachusetts Institute of Technology

Date/Time: Thursday, June 19, 2003, 10:00-11:00 am

Location: Building 980 Room 95

Brief Abstract: We consider the problem of protein folding in the HP model. This problem is combinatorially equivalent to folding a string of 0's and 1's on a lattice so that the string forms a self-avoiding walk and the number of pairs of adjacent 1's is maximized.

Approximation algorithms for this problem on the 2D and 3D square lattices were first introduced by Hart and Istrail in 1995. They presented simple linear-time algorithms with approximation guarantees of 1/4 and 3/8, respectively, for the 2D and 3D problems. We present a simple linear-time time 1/3-approximation algorithm for the problem on the 2D square lattice. We also show that the upper bound used in this algorithm and in Hart and Istrail's earlier algorithm cannot be used to obtain an approximation factor better than 1/2.

CSRI POC: Bill Hart, (505) 844-2217

Title: High-Performance I/O for Data-Intensive Computational Grid Applications

Speaker: Ron Oldfield, Ph.D. Candidate, Dartmouth College

Date/Time: Tuesday, January 21, 2003, 10:30-11:15 am

Location: Building 980 Room 95

Brief Abstract: High performance computing increasingly occurs on "computational grids" composed of heterogeneous and geographically distributed systems of computers, networks, and storage devices that collectively act as a single "virtual" computer. A key challenge in this environment is to provide efficient access to data distributed across remote data servers. Our parallel I/O framework, called Armada, allows application and data-set providers to flexibly compose graphs of processing modules that describe the distribution, application interfaces, and processing required of the dataset before computation. In this talk, we describe the Armada framework and present an algorithm used to restructure application graphs to increase parallelism and to improve network performance in the context of a wide-area computational grid.

CSRI POC: Neil Pundit, (505) 845-7601

Title: Continuous, Macroscopic Fibers from Liquid Crystalline Mesophases of Single-Wall Carbon Nanotubes in Superacids

Speaker: Matteo Pasquali, Rice University

Date/Time: Monday, November 3, 2003, 10:00-11:00 am

Location: Building 980 Room 95 (Sandia NM)

Brief Abstract: Single Wall Carbon Nanotubes (SWNTs) can be dispersed at high concentration in superacids; the direct protonation of SWNTs sidewalls eliminates wall-wall van der Waals interactions and promotes the dispersion process. At very low concentration, SWNTs in superacids dissolve as individual tubes, which behave as Brownian rods. At higher concentration, SWNTs form a highly unusual nematic phase consisting of spaghetti-like self assembled supermolecular strands of mobile, solvated tubes in equilibrium with a dilute isotropic phase. At even higher concentration, the spaghetti strands self-assemble into a polydomain nematic liquid crystal. Upon the introduction of small amounts of water, the liquid crystal phase separates into needle-shaped strands (~20 microns long) of highly aligned SWNTs, termed alewives. Under anhydrous conditions, the liquid crystalline phase can be processed into highly aligned fibers of pure SWNT without the aid of any surfactants or polymers.

This novel-processing route opens the way to mass production of well-ordered, macroscopic objects comprised solely of SWNTs. Developing effective manufacturing processes will require new computational and theoretical modeling tools, which can identify optimal acid solvents, coagulation configurations, and annealing processes.

CSRI POC: Lawrence C. Musson, (505) 284-3647

Title: Reduced-order Quadratic Approximations and Interior-point Algorithm Development for Sequential Approximate Optimization

Speaker: Dr. Victor Perez, University of Notre Dame

Date/Time: Wednesday September 24, 2003, 2:30-3:00 pm

Location: Building 980 Room 95

Brief Abstract: This presentation focuses on the development of methodologies that help designers reduce the cost of using optimization to manage the simulation-based design process. Two topics will be presented: the development of an interior-point approach for trust-region-managed, sequential, approximate optimization and the construction of sequential, reduced-order, quadratic approximations. The interior point approach insures that feasibility is maintained throughout the optimization process. This facilitates the delivery of a consistent and feasible design when subject to reduced design cycle time constraints. In order to deal with infeasible starting points, probability-one homotopy methods are used to relax constraints and push designs toward feasibility. The method presented to construct reduced-order of sequential, local quadratic approximations attacks the "curse of dimensionality" by making the required number of sampling points linearly dependent on the number of design variables.

CSRI POC: Mike Eldred, (505) 844-6479

Title: Combinatorial Scientific Computing

Speaker: Ali Pinar, LTE

Date/Time: Thursday, September 18, 2003, 10:00- 11:00 am (PDT)

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

Brief Abstract: Although scientific computing is generally viewed as the province of differential equations and numerical analysis; combinatorial algorithms have long played a crucial role. In this talk, I will address several areas in which combinatorial insights have been successfully applied. I will briefly mention my work on communication and load-balancing algorithms for parallel computing, improving memory performance of sparse-matrix operations, and constructing a sparse null-space basis for a sparse matrix.

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

Title: Solving Transport Equations Subject to Constraints

Speaker: Jerome Pousin**,** Institute for Applied Sciences

Date/Time: Tuesday, September 9, 2003, 10:30-12:30 pm (MT)

Location: Building 980 Room 95 (Sandia NM)
Building 905 Room 210 (Sandia CA)

Brief Abstract: In the context of the Low Mach number equations for reacting flows, the continuity equation (which is a Transport equation) must be solved under the constraint g(rho,P,T) = 0 due to the state law. Numerical resolution of the low Mach number equations consists in solving the continuity equation ignoring the constraint g(rho,P,T) = 0, then a step of projection more or less sophisticated is applied. Such a strategy is not riskless, in particular because it can lead to projections on a manifold larger than the constraint one.

During this presentation, by analyzing the academic case of the weakly compressible Stokes equations for isothermal flows, in a first part I will remind the functional framework for recovering the pressure. In a second part, I will propose some mathematical formulations for solving the Transport equation under constraint. Finally I will discuss the numerical efficiency of such formulations.

CSRI POC: Philippe Pebay, (925) 294-2024

Title: Numerical algorithms and application design in UG

Speaker: Volker Reichenberger, University of Heidelberg

Date/Time: Monday, August 4, 2003, 10:45-11:15 am

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

Brief Abstract: The numerical simulation framework UG supports unstructured grids, adaptivity, multigrid methods and parallel computing for the finite element and finite volume solution of partial differential equations. Object-orientation and modularization shield the complexity from the user and allow to put the development emphasis on the problem class level, i.e. discretization methods and application-specific functionality. We explain how the numerical algorithms are modularized and implemented and then explain how this approach can be extended to complex application classes. We explain the structure of a framework for porous media computations and show results of simulations for transport and multiphase flow problems.

CSRI POC: Carter Edwards, (505) 284-4640

Title: Uncertainty Management in MDO using Evidence Theory

Speaker: John E. Renaud, University of Notre Dame

Date/Time: Wednesday September 24, 2003, 2:00-2:30 pm

Location: Building 980 Room 95

Brief Abstract: This research focuses on how epistemic uncertainty can be quantified in multidisciplinary systems and accounted for in an optimization under uncertainty (OUU) process. Epistemic uncertainties associated with the disciplinary design tools (model form uncertainty) and with input parameters (parametric uncertainties) are considered. Evidence theory is used to quantify the epistemic uncertainties in terms of the uncertain measures of belief and plausibility. Stochastic constraints are formulated using these measures of belief or plausibility. The designer seeks the stochastic optimum design under epistemic uncertainty.

The measures of uncertainty provided by evidence theory are discontinuous functions. Such non-smooth functions cannot be used in traditional gradient-based optimizers because the sensitivities of the uncertain measures are not properly defined. In this research surrogate models are used to represent the uncertain measures as continuous functions. A formal trust region managed sequential approximate optimization approach is used to drive the optimization process. The trust region is managed by a trust region ratio based on the performance of the Lagrangian. The Lagrangian is a penalty function of the objective and the constraints. The methodology will be illustrated in application to multidisciplinary test problems.

CSRI POC: Mike Eldred, (505) 844-6479

Title: Modeling Phase Transformation in Shape Memory Alloy Wires

Speaker: Daniel R. Reynolds, Interview Candidate, Rice University

Date/Time: Monday, March 20, 2003, 11:00-12:00 noon (MST)

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

Brief Abstract: Smart Materials present a new and promising field of research and application in the fields of Materials Science and Mathematics. These materials are characterized by their unique ability to undergo dramatic changes to their structure upon application of relatively small thermal or electromagnetic loading. For this reason, smart materials such as Shape Memory Alloys and Ferro magnets have become promising candidates for various applications, including vibration damping, nanomachinery and computer hardware. The behavior of these materials has been of great interest in the materials science community for some time, however a complete mathematical description has only recently been more fully examined.

In this talk I introduce a nonlinear thermodynamic model governing phase transformations in shape memory alloy wires. I first introduce some of the unique thermodynamic characteristics of these alloys. I then discuss my research constructing a continuum-level model governing their thermodynamic behavior. This model is first-principles based application of the balance laws of physics, and is rooted in Landau's theory of continuum thermodynamics. I then provide a mathematical approach to solving the resulting system of nonlinear partial differential equations, based on Continuous Time Galerkin methods and a variant on Newton's method. I conclude the talk by presenting results of computational experiments showing first the validity of the model and then its application toward vibration damping.

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

Title: Fairness in Job Scheduling on CPlant

Speaker: Gerald Sabin, CSRI Summer Student, Ohio State University

Date/Time: Wednesday, August 27, 2003, 10:00-11:00 am

Location: Building 980 Room 95

Brief Abstract: The current Cplant scheduling policy attempts to approximate user fairness based on the fairshare priority policy. We evaluate the current schemes ability to achieve this fairness metric, as well as some of the currently proposed modifications. The modifications range from minor changes (changing when/how starvation works and limiting the runtime of an individual job) to more fundamental changes (full blocking/conservative backfilling). Many of the changes show a substantial improvement in the fairness of the scheduler with minor differences in loss of capacity (utilization).

CSRI POC: Vitus Leung, (505) 844-1896

Title: Power laws, with applications to internet traffic, finance, and other areas: heavy tails and long range dependence

Speaker: Gennady Samorodnitsky, Cornell University

Date/Time: Tuesday, August 5, 2003, 11:00 am – 12:00 noon

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

Brief Abstract: The "usual" stochastic models have exponentially fast decreasing tails and short memory, which is ordinarily also expressed in correlations that decrease exponentially fast with lag. In many modern applications one or both of these features are believed to be absent.

We will discuss stochastic models with heavy tails and/or long range dependence, where exponential decay is replaced by power-like decay in either space or time or both. Many of our examples are financial or teletraffic in nature, but other applications will be discussed as well.

CSRI POC: Steve Wojtkiewicz, (505) 284-5482

Title: Multiscale Methods in Engineering

Speaker: Guglielmo Scovazzi, Stanford University/The University of Texas, Austin

Date/Time: Tuesday, July 1, 2003, 10:00-11:00 am

Location: Building 980 Room 95

Brief Abstract: I will present an introduction to variational multiscale methods and describe their potential for better representing phenomena involving scales differing by orders of magnitude. Theoretical analysis of physical models with emphasis on the Navier-Stokes equations will be presented, and results on numerical testing of the simplified Burgers' equation model will be discussed.

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

Title: Valgrind: a supervised execution framework for x86-linux (1:00 pm)
Challenges in porting Valgrind away from Linux-on-x86 (2:30 pm)

Speaker: Julian Seward, Valgrind

Date/Time: Monday, June 30, 2003
1:00 pm Valgrind: a supervised execution framework for x86-linux (PDT)
2:00 pm Q&A/Break
2:30 pm Challenges in porting Valgrind away from Linux-on-x86
3:30 pm Q&A

Location: Building 980 Room 24 (Sandia NM)
Building MO52 Room 165 (Sandia CA)
LANL: contact Mike McKay (lmdm@lanl.gov)

Brief Abstract: Valgrind was originally released as a memory debugging tool for Linux on x86s, and seems to have been quite successful as such. Since then it has evolved significantly. The instrumentation phases have become much more modular, so that building a new debugging or profiling tool based on Valgrind is the work of an afternoon.

In this talk I'll give an overview of memory debugging with Valgrind, then look at some of the other debugging and profiling tools that we've constructed in our new modular framework. Some of these tools have capabilities we've never seen anywhere else.

Valgrind, as it stands now, comprises three major components. The first, and most obvious, is an instrumenting x86-to-x86 dynamic translation system -- in short, a simulated CPU. The second part is a modular structure allowing different instrumentation passes to be easily added, without disturbing the core machinery. The third component provides a simulation of the environment in which the instrumented program is to run -- that is, a simulation of Unix signals, system calls and POSIX threads.

When this project began, we believed the CPU simulation to be the really hard part. We were wrong. Building and maintaining the environment simulation continues to prove the most difficult, and out of the three components, is the biggest unknown in porting Valgrind to other platforms. In this talk I'll show the design constraints arising from the three components and how these interact with different proposals for porting. I'll finish up with an outline of a proposal we feel gives significant portability without the need for enormous engineering effort.

CSRI POC: Zhaofang Wen, (505) 284-0206

Title: Geometric proofs of the existence of Parter and Fiedler vertices

Speaker: Brian Shader, University of Wyoming

Date/Time: Monday, June 9, 2003, 9:30-10:30 am

Location: Building 980 Room 95

Brief Abstract: Symmetric matrices A that have an eigenvalue k of multiplicity at least 2, and whose graph is a tree are studied. We study two types of vertices: a Fiedler-vertex is an index i such that each eigenvector of A corresponding to k has with coordinate 0, and a Parter-vertex is an index i such that m_A(k)+1=m_{A(i)}(k), where m_B(k) denotes the multiplicity of k as an eigenvalue of B, and A(i) is the principle submatrix of A obtained by deleting row and column i. We give geometric proofs of a characterization of Parter-vertices in terms of F-vertices, the existence of a Fiedler-vertex, and the existence of a Parter vertex.

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

Title: Sweatshop: Ironing and Cutting (Parameterization of Surface Meshes)

Speaker: Alla Sheffer, Technion University, Haifa, Israel

Date/Time: Tuesday, February 25, 2003, 1:30-2:30 pm

Location: Building 980 Room 24

Brief Abstract: Many applications using triangulated surface models benefit from a mapping between the mesh and the plane. Mapping of a texture onto a model is an essential part of model representation in computer graphics. Providing a parameterization simplifies surface reconstruction from scanned data and 3D images. It improves the efficiency of surface mesh generation and the quality of the resulting mesh.

In the first part of the talk I will describe a new formulation of the problem and introduce a parameterization method based on it. The Angle Based Flattening (ABF) method formulates the mapping problem in terms of the flat mesh angles, and solves it in the angle space. The ABF algorithm provides a quasi-conformal mapping and computes the optimal planar domain boundary in terms of minimal distortion. The method is guaranteed to compute a continuous mapping and the numerical solution is proven to converge.

I will then discuss a new post-processing algorithm, aimed at reducing length distortion of an existing parameterization and apply it to ABF results. The post-processing is based on computing a mapping from the plane to itself, which has length distortion very similar to that of the ABF parameterization. By applying the inverse mapping to the result of the initial parameterization, we obtain a new parameterization with low length distortion. We notice that the procedure for computing the inverse mapping can be applied to any other (convenient) mapping from the three-dimensional surface to the plane in order to improve it.

Time permitting, I will get to the "cutting" part of the talk, and describe an algorithm for cutting the surfaces, to enable parameterization of surfaces with high curvature as well as closed surfaces. The addition of seams reduces the surface curvature and hence reduces the metric distortion produced by the mapping. The downside of cutting seams in the surface is the mapping discontinuity they generate. The method minimizes the absolute length of the seams or alternatively minimizes their visual impact by placing them in less visible regions of the model.

CSRI POC: Steve Owen, (505) 284-6599

Title: Compiler-Assisted Checkpointing for MPI Programs

Speaker: Alison N. Smith, Department of Computer Sciences
University of Texas at Austin

Date/Time: Tuesday, April 29, 2003, 2:00 - 3:00 pm (MST)

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

Brief Abstract: This talk addresses issues of fault tolerance in large-scale supercomputing clusters. In particular, we discuss ways that compiler technology can transparently support efficient checkpointing of MPI programs. Our main goal is to strategically place checkpoints to reduce bandwidth contention among processes and to reduce checkpoint size within each process. In contrast to other proposed techniques, our solution ensures no extra messages or useless checkpoints, allowing for low failure-free overhead. This talk discusses both our on-going efforts and future plans.

CSRI POC: Patty Hough, (925) 294-1518

Title: Discovery and Visualization of DNA Sequence Signals

Speaker: Mark A. Smith, Harvard University

Date/Time: Tuesday, June 3, 2003, 10:30-11:30 am

Location: Building 980 Room 95

Brief Abstract: Although we have read out the genome sequences of many organisms, we are still far from knowing all their words, punctuation, and grammar, much less appreciating their full meaning. Protein coding regions are reasonably well understood and delineated, but the many levels of gene regulation encoded in the sequences are less well understood. Furthermore, chromosomes are physical objects, which are involved, in many biological processes such as crossover, segregation, transposition, and packaging, and the primary DNA sequence impinges on all these processes.

An essential step in comprehending any text is to isolate the meaningful words and their relationships. To this end, I will describe some of our pattern discovery efforts applied to the Drosophila genome. Pattern discovery is the result of unsupervised learning, which amounts to clustering which in our case, boils down to counting. Our basic approach is binning the occurrences of DNA words in various ways with respect to annotated features. The significance of any such counting must then be assessed with respect to some hypothetical null background model. We maintain that any word with a high enough statistical significance must necessarily have a biological, biochemical, or evolutionary significance, and although we may not know what that significance is, the word and its context provides potential targets for experimental studies.

Specifically, we analyzed the sequences surrounding the 5' ends of untranslated regions as determined by the mRNA annotations in GenBank. Although this data is noisy, it is remarkable how looking at large enough data sets reveals extremely significant patterns. In addition to observing well-known elements such as the TATA box, we discovered a number of formerly unknown promoter proximal elements. All of this can be done without resorting to expression data, grouping genes by function, performing phylogenetic comparisons, or constructing sophisticated sequence statistics models. These results suggest further detailed studies of spelling variations as well applications to promoter annotation and prediction. Finally, graphical display of the histograms of any short word reveals many detailed patterns, all of which would be explained by a truly complete understanding of a genome sequence.

Joint work with Temple F. Smith, Yaneer Bar-Yam, and William M. Gelbart.

CSRI POC: Mark D. Rintoul, (505) 845-9592

Title: Rosetta Structure

Speaker: Charlie Strauss, Los Alamos National Laboratory

Date/Time: Wednesday, January 29, 2003, 10:30-11:30 am

Location: Building 980 Room 95

Brief Abstract: Charlie Strauss from Los Alamos National Lab will speak on the Rosetta Structure prediction algorithm and its applications to De novo structure determination and functional annotation.

CSRI POC: Steve Plimpton, (505) 845-7873

Title: Materials solutions for the magnetic random access memory (MRAM) based on electronic structure calculations

Speaker: Roland Stumpf, Interview Candidate, Org. 01114

Date/Time: Wednesday, January 29, 2003, 9:00-10:00 am

Location: Building 980 Room 95

Brief Abstract: MRAM has the potential to replace many microchip based memory architectures. Motorola's MRAM project is based on the resistance of a tunnel junction between a fixed and a free magnetic layer. The fixed layer is a synthetic anti-ferromagnet (SAF) consisting of a non-magnetic spacer layer sandwiched between two magnetic layers. The design goals for the SAF are thermal stability, strong AF coupling, well defined microstructure, low stress, and manufacturability. First principles calculations help to choose the best materials combination. The determination of the energetics of intermixing addresses the thermal stability issue. The magnetic exchange coupling between the two magnetic layers of the SAF is given by the calculated energy difference between the two possible magnetic configurations. I show how thickness variations and the size of the spin splitting in the magnetic layers determine the nature of the coupling. Based on this study I propose an improved SAF design with a high coupling strength, a low tendency to intermix, and a high tolerance with respect to thickness variations in the spacer layer.

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

Title: Prophesy: Analysis and Modeling of Parallel and Distributed Applications

Speaker: Valerie Taylor, Texas A&M University

Date/Time: Monday, September 8, 2003, 2:00-3:00 pm

Location: Building 921 Room 137 (Sandia-CA)

Brief Abstract: Efficient execution of applications requires insight into how the system features impact the performance of the application. This insight generally results from significant experimental analysis and possibly the development of performance models. Prophesy, is a system that focuses on the performance analysis and modeling of parallel and distributed applications. Prophesy includes three major components: automatic instrumentation, databases for archiving performance data, and a model builder. The model builder can use one of three techniques (curve fitting, parameterization, or coupling) to develop performance models. This talk will describe the Prophesy system and include some examples to demonstrate the use of Prophesy.

This is a joint project with Rick Stevens at Argonne National Laboratory. The project is sponsored in part by NSF and NASA Ames.

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

Title: Geometric Aspects of the Discretization of Maxwell's Equations

Speaker: Fernando L. Teixeira, Ohio State University

Date/Time: Thursday, February 20, 2003, 10:00-11:00 am

Location: Building 980 Room 95 (Sandia – NM)
Building 921 Room137 (Sandia – CA)

Brief Abstract: Finite-difference (FD), finite element (FE), and finite volume (FV) electromagnetic field simulations in complex unstructured grids are usually plagued by deleterious phenomena such as spurious (ghost) modes and unconditional late-time instabilities whose origin is sometimes not well understood. In this talk, we first review the terminology and basic concepts of the spatial discretization of electromagnetic theory using discrete differential forms (cochains). Differential forms provide natural bridge between the continuum and the lattice versions of the theory, allowing for a natural factorization of the field equations into topological part (i.e., invariant under homeomorphisms) and a metric part. We describe the exact discretization of the topological part and the nature of the approximation for the metric part (discrete Hodge operator). The various potential sources of inconsistency in the discretization process for FD, FE, and FV algorithms are identified, distinguished, and discussed. We then describe the use of Whitney forms as the basic interpolants for simplicial grids, and the use of barycentric subdivisions to define Whitney forms in non-simplicial grids. In the last part of the talk, we use this geometric backdrop to provide a unified viewpoint of the various formulations of the perfectly matched layer (PML) absorbing boundary condition for unbounded domains.

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

Title: Numerical Methods for Stochastic Partial Differential Equations

Speaker: Raul Tempone, ICES, University of Texas at Austin

Date/Time: Thursday, October 23, 2003, 2:00-3:00 pm

Location: Building 980 Room 95

Brief Abstract: We describe and analyze two numerical approximations for the statistical moments of the solution of a model linear elliptic partial differential equation with stochastic coefficients, illustrating on the particular case of the expected value and estimating numerical and modeling errors.

The first is the Monte Carlo Galerkin Finite Element Method (MCGFEM) which generates iid standard Galerkin finite element approximations of the solution by sampling the stochastic coefficients of the equation.

The second is motivated by the fact that the solution of the parametric elliptic problem is analytic with respect to the parameter variable. We propose a tensor product approximation, the p-h version of the Stochastic Galerkin Finite Element Method (SGFEM), that uses a p version on the stochastic parameter set and an h version on the physical domain.

For both methods, we present a priori error estimates in the H^1 and L^2 norms for the convergence of the numerical approximations toward the expected value of the solution.

Lastly, we address the computational work required by the different numerical approximations to achieve a prescribed accuracy. We use the available a priori error estimates for the Monte Carlo Galerkin Finite Element Method (MCGFEM), k-h SGFEM, and p-h SGFEM methods to compare the asymptotic computational work required by each method to achieve a given accuracy. This comparison suggests intuitive conditions for an optimal selection of these methods.

CSRI POC: John Red-Horse, (505) 845-9190

Title: An Initial Boundary Value Problem for Testing Dynamical Cores of Atmospheric General Circulation Models

Speaker: Stephen Thomas, National Center for Atmospheric Research

Date/Time: Thursday, March 6, 2003, 11:00-12:00 pm

Location: Building 980 Room 95

Brief Abstract: Traditionally, climate model dynamical cores have been based on the spectral transform method because the global spherical harmonic basis functions provide an isotropic representation on the sphere. It is trivial to implement semi-implicit time stepping schemes, as the spherical harmonics are eigenfunctions of the Laplacian on the sphere and the resulting Helmholtz problem is diagonalized in spectral space. Distributed-memory parallel computation of the spectral transform relies on a data transposition, and scalability is limited by latency and bisection bandwidth. Spectral elements maintain the accuracy and exponential convergence rate exhibited by the spectral transform method but the computations are local. However, a semi-implicit scheme based on a preconditioned Krylov iterative solver is hampered by global reductions.

In this talk we describe validation experiments with a 3-D climate spectral element dynamical core developed at NCAR. The Held-Suarez (1994) idealized physical forcings are often used to evaluate numerical schemes for atmospheric general circulation models. We present results for this 1200-day integration using the spectral element model with explicit time stepping and time-split diffusion. Spectral transform and spectral element model results for a new test proposed by Polvani et al (2002) will also be presented.

CSRI POC: Bill Spotz, (505) 845-0170

Title: Lustre Lite (Scalable Parallel File System)

Speaker: Sonja Tideman, Org. 09223

Date/Time: Monday, May 12, 2003, 1:00-2:00 pm

Location: Building 980 Room 95

Brief Abstract: Lustre is a tri-Lab endorsed scalable parallel file system being built by Cluster File System, a California company co-founded by ex-CMU Professor Peter Braam. It will be produced under open source and it is expected to meet the tri-Lab requirements. It is a parallel file system option for Sandia's Red Storm.

A Beta version of Lustre Lite has been available for about a month, and Sonja Tideman has been experimenting with it on a 16-node cluster (4 nodes used for IO). She will describe the features and performance against a common benchmark. Our Principal Investigator for Scalable IO, Lee Ward, has been guiding this effort.

CSRI POC: Heidi Kolden, (925) 294-3683, Neil Pundit, (505) 845-7601

Title: Modeling, Controlling and Optimizing Nonlinear Dynamical Systems Behavior

Speaker: Marc Turcotte, Ph.D., Interview Candidate

Date/Time: Thursday, January, 30, 2003, 9:00-10:00 am

Location: Building 980 Room 95

Brief Abstract: Nonlinear dynamical systems defined on complex networks are ubiquitous in nature and engineering yet their properties have only become much clearer in the recent past and continue to be the subject of intense scientific investigation. In this talk, complex systems will be introduced by virtue of an analogy with physics and examples from other sciences will be shown (biology, social sciences, medicine). Topics in control theory and relevant topics in multi-objective optimization will be covered.

Work done by the presenter will be set in context by and integrated with current results quoted from the research literature. Topics considered range from evolution, human cooperation, epidemics propagation dynamics and some possible applications in radiation oncology for treatment planning optimization of cancerous tumors.

CSRI POC: Scott Mitchell, (505) 845-7954

Title: Strain, suboxides and the band offset between Si(100) and SiO-2

Speaker: Dr. Blair Tuttle, Penn State University - Erie, Behrend College

Date/Time: Wednesday, July 16, 2003, 2:00 - 4:00 pm (MDST)

Location: Building 980 Room 95 (Sandia NM)
Building 960 Room 107 (Sandia CA)

Brief Abstract: Interfaces between Si(100) and Si0_2 are important for many semiconductor devices including ubiquitous field effect transistors. The band offset at the interface is a critical parameter controlling leakage currents. As devices are scaled to smaller dimensions, the microscopic character of the interface is becoming more important. In this talk, I will present recent calculations, based on density functional theory, which elucidate how interfacial structure, including strain and sub-oxide layers, effect the electronic structure of the Si(100)-SiO_2 interface.

CSRI POC: Harry P. Hjalmarson, (505) 844-8888

Title: Parallel Adaptive Refinement for 3D Unstructured Grids

Speaker: Jacob Waltz, Los Alamos National Laboratory

Date/Time: Thursday, February 13, 2003, 10:00-11:00 am (Mountain Time)

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

Brief Abstract: A parallel adaptive refinement algorithm for three-dimensional unstructured grids is presented. The algorithm is based on a hierarchical h-refinement/derefinement scheme for tetrahedral elements. The algorithm has been fully parallelized for shared-memory platforms via a domain decomposition of the mesh at the algebraic level. The effectiveness of the procedure is demonstrated with applications, which involve steady and unsteady flows. Scaling studies of the algorithm are included.

CSRI POC: Curtis C. Ober, (505) 844-1846

Title: Modeling Local Search Algorithm Performance in Scheduling

Speaker: Jean-Paul Watson**,** Colorado State University

Date/Time: Monday, February 24, 2003, 3:00-4:00 pm

Location: Building 980 Room 95

Brief Abstract: Local search algorithms are among the most effective techniques for generating high-quality solutions to notoriously difficult schedule optimization problems. Yet, despite widespread success, very little is known about why these algorithms work so well and under what conditions they excel. In this talk, I will examine performance models of local search algorithms for the widely studied job-shop scheduling problem. I consider two types of performance models: static models based on search space features and dynamic models based on run-time behavior. I show that highly accurate static and dynamic performance models need not be complex. In particular, two well-known local search algorithms, tabu search and iterated local search, can be accurately modeled using simple variants of a one-dimensional random walk. I conclude by discussing the relationship between the static and dynamic models, and illustrate that the structure of these models is largely a function of the representation used to encode solutions: i.e., the binary hypercube. These models provide a significant step toward "demystifying" the behavior of local search algorithms and identify new paths for further improving the performance of local search.

CSRI POC: Bill Hart, (505) 844-2217

Title: Discontinuous Galerkin Methods for Modeling Flow and Reactive Transport

Speaker: Mary F. Wheeler, The University of Texas at Austin

Date/Time: Thursday, April 17, 2003, 10:00-11:00 am (MST)

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

Brief Abstract: In this presentation, we consider several primal formulations of discontinuous galerkin methods for modeling flow and reactive transport in porous media which include Baumann-Oden-babuska, the symmetric and nonsymmetric interior penalty methods, SIPG, NIPG, and the incomplete interior penalty method IIPG recently introduced by Sun and Wheeler. Error estimates and estimators for adaptivity and computational results are provided. Extensions to the coupling flow and mechanics in porous media such as Biot's consolidation are also discussed.

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

Title: Estimating cardinalities for tables determined by multivariate predicates

Speaker: Kwong Hiu Yung, Interview Candidate, Stanford University

Date/Time: Tuesday, January 21, 2003, 9:15-10:15 (MST)

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

Brief Abstract: To improve query optimization, more accurate cardinality estimates of intermediate results are calculated using feedback from previous queries. Actual cardinalities of intermediate results are recorded for previously executed queries and then applied to optimization of future queries, often with similar sub-structure. Given marginal histograms and known bin counts, the remainder of the multivariate histogram is estimated using the maximum-entropy distribution. This approach to learning from past queries is tested against the TPC-H benchmark.

A cost-based optimizer chooses the best physical query plan based on cardinality estimates of intermediate tables. Because database systems maintain only univariate histograms, query optimizers often assume independence and uniform joins to estimate the cardinality of results satisfying multivariate predicates. These simplifying assumptions fail to capture important correlations between variables, and so the cardinality estimates are often skewed.

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

Chapter 6. Fellowships

The Computer Science Research Institute supported two students during CY2003 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 researchers 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, MS 1110
P.O. Box 5800
Sandia National Laboratories
Albuquerque, NM 87185-1110

dewombl@sandia.gov
(505) 845-7471 (voice)
(505) 845-7442 (fax)

Barbara J. DeLap

Computer Science Research Institute, MS 1110
P.O. Box 5800
Sandia National Laboratories
Albuquerque, NM 87185-1110

bjdelap@sandia.gov
(505) 845-7002 (voice)
(505) 284-4520 (fax)

Deanna R. Ceballos

Computer Science Research Institute, MS 1110
P.O. Box 5800
Sandia National Laboratories
Albuquerque, NM 87185-1110

drcebal@sandia.gov
(505) 845-7753 (voice)
(505) 284-4520 (fax)

Appendix A. Distribution:

External:

Dr. Dimitri Kusnezov
Director, Office of Advanced Simulation & Computing
NA-114/Forrestal Building
U. S. Department of Energy
1000 Independence Avenue, SW
Washington, DC 20585

Dr. Njema Frazier
Office of Advanced Simulation and Computing
NA-114/Forrestal Building
U.S. Department of Energy
1000 Independence Avenue, SW
Washington, DC 20585

Internal:

1	MS 0139	Art Hale, 01900
1	MS 0139	Robert K. Thomas, 01904
1	MS 0151	Rick Stulen, 01000
1	MS 0310	Mark Danny Rintoul, 01410
1	MS 0321	William J. Camp, 01400
5	MS 1110	David E. Womble, 01400
1	MS 1110	S. Scott Collis, 01414
2	MS 9960	Central Technical Files, 8945-1
2	MS 0899	Technical Library, 4536