Laboratory Directed Research and Development in High-Performance Computing and Simulation Science at Oak Ridge National Laboratory

ORNL uses a small portion (<3%) of its overhead budget to perform proof-of-principle type research and to develop capabilities that ensure the Laboratory stays on the cutting edge of science and technology in certain areas. This internal R&D program, formally called Laboratory Directed R&D (LDRD), has two components: Seed Money and Director's R&D. Seed Money provides an avenue of support for innovative, risky ideas that "bubble up" during the course of normal programmatic activity. The projects are relatively small and span 12 to 18 months in duration (Table 1). Director's R&D is for strategic purposes to develop capabilities that are consistent with the Laboratory's R&D agenda. Director's R&D projects are much larger than Seed Money projects (Table 1) and are typically funded for two years. The LDRD projects funded in high-performance computing and simulation science since FY 1998 are listed in Tables 2 and 3. Summaries and abstracts of this research are included below.

Table 1. ORNL LDRD Program

	Seed Money Fund	Director's R&D Fund
Purpose	Supports risky ideas	Strategic
Established	1974	1983
Funding cycle	Continuous	Annual
Proposal review	R&D Staff Members	Senior Management
Project budget	≤ $125,000	≤ $800,000
Project duration	12 to 18 months	24 to 36 months
LDRD outlay	25%	75%

Table 2. Seed Money Projects in High-Performance Computing: FY 1998–2002

Project Title	Project Leader (Division)	Year	Total Budget ($K)
First-Principles Investigation of Metal-Oxide Cathodes for Rechargeable Lithium Batteries	W. A. Shelton (CSMD)	1998-1999	101
Calculation of Positron Wave Functions, Lifetimes, and Momentum-Dependent Annihilation Probabilities	D. Nicholson (CSMD)	1999	15
Protein Structure Determination by Combining NMR Spectral Data and Protein Threading	Y. Xu (LSD)	1999-2000	100
Improved Methods for Methane Hydrate Resource Characterization	D. B. Reister (CSMD)	2000-2001	98
Correlated Structure and Dynamics in Quantum Dots at Finite Temperature via Auxiliary-Field Monte Carlo Method	J. C. Wells (CSMD)	2000-2001	100
Molecular Modeling and Simulation of the Formation and Decomposition of Methane Hydrates	P. T. Cummings (CSD)	2000-2001	100
A Collaborative Problem-Solving Environment for Welding Process Modeling	S. S. Babu (M&C)	2000-2001	95
Development of Spatial Relationships Between Linear Graphical Information-Systems Data of Different Scales	B. L. Bhaduri (CSED)	2001	15
Advancing Protein Structure Prediction by Integrating Two Revolutionary Technologies	Y. Xu (LSD)	2001-2002	100
Computing Transition States on High Dimensional Potential Surfaces with Application to Chemistry in Nanospaces	G. Ostrouchov (CSMD)	2001-2003	100
Simulation of Biomolecular Complexes: Advanced Computational Sciences at the Molecular Level	B. E. Hingerty (LSD)	2001-2002	60
Development of a Generic Computational Method for Biological Data Clustering	D. Xu (LSD)	2001-2002	100

Table 3. Director's R&D Projects in High-Performance Computing: FY 1998–FY 2002

Project Title	Project Leader (Division)	Year	Total Budget ($K)
Development and Application of Fast Computational Protein Folding Algorithms Using Massively Parallel Supercomputers	P. T. Cummings (CSD)	1998-2000	374
Regional Climate Modeling	J. B. Drake (CSMD)	1998-1999	438
Advanced Dynamic Simulation for Development and Control of Automotive Catalytic Converters	C. S. Daw (ESTD)	1999-2000	821
Continuum/Mesoscale Simulations of Failure in Solids	L. J. Gray (CSMD)	1999-2000	743
Development of Software Enabling Technologies for Terascale Computing	G. A. Geist (CSMD)	2000-2001	708
Information Systems to Support Experimental and Computational Research in Systems Biology	E. C. Uberbacher (LSD)	2000-2001	751
Integrated Respiratory System Model for the Virtual Human	R. C. Ward (CSED)	2000-2001	717
Computation with Arrays of Quantum Dots	J. Barhen (CSMD)	2000-2001	674
Computational Inference of Gene Function and Pathways: A Data-Mining Approach	Y. Xu (LSD)	2001-2002	625
Computational Nanoscience at the Terascale	D. J. Dean (PD)	2001-2002	600
Advanced High-Speed Networking Technologies to Support Terascale Computations	N. S. Rao (CSMD)	2001-2002	440
Multiscale Modeling Technique for Magnetic Nanostructures	T. C. Schulthess (CSMD)	2001-2002	538
A Scalable Virtual Environment for Scientific Computing	B. A. Worley (CSED)	2001-2002	315
Synthesis of High-Performance Algorithms for Electronic and Nuclear Structure Calculations	D. E. Bernholdt (CSMD)	2001-2003	520
Cellular Algorithms for Next Generation High-Performance Cellular Architectures	G. A. Geist (CSMD)	2001-2003	500
Scalable Tools for Petascale Distributed Data Analysis	G. Ostrouchov (CSMD)	2001-2003	590

Summaries and Abstracts of Seed Money Projects in High-Performance Computing and Simulation Science

Final Report Executive Summary

First-Principles Investigation of Metal Oxide Cathodes for Rechargeable Lithium Batteries *(Seed Money)*

W. A. Shelton and T. Kaplan
Computer Science and Mathematics Division

The development of rechargeable lithium and lithium-ion batteries is crucial to the advancement of portable electronic devices and the realization of electric vehicles. Lithium and lithium-ion batteries have superior energy, power densities, and cycle life as compared to other important battery technologies, such as nickel–metal hydride, nickel-cadmium, and lead-acid rechargeable systems. In fact, it is the electronic structure of these materials that gives rise to the important technological properties. The development of advanced lithium battery systems has rested largely on empiricism that is a time- consuming and economically expensive process. This type of approach, though initially successful, eventually yields minimal improvements. A simulation method capable of treating the effects of induced disorder on the electronic structure would provide the necessary information to accelerate the development of this important technology. Our goal is to extend first-principles, electronic-structure methods based on the coherent potential approximation (CPA) to these systems in order to understand the effects of disorder induced by (1) lithium insertion and extraction in the metal oxide cathodes during battery operation and (2) alloying on the transition metal site, which is used to try to improve battery performance.

We have developed a screened version of the Korringa, Kohn, and Rostoker coherent potential approximation (KKR-CPA) method. The KKR-CPA is a density functional theory, multiple scattering–based method that can treat multiple-atom, multiple-component disorder for all types of Bravais lattices. The screened KKR-CPA can be viewed as a tight binding–like representation that produces a sparse matrix representation. To date, we have performed test calculations containing over 1000 atoms on a single workstation containing 500 megabytes of random access memory. Equally important, this method incorporates the charge-correlated CPA, which includes certain short-range correlations that are ignored in the standard CPA formulation. This is the first and only computational method able to treat multiple-atom, multiple-component disordered systems for all Bravais lattices. Because this method can be used to treat large-scale general lattices, it can be applied to a number of other interesting scientific investigations including island formation on catalytic surfaces, surface magnetism, and possibly biological systems.

This project has led to some initial follow-on funding as well as a number of funding opportunities. The DOE High-Performance Computing and Communications Program Grand Challenge project, Materials, Methods, Microstructure and Magnetism, is providing $100,000. An ongoing 2% initiative involving the Metals and Ceramics and Solid State divisions on surface magnetism will provide funds to perform simulations on the Sr_2RuO_4 system. In addition, our proposal on metal-particle coarsening in catalytic materials involving metal-island formation has been chosen as one of the primary areas of the Engineering Simulation Initiative. We are also pursuing funding opportunities on lightweight batteries through the Office of Advanced Automotive Technologies, part of the Office of Transportation Technologies in DOE's Office of Energy Efficiency and Renewable Energy. This work has led to several presentations and submittal of a paper to *Physical Review Letters*.

Final Report Executive Summary

Calculation of Positron Wave Functions, Lifetimes, and Momentum-Dependent Annihilation Probablilities *(Seed Money)*

D. M. Nicholson
Computer Science and Mathematics Division

Positrons have been used at ORNL to characterize several systems. Interesting results have been obtained for the bulk amorphous alloy $Ni_{0.4}Pd_{0.4}P_{0.2}$, Al_2O_3 coatings, and the ordered intermetallic alloys NiAl and FeAl. As part of this effort, the Laboratory has recently developed the capability to measure the momentum-dependent annihilation probability. This is a new technique that provides information about the

atomic environment of the trapped positron. The goal of this project was to modify the locally self-consistent, multiple-scattering (LSMS) electronic structure code so that positron properties could be calculated and used to interpret the experimental results. Positron experiments in combination with modeling can provide information about free volume in metals. The free volume can in turn tell about the active diffusion mechanisms.

The LSMS code was modified and tested, and it can now be used to calculate positron wave functions for systems too large and complex to be handled by other codes. The calculation of positron states involves the lowest energy free-particle states at the bottom of the positron band. For electrons, this type of state has little impact on electronic properties, which are governed by the whole band and in particular by states near the Fermi level. To improve the accuracy with which the LSMS code modeled these low-energy states, the LSMS approach was modified by shifting the potential beyond the local interaction zone by a constant in order to get detailed agreement with test results. The LSMS code can now calculate the effective positron potential, the positron density of states, positron density, and positron lifetime. The code was applied to NiAl and FeAl with different vacancy configurations, and the results were conveyed to B. Somieski at Unversität Halle to help interpret his experiments. It was also applied to amorphous $Ni_{0.4}Pd_{0.4}P_{0.2}$, where it revealed that positrons are trapped near phosphorus atoms. This result implies that diffusion is controlled by free volume adjacent to phosphorus sites.

The work on $Ni_{0.4}Pd_{0.4}P_{0.2}$, was presented in invited talks at the Materials Research Society (MRS) meeting in November 1998, at Lawrence Livermore National Laboratory in December 1998, and at the Workshop on Thermodynamics and Structural Properties of Alloyed Materials in June 1999. It will appear as a refereed article in *Modeling and Simulation in Materials Science and Engineering*. A white paper on bulk amorphous metals was sent to Bob Gottschall at the Office of Basic Energy Sciences in DOE's Office of Science. The capability to calculate positron lifetimes bolstered the strength of the white paper.

Final Report Executive Summary

Protein Structure Determination by Combining NMR Spectral Data and Protein Threading

(Seed Money)

Y. Xu
Life Sciences Division

The goal of this project was to investigate new methods for protein structure determination by combining sparse nuclear magnetic resonance (NMR) data and protein threading. As one of the two main experimental approaches for protein structure determination, NMR does not require the crystallization of proteins and, hence, can be potentially used as a high-throughput method. However, current NMR techniques are limited to small proteins (< 30 kDa). The problems with larger proteins are the increased number of resonances and line broadening, which result in spectral crowding and reduction in the fraction of spectral peaks that can be identified and assigned. Protein threading is becoming one of the main computational methods for protein structure prediction, as more and more experimental structures have been solved and deposited into the protein databank. Potentially it applies to 70–90% of the proteins. However, its prediction accuracy is far from being on the same level as that obtained with NMR, and it is limited to backbone structure predictions. Because information derived from NMR experiments complements that derived from threading computation, effectively combining these methods could lead to new paradigms for structure determination that would require significantly less NMR experimental time with lower costs and might be applicable to larger proteins.

The project focused on two paradigms for combining sparse NMR data with protein threading: (1) application of sparse NMR data as a constraint on protein threading to improve the threading accuracy; and (2) development of methods for compensating for missing NMR data using structural information from protein threading. Significant progress was made in both areas. We developed a rigorous computational framework for NMR-constrained threading, the first such paradigm reported in the literature. Using this framework, we demonstrated that a small number of nuclear Overhauser effects (NOEs) can significantly improve the threading accuracy, in both fold recognition and threading alignments. Also, it can extend the scope of threading from structural homologs to structural analogs—a surprising discovery! In the second area, we have convincingly demonstrated that the structural information provided by an approximate backbone structure (with 4–5 Å rmsd) from protein threading can greatly reduce the required number of NOEs typically needed for a ~2-Å NMR structure. This implies that fewer NMR experiments can potentially yield an accurate structure with the help from threading. Active collaborations are under way with a number of NMR labs to further develop the technology.

Although this work is at a very early stage of development, it has gained much recognition. It resulted in four invited oral presentations at national and international conferences and the publication of two refereed papers, one an invited paper in the *Journal of Computational Biology* [Vol. 7, No. 3/4, pp. 449–67 (2000)]. As a result of this project, we have started collaborating with top NMR researchers, including Dr. Michael Summers of the University of Maryland at Baltimore County, Dr. Werner Braun of the University of Texas, and Dr. X. Gao of the University of Houston. A proposal submitted to the DOE Structural Biology Program has been funded to further develop this technology in collaboration with Dr. Michael Summers. Another proposal has been submitted to the NIH NIGMS as a joint project with NMR expert Prof. Engin Serpersu at the University of Tennessee. Another NIH proposal will be submitted to the Structural Genomics review panel after some preliminary results are collected from a joint project with Dr. Gao.

Final Report

Improved Methods for Methane Hydrate Resource Characterization *(Seed Money)*

David B. Reister and N. S. V. Rao
Computer Science and Mathematics Division

The goal was to develop better methods for using measured data to improve estimates of hydrates in ocean sediments. There are two main methods for making measurements: regional seismic and ocean drilling. Seismic data can be collected for large volumes of sediments at a reasonable cost. High cost ocean drilling provides detailed data along holes that penetrate the sediments at a few isolated locations. The ocean drilling data produces several independent estimates of the hydrate concentration in the sediments. Using advanced information fusion techniques, the multiple estimates are combined into a single best estimate. The existing models that used measured seismic velocities to estimate hydrate concentration are heuristic and inaccurate. Our primary scientific accomplishment has been to develop a new method for using measured P-wave and S-wave velocities to estimate gas hydrate concentration. Our method is built on the firm foundation of rock physics and uses the Hashin-Shtrikman (HS) lower bound to provide both an upper and lower bound for mixtures of water and hydrate in unconsolidated marine sediments. We bridge from the upper to the lower bound by using both the HS upper bound and the Voigt upper bound.

Final Report Abstract

Correlated Structure and Dynamics in Quantum Dots at Finite Temperatures via Auxiliary-Field Monte Carlo Method *(Seed Money)*

Jack C. Wells,[1] David J. Dean,[2] and Michael R. Strayer[2]
[1]*Computer Science and Mathematics Division*
[2]*Physics Division*

The focus of this project is the theoretical description and numerical simulation of the correlated structure and properties of semiconductor quantum dots at finite temperature. We approached this problem via the Auxiliary-Field Monte-Carlo (AFMC) method and the finite-temperature mean-field theory for many-body quantum mechanics. This approach is ideally suited for implementation on high-performance parallel computers.

Final Report Abstract

Molecular Modeling and Simulation of the Formation and Decomposition of Methane Hydrates *(Seed Money)*

Peter T. Cummings and Ariel A. Chialvo
Chemical Sciences Division

The broad goal of this research was to answer two fundamental questions from a molecular-based viewpoint (a) How can we predict phase stability in multicomponent systems containing methane hydrates?, and (b) How does the pore size and surface chemistry of a porous solid matrix affect the phase stability of methane hydrate systems?

In order to tackle this broad goal successfully we proposed to demonstrate our modeling capabilities by performing several groundbreaking calculations that will be publishable in high-quality journals. Particularly, our goals for this project were (c) the determination of thermodynamic and structural properties of methane and carbon dioxide hydrates, and (d) the determination of relative stability between of methane and carbon dioxide hydrates.

Toward the achievement of these goals we studied, by isobaric-isothermal molecular dynamics of realistic host-guest representations, the structure and the thermodynamic properties of methane and carbon dioxide hydrates. The simulation results were then used to interpret the limitations and main implications behind the most popular macroscopic modeling tool, i.e., the so-called van der Waals-Platteeuw theory (vdWP) and its modifications. In addition, we developed a simulation methodology to analyze the feasibility of methane replacement by carbon dioxide sequestration as an alternative way to exploit methane by simultaneously mitigating the buildup of carbon dioxide in the atmosphere.

Progress Report Abstract

A Collaborative Problem Solving Environment for Welding Process Modeling

(Seed Money)

S. S. Babu,[1] M. Summers,[2] and G. Sarma[2]
[1]*Metals and Ceramics Division*
[2]*Computer Science and Mathematics Division*

The objective of this project is to develop a problem-solving environment (PSE) that will permit advanced welding process models to be constructed by integrating distributed sub-models. This research will allow all the physical processes that affect a weld to be accounted for, resulting in better predictive capabilities. This requires models to account for the interaction of complex physical processes, such as heat transfer, fluid flow, thermo-chemical reactions, thermo-mechanical effects, and phase transformations. In the research so far, we have developed sub-process models for describing heat transfer in simple geometry and microstrucutre prediction in low alloy steel welds. We have also developed a framework to integrate these sub-process models as a collection of INTERNET web services. Each of sub-model is implemented as one or more a java Servlets within a web servers.

Final Report Abstract

Development of Spatial Relationships Between Linear GIS Data Sets of Different Scales *(Seed Money)*

Budhendra Bhaduri, Phillip Coleman, and Edward Bright
Computational Sciences and Engineering Division

The main objective of this proposed research was to build an GIS-based tool to establish spatial and network relationships between two national hydrologic data sets that can serve as a prototype for data linkage that can be applied to similar linear data sets such as roads, railroads, and utility networks. This linkage is expected to greatly bridge the lag between model and data developments. Given the volume and complexity of national GIS data sets, such a process is computationally intensive and requires an innovative approach of

using GIS techniques to make the process time and cost efficient. In this work, we discuss the enhancement of River Reach File 1 (RF1) data to the National Hydrography Data (NHD) set, and how a tool or "crosswalk" can be built between these two data sets to allow models (that are currently designed to use RF1) use the spatially refined information offered by NHD to improve a regional regression model. Although this research is specifically focused on hydrologic data, a successful development of a "crosswalk" tool will have a much broader impact and can be modified for application on other similar liner GIS data sets including transportation or utility features.

Progress Report Abstract

Advancing Protein Structure Prediction by Integrating Two Revolutionary Technologies *(Seed Money)*

Michael Unseren[1] and Ying Xu[2]
[1]*Computer Science and Mathematics Division*
[2]*Life Sciences Division*

The proposed research seeks to develop ground breaking algorithms and computer algorithms for predicting the 3D structure of proteins given their amino acid sequences. This will be accomplished by integrating two distinct protein structure prediction methods: threading and ab initio folding. Specifically, we will integrate two ORNL-based methodologies: (i) TRUST, the best continuous global optimization algorithm available, and (ii) PROSPCT, the best computer program available for threading. The idea is to use the optimal backbone structures obtained by PROSPECT as a starting point for a TRUST-based ab initio folding. Present ab initio folding methods are computationally too intensive for practical applications and are often intractable for protein sequences. Our approach will significantly reduce the size of the conformation space that must be searched. After obtaining a refined backbone structure prediction via integration, we will integrate TRUST and an established method for predicting the 3D structures of the sidechains that is based on rotamer libraries. TRUST-based ab initio folding algorithms will be developed using the rotamer library predictions as a starting point. The planned research consists of two tasks: (i) Task 1 is to refine prediction for the backbone 3D structure by integrating PROSPECT with TRUST; and (ii) Task 2 is to refine prediction of side chain 3D structures by integrating TRUST and prediction methods based on rotamer libraries.

Progress Report Abstract

Computing Transition States on High Dimensional Potential Surfaces with Application to Chemistry in Nanospaces *(Seed Money)*

George Ostrouchov, Bryan Hathorn, and Charles K. Bayne
Computer Science and Mathematics Division

The proposed research is divided into two tasks. The first task examines the model systems to represent the chemical dynamics of molecular changes that will be applied to small polymer chains. The second task is to develop computational methods using response surface methodology to calculate the optimum potential energy surfaces for small polymer chains. Scaling up this work to larger systems will provide a variety of interesting research problems in sparse matrix computation (solving a series of large sparse linear algebra problems with a gradually varying sparsity pattern). Because the number of possible paths being followed at each stage of the computation may be large, the larger problem will be interesting from a parallel computing standpoint and provide a high-impact scalable application to solve on a high performance parallel computer. Here we report on the progress made during the first three months of the project, representing about one quarter of the total funding.

New Project Abstract

Simulation of Biomolecular Complexes: Advanced Computational Sciences at the Moleculer Level *(Seed Money)*

Brian E. Hingerty,[1] Tuan Vo-Dinh,[1] Suse Broyde,[2] and Monique Cosman[3]
[1]*Life Sciences Division*
[2]*Biology Department, New York University*
[3]*Lawrence Livermore National Laboratory*

We will use advanced theoretical modeling to study the interactions between bioreceptors (DNA, protein, antibodies) with target species at the molecular level. We will use *a unique, advanced computational program, DUPLEX*, developed by B. Hingerty to computationally determine specific structures of RNA, DNA, damaged DNA, and DNA-protein complexes. We will expand our extensive previous experience with carcinogen-DNA adducts to model target bioreceptors. Experimental studies will be done in collaboration with T. Vo-Dinh and his coworkers at ORNL using *unique nanosensors* for the real-time detection of a wide range of biotargets. These nanosensors will take advantage of near-field spectroscopy to monitor the interactions of various bioreceptor molecules (DNA, antibodies, enzymes, RNA) with their respective analyte at the molecular level. From these fundamental experiments, values of association constants, disassociation constants, binding affinities, binding speeds, etc. will be determined, and used to evaluate simulation models.

New Project Abstract

Development of a Generic Computational Method for Biological Data Clustering *(Seed Money)*

Dong Xu, Victor Olman, and Ying Xu
Life Sciences Division

The proposed research will seek to develop a general-purpose clustering method for various biological data. Currently clustering for biological data is done using a large number of *ad hoc* clustering methods, and there is no optimal solution for clustering in general. It is important to have a generic computational method for biological data clustering, as more and more data being generated. We plan to develop a general clustering method based on the representation of multi-dimensional data as a minimum spanning tree. Such a method potentially works well on different types of data and simplifies a multi-dimensional clustering problem to a tree-partitioning problem without losing essential information for the purpose of clustering. Our method will include general clustering algorithms dealing with complicated issues related to different forms of data. We will test the method on the construction of protein trees and the analysis of two-hybrid data. This proposal constitutes high-risk research since no general-purpose clustering method has ever been developed. However, based on our limited preliminary studies, this approach is very promising.

Summaries and Abstracts of Director's R&D Projects in High-Performance Computing and Simulation Science

Final Report Executive Summary

Development and Application of Fast Computational Protein Folding Algorithms Using Massively Parallel Supercomputers *(Director's R&D)*

P. T. Cummings,[1] P. F. LoCascio,[2] and K. A. Dill[3]
[1]*Chemical Technology Division*
[2]*Life Sciences Division*
[3]*University of California, San Francisco*

Our objective was to develop a world-class capability at ORNL to develop and apply fast *ab initio* computational protein folding algorithms, with the goal of first-principles prediction of the 3-D structures of proteins from their amino acid sequence. The proposed work involved two tasks: (1) parallelization of Dill's Geocore protein folding algorithm to significantly increase the speed of calculations, and (2) improvement of the Geocore protein folding algorithm.

Our first objective was to parallelize the serial Geocore algorithm. Since Geocore is a branch and bound search algorithm on a discrete set of possible states, we successfully parallelized Geocore by developing a mechanism for distributing subtree searches over many processors and having a robust mechanism for assigning work to idle processors. The parallelization efficiency was almost perfect. On the 1024-node Intel Paragon at ORNL, we achieved a speedup of 1011. Such a speedup has a dramatic effect: It means that a calculation that takes 5 days on a single processor can be completed in 18 minutes. As a result, it becomes possible to optimize the energy parameters in the Geocore model, since the time required to predict the structure of a protein with a given set of energy parameters is reduced to a manageable level. Our second objective was to develop an improved Geocore algorithm. We began by testing Geocore on a larger number of proteins and peptides. We studied 18 short molecules for which there are structures in the Protein Data Bank. Next, we added to Geocore the capability to directly handle secondary structures. Since this was a major enhancement, we call the upgraded algorithm Geocore 2. Geocore 2 addresses one of the main problems in computational protein folding: to devise methods that can go beyond helix formation and that can assemble secondary structures into tertiary conformations. Finally, we added flexibility in side chain packing and revised the energy functions in Geocore to improve structure prediction.

This project resulted in two journal publications, two presentations at national meetings, and four invited talks. Two proposals submitted to DOE were not funded. When the project was proposed in 1997, there existed very promising funding opportunities for *ab initio* computational protein folding within DOE in the DOE-SC OBER Structural Biology Program. However, OBER has now steered its emphasis toward high-throughput bioinformatics-based approaches. Although our work has not led to follow-on funding for *ab initio* computational protein folding, Phil LoCascio is applying some of the expertise gained in this project to bioinformatics in his new position of computational biology group leader in the Life Sciences Division.

Final Report Executive Summary

Regional Climate Modeling *(Director's R&D)*

J. B. Drake,[1] J. A. Kohl,[1] G. Ostrouchov,[1] W. M. Putman,[1] W. M. Post,[2] A. W. King,[2] M. J. Sale,[2] Y. S. Bao,[2] F. M. Hoffman,[2] T. J Wilbanks,[2] and W. W. Hargrove[3]
[1]*Computer Science and Mathematics Division*
[2]*Environmental Sciences Division*
[3]*Computational Sciences and Engineering Division*

The goal of this project was to establish the connection between advanced general circulation climate prediction models and regional-scale information needs for a scientifically grounded assessment of the regional consequences of climate change and climate variability. This required an understanding of the coupling of large-scale climatic processes of the global circulation with local ecological and hydrological processes. First we gathered and processed the long-term record of stream flow from gauging stations in the Southeast region of the United States. The development of ways to benchmark a regional climate model's hydrological results against this observational record is important for understanding the range of reliable

predictions. Second, we developed methods for producing regional downscaled climate predictions. We pursued physically based approaches as well as statistical methods based on observations. We developed a new statistical downscaling technique based on a combination of empirical othogonal function projection, canonical correlation analysis, and the theory of nonlinear dynamical systems. Finally, we made a regionally downscaled prediction of the change in temperature, precipitation, and stream flow from the projected climate change over the next 120 years with atmospheric CO_2 increasing at a rate of 1% per year. The accomplishments of investigators involved with this project have allowed ORNL to play a significant role in the planning and preparation of proposals for a network of regional climate centers to assist with ongoing assessment efforts. Some of the key technological and scientific expertise for regional climate research is now part of the ORNL portfolio. In addition, this project brought to ORNL mesoscale climate and weather modeling capability, regional climate diagnostic capability, and an assembled set of tools and data to enable proposed investigations.

Final Report Executive Summary

Advanced Dynamic Simulation for Development and Control of Automotive Catalytic Converters *(Director's R&D)*

C. S. Daw[1] and W. A. Shelton[2]
[1]*Engineering Science and Technology Division*
[2]*Computer Science and Mathematics Division*

The project goal was to develop unique capabilities at ORNL for simulating the dynamic performance of catalytic reactors used to reduce automotive and truck emissions. In particular, we focused on simulating advanced catalyst materials and reactor configurations proposed as candidate technologies for removing nitrogen oxides, hydrocarbons, and carbon monoxide from advanced lean-fueled internal combustion engines (i.e., engines designed to operate with a large excess of oxygen). Such engines are now considered by DOE and industry to be the most feasible near-term approach for reducing greenhouse gases and fuel consumption in the transportation sector. Implementation of these advanced engine designs depends critically on developing aftertreatment technologies for reducing nitrogen oxide emissions in lean exhaust. It is clear that the new emissions standards for both cars and trucks cannot be met by existing technology, and thus new types of catalytic reactors will be required. Development of new catalyst formulations, converter designs, and associated control strategies is currently almost entirely empirical and very inefficient. Thus, it is expected that improved simulation capabilities could significantly reduce the cost and time involved. The biggest stumbling block to such simulations is the accurate integration of the dynamics over the large range of length and time scales that are important to the relevant chemistry and transport phenomena. Our approach was to construct a coupled flow and kinetics model that starts at the intermediate scales and then bridges both the large and small scales in a way that preserves the essential dynamics under realistic driving conditions. We also emphasized the capabilities for experimentally validating the model components and exploiting parallel construction for high-performance computing.

At the initiation of this project, the primary targeted opportunity for follow-on funding was the DOE-SC Scientific Simulation Initiative (SSI). One of the three main focus areas proposed for SSI was the Combustion Simulation and Modeling Initiative (CSMI). At that time, ORNL was the only national laboratory that was specifically pursuing the simulation of aftertreatment chemistry and materials in conjunction with CSMI. In FY 2000, SSI and CSMI were removed from the DOE-SC budget, and a major redirection of effort was required on our part to identify an alternate follow-on target. We accordingly refocused our research activities to include more emphasis on measurement of key physical parameters and the integration of our catalyst submodel into more global systems models. The result of this redirection included the construction of the new Supported Catalyst Infrared Thermography reactor in the Advanced Propulsion Technology Laboratory and additional validation of our 1-D single-channel model against large-scale experimental performance measurements of a diesel oxidation catalyst. In addition, we successfully completed most of our original plans, including (1) 1-D, single- and multi-channel transient catalytic reactor models for typical catalyst monoliths; (2) a single-particle reaction model that can be used to relate detailed catalyst washcoat properties to the propagation of unsteady reaction fronts; and (3) a new Monte Carlo approach for computing global reaction kinetics from molecular-scale simulations.

We now believe that ORNL is in a unique position to respond to several important aftertreatment simulation needs that have been identified by DOE-EE's OTT and industry. We have received funding of over $300,000 from OTT to begin adapting our models to NO_x traps and global simulation of the full engine catalyst system and developing submodels that couple atomistic catalytic surface dynamics to the global scale. Aftertreatment modeling proposals requesting over $900,000 for FY 2002 have been submitted to

DOE-EE and OTT. This total includes a $500,000 request to the DOE-EE Energy Efficiency Science Initiative. ORNL has assumed a major role in coordinating the Aftertreatment Systems Simulation Initiative for the DOE Diesel Crosscut Team. In this role, ORNL is currently the only national laboratory member of the crosscut subcommittee that is charged with developing a joint DOE/industry plan for lean exhaust aftertreatment simulation. Two diesel engine companies, Cummins and Detroit Diesel Corporation, have requested that ORNL join in individual aftertreatment simulation CRADAs with them. We are also assisting OTT in defining a new Engineering Simulation Initiative that is currently planned as a line item in the FY 2002 budget.

Final Report Executive Summary

Continuum/Mesoscale Simulations of Failure in Solids *(Director's R&D)*

L. J. Gray,[1] T. Kaplan,[1] J. D. Richardson,[1] A.-V. Phan,[1] T. Besmann,[2] and J. A. Haynes[2]
[1]*Computer Science and Mathematics Division*
[2]*Metals and Ceramics Division*

The goal of this project was to develop an advanced computational capability, based on recent mathematical advances, to model material evolution at the continuum and mesoscale levels. The project was to focus on failure mechanisms in solids. Breakdown of TBCs was chosen as a test application. Cracking and debonding of TBCs is a critical failure mechanism of turbine blades in land-based gas turbine power generators and aircraft engines. Experiments to characterize the breakdown process were to be carried out for void-forming coatings and then compared to simulations.

Our efforts to develop an advanced computational capability divided into two areas: new mathematical methods and code development. This project led to three important breakthroughs in mathematical methods. (1) We developed a new crack tip element that has been shown to be at least an order of magnitude more accurate than the quarter-point element that has been the standard element used in finite and boundary element analyses for the past 25 years. (2) We have successfully derived a closed-form expression for the Green's function of a graded material wherein the material properties vary exponentially. With this Green's function, an analysis for graded materials can be formulated as simply as for homogeneous media in terms of boundary integral equations. This is the first serious application of this method to nonhomogeneous media. (3) We have developed an advanced fast multipole method that enables us to simulate extremely large and complex systems.

We have achieved significant results in code development as well. The boundary integral stress codes have been successfully coupled with the level set boundary evolution method. This code has already provided the most accurate simulation of the interface instability observed experimentally in silicon crystal growth in a system with an applied external stress. We also developed an advanced 2-D crack propagation code. This code, which can treat many different material regions, includes body forces, incorporates the new crack tip element, and is based on the symmetric Galerkin boundary integral formulation. Boundary integral methods are the most accurate for treating cracks and by far the most efficient method for tracking crack propagation. Significant advances have been made in developing the 3-D version of the crack propagation code, although further work is needed. In the experimental part of this project, interfacial and internal void- forming films were grown and thermally cycled to simulate conditions that induce cracking in TBCs. However, these films showed almost no cracking. While yielding an interesting result, adequate examples for testing the crack propagation code were not produced.

This project has produced a number of important scientific results. They are reported in seven journal articles (two in press, one submitted, and four in preparation), four oral presentations at national and international meetings (two invited and two contributed), and three seminars (two at universities and one at a research laboratory). In addition, we have been invited to write a book chapter on fracture and boundary integral methods in a special volume honoring Prof. Frank Rizzo. Finally, we have submitted four proposals for follow-on funding, all of which are pending.

Final Report Abstract

Development of Software Enabling Technologies for Terascale Computing *(Director's R&D)*

Stephen L. Scott, Al Geist, Ray Flanery, and Jack Dongarra
Computer Science and Mathematics Division

The purpose of this project was to provide enabling technology to the scientific simulation community so that they may reap the benefits of the rapidly improving high-performance computing and communication technologies. This was done through the rapid prototyping and deploying of advanced technology capabilities for terascale computing. Furthermore, it was the intent of this project to establish ORNL in a leadership role in several key emerging areas in software for terascale computing. One such area is in software tools for meta-cluster management, administration, and use of federated clusters operating within the computation GRID environment.

Final Report Abstract

Information Systems to Support Experimental and Computational Research in Systems Biology *(Director's R&D)*

J. Snoddy,[1] E. Baker, [1] S. Petrov,[1] D. Schmoyer,[2] M. Land,[1] B. Jackson,[2] D. Johnson,[1] and E. Uberbacher[1]
[1]*Life Sciences Division*
[2]*Computer Science and Mathematics Division*

In order to promote a research capability to understand complex biological systems, the ability to acquire, manage, and interpret the complex information of biology is a prerequisite. This project's goal was to design and build a series of computer-based analysis tools and data management systems to aid in the study of biological systems, to automate the routine operations that are needed in large-scale data-driven research projects, and the manage and share the resulting data. These information systems enable mammalian geneticists to analyze mouse data from both data-driven and hypothesis-driven research and connect genome-scale data or other data-driven approaches with more focused, smaller-scale hypothesis-driven research into complex mammalian biosystems. The combined capabilities support biological inference and decision-making by biologists.

Final Report Abstract

Integrated Respiratory System Model for the Virtual Human *(Director's R&D)*

R. C. Ward,[1] K. L. Kruse,[1] P. T. Williams,[1] N. B. Munro,[2] C. E. Easterly,[2]
G. O. Allgood,[3] S. S. Gleason,[3] L. M. Hively,[1] R. J. Toedte,[4] and J. J. Dongarra[4][1]*Computational Sciences and Engineering Division*
[2]*Life Sciences Division*
[3]*Engineering Science and Technology Division*
[4]*Computer Science and Mathematics Division*

The goal of the project was to develop computational, imaging, experimental, and data analysis capabilities that would facilitate prediction of pulmonary diseases or injuries from lung sounds. These capabilities include simulation of the propagation and attenuation of lung sounds in the thorax, generation of lung sounds in model airways, and comparison with auditory recordings of breath sounds available to the team. The ORNL MicroCAT was used to obtain images for geometric airway models. A computational environment, using terascale computational and distributed, agent-based resources, was developed to integrate simulation of lung sound generation and propagation.

Progress Report Abstract

Computation with Arrays of Quantum Dots *(Director's R&D)*

J. Barhen,[1] Y. Braiman,[1] D. J. Dean,[2] L. Maya,[3] V. A. Protopopescu,[1]
N. S. V. Rao,[1] M. R. Strayer,[2] T. G. Thundat,[4] and J.C. Wells[1]
[1]*Computer Science and Mathematics Division*
[2]*Physics Division*
[3]*Chemical Sciences Division*
[4]*Life Sciences Division*

The goal of this proposal is to develop a quantum-dot array for carrying out innovative computations to address the information processing needs of future intelligent systems. The effort includes the actual fabrication of 2-nm gold clusters, device architecture, device simulation, development of a computational model, and novel applications. Innovative and unconventional paradigms underlie the different stages of the project. Regular array geometry will be achieved by the development of techniques for the programmed assembly of appropriately functionalized gold clusters to preselected locations along stretched strands of engineered DNA sequences. Our applications include the implementation of neuromorphic algorithms for pattern recognition.

Progress Report Abstract

Computational Inference of Regulatory and Metabolic Networks *(Director's R&D)*

Ying Xu,[1] Frank Larimer,[1] Victor Olman,[1] Dong Xu,[1] Jay Snoddy,[1] Ed Uberbacher,[1]
Denise Schmoyer,[2] Barbara Jackson,[2] and Jeff Becker[3]
[1]*Life Sciences Division*
[2]*Computer Science and Mathematics Division*
[3]*Department of Biochemistry, University of Tennessee at Knoxville*

Biological processes (e.g., signal transduction) are conducted through a sequence or network of interacting genes/proteins. Knowing the details of such processes and their pathways provides a foundation for understanding how a biological function is achieved at both the molecular and organism level. Thus, the importance of this proposed study lies in developing a new capability for functional studies of genes/proteins in the post-genome sequencing era that can be used to provide practical models of function and phenotype. It is generally recognized by the scientific community and major funding agencies for biological research that the next phase of biology will involve learning to understand and model these systems of genes and proteins. However such systems are very complex and current methods for elucidating the detailed relationships between genes in pathways and networks are limited and subject to considerable uncertainty.

Progress Report Abstract

Computational Nanoscience at the Terascale: Self-Assembled Monolayers and Molecular Electronics *(Director's R&D)*

D.J. Dean,[1] W.H. Butler,[2] P.T. Cummings,[3] S. Cui,[3] P. Krstic,[1]
D.R. Schultz,[1] M.R. Strayer,[1] J. C. Wells,[4] and X. Zhang[4]
[1]*Physics Division*
[2]*Metals and Ceramics Division*
[3]*Chemical Sciences Division*
[4]*Computer Science and Mathematics Division*

We propose a computational simulation of the chemical assembly, structural properties, and conductance characteristics of self-assembled monolayers (SAMs) of organic molecules that may be used in emerging revolutionary computational devices. The necessary science for a wall-to-wall simulation over multiple scales requires a vertically integrated and interdisciplinary effort encompassing molecular dynamics simulations, molecular structure computation (such as Density Functional Theory (DFT)), and conductance calculations based on both DFT and approaches with full dynamic electron correlation. These simulations,

relying on ORNL's terascale computational facilities, and the corresponding code developments will allow us to predict device properties for molecular-scale electronics.

Progress Report Abstract

Advanced High-Speed Networking Technologies to Support Terascale Computations *(Director's R&D)*

N. S. Rao, J. Barhen, W. C. Grimmell, T. Dunigan, and W. R. Wing
Computer Science and Mathematics Division

Our objective is to initiate and develop a research program targeting high-performance networks to support terascale computing and simulation applications. These applications demand quality-of-service unprecedented and unavailable in current networks. We propose state-based dynamic path planning and routing technologies for the next generation of high-performance networks to ensure the required levels of end-to-end performance. Based on in-situ network measurements and/or estimates provided by the instrumentation routines, we estimate the "dynamic state" of the network. The state of the network is then utilized to transfer data on demand over the high-performance networks with the required quality-of-service.

Progress Report Abstract

Multiscale Modelling Technique for Magnetic Nanostructures *(Director's R&D)*

T. C. Schulthess,[1] W. A. Shelton,[1] and G. M. Stocks[2]
[1]*Computer Science and Mathematics Division*
[2]*Metals and Ceramics Division*

We are developing a tool-set for computational magnetism at the nano-scale, where we make use of object oriented and generic design and implementation techniques. The aim is to demonstrate how efficient, generic, and scalable tools can be developed from which applications to solve relevant research problems in nano-magnetism can be built in a timely manner. The tools are used as they are developed to build computer models for ongoing research projects at ORNL.

Progress Report Abstract

A Scalable Virtual Environment for Scientific Computing *(Director's R&D)*

B. A. Worley and Ray Flanery
Computational Sciences and Engineering Division

Terascale computer simulations often produce very large amounts of data that overwhelm current data management, access, storage, networking, and visualization technologies. Furthermore, the interpretation of these vast volumes of data often requires the experience and expertise of several individuals. Thus, collaborative visualization in a shared immersive environment is emerging as a new and effective way of extracting scientific insights from large numerical data sets. The research purpose of this project is to take advantage of the CAVE immersive system at ORNL to initiate a research program in the area of the collaborative and immersive visualization of large data sets. The first year accomplishments were research and understanding of scalability and performance issues for immersive environments, putting together the CAVE immersive environment, and developing the initial prototype of scalable immersive visualization for large scientific data sets. The second and final year goal is to make the visualization system resource-aware so that the tasks of data extraction, processing, rendering, and communication across the network can be optimized to enhance performance, and hence the illusion of immersion.

New Project Abstract

Synthesis of High-Performance Algorithms for Electronic and Nuclear Structure Calculations *(Director's R&D)*

D. E. Bernholdt,[1] James B. White III,[1] David J. Dean,[2] and Michael R. Strayer[2]
[1]*Computer Science and Mathematics Division*
[2]*Physics Division*

The development of high-performance parallel programs for scientific applications is usually very complex, and it is being made more so by the increasing size and complexity of modern parallel computers. This proposal provides a means to address the issue of software complexity while at the same time automating the creation of highly-scalable algorithms capable of taking into account characteristics of the specific problem and hardware platform. We will develop a suite of tools to translate a high-level description of the equations to be solved into highly optimized source code, which can then be compiled and run as usual. Working with a high-level representation of the problem allows significant optimization of the algebraic expressions themselves, trade-offs between storage space and computation time, management of storage for out-of-core problems, and distribution of the data across the parallel computer. We will apply these techniques to produce a *Tensor Contraction Engine* targeted at a class of problems in computational chemistry and computational nuclear physics. This proposal is intended to complement a proposal recently submitted to the National Science Foundation by the same team. LDRD funding will allow more active involvement by the ORNL investigators than is possible with the NSF proposal alone, thus allowing us to develop an in-house expertise in this area, which can later be applied to other computational domains in collaboration with other ORNL researchers.

New Project Abstract

Cellular Algorithms for Next Generation High-Performance Cellular Architectures *(Director's R&D)*

Al Geist,[1] Manish Gupta,[2] Jay Snoddy,[3] and Bill Shelton[1]
[1]*Computer Science and Mathematics Division*
[2]*International Business Machines Corporation*
[3]*Life Sciences Division*

The world of high performance computing is rapidly changing and a new type of highly scalable computer called *cellular architecture* is emerging as the next generation of computer for scientific simulation. In order to position ORNL for such 100 Tflops systems we must begin today to develop new approaches to solving computational problems that are able to scale to tens of thousands of processors and at the same time be tolerant of failures of some of these processors. This proposal is about the development of a new programming paradigm we call *cellular algorithm*, which is suited to these next generation computers. To validate this concept our research includes collaboration with application scientists from materials and biology as well as the cellular architecture group in IBM. They will be applying cellular algorithms to important problems in biology, materials and system monitoring.

ORNL has experienced tremendous growth in computing power in the past year and a half going from 150 Gflops to over 5000 Gflops by this winter. During this period, ORNL has become the largest unclassified computing center in DOE. But if we are going to maintain this position we must begin preparing for the next generation of computers that will take us from 10 Tflops to over 100 Tflops. These next generation computers are being designed with tens of thousands of processors where each processor forms a "cell" with memory, communication, and I/O built into it. These types of computers are being called ***cellular architecture*s** and are the most promising computers to get to 100 Tflops. Present software is simply not scalable enough to work on such computers and no one in the world yet understands how to use or develop programs for cellular architectures. The opportunity exists for ORNL to invest in becoming the world expert on such computers. Then ORNL would become the place where scientists from many different fields would want to come to learn how to run their applications on 100 Tflops computers.

New Project Abstract

Scalable Tools for Petascale Distributed Data Analysis *(Director's R&D)*

George Ostrouchov, Nagiza Samatova, David Erickson, and Ross Toedte
Computer Science and Mathematics Division

The vision of this proposal is to enable scientists to perform complex analyses of distributed data on a computational grid with scalable tools that smoothly operate at the petascale. To begin realizing this vision, we focus this proposal on scalable dimension reduction computations and visualization for astrophysics and climate simulation data. The broader purpose is to continue developing other scalable distributed data analysis tools as enabling technologies and to involve other applications, such as computational biology, neutron science, material science, and combustion analysis through follow-on funding.

Dimension reduction techniques are fundamental to discovery and visualization of structure in high-dimensional data. These computationally intensive methods are used in a number of applications under different names, however most are based on a variation of the same underlying computation. Pushing the frontier of dimension reduction tools to distributed data will enable progress across many applications and lead to opportunities for funding in these areas. Astrophysics and climate simulation are sources of virtually unlimited high-dimensional data and thus provide an ideal test bed for developing these tools. However, dimension reduction tools are equally applicable to data collected from other physical or biological systems.

The goal is to establish ORNL as leader in key technologies of the emerging area of scalable distributed data analysis. Our team and the resulting set of innovative software tools for scalable distributed dimension reduction will put ORNL in a strong position to pursue further funding in high-performance computing for data analysis, to influence program development at DOE, and to become a favorite resource for many application areas that generate distributed data.

Oak Ridge National Laboratory is managed by UT-Battelle for the Department of Energy. ORNL is the largest DOE science and energy laboratory, conducting basic and applied research to deliver transformative solutions to compelling problems in energy and security.Â This research used resources of the Compute and Data Environment for Science (CADES) at the Oak Ridge National Laboratory 16:51. The Sound of Science - TCR: Printing the future of nuclear.Â Suzanne is an high performance computing engineer at Oak Ridge National Laboratory. Her daughter, Alex, had the opportunity to ask her questions about what she does at the lab. 5:42. Exascale Day gives the high-performance computing (HPC) industry a chance to reflect on the progress made in bringing exascale systems to fruition and the new research areas that they will enable. In honor of the day, AMD and HPCwire hosted a discussion on Exascale Computing with several people involved with exascale efforts at Oak Ridge National Laboratory (ORNL). Last year, the lab announced it was building the Frontier supercomputer.Â In the conversation, Bronson Messer, Distinguished Scientist and Director of Science for the Oak Ridge Leadership Computing Facility (OLCF) at ORNL, noted the importance of such systems used at the lab in the OLCF. â€œThe great thing about supercomputing is thereâ€™s no one killer app. Oak Ridge Leadership Computing Facility. ___ Spallation Neutron Source. High Flux Isotope Reactor. Manufacturing Demonstration Facility. ___ Center for Nanophase Materials Sciences. Center for Structural Molecular Biology. National Transportation Research Center. Science and Discovery. ___ Biology and Environment.Â Predicting biofuel performance. The Co-Optima research team focused on identifying factors that influence how a fuel will perform in a spark ignition engine. Read More. The Big Impact.Â We apply expertise in advanced materials, supercomputing, neutrons, and nuclear science to national priorities in energy, security, and scientific discovery. Learn more. DOE/Oak Ridge National Laboratory. SummaryÂ By relying on its 299,008 CPU cores to guide simulations and allowing its new NVIDIA GPUs to do the heavy lifting, Titan will enable researchers to run scientific calculations with greater speed and increased fidelity. "The order of magnitude performance increase of Titan over Jaguar will allow U.S. scientists and industry to address problems they could only dream of tackling before," said Buddy Bland, Titan project manager at DOE's Oak Ridge Leadership Computing Facility.Â The Oak Ridge Leadership Computing Facility supports national science priorities through deployment and operation of advanced supercomputers as part of DOE's commitment to providing scientists with world-leading research tools. MSR development efforts in industry, national laboratories, and universities are growing, due to the urgent need for safe, clean sources of power generation. This research is encouraged by the successful MSRE experiment at Oak Ridge National Laboratory (ORNL) in the 1960s. This section includes the following items: Introduction. Oak Ridge National Laboratoryâ€”David Holcomb and Kun Chen. CRADA. Massachusetts Institute of Technologyâ€” Charles Forsberg.