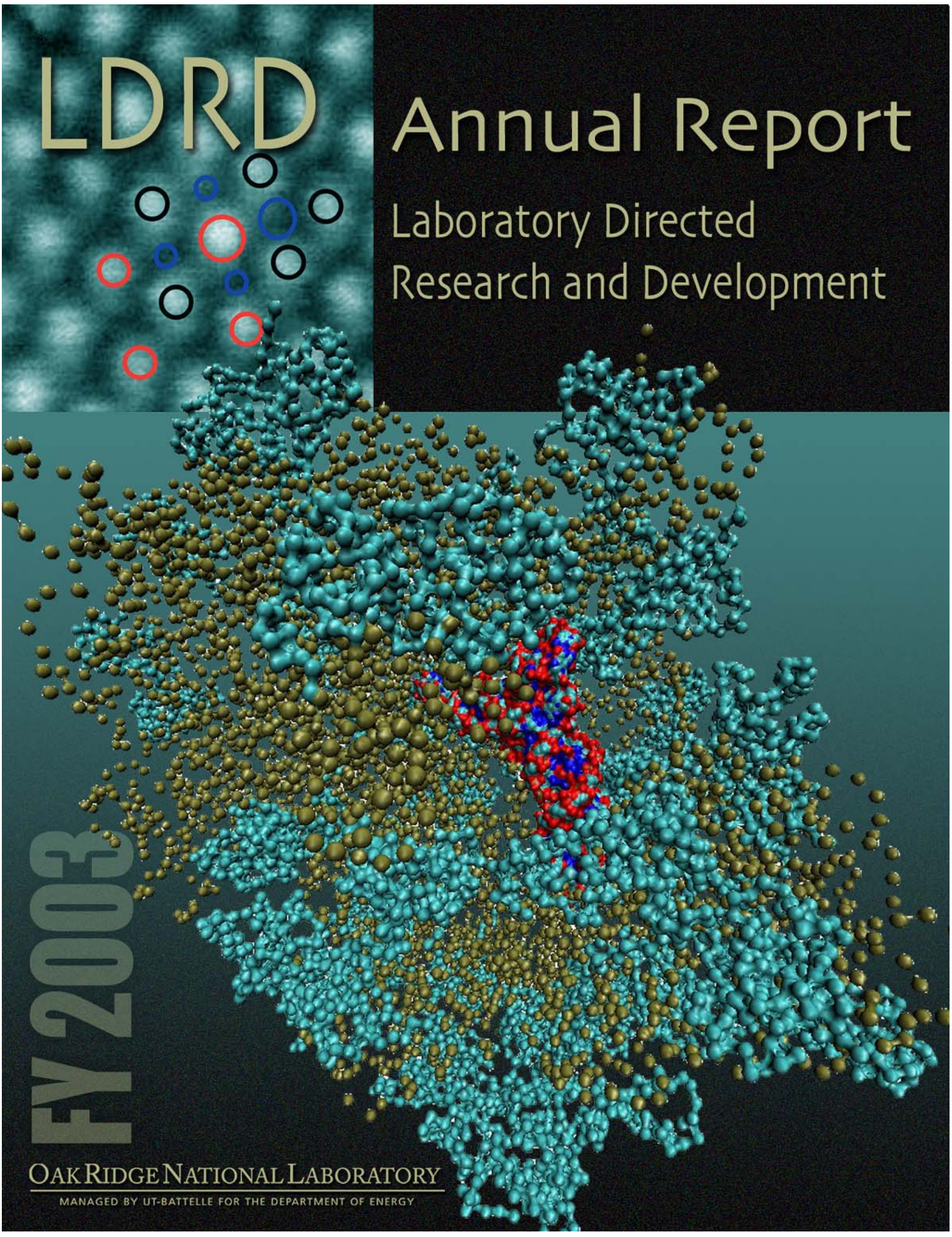
LDRD Annual Report

Laboratory Directed Research and Development



FY 2003

OAK RIDGE NATIONAL LABORATORY

MANAGED BY UT-BATTELLE FOR THE DEPARTMENT OF ENERGY

ORNL/PPA-2004/1

Oak Ridge National Laboratory

LABORATORY DIRECTED RESEARCH AND DEVELOPMENT PROGRAM

FY 2003 ANNUAL REPORT

March 2004

Prepared by
OAK RIDGE NATIONAL LABORATORY
P.O. Box 2008
Oak Ridge, Tennessee 37831-6285
managed by
UT-BATTELLE LLC
for the
U.S. DEPARTMENT OF ENERGY
under Contract DE-AC05-00OR22725

CONTENTS

INTRODUCTION

The Oak Ridge National Laboratory (ORNL) Laboratory Directed Research and Development (LDRD) Program reports its status to the U.S. Department of Energy (DOE) in March of each year. The program operates under the authority of DOE Order 413.2A, "Laboratory Directed Research and Development" (January 8, 2001), which establishes DOE's requirements for the program while providing the Laboratory Director broad flexibility for program implementation. LDRD funds are obtained through a charge to all Laboratory programs.

This report describes all ORNL LDRD research activities supported during fiscal year (FY) 2003 and includes final reports for projects completed during the year and shorter progress reports for projects that were active, but not completed, during this period. The *FY 2003 ORNL LDRD Self-Assessment* (ORNL/PPA-2004/3) provides financial data about the FY 2003 projects and an internal evaluation of the program's management process. It also contains information about new projects started during the first five months of FY 2004.

ORNL is a DOE multiprogram science, technology, and energy laboratory with distinctive capabilities in materials science and engineering, neutron science and technology, energy production and end-use technologies, biological and environmental science, and scientific computing. With these capabilities ORNL conducts basic and applied research and development (R&D) to support DOE's overarching national security mission, which encompasses science, energy resources, environmental quality, and national nuclear security. As a national resource, the Laboratory also applies its capabilities and skills to the specific needs of other federal agencies and customers through the DOE Work For Others (WFO) program. Information about the Laboratory and its programs is available on the Internet at <http://www.ornl.gov/>.

LDRD is a relatively small but vital DOE program at the Laboratory that allows ORNL, as well as other multiprogram DOE laboratories, to select a limited number of R&D projects for the purpose of

- maintaining the scientific and technical vitality of the Laboratory,
- enhancing the Laboratory's ability to address future DOE missions,
- fostering creativity and stimulating exploration of forefront science and technology,
- serving as a proving ground for new research, and
- supporting high-risk, potentially high-value R&D.

Through LDRD the Laboratory is able to improve its distinctive capabilities and enhance its ability to conduct cutting-edge R&D for its DOE and WFO sponsors.

To meet the LDRD objectives and fulfill the particular needs of the Laboratory, ORNL has established a program with two components: the Director's R&D Fund and the Seed Money Fund. As outlined in Table 1, these two funds are complementary. The Director's R&D Fund develops new capabilities in support of the Laboratory initiatives, while the Seed Money Fund is open to all innovative ideas that have the potential for enhancing the Laboratory's core scientific and technical competencies. Provision for multiple routes of access to ORNL LDRD funds maximizes the likelihood that novel and seminal ideas with scientific and technological merit will be recognized and supported.

The ORNL LDRD program recognizes that the R&D staff generally initiate new ideas, that peer review of proposals is an essential component, and that management has a role in fostering projects that are aligned with the strategic directions of the Laboratory. All proposals to the ORNL LDRD program undergo a rigorous review including peer review and must be approved by the Deputy Director for Science and Technology and concurred upon by DOE before funding can be provided.

Table 1. ORNL LDRD Program

	Seed Money Fund	Director's R&D Fund
Purpose	Supports core competencies	Supports laboratory initiatives
Year established	1974	1983
Funding cycle	Continuous	Annual
Proposal review	Proposal review committee	Initiative review committees
Project budget	≤$125,000	≤$800,000
Project duration	12 to 18 months	24 to 36 months
LDRD outlay	22%	78%

Director's R&D Fund

The Director's R&D Fund is the strategic component of the ORNL LDRD program and the key tool for addressing the R&D needs of the Laboratory initiatives. The initiatives, which are the focus of the strategic plan developed by ORNL in collaboration with DOE, are the critical areas that the Laboratory must concentrate on if it is to be prepared to meet future national requirements for science and technology.

The review and selection of Director's R&D Fund proposals is an annual process. In April, a call for proposals is issued to the research staff by the Deputy Director for

Science and Technology. In the call, ORNL's senior R&D managers identify specific research thrust areas that are critical to accomplishing the Laboratory's major initiatives. A committee is established for each thrust area to review the proposals and ongoing projects associated with that area.

The Director's R&D Fund has two rounds of review. In the first round, preliminary proposals are submitted to the appropriate review committee, which evaluates the ideas and selects a subset for development into full proposals. In the second round, each full proposal is peer reviewed for technical merit by two subject matter experts and presented to the review committee during the oral presentation sessions. Ongoing projects for which second- or third-year funding is being requested are also reviewed at this time. Following the presentations, each committee develops a funding recommendation for the suite of new proposals and ongoing projects in its charge. The Deputy Director for Science and Technology then reviews the committee recommendations, develops an overall funding strategy, and presents it for approval to the Leadership Team, ORNL's executive committee headed by the Laboratory Director. Based on the Deputy Director's recommendation, about 10% of the fund's annual allocation is usually held in reserve to support research projects of new R&D staff members who are being recruited to address strategic Laboratory needs.

In FY 2003, $12.5 million was allocated to the Director's R&D Fund to support 54 projects, 21 of which were new starts (see Table 2). Projects may continue for a maximum of 3 years, although most have a duration of 2 years.

Table 2. FY 2002 ORNL LDRD Breakdown by Fund

	Seed Money Fund	Director's R&D Fund
Costs	$3.56 million	$12.5 million
Number of projects	69	54
Number of new starts	40	21
Number of continuing projects	29	33
Average total project budget	~$89,000	~$488,000
Average project duration	15 months	25 months

Director's R&D Fund Research Thrust Areas

In requesting proposals for FY 2003, the Laboratory's senior R&D managers developed thrust areas to support the research needs of the Laboratory initiatives in advanced materials, complex biological systems, energy and environmental systems of the future, national security, neutron sciences, and terascale computing and simulation science. These thrust areas are briefly described below. Detailed descriptions of the Laboratory initiatives and the LDRD research thrust areas are provided in the *ORNL Institutional Plan FY 2003–FY 2007* (ORNL/PPA-2002/2). The levels of investment in the thrust areas are summarized in Fig. 1; an index of projects by initiative is provided on p. 433.

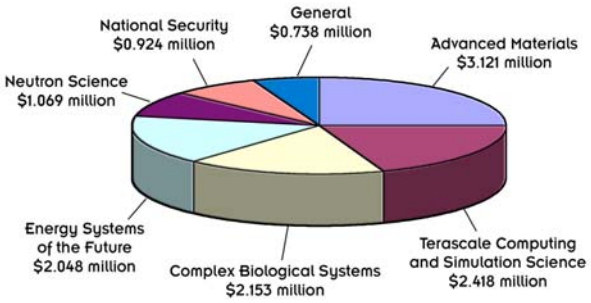


Fig. 1. Level of Director's R&D Fund investment in the Laboratory-wide initiatives for FY 2003.

Advanced Materials

The focus of the advanced materials research thrust is nanoscale science, engineering, and technology, an area that is expected to provide breakthrough opportunities in many fields central to DOE's missions, including materials science, biotechnology, processing science, energy, environmental technology, information technology, and national security. A high priority was given to the development of capabilities for the controlled synthesis of nanomaterials and nanostructures, resulting in functionality for nanoscale systems.

Complex Biological Systems

The research thrust in complex biological systems builds on ORNL's expertise in the life and environmental sciences, chemical and analytical sciences, instrumentation sciences, and computational sciences. In FY 2003, this thrust aimed at the development of new molecular and computational capabilities and approaches in response to DOE's proposed Genomes to Life Program and genomics tools and computational strategies to further our understanding of the natural and pathological functioning of cells from microbes to higher organisms (humans, mouse, fish, plants, etc.).

Energy and Environmental Systems of the Future

This research thrust area seeks to build new R&D capabilities that promote the development of new energy services that are environmentally responsible and economically competitive. It is focused on meeting the increasing demand for energy throughout the world while

recognizing that a very large proportion of today's energy needs are met through resources and technology systems that are nonsustainable. Thus, ways to increase the availability of energy, safely and affordably, while shifting to significantly different energy paths with minimal effects on human health and the environment will be required. Making this shift will require a thorough understanding of a broad range of science, technology, economic, and policy issues that will drive the development of future energy systems.

National Security

As a DOE multiprogram laboratory, ORNL possesses critical capabilities that can provide federal and state authorities with the technologies and expertise required to protect the public. The objective of this thrust was to develop leading-edge capabilities to increase national security. In FY 2003 this thrust addressed (1) sensors and detectors for chemical, biological, and nuclear threats; (2) computer and computational science for homeland security; and (3) transportation security.

Neutron Sciences

The Spallation Neutron Source (SNS) and the upgrades at the High Flux Isotope Reactor (HFIR) that will be completed as part of the Laboratory initiative in neutron sciences offer a unique opportunity for ORNL and the nation to demonstrate world leadership in this area. The success of these projects (and their impact on ORNL) depends critically on the ability to develop the full potential of these new scientific tools and to integrate neutron sciences into research programs across the Laboratory. The intent of this research thrust area is to stimulate new research directions, proof-of-principle experiments, and technical innovations essential to achieving leadership in neutron science. This thrust area focuses on three research themes that are essential to leadership in neutron sciences: novel applications of neutron scattering, neutron physics, and novel instrumentation concepts.

Terascale Computing and Simulation Science

This research thrust recognizes that new computational tools and codes are needed to enable the creation of realistic simulations of physical situations and to provide new insights into a host of scientific problems. Research, development, and deployment of mathematical models, computational methods, numerical libraries, and scientific codes are needed to take full advantage of the capabilities of ORNL's terascale computers for solving strategic or critical problems in materials sciences, chemistry, combustion, accelerator design, high-energy physics, nuclear physics, fusion, biology, and global systems.

Seed Money Fund

The Seed Money Fund complements the Director's R&D Fund by providing a source of funds for innovative ideas that have the potential of enhancing the Laboratory's core scientific and technical competencies. It also provides a path for funding new approaches that fall within the distinctive capabilities of ORNL but outside the more focused research thrusts of the major Laboratory initiatives. Successful Seed Money Fund projects are expected to generate new DOE programmatic or Work-for-Others sponsorship at the Laboratory.

Proposals for Seed Money Fund support are accepted directly from the Laboratory's scientific and technical staff (with management concurrence) at any time of the year. Those requesting more than $20,000 ($125,000 is the maximum) are reviewed by the Proposal Review Committee (PRC), which is comprised of 11 scientific and technical staff members representing each of the Laboratory's research divisions and the LDRD manager (chair). To assist the Committee, each proposal is also peer reviewed by two Laboratory staff members selected by the LDRD manager. Proposals requesting $20,000 or less are not reviewed by the PRC but are peer reviewed by a research staff member selected by the LDRD manager. All Seed Money Fund proposals receiving a favorable recommendation are forwarded to the Deputy Director for Science and Technology for approval and require DOE concurrence.

In FY 2003, $3.6 million of the LDRD program was apportioned to the Seed Money Fund to support 69 projects of which 40 were new projects (see Table 2). The distribution of Seed Money Fund support by science and technology area is shown in Fig. 2. The assignment of projects to specific areas is not meant to be definitive as many projects are crosscutting and could be assigned to more than one category.

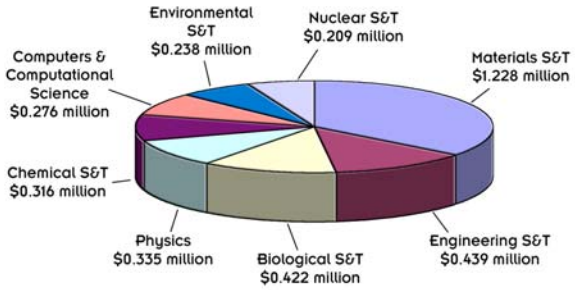


Fig. 2. Distribution of Seed Money by science and technology area for FY 2003.

Report Organization

The report is divided into eight broad areas of science and technology that underlie the programmatic and Work for Others activities at ORNL. Within each of these areas, the projects are grouped by fund (i.e., either the Director's R&D Fund or the Seed Money Fund). Within each fund grouping are the final reports and progress reports in that order. A list of Director's R&D Fund projects by initiative an author list, and a project list are included at the back of this document.

MATERIALS SCIENCE AND ENGINEERING

Director's R&D Fund

Creating Oxygen-Rich Nanoclusters for High-Temperature Strengthening of Structural Alloys

D. T. Hoelzer,[1] P. J. Maziasz,[1] S. S. Babu,[1] M. K. Miller,[1] C-L. Fu,[1] E. A. Kenik,[1]
I. G. Wright,[1] E. D. Specht,[1] and J. L. Robertson[2]
[1]*Metals and Ceramics Division*
[2]*Condensed Matter Sciences Division*

The goal of this project was to develop the understanding of a new phenomenon involving the formation of a high-number density of uniformly dispersed nanoscale oxygen-rich clusters in an iron-based alloy. The technological impact of forming the nanoscale oxygen-rich clusters is that the high-temperature strength and creep resistance of the iron alloy is significantly enhanced. The scientific impact challenges our current understanding of the stability of oxide phases in metallic systems. An experimental and theory/modeling effort was developed to investigate the evolution from the meso-scale oxide structure to the nanoscale clusters and to expand the current scientific knowledge of non-equilibrium thermodynamics, phase formation and stability, and nanoscale strengthening mechanisms at high temperatures. This understanding may also lead to new alloy design and alternative processing methods to produce a range of advanced structural alloys with superior high-temperature performance.

Introduction

New structural materials possessing significantly improved high-temperature, high-strength properties are required to meet future goals for higher thermal efficiencies in energy generation systems. One approach for achieving these requirements is to disperse oxide particles in the alloy matrix, which improves the high-temperature creep resistance by acting as hard obstacles that hinder the motion of dislocations and grain boundaries. Historically, the mechanical alloying (MA) approach has been used to disperse oxide particles in the matrix of many technologically important alloy systems, such as aluminum, iron, and nickel, to increase the high-temperature mechanical strength properties. These advanced materials are usually referred to as oxide dispersion strengthened (ODS) alloys.

Mechanical alloying has been used for more than 30 years to make commercial ODS alloys and for basic research studies.[1] This process involves ball milling metallic or pre-alloyed powders with the dispersoid powder, such as Y_2O_3, and then consolidating the powders into a bulk shape. In the past, the general understanding of MA is that it essentially "mixes" the alloy and dispersoid powders by mechanical means. Very few studies focused on the structural and chemical evolution of the dispersoid during ball milling. However, in the late 1990s, a new phenomenon involving a high-number density of nanosize clusters containing Y, Ti, and O atoms was discovered in a mechanically alloyed (MA) Fe-12Cr-3W-0.4Ti-$0.25Y_2O_3$ (wt %) iron alloy, referred to as 12YWT, that showed a remarkable improvement in high-temperature, high-strength properties.[2] This discovery has stimulated the importance of understanding the structural and chemical evolution of the initial oxide dispersoid to nanoclusters.

We believe that a study must comprise both experiment and theory to develop the understanding of nanosize clusters formation. Our primary focus was to investigate the structure and stability of the nanoclusters at high temperatures and to develop the understanding of how the stable yttria-oxide evolves into new nanoclusters in a model iron alloy during mechanical alloying (MA). This research project expanded the current scientific knowledge of non-equilibrium thermodynamics, phase formation and stability, and nanoscale strengthening mechanisms at high temperatures. We also foresee developing the knowledge for improving the MA processing method and for developing non-MA processing methods for achieving high-number densities of uniformly dispersed nanoclusters.

Technical Approach

Stability and Structure of Nanoclusters

Determining the salient properties of the nanoclusters was paramount to understanding the mechanisms for their formation. The approach was to investigate the nanoclusters, initially discovered in the 12YWT alloy, using advanced analytical techniques to obtain information regarding their unique structure, such as the crystal

structure, composition, and interfacial structure. These studies mainly included the atom probe tomography (APT) and advanced electron microscopy techniques but also included several attempts using different X-ray scattering techniques. The initial discovery of the nanoclusters was made on the as-processed 12YWT alloy, which was reportedly exposed to processing temperatures up to 1150ºC. This result suggested that the nanoclusters were stable at high temperature. In order to more clearly define the high-temperature stability of the nanoclusters, heat treatments were conducted on the 12YWT at temperatures as high as 1300ºC, which is higher than 85% of the melting point of this alloy.

Formation of Nanoclusters

Our experimental approach to forming nanoclusters in an iron-based alloy was largely based on information available regarding the composition and processing history that went into developing the 12YWT ferritic alloy. This information included the composition, ball milling conditions, and consolidation temperature, which was stated to be 1150ºC. High-quality, gas-atomized pre-alloyed Fe-14%Cr (wt %) powders containing systematic additions of 3%W, 0.4%Ti, and 3%W+0.4%Ti were procured for ball milling and powder consolidation studies. Four different alloy compositions were ball milled with the same concentration of Y_2O_3 used in 12YWT, which was 0.25%Y_2O_3. A fundamental study was first conducted on the ball milled powders using X-ray diffraction (XRD) and transmission electron microscopy (TEM) to study the effects of ball milling parameters and alloy composition on the structural evolution of Y_2O_3. The ball milled powders were then consolidated by hot extrusion at two temperatures; the first extrusion was conducted close to the reported consolidation temperature of the 12YWT alloy (1175ºC), and the second extrusion was conducted at a considerably lower temperature of 850ºC. Issues that were being explored centered mainly of the long-range diffusion of Y, Ti, and O in the iron matrix.

Theoretical Study

First-principles local-density-functional (LDF) theory was used to develop an understanding of conditions that favored nanocluster formation by focusing on the relationship among energetics, electronic structure, and bonding geometry of oxygen. Data collected from the APT studies provided significant information concerning the possible role of oxygen supersaturation as a mechanism for nanocluster formation. The goal of the theory was to answer why the unique combination of Y, Ti, and O is required for cluster formation. The theory examined the conditions under which the oxygen atoms are soluble in the Fe matrix in the presence of dissolved Ti and/or Y atoms without forming many stable oxide compounds. The model was successfully developed to also include the additional role of vacancies on the solubility of O in Fe.

Results and Accomplishments

Stability and Structure of Nanoclusters

Improving the maximum separation envelope (MSE) code was an important development in this project for analyzing the APT data on the nanoclusters.[3] This code is used for determining the atoms associated with a precipitate based on the principle that solute atoms in a precipitate are closer together than those in solid solution in the surrounding matrix because of differences in chemical bond strength. This permitted a more precise analysis of their size, number density, and composition. The composition of nanoclusters was consistently found to have a stoichiometric ratio of Ti and Y to O atoms close to 1:1 and to contain ~10–15 at % Fe atoms. However, the Ti, Y, and O atoms are not uniformly distributed in the nanoclusters.The radial concentration profile from the center of mass of the nanoclusters showed that their core is populated with Ti and O atoms, while the interface is enriched with Y atoms. This analysis also showed a diffuse interface, which could account for the apparent stability of the nanoclusters. Advanced electron microscopy techniques confirmed that the nanoclusters do not possess a unique crystal structure but that the atoms associated with them are clustered on the bcc lattice sites in the Fe matrix. However, the composition analysis of the matrix revealed an unusually high oxygen concentration that was more than 200 times the equilibrium solubility limit in bcc Fe at 800ºC.[4] These results led to questions about the structure and formation of nanoclusters that were addressed by theory and modeling.

Experiments consisting of creep testing for long times with an applied stress at temperatures up to 850ºC and for short-time exposures at temperatures up to 1300ºC revealed a remarkable degree of stability for the nanoclusters. In one significant creep test, the sample of 12YWT failed after ~14,500 h (~1.6 year) at 800°C and applied load of 138 MPa. This level of performance is several orders of magnitude better than that of any previously developed MA ODS Fe-Cr alloy. Surprisingly, as shown in Fig. 1, the APT analysis revealed essentially no change in the size, number density, and composition of the nanoclusters compared to the as-processed sample. Even exposures of 12YWT to extremely high temperatures of 1300ºC, which is ~86% of the melting point for this alloy, for up to 10 h did not significantly change the composition or size of the nanoclusters and only slightly changed their number density. Figure 2 shows a compilation of the data obtained for all of the nanoclusters analyzed in 12YWT after exposure to high-temperature environments.

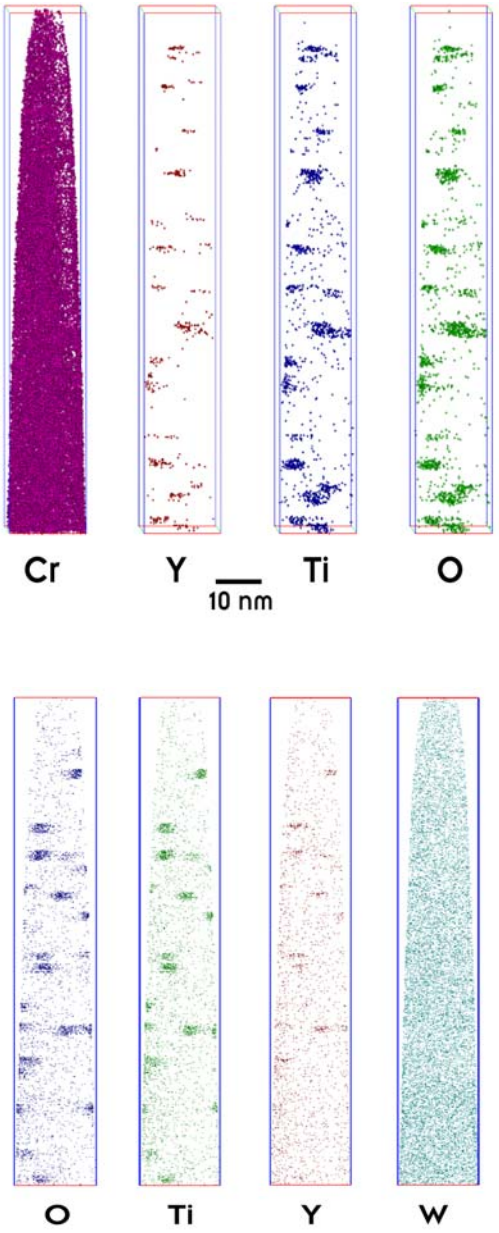


Figure 1. Three-dimensional atom probe tomography images of nanoclusters in 12YWT in (a) as-processed conditions and (b) after annealing at 1300ºC for 1hour.

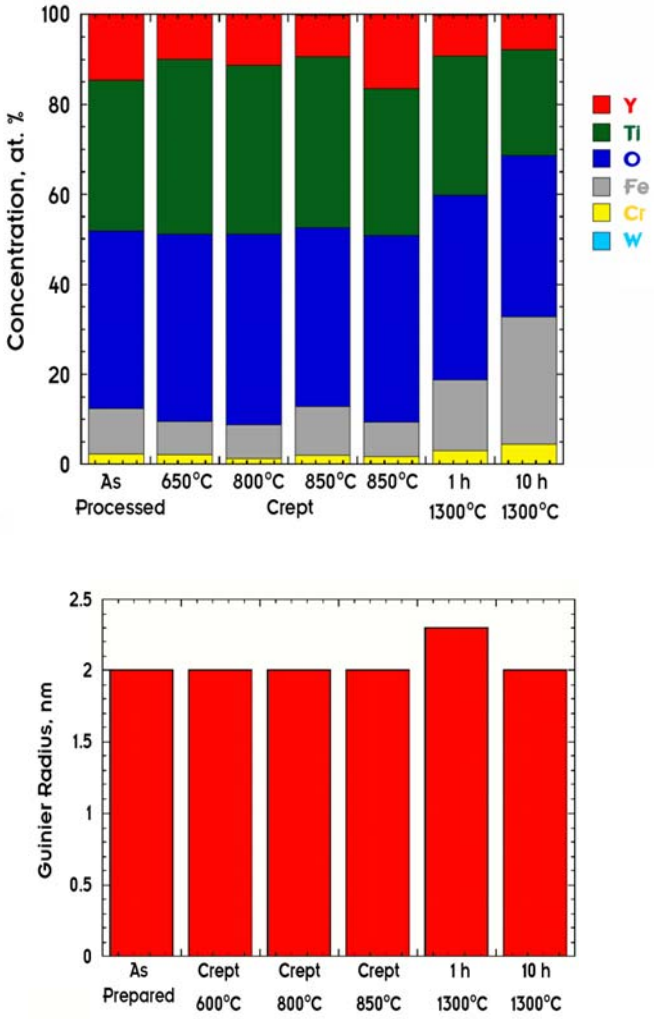


Figure 2. Atom probe tomography data showing essentially no changes in the (a) composition and (b) size of the nanoclusters in 12YWT upon exposure in long-time creep tests up to 850ºC and short-time high-temperature exposure at 1300ºC.

Results obtained in this study regarding the structure and stability of nanoclusters challenge our understanding of some key areas in materials science. The most significant contribution to the materials science community is the result showing the amazing stability of the nanoclusters at extremely high temperatures of 1300ºC for up to 10 h. The results suggest that the nanoclusters are a metastable phase occupying the bcc lattice sites in the Fe matrix with a size range of less than ~4-nm diameter. According to thermodynamic laws, these characteristics should result in rapid coarsening of size and in the transformation to an oxide compound at these high temperatures. Thus, this study has revealed a new concept of stable nanoclusters as a method for high-temperature strengthening of advanced structural materials.

Formation of Nanoclusters

Our fundamental study on ball milled powders of the model alloys demonstrated that the Y_2O_3 particles are reduced from an initial size of 17–31 nm to below ~2 nm and are forced into a nonequilibrium suspension in the bcc Fe matrix. The lower size limit was defined by the spatial resolution of the electron microscopy techniques that were employed. The composition of the powders had a minor effect on differences in the magnitude of internal lattice strain and final nanosize crystallite induced by ball milling, and had no significant effect on the evolution of the Y_2O_3 particles during ball milling.

A significant breakthrough was achieved in the investigation of consolidated powders. Our experimental results showed that the formation of nanosize particles occurs at low consolidation temperatures rather than at high temperatures. We first attempted to reproduce the extrusion conditions used for making the MA 12YWT alloy. However, extruding the four model alloys at 1175ºC resulted in the formation of a low number density of coarse oxide particles. The presence of Ti in two of the model alloys resulted in the formation of ternary Ti-, Y-, and O-based oxide particles that were slightly smaller in size and higher in number density compared to those in the two model alloys that did not contain Ti. These results indicated that the diffusion of Y and Ti atoms was too extensive at 1150ºC and higher, which favors large particles and a low nucleation rate. An extrusion temperature of 850ºC was then chosen based on the criteria of lowering the diffusion of Y and Ti atoms in the Fe matrix. This approach was successful in forming nanosize particles in the iron alloy containing both W (tungsten) and Ti, which is referred to as 14YWT. Figure 3 shows composition maps of the distribution of Fe (Fig. 3a) and Ti (Fig. 3b) atoms obtained by energy-filtered transmission electron microscopy (EFTEM). The darkly imaging particles in the Fe map are depleted in Fe and are enriched in Ti, which are brightly imaging particles in the Ti map. The formation of the particles was both inter- and transgranular with sizes as small as the detection limit of ~2 nm. The APT has revealed that the particles are enriched in Y, Ti, and O atoms similar to the nanoclusters discovered in the 12YWT alloy. Contrary to what was previously known about the processing conditions for the 12YWT, the results of this study establish low-temperature consolidation temperatures that minimize diffusion of critical elements such as Ti and Y as being significant for forming nanoclusters in iron-based alloys.

Theoretical Study

A theoretical model of the mechanism for forming Ti-Y-O-Fe-enriched nanoclusters in iron was developed in this study. The model is consistent with experimental results and describes the balance between the increase in strain energy and the gain in electronic bonding energy with Y, Ti, and O atoms in the bcc-Fe lattice. Critical factors for nucleating the nanoclusters have been identified from the model. It is believed that validation of the factors in a carefully designed experiment can form the basis of a patent for developing nanoclusters in iron-based as well as other technically important alloy systems.

Summary and Conclusions

In summary, we determined that a very high number density of "nanoclusters" that formed in the mechanically

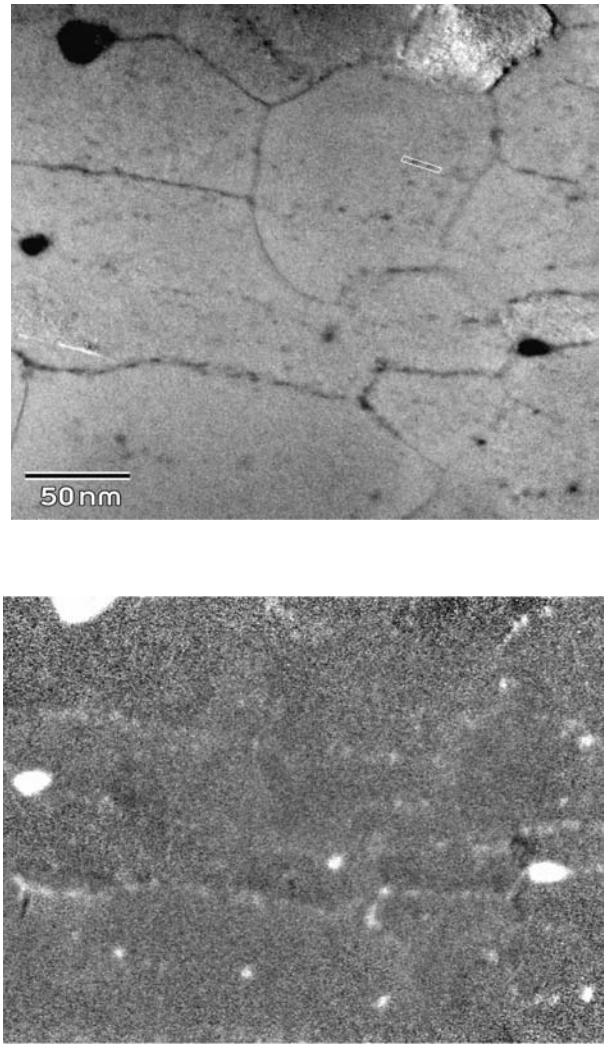


Figure 3. Energy filtered TEM images showing nanoclusters in the 14YWT alloy that was extruded at 850ºC. The nanoclusters are dark in the Fe jump ratio map (a) and bright in the Ti jump ratio map (b). The intensity histogram of two nanocluster particles (line in 3a) indicates that they are 3 nm in diameter at full-width, half-maximum (c).

alloyed 12YWT steel is stable in size and composition at extremely high temperatures of 1300ºC. Results were obtained showing the unique structure and composition of the nanoclusters. Based on the physical characteristics of the nanoclusters, their remarkable stability in iron cannotbe explained by current theories in materials science. We also achieved the objective of integrating theory and modeling with experiment to develop the understanding of nanocluster formation in iron alloys. A model based on forst-principles theory was developed showing that the formaiton of nanoclusters requires enhanced oxygen solubility in iron and that this can be achieved by Ti and Y additions and on vacancies in iron. Ball milling of the pre-alloyed iron powders produces favorable conditions for enhancing the oxygen solubility and reduces the particle size of Y_2O_3, forcing it into a nonequilibrium suspension in the iron matrix. Finally, the nanoclusters form by precipitating from the iron matrix during consolidation of the ball milled powders. However, it was discovered that a low consolidation temperature, where long-range diffusion of Ti and Y atoms is negligible, resulted in the formation of nanoclusters. The results of this investigation provide a new approach to dramatically improving the high-temperature strength of iron-based alloys by dispersion of a high-number density of nanoclusters.

The results of this project showing the improved creep strength of the 12YWT and the stability of the nanoclusters at high temperature led to a collaboration with research scientists from the French Commissariat al' Energie Atomique (CEA) that resulted in a successful proposal to the International Nuclear Energy Research Initiative (I-NERI), Office of Nuclear Energy for $1.2M over three years. This project will study the deformation and fracture modes, corrosion behavior, and stability of oxide- and nanoclusters-dispersion strengthened steels under long-term high-temperature aging and understand the effects of displacement damage on the stability of nanoclusters and oxide particles using multi-MeV heavy ions.

There is growing interest in the nanocluster dispersion strengthened steels for development of advanced nuclear technologies embodied in the overall concept of Generation IV reactors and in several DOE Fossil Energy Vision 21 programs for extreme high-temperature materials for use in ultra-super-critical steam power plants.

Our effort on this project has recently stimulated a Basic Energy Science (BES) proposal entitled, "Self-Assembly of Stable Nanoclusters in Metallic Matrices," which is being led by C. T. Liu in ORNL's Metals and Ceramics Division. This proposal, which requests $3M for three years, focuses on fundamental issues of the formation mechanism and thermal stability of stable nanoclusters, including atomic-scale characterization using advanced analytical techniques; theoretical modeling by ab initio calculations and kinetic Monte Carlo simulations; and experimental studies of nanocluster formation and thermal stability applying physical metallurgy analysis, alloying designs, etc. The proposed study would also investigate hardening mechanism at ambient and elevated temperatures including atomic-scale resolution of dislocation line curvature.

References

[1]C. Suryanarayana, *Progress in Materials Science* **46**, 1 (2001).

[2]D.J. Larson, P.J. Maziasz, I-S. Kim, and K. Miyahara, *Scripta Mater.* **44**, 359 (2001).

[3]M.K. Miller, in *Atom Probe Tomography: Analysis at the Atomic Level*, Kluwer Academic Publishers Group, 2000.

[4]Swisher and Turkdogan, *Trans. Met. Soc. AIME* **239**, 426 (1967).

Combined Neutron and X-Ray Diffraction

J. W. L. Pang
Metals and Ceramics Division

The goal of this project was to demonstrate how neutron and synchrotron X-ray microbeam diffraction can be combined for the investigation of deformation mechanisms over different length scales. A series of in situ and ex situ experiments on pure Ni samples have been carried out using both techniques. Experiments designed to investigate the dependence of deformation on grain size highlighted the distinctive strengths of various X-ray and neutron techniques and illustrated how they can be applied to the study of deformation. Inhomogeneous developments of strain and rotational behaviors are found on both the micro and meso scale. Differences in the observed behavior between samples provides direct experimental evidence of grain-size and grain-boundary effects on deformation and the need to use neutron and X-ray diffraction to advance our understanding of deformation mechanisms.

Introduction

A sophisticated understanding of plastic deformation is required to predict materials behavior in service and to accelerate the development of novel new materials. For example, grain size, preferred orientation, residual stress and a host of other properties result from inhomogeneous plastic deformation. Theoretical progress in deformation modeling has been rapidly moving forward over the last two decades. Deformation is length-scale sensitive, and what is needed for comparison with theoretical models is a way to study deformation of polycrystalline materials that focuses on the three-dimensional (3-D) microscopic and mesoscopic carriers of deformation, namely, grain conglomerates, grains, subgrains, dislocation cells, and second-phase particles. A synergistic combination of techniques that cover the appropriate length scales will have a significant impact on our basic understanding of deformation and will guide the development of computational intensive models. This project focused on the exploitation of novel neutron and X-ray diffraction techniques to provide unprecedented experimental information to guide and test advanced models.

Technical Approach

There are numerous microscopic variables which affect deformation behavior. In our study, we examined the grain-size dependence of deformation. Our experimental approach was based on linking materials physics over the very different numbers of grains, and the radically different length scales that can be probed by the various 3-D techniques. Neutron penetration depths, for most elements, are on the order of centimeters, and volume-averaged internal strains in the bulk can be measured non-destructively. The sensitivity of neutron diffraction to microstrain (intergranular strain) and the application of neutron diffraction to study deformation was proven in the last decade.[1–2]

The recent development of polychromatic X-ray microdiffraction with submicron spatial resolution has revolutionized the measurement of mesoscopic structure. Individual grains, their substructures, and grain boundaries can all be studied nondestructively and in three dimensions based on white beam Laue technique.[3–4] Detailed boundary conditions and variation of deformation within a grain and across grain boundaries can be determined, which was not possible before.

Our strategy is to use neutron powder measurements to understand the statistical behavior of bulk polycrystalline grains. Based on the guidance from neutron powder measurements, X-ray microbeam Laue measurements are applied to study specific deformation features with subgrain resolution. Our experimental efforts demonstrated how these techniques can be applied to study length-scale-sensitive materials problems.

The material chosen was 99.97% pure nickel (Ni-270); this material has much reduced effects of impurity and precipitate particles compared to most alloys. The samples were fully dense, and porosity was not an issue. Two sets of samples of grain sizes 20 μm and 200 μm were prepared. A coupon from a Ni-270 extruded rod was first subjected to 5% reduction by cross rolling. The 20-μm grain size was produced by annealing at 550°C for 2 hours; 200-μm grain size was achieved by an additional annealing at 1200°C for 2 hours. Uniaxial tests were performed on both sets of samples and the evolution of microstrain, and texture was measured at various levels of plastic deformation.

The first measurements were on the 99.97% pure Ni tensile sample with 200-μm grains. The sample was

mechanically polished followed by electropolishing. Microbeam Laue measurements were carried out on the beamline ID-34E at the Advanced Photon Source, Argonne National Laboratory. Figure1(a) illustrates the experimental setup. The sample was probed by a polychromatic X-ray beam with an energy range of ~8–25 keV. Depth resolution was achieved using the differential aperture microscopy method developed by Larson et al.[4] The volumetric resolution of the measurements is 0.5(W) × 0.5 (H) × 1.0 (D) μm^3, and Laue images were collected down to 30 μm beneath the surface. After the near-surface grain pattern was measured, a 3-D grid of 5 × 5 × 30 volume elements was collected.

Measurements focused on three grains joined by a common triple junction. The local orientations along 25 lines through the sample were probed. The 5 × 5 transverse grid of the measurements are indicated in Fig. 1(b). Local orientations along the lines into the sample were probed at various plastic strains. After each deformation step, the positions of the measurements were referenced back to the triple junction, which could be located to within ±1 μm. Because the measurements resolve volumes within neighboring grains, effects of sample heterogeneity can be correlated to local deformation. Orientations are determined with a precision of <0.02° at 0% strain, comparing to ~1° for orientation imaging microscopy. As strain increases, dislocation density increases and the angular resolution drops to ~0.2° at 14% strain. The data acquisition time required for each line scan was about 20 minutes. Experiments that demand submicron resolution such as grain boundary studies and stringent angular resolution measurements can now be performed. The main drawback at present is the relatively slow count rate for depth-resolved measurements; it is not practical to conduct an extensive depth-resolved study on a large cross-sectional area. The strain resolution was determined by performing in situ loading measurements using a monochromatic beam on another Ni tensile sample. The uncertainty of the elastic stress tensor for measurements within the elastic regime was found to be ~1 part in 10^4.

Neutron diffraction measurements were carried out on the spectrometer E3 at the NRU Reactor, Chalk River Laboratories, Canada. The experimental setup is shown in Fig. 2. Uniaxial tension was applied to samples with grain sizes of 20 μm and 200 μm. The accumulated strains and texture changes within the grains over the sampling volume were then measured as a function of (*hkl*) reflections and sample orientations with respect to the tensile axis. Measurements were performed on all the undeformed samples to ensure accurate determination of strain and texture evolution introduced by the applied tension instead of any previous processing procedures. The neutron scattering factor for Ni is large, and fast data acquisition with high precision was achieved.

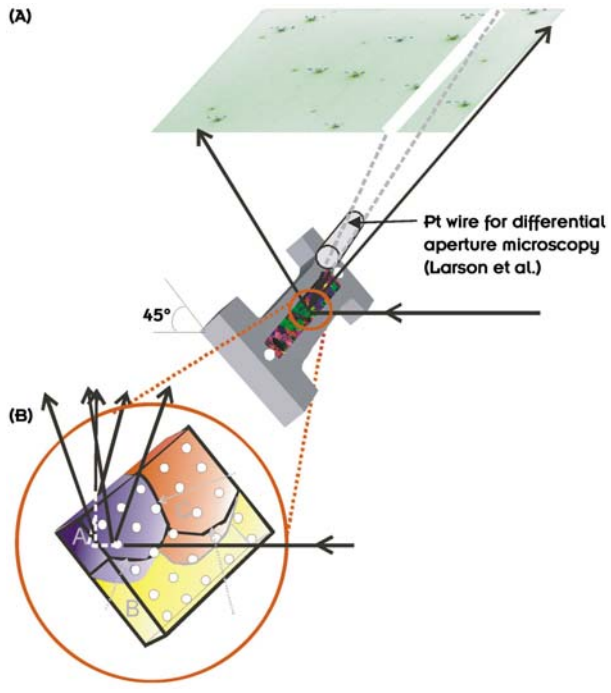


Fig. 1. (a) Schematic of the experimental setup of the measurements on beamline ID-34E; (b) the three grains on which the measurements were focused on.

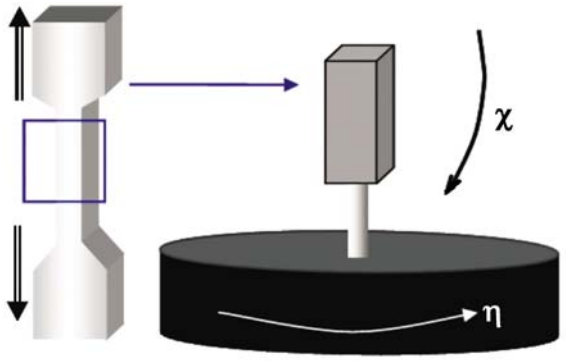


Fig. 2. The tensile sample was put on an Eulerian cradle for the neutron diffraction measurements on E3 at the NRU Reactor. Information on the intergranular strain variations as a function of angular positions were determined by measuring over a quadrant of a hemisphere on a (15°, 15°) grid.

Knowledge of the variations in strains and misorientations at the subgrain level is essential to complement neutron diffraction measurements where an ensemble average is probed. It is important to note however that deformation studies on many bulk polycrystals are difficult or impossible with x-ray microbeams, because

of the relatively short penetration depth (50 μm for Ni). In order to make measurements with submicron resolution, the deformed samples were cut by electron discharge along the tensile axes to produce samples ~200 μm thick. The samples were then mechanically ground and chemically etched to reduce the thickness down to ~50 μm. Selective regions of the samples were then measured by X-ray microbeam, as described above.

Results and Accomplishments

The average flowcurves of the 20-μm and 200-μm grain-sized samples are plotted in Fig. 3. The yield stress of the fine grained samples is 60% greater, and the hardening behavior is almost identical for both samples. Large residual tensile strains were found by neutron diffraction for the (002) reflection along the tensile axis for both grain sizes. Residual strains for other reflections are within 0.03% in all measured sample directions. The significant observation is the difference in strain magnitudes between the two grain sizes. Figure 4 shows the maximum amplitudes of the residual strains for (002) at plastic strain of 0.3, 1, 5, and 15%. Samples with larger grain size consistently show smaller lattice strains. It was found that the lattice strains for the large-grained samples reached to plateau at around 5% plastic strain, whereas lattice strains continue to increase for the 20-μm grain samples.

Texture evolution from 0% to 15% strain was determined, and the difference between the two sets of samples was small. However it is believed that the difference will be more significant when rotational incompatibility among grains increases with deformation. These results show convincingly that, with well-designed experiments, neutron diffraction can be used to study the effects of microstrucutral parameters on deformation and well-characterized samples.

A self-consistent (SC) model[5] has been applied to simulate the deformation process. The input parameters were obtained from the macroscopic flow curves (Fig. 3). The predicted strain magnitudes are compared to the experimental results in Fig. 4. Because the SC model is not length-scale sensitive, the model can only capture the effects introduced by yield-strength differences. The results clearly show that the model overestimates the strain magnitudes and cannot capture the differences observed between the grain sizes.

Residual strain profiles across a twin boundary and a high-angle boundary in the 50-μm-thick slices cut out from the 1% deformed samples were measured by microbeam Laue diffraction and the results are shown in Fig. 5. These measurements clearly indicate variations in strains across grain boundaries. The important observation is the strain discontinuity observed in the high-angle boundaries but not the twin boundaries for both grain sizes. It was also

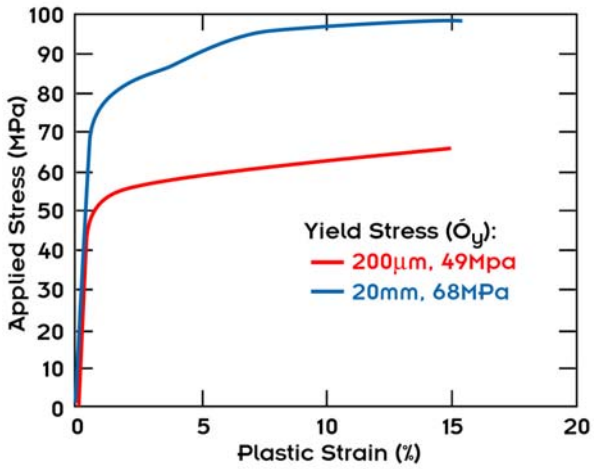


Fig. 3. The flowcurves for the Ni-270 samples of grain sizes 20 μm and 200 μm.

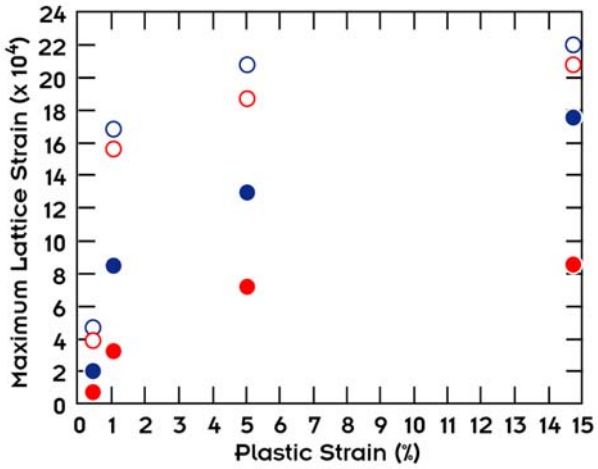


Fig. 4. Maximum lattice strain for reflection (002) determined by neutron diffraction as a function of plastic deformation. Predictions determined by SC model are plotted together with the experimental data.

found that the strain gradients across the boundaries are sharper in the large-grained sample. More measurements will be needed to develop statistically significant results. However the results thus far provide direct submicron evidence to validate the general belief that grain boundary behavior is a function of grain size and is dependent on the grain boundary types. Companion misorientation measurements were made, and similar discontinuities across grain boundaries were found for the high-angle grain boundaries.

These experimental results will be used to test a length-scale-sensitive deformation model which is based on crystal plasticity with finite element discretization.[6] If the assumptions in the model are correct, the model should capture both the microscopic and mesoscopic information from the experiments.

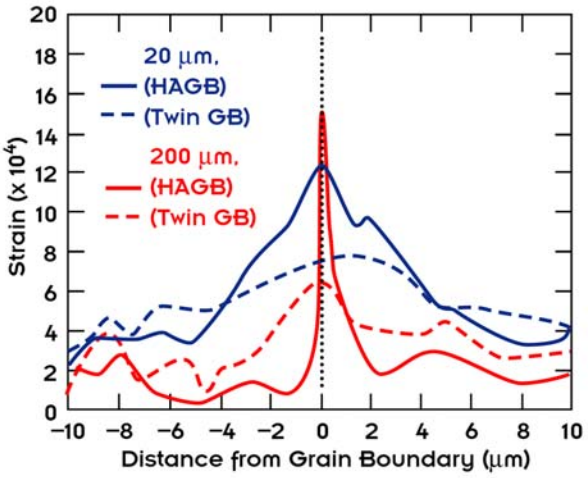


Fig. 5. Lattice strains in the loading direction across grain boundaries for 1% deformed samples. HAGB indicates high-angle grain boundaries.

Summary and Conclusions

This project demonstrates the complementary capabilities of neutron diffraction and X-ray microbeam Laue techniques in the study of deformation mechanisms. A synergistic combination of these techniques can provide experimental information which will make significant contribution to the understanding of deformation mechanisms and ultimately advance our ability to predict materials performance.

For the last two years, there has been a continuing emphasis on the promotion of the use of neutron for materials research in the anticipation of the completion of the Spallation Neutron Source (SNS) in 2006. This project was designed to use neutron scattering, our strong in-house capability in advanced materials characterization, and world-leading expertise in microbeam Laue technique to examine deformation. Another core competency at ORNL is parallel computing. ORNL has an active community in parallel processing and recently was selected to host the next-generation parallel processing machine designed to return computational preeminence to the United States. Computer-intensive modeling efforts are under way. This research and emerging new computational tools place ORNL in a unique position to advance deformation modeling.

This project is relevant to DOE's Basic Energy Science program to which a follow-on proposal has been submitted. It will provide unique insights into the deformation behavior of polycrystalline materials, which is one of the major considerations in the development of new structural materials. The length scale of the results will be of direct relevance to many developing deformation models.

References

[1]T. M. Holden, R. A. Holt, and J. W. L. Pang, "Intergranular stresses in ZIRCALOY-2," *Metallurgical and Materials Transactions a-Physical Metallurgy and Materials Science* **33**(3), 749–755 (2002).

[2]J. W. L. Pang et al., "The generation of intergranular strains in 309H stainless steel under uniaxial loading," *Acta Materialia* **48**(5), 1131–1140 (2000).

[3]R. I. Barabash et al., "Application of white x-ray microbeams for the analysis of dislocation structures," *Review of Scientific Instruments* **73**(3), 1652–1654 (2002).

[4]B. C. Larson et al., "Three-Dimensional X-ray Structural Microscopy with Submicrometre Resolution," *Nature* **415**(6874), 887–890 (2002).

[5]P. A. Turner and C. N. Tome, "A Study of Residual-Stresses in Zircaloy-2 with Rod Texture," *Acta Metallurgica Et Materialia* **42**(12), 4143–4153 (1994).

[6]G. B. Sarma and P. R. Dawson, "Effects of interactions among crystals on the inhomogeneous deformations of polycrystals," *Acta Materialia* **44**(5), 1937–1953 (1996).

Electrical Conductivity at the Nanoscale

A. P. Baddorf,[1] C. L. Fu,[2] S. V. Kalinin,[1] J. Shen,[1] H. H. Weitering,[1,3] J. F. Wendelken,[1] and X. Zhang[4]
[1]*Condensed Matter Sciences Division*
[2]*Metals and Ceramics Division*
[3]*The University of Tennessee*
[4]*Computer Science and Mathematics Division*

This research focused on measurement and calculation of conductivity in nanoscale objects, specifically objects made with transition metal oxides. Nanoscale transition metal oxides are complex because the structural, electronic, and magnetic properties are strongly coupled, but not always linearly. The result is ferroelectric, superconducting, and magnetoresistive materials with great potential as optical switches, magnetic sensors, and transducers where properties can be "tuned" by the introduction of reduced dimensions, doping, or strain. This project focused on low-dimensional systems of ruthenium oxides, with their metal-insulator switching properties, and on manganese oxides, with potential applications for their colossal magnetoresistance. The goal was to both understand transport in these systems and to develop the tools to more generally study electrical conductivity at the nanoscale.

Introduction

Electronic transport in nanoscale materials, such as ultra thin films, interfaces, or surfaces, challenges our understanding of the mechanisms of conductivity. In reduced dimensions, quantum interference of scattered electrons gains importance and nanoscale films, interfaces, or surfaces can exhibit qualitatively different transport from bulk materials. One exciting class of materials receiving increasing attention is the transition metal oxides, which display a fascinating and complex interaction between atoms, electrons, and spins, to a degree unseen in any other materials. Many are highly correlated electron systems that possess several competing quantum mechanical ground states. This competition often makes these materials "bad metals" but also produces a number of highly useful physical phenomena, including superconductivity, colossal magnetoresistance, and ferroelectricity. Electrical transport properties are especially intriguing as they can be tuned over many orders of magnitude by small changes in composition, temperature, strain, or by application of an external magnetic field. This sensitivity makes transition metal oxides excellent candidates for sensors and switches. Technological applications of these materials requires a greater insight into their mechanisms and the development of tools, both experimental and theoretical, for their study.

Technical Approach

This proposal sought novel approaches to measuring and calculating electronic transport of transition metal oxides in low dimensions (i.e., surfaces, thin films, or wires). Where possible, these measurements were made in ultra high vacuum conditions. This involved either growing thin films and measuring in situ, or cleaving bulk samples in vacuum. The growth and characterization of these materials involved an array of tools, including pulsed laser deposition (PLD) growth, low energy electron diffraction (LEED) and scanning tunneling microscopy (STM) for structure, and X-ray photoemission (XPS) and Auger electron spectroscopy (AES) for chemical composition.

Our experimental plan incorporated four approaches to conductivity measurements. The first is most direct: a macroscopic four-point probe with probe spacing around 100 mm. The second approach was to use a microscopic four-point probes develop at the Technical University of Denmark. The third approach is was to apply scanning probe microscopy (SPM) both to obtain local spectroscopy measurements with nanometer resolutions and to use the probe tip as a local voltage probe. Part way through the project, we added a fourth experimental tool, reflection electron energy loss spectroscopy (EELS) with advantages for a specific system discussed below.

The theoretical objective of this project was to model the electron transport properties in order to understand the underlying mechanisms and to explain and suggest future measurements. First-principles methodology was employed, since this approach provides the deepest insight into the fundamental physics and the greatest information on the transport mechanisms. The approach involved a two-step process. First, the ground-state electronic

structure in the film or surface was calculated, since electron transport is inherently linked to the availability and mobility of electrons near the Fermi energy. Electronic structure results were subsequently employed as input for transport calculations.

Results and Accomplishments

Our research has discovered nanoscale effects on conductivity in three types of transition metal oxide systems. The first was evidence that the charge induced by polarization of a ferroelectric material can lead to a metallic surface on the insulating substrate. First-principles local-density-functional calculations for the electronic structure of a film of $BaTiO_3$ show that Ti-3*d*- and O-2*p*-related surface states are present. In the cubic phase, the O surface states are fully occupied and the Ti surface states are fully empty, resulting in an insulating surface. When the bulk material is polarized into the tetragonal state (either by changes in temperature or electric field), the O surface states on one face of the film supply electrons to the Ti surface states on the opposite face. The result is, while the bulk remains insulating, both faces of the film become conducting: on one face through the O surface states, and on the other through the Ti surface states. This is illustrated in Fig. 1, which shows the density of states on O and Ti in the bulk and at the two surfaces.[1] This result could lead to development of highly sensitive thermal and electric field sensors.

A second area of success has been the establishment of experimental capabilities for in situ growth and measurement of conductivity in transition metal oxides that exhibit colossal magnetoresistance. Colossal magnetoresistance (CMR) is a very large change in conductivity in a magnetic field and may replace giant magnetoresistance materials now used in virtually all hard disk drives. Combining pulsed laser deposition with concurrent electron diffraction, we have grown ultra thin, single-crystal films of manganese oxides on insulating $SrTiO_3$. In one example, the effects of strain and nanoscale confinement on the magnetic and resistivity transition temperatures for dimensions down to 7 nm have been measured for $La_{0.7}Sr_{0.3}MnO_3$ films. Results are shown in Fig. 2.

The strong spin-charge-lattice interaction in manganites often leads to a striking phenomenon called electronic phase separation, which indicates the coexistence of a range of exotic electronic and magnetic phases despite the single-crystal structure.[2] The role of electronic phase separation in the related physical properties such as CMR is a hotly debated issue in the field of condensed matter physics.

Using in situ scanning tunneling microscopy/spectroscopy, we have observed both large-scale (greater than a few tens nanometers) and nanoscale electronic phase

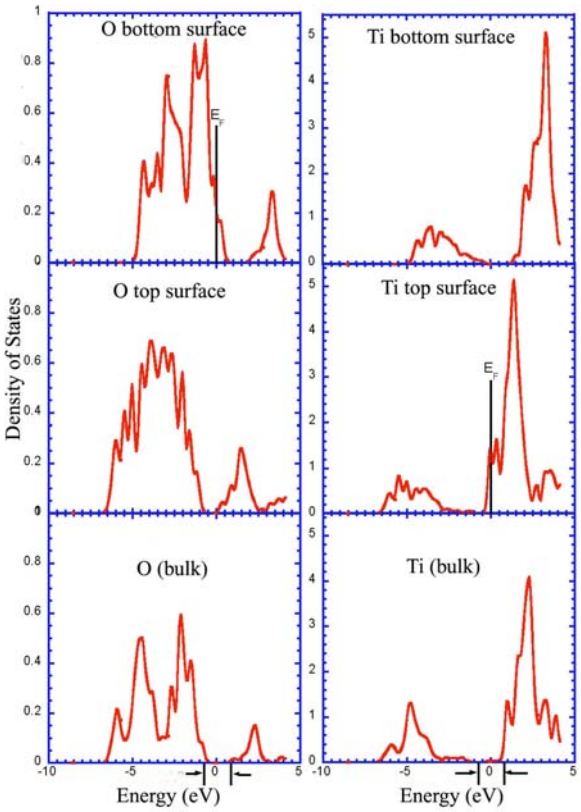


Fig. 1. First principles calculation of the layer-projected density of electronic states of O and Ti atoms in a ferroelectric $BaTiO_3$ slab. The ferroelectric distortion toward the top surface breaks the symmetry and the polarization adds electron density into conducting surface

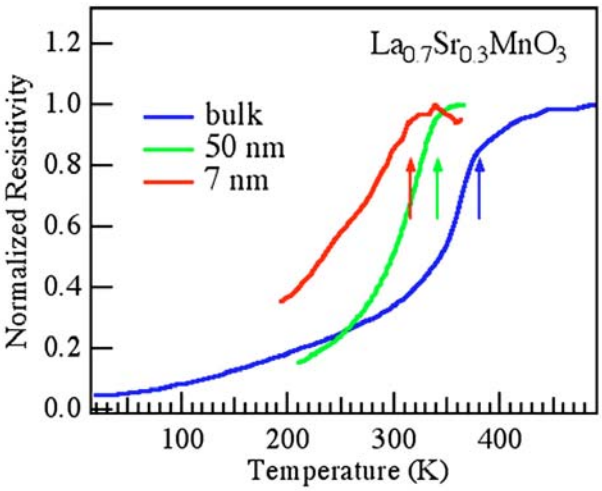


Fig. 2. In situ measurements of the resistivity of $LaCaMnO_3$ films grown using pulsed laser deposition show the metal-insulator transition decreases for thinner films. This is correlated with a decrease in the magnetic phase transition in the same films and reveals the results of strain and confinement.

separation in epitaxially grown thin films of $(La_{5/8-0.3}Pr_{0.3})Ca_{3/8}MnO_3$.[3] Images in Fig. 3 show that the topography (a) is relatively featureless, but the conductivity map (b) varies considerably across the same region. The bottom panel (c) illustrates the different metallic and insulating characteristics in the different regions. While the large phase separation domains are present only below the Curie temperature, the nanoscale phase separation clusters exist at temperatures both below and above the Curie temperature, which implies that the small clusters may originate from doping-related disorder. This is consistent with theories showing that quenched disorder is the driving force for the appearance of large-scale phase separation in magnates.[4]

A third type of nanoscale transition metal oxide system studied in this project is the surface of ruthenates. Sr_2RuO_4 is a superconducting relative of the high-T_c copper oxides, but the isoelectronic material Ca_2RuO_4 is insulating. By examining compounds of the alloy $Sr_xCa_{1-x}RuO_4$, a temperature-dependent metal-insulator transition is observed. In particular, $Ca_{1.9}Sr_{0.1}RuO_4$ exhibits a Mott transition below 154 K in bulk single crystals.[5] Using electron energy loss spectroscopy, which is sensitive only to the surface, we have been able to distinguish the transition at the surface from that of the bulk material. The surface has a surprisingly lower metal-insulator transition temperature ~130 K (Fig. 4). The lowering of transition temperature is intimately related to different lattice distortions at the surface (a relatively smaller orthorhombic distortion but larger RuO_6 compression than in the bulk), indicating a strong electron-phonon coupling in addition to electron-electron correlation. These results imply a coexistence of a metallic surface on a nonmetallic bulk.

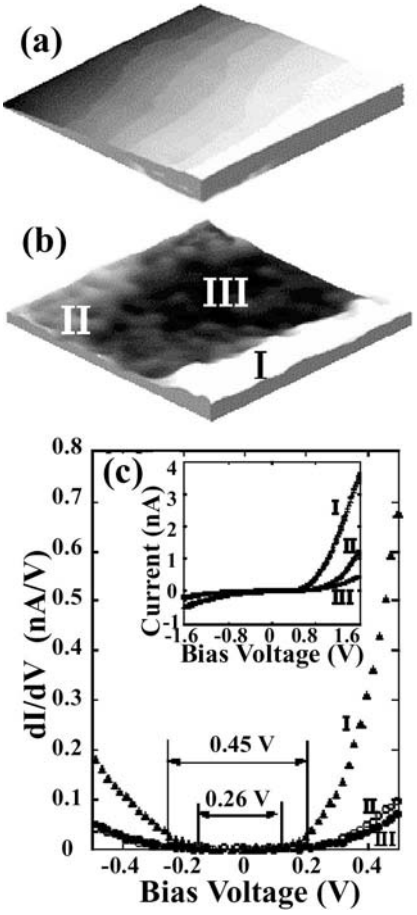


Fig.3. 240 nm × 240 nm topography image (a) and corresponding current image (b) obtained simultaneously from an in situ–grown $(La_{5/8-0.3}Pr_{0.3})Ca_{3/8}MnO_3$ film at T = 60 K. The current image is shown for a voltage bias of 1.6 V. (c) Derivative dI/dV vs V curves obtained from three different areas marked as I, II, and III. The inset shows corresponding I/V curves. The variation in these curves reveals different conductivity characteristics within the film caused by nanophase electronic separation.

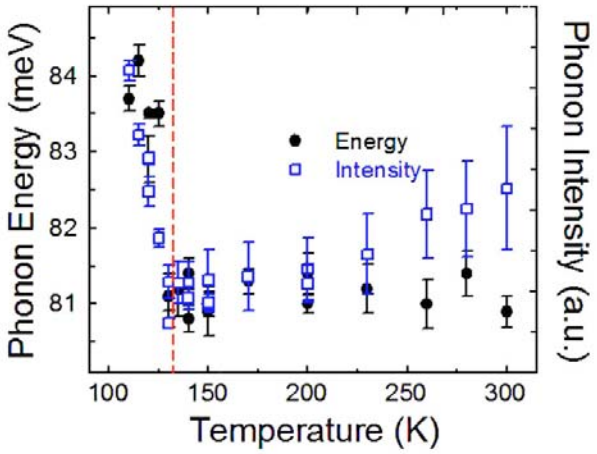


Fig. 4. Electron energy loss spectra reveal a surface Mott transition at ~130 K, lower than the bulk transition temperature of 154 K. The transition is reflected in the phonon energies (left axis) and intensities (right axis).

Summary and Conclusions

Transition metal oxides continue to show a number of novel electrical conductivity phenomena, particularly at dimensions confined to the nanoscale. We have observed a number of metal-insulator transitions on the surface of a ferroelectric, $BaTiO_3$, where the bulk remains insulating, and on the surface of a ruthenate, $Ca_{1.9}Sr_{0.1}RuO_4$, where the bulk becomes an insulator at higher temperatures than the surface. These materials should be pursued as the basis for sensor devices since the strong metal-insulator transition is readily interfaced to electronic devices.

We have also provided a great step toward understanding and controlling electronic transport in colossal magnetoresistance materials. Direct identification of electronic phase separation in $(La_{5/8-0.3}Pr_{0.3})Ca_{3/8}MnO_3$

can identify the mechanism behind colossal magnetoresistance and lead to applications such as hard drive read heads.

More generally, we have developed techniques and expertise for measuring electrical transport in nanoscale materials, using both direct electrical contacts and by analysis of energy losses in scattered electrons. This project has also advanced calculational capabilities, applying first-principles theory to a class of materials, ferroelectrics, which has previously been elusive. In developing these techniques, we have established collaborations with world-recognized leaders in the field of nanoscale conductivity, including Andy Millis, University of Columbia, Allan MacDonald, University of Texas, and S. Kalinin, now a Wigner fellow at ORNL. These techniques and collaborations promise to be of great benefit to future research, to development of the Center for Nanoscale Materials Science at ORNL, and to the interests of the Defense Advanced Research Projects Agency's Defense Sciences Office.

References

[1]M. Krcmar and C. L. Fu, "Structural and electronic properties of $BaTiO_3$ slabs: Mechanism for surface conduction," *Phys. Rev. B* **68**, (2003) 115404.

[2]N. Mathur and P. Littlewood, "Mesoscopic Texture in Manganites," *Phys. Today* **56**, 25 (2003).

[3]J. X. Ma, Z. Gai, M.A. Torija, E. W. Plummer, and J. Shen, "Observation of Two Length Scales for Phase Separation in $(La_{5/8-0.3}Pr_{0.3})Ca_{3/8}MnO_3$," submitted to *Phys. Rev. Lett*. (2003).

[4]A. Moreo, M. Mayr, A. Feiguin, S. Yunoki, and E. Dagotto, "Giant Cluster Coexistence in Doped Manganites and Other Compounds," *Phys. Rev. Lett.* **84**, 5568 (2000).

[5]S. Nakatsuji and Y. Maeno1, "Quasi-Two-Dimensional Mott Transition System $Ca_{2-x}Sr_xRuO_4$," *Phys. Rev. Lett.* **84**, 2666 (2000).

Self-Organized Copolymer and Nanoporous Oxide Thin-Film Templates for Controlled Synthesis and Periodic Replication of Nanoscale Materials

S. Dai,[1] Z. W. Pan,[1] D. B. Beach,[1] M. Dadmun,[2] D. H. Lowndes,[3] V. I. Merkulov,[4] S. J. Pennycook,[3] and C. Tsouris[5]

[1]*Chemical Sciences Division*
[2]*Deparment of Chemistry, University of Tennessee*
[3]*Condensed Matter Sciences Division*
[4]*Engineering Science and Technology Division*
[5]*Nuclear Science and Technology Division*

This project focused on (1) the use of ordered mesoporous oxide thin film with controlled orientations and locations to synthesize well-aligned arrays of one-dimensional (1-D) nanostructures and (2) the development of new synthetic techniques to rationally synthesize 1-D nanostructures. The project demonstrated that highly aligned arrays of multi-wall carbon nanotubes could be grown in a high yield on sol-gel-prepared mesoporous silica film doped with iron nanoparticles. By combination of the sol-gel and shadow-mask techniques, a simple and low-cost methodology was developed to grow micropatterns of aligned carbon nanotube towers. In addition, single-crystalline quasi-aligned germanium nanowires were grown on the sol-gel-prepared silica film doped with gold nanoparticles. This project also demonstrated that molten gallium could be used as an efficient catalyst for the growth of 1-D nanomaterials. For example, carbon nanotubes with large inner/outer diameter ratio were first synthesized by using gallium as the catalyst. Complex SiO_x nanowire assemblies with a variety of morphologies were synthesized by using gallium as the catalyst and SiO powder and SiH_4 gas as the source materials. The gallium-catalyzed nanowire growth exhibits many interesting new nanowire growth phenomena that enrich current nanowire growth mechanism.

Introduction

One-dimensional (1-D) nanostructures have attracted much attention in the past decade because of their unique electrical, magnetic, optical, and mechanical properties, as well as their potential applications.[1] The key to developing advanced 1-D nanostructure-based nanodevices is the efficient synthesis of nanostructures with controlled size distribution, growth direction, and morphology. Thus, controlled growth of 1-D nanostructures is essential for both basic research and technical applications. One objective of this project is to develop a novel methodology for using ordered mesoporous oxide as generic nanoscale reactors for manufacturing and replicating technologically important 1-D nanostructures, for example, aligned carbon nanotubes and semiconductor nanowires. The other objective of this project is to develop a new synthetic route for rationally growing new kinds of 1-D nanostructures that are difficult to be synthesized by traditional methodologies.

Technical Approach

(a) Vertically aligned carbon nanotubes and their patterns grown on sol-gel-prepared mesoporous silica film doped with iron nanoparticles

We have successfully synthesized highly aligned arrays of multi-wall carbon nanotubes on iron-doped mesoporous silica thin films by chemical vapor deposition (CVD). The silica films were prepared by a sol-gel method coupled with surfactant or block copolymer template synthesis. Two kinds of solution, Solution I and Solution II, were used for preparation of the silica films. Solution I was prepared by mixing tetraethoxysilane (TEOS, 3 mL), $Fe(NO_3)_3$ aqueous solution (1.5 *M*, 3 mL), and ethanol (10 mL) by means of magnetic stirring for 1 h. For Solution II, 0.8 g of Pluronic P-123 triblock copolymer was added into Solution I to increases the viscosity of the solution, which facilitates the formation of a very thin and uniform film on silicon wafer without cracking.

Two kinds of film-like iron/silica substrates were prepared for nanotube growth: (1) thick film (with thickness of ~30 to 50 μm) without patterning and (2) thin film (with thickness of ~ 1 μm) patterned with transmission electron microscope (TEM) grids. For the nonpatterned thick film, the solution was dropped onto the surface of a

silicon wafer to form a film of solution with a thickness of about 30 to 50 µm. After gelation, the gel was dried overnight at 80°C to remove the excess water and solvent, during which time the film cracked into small pieces and detached from the wafer surface to form film-like freestanding substrates.

For the patterned thin film, the solution was coated on silicon wafers by dip coating to form a very thin film of solution. When the film was still in the liquid state, copper or gold transmission electron microscope (TEM) grids with square, hexagon, or other kinds of openings were placed on the film to act as a shadow mask. During the process of gelation, Solution I tended to break down in each opening of the TEM grid because of the development of holes in the film, resulting in the formation of a ring-like solid catalyst film along the edge of each opening [Fig. 1(a)]. For Solution II, however, a stable and uniform solid film was formed in each hole of the grid after gelation [Fig. 1(c)], due to the effect of triblock copolymer. These two kinds of catalyst film configurations will result in the formation of two kinds of micropatterns of nanotube towers, that is, solid towers [Fig. 1(d)] for the uniform film and hollow towers [Fig. 1(b)] for the ring-like film. In addition, our results show that the morphologies of the nanotube patterns can be easily controlled by using TEM grids with different openings and by tuning the parameters in preparing the patterned substrates.

Both kinds of substrates were then loaded into a tube furnace and calcinated at 450°C for 10 h under vacuum condition, followed by reduction at 600°C for 5 h in hydrogen atmosphere. At this stage, large quantities of iron/silica nanoparticles formed on the film surfaces due

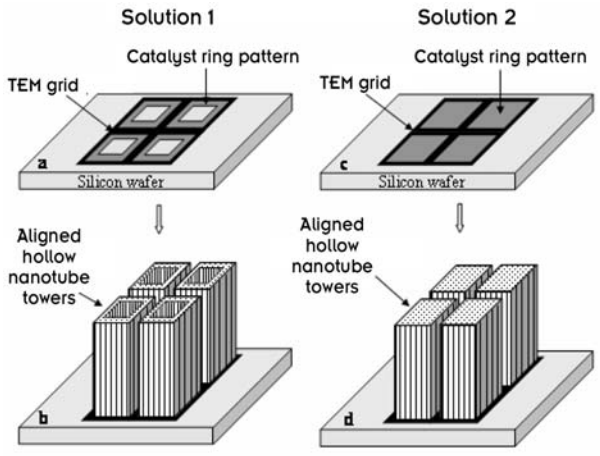


Fig. 1. Schematic diagram of the ring-like (a) and uniform film-like (c) catalyst patterns formed inside the TEM grids with square openings by using Solution I and Solution II as the catalyst precursors, respectively. The ring-like catalyst pattern results in the formation of micropattern of hollow tube towers (b), while the uniform film-like catalyst pattern generates micropattern of solid tube towers (d).

to the transportation of Fe species from within bulk substrate and subsequent aggregation, acting as catalysts for nanotube growth. Finally, acetylene (10 sccm) was introduced into the reaction chamber and aligned and/or patterned nanotubes were grown on the substrates at 700°C for 1 to 5 h.

(b) Single-crystalline quasi-aligned germanium nanowires grown on sol-gel-prepared mesoporous silica film doped with gold nanoparticles

The above mentioned sol-gel technique can be adapted to prepare silica film doped with gold nanoparticles for growing Ge nanowires by substituting $HAuCl_4$ for $Fe(NO_3)_3$. The procedure for substrate preparation is similar to that described above. Prior to being loaded into furnace for nanowires growth, the substrates were annealed at 400°C in air for 2 h, during which time large quantities of Au nanoparticles with diameters of 50–100 nm were formed and evenly distributed on the substrate surface. These Au particles act as seeds for Ge nanowires growth. The Ge nanowires growth was then carried out in a tube furnace by heating Ge powder at 900°C under an Ar flow of 50 sccm and a pressure of 250 Torr for 10 to 60 min.

(c) Gallium-catalyzed growth of carbon nanotubes and SiO_x nanowire assemblies

The vapor-liquid-solid (VLS) crystal growth mechanism, proposed by Wagner and Ellis in 1964 for Si whisker growth,[2] has been widely used to guide the growth of various kinds of 1-D nanostructures including carbon nanotubes and nanowires of semiconductors and oxides. In this mechanism, a small size (5–100 nm in diameter), high-melting-point metal (such as Au and Fe) was used as the catalyst to direct the nanowires' growth direction and defines the diameter of the crystalline nanowire. In this project, we showed that a large size (up to tens of micrometers in diameter) and low-melting-point Ga droplet can be used as an effective catalyst for large-scale growth of multi-wall carbon nanotubes and aligned SiO_x nanowires. Many interesting growth phenomena were observed in our studies.

In the Ga-mediated growth of carbon nanotubes, GaN powder was used as a source of Ga and C_2H_2 was used as the carbon source. The experiment was conducted in a tube furnace at 1150°C under 200 Torr with a C_2H_2 flow rate of 10 sccm and carrier gas Ar flow rate of 50 sccm for 1 h. The carbon nanotubes were grown on the alumina substrates located at the downstream part of the furnace.

In the Ga-catalyzed SiO_x nanowires growth, GaN was also used as the Ga source, and the Si source was provided by either heating SiO powder at 1350°C (the GaN powder was put at the upstream part where the temperature is about 1150°C) or thermal pyrolysis of SiH_4 gas at 1150°C. In

both cases, the growth time is 1 h and the products were deposited at the alumina substrates located at the downstream low temperature part.

For all above syntheses, the products were characterized by scanning electron microscope (SEM), transmission electron microscope (TEM), high-resolution TEM (HRTEM), and energy-dispersive X-ray (EDX) spectroscopy.

Results and Accomplishments

(a) Vertically aligned carbon nanotubes and their patterns[3,4]

Figure 2 shows the SEM and TEM images of the aligned carbon nanotubes and nanotube patterns. For the film-like nonpatterned thick substrates, the nanotubes grow out perpendicularly from all substrate surfaces to form aligned nanotube arrays with the same density and length. This growth feature is clearly displayed in Fig. 2(a), which shows the nanotube arrays growing out perpendicularly from the four adjacent surfaces of a substrate corner. The edge of the substrate is very clear, and the angle between two adjacent arrays is about 90°. Figure 2(a) also shows that the nanotube arrays grown on different surfaces have the same length and nanotube density. High-magnification SEM examinations reveal that the nanotubes are vertical to the substrate surface and remain aligned as their length increases. The length of the nanotube arrays increases with the period of growth at a rate of ~15 μm/h. Figure 2(b) is a TEM image of a bundle of aligned nanotubes, showing that the nanotubes are of uniform diameter, with outer and inner diameters in the range of 20 to 30 nm and 5 to 10 nm, respectively. No catalyst particles were observed inside the nanotubes. HRTEM studies show that the nanotubes are well graphitized and typically consist of 10 to 30 concentric graphite layers [inset in Fig. 2(b)].

By using the patterned substrates prepared by our unique and simple combined approach, we can grow nanotube patterns with a variety of morphologies and configurations. As viewed from the top, patterned nanotube towers in the shape of squares [Fig. 2(c)], hexagons [Fig. 2(d)], letters [Figs. 2(e) and 2(f)], and Arabic numerals [Fig. 2(g)] were obtained by using the corresponding TEM grids as the shadow masks. More importantly, patterns with complex morphologies, which are difficult to achieve by other techniques, can be readily prepared by our method. For example, by using the ring-like catalyst film pattern [see Fig. 1(a)] as the substrate and the TEM grids absent during nanotube growth, a nanotube network can be formed along the trace of the TEM bars [Fig. 2(h)] which replicates the structure of the original TEM grid very well. On the other hand, if the TEM grids remained in place during carbon deposition, micropatterns of hierarchically ordered carbon tubes (i.e., ordered microtubes composed of aligned carbon nanotubes) were obtained on the ring-like catalyst film pattern [Figs. 2(i) and 2(j)]. Note that all the towers shown in Fig. 2(c–j) are composed of highly aligned carbon nanotubes, which grow perpendicularly from the substrate surface and have the microstructures as shown in Fig. 2(b).

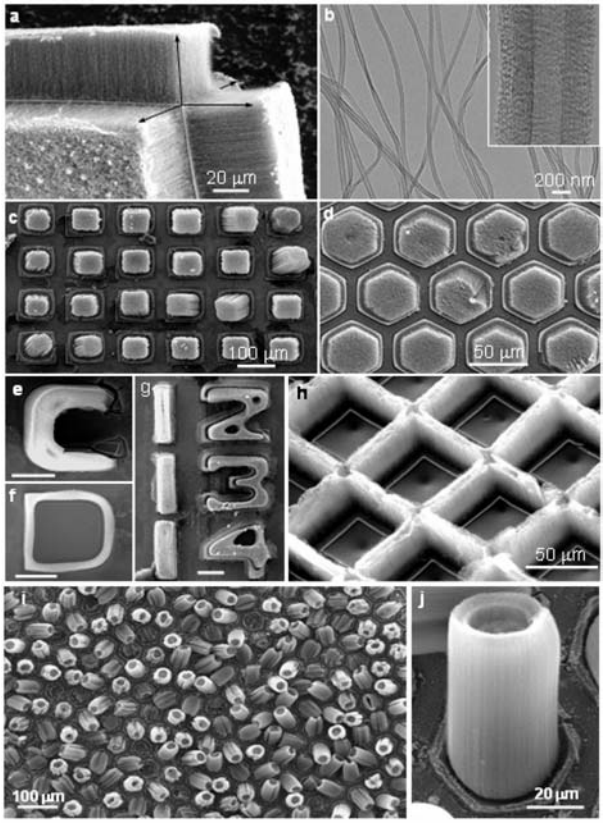


Fig. 2. SEM and TEM images of aligned carbon nanotubes and nanotube patterns. (a) A close-up view of a substrate corner, showing aligned nanotubes growing out perpendicularly from four adjacent surfaces of the corner. The growth direction of the nanotubes on each surface is indicated by a black arrow. (b) TEM image of a bundle of aligned carbon nanotubes. Inset, HRTEM image of a carbon nanotube, showing well-ordered graphitic lattice fringes. (c) Micropatterns of regularly arranged square towers obtained by using a 200 mesh copper grid as the mask after 4 h of growth. (d) Micropatterns of regularly arranged hexagonal towers obtained by using a 300 mesh copper beehive grid as the mask after 2.5 h of growth. (e) and (f) Letters consisting of vertically aligned carbon nanotubes. (g) Arabic numerals consisting of vertically aligned carbon nanotubes. Scale bar in e–g, 50 μm. The substrates corresponding to images c–g were prepared by using Solution II (with copolymer) as the catalyst precursor. (h) A nanotube network grown along a catalyst network that initially existed under the bar of a TEM grid. The grid was removed before carbon deposition. (i) Low-magnification SEM image of micropattern of hierarchically ordered carbon tube towers formed by using a 400 mesh copper beehive grid as shadow mask after 3 h of growth. (j) High-magnification image of one hierarchically ordered hollow tower. The substrates corresponding to images c–g were prepared by using Solution I (without copolymer) as the catalyst precursor.

(b) Single-crystalline quasi-aligned germanium nanowires[5]

Figure 3 shows the SEM and TEM images of Ge nanowires grown on film-like gold/silica substrates. After the growth, the substrates were covered with brown nanowire materials in a high-yield, as shown in Fig. 3(a). The nanowires are straight with length up to tens of micrometers. Similar to the growth behavior of aligned carbon nanotubes shown in Fig. 2, the Ge nanowires grown on the gold/silica substrate tend to grow perpendicularly to the substrate surface, and in some case, quasi-aligned nanowire array can be obtained [Fig. 3(b)]. The nanowires shown in Fig. 2(b) are very straight and uniform in diameter (~50–80 nm). Each nanowire is tipped with an Au nanoparticle with diameter comparable to the connected nanowire [inset in Fig. 3(b)], showing typical VLS growth feature.

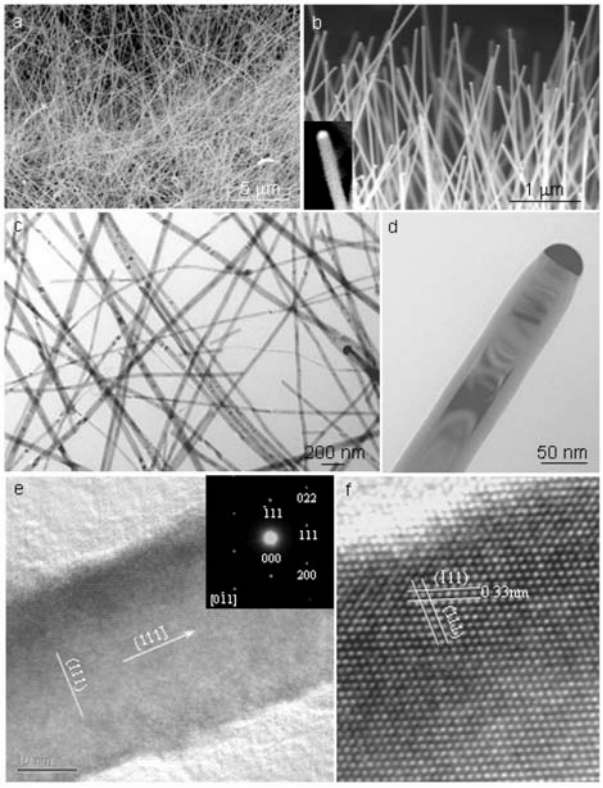


Fig. 3. SEM and TEM images of Ge nanowires grown on sol-gel prepared gold/silica substrate. (a) SEM image showing straight Ge nanowires grown in high-yield and high purity. (b) Quasi-aligned straight Ge nanowires with uniform diameter. The inset shows a nanowire tipped with an Au nanoparticle. (c) Low-magnification TEM image showing straight and uniform Ge nanowires. (d) TEM image of an individual Ge nanowire with a hemispherical Au nanoparticle located at the tip, showing a core-shell structure. (e) HRTEM image of a Ge nanowire with diameter of ~40 nm. The inset shows a corresponding electron diffraction pattern. (f) An atomic resolution image taken from the edge part of the nanowire in (e).

The uniformity of the Ge nanowires was further proved by TEM investigation. Figure 3(c) is a low-magnification TEM image of the Ge nanowires, showing straight and uniform nanowires. Figure 3(d) is a TEM image of an individual Au-tipped Ge nanowire, showing a core-shell structure. EDX analysis shows that the crystalline core is Ge and the amorphous sheath has an approximate composition of GeO_2. Figure 3(e) is a HRTEM image of a Ge nanowire with diameter of ~40 nm. The lattice fringes perpendicular to the wire axis correspond to the (111) lattice planes [spacing, 0.33 nm; see the atomic resolution image in Fig. 3(f)] of Ge, indicating a [111] growth direction. This direction was confirmed by the selected area electron diffraction pattern [inset in Fig. 3(e)] recorded perpendicular to the nanowire long axis, which could be indexed as the $\langle 0\,\bar{1}\,1\rangle$ zone axis of crystalline Ge.

(c) Gallium-mediated growth of carbon nanotubes[6]

Figure 4 shows the SEM and TEM images of the carbon nanotubes catalyzed by molten Ga. The Ga-catalyzed nanotubes tend to grow in groups, in which hundreds of nanotubes radially grow upward around an island with a diameter of ~10 μm, as shown in Fig. 4(a). High-magnification SEM observations and EDX analyses

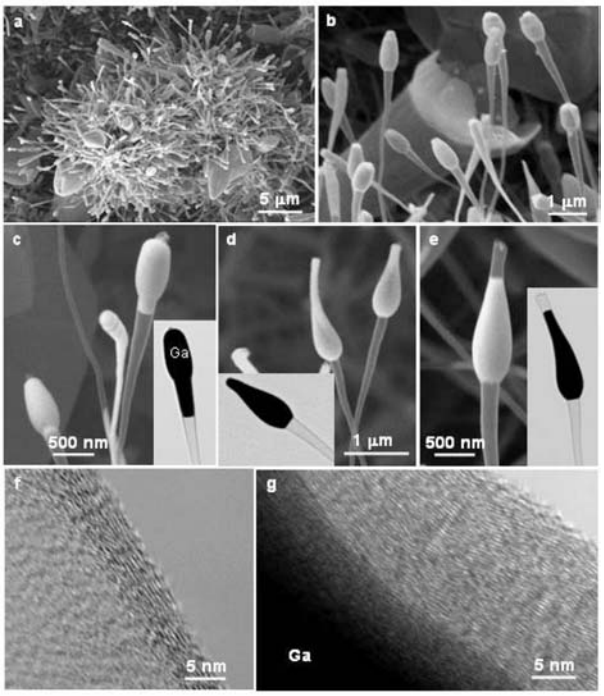


Fig. 4. SEM and TEM images of Ga-mediated carbon nanotubes. Low- (a) and high-magnification (b) SEM images of pin-like nanotubes. (c–e) SEM images of three typical tube tip morphologies. The insets are the corresponding TEM images of these three typical tip morphologies. HRTEM images taken at one side of a hollow tube body (f) and the joints of the tip and tube (g).

show that the nanotubes have a pin-like morphology composed of an oval-shaped tip filled with liquid Ga and a tapered hollow body [Fig. 4(b)]. The diameters of the tapered tubes vary from ~50 nm at the thinner parts to ~200 nm at the thicker parts, and the length of the tubes is in the range of 4 to 8 μm. The oval-shaped tips have a maximum diameter in the range of 200–500 nm and length of up to 2 μm. The wall of the nanotube is so thin that the hollow structure of the tube, the tube walls, and the Ga level inside can be clearly seen even with the SEM operated at 5 kV. Three kinds of tip morphologies can be distinguished: the first has a flat tip front [Fig. 4(c)]; the second has a Ga-filled sharp end [Fig. 4(d)]; and the third has a short open tube that extrudes from an oval-shaped ball [Fig. 4(e)]. Figure 4(f) shows a HRTEM image of one side of a hollow tube, and Fig. 4(g) shows the joint part of the tip and tube body (the black part is Ga). It is clear that the tubes are multi-walled carbon nanotubes. Both SEM and TEM results indicate that the nanotubes present here have a very large inner-outer diameter ratio with a value of >0.9, which is much larger than that of conventional nanotubes (usually <0.5).

We heated the Ga-filled carbon nanotubes in situ in TEM in the temperature range of 30–850°C and found that below 550°C, the Ga inside the tube tip expanded almost linearly along the hollow tube with the temperature, with an average increase of ~20 nm/100°C. This means that the Ga-filled nanotubes might work as a nanothermometer in the temperature range of 30–550°C, which might be suitable for use in a wide variety of microenvironments.

(d) Gallium-catalyzed growth of SiO_x nanowire assemblies[7,8]

Figure 5 shows the SEM and TEM images of Ga-catalyzed SiO_x nanowire assemblies by using SiH_4 as the Si source. Electron microscopy observations reveal that the morphologies of the products are very sensitive to the growth temperature. From the highest deposition temperature of 1100°C to the lowest deposition temperature of 950°C (covering a substrate length of ~45 mm), the morphology, distribution, and density of the products vary gradually with their positions on the alumina substrate and five distinctive deposition zones can be identified based on the morphologies of the products. From 1100°C to 950°C, the morphologies of the products change from fishbonelike (Zone I, 1100–1050°C) [Fig. 5(a)] to gourdlike (Zone II, 1050–1030°C) [Fig. 5(c)], spindlelike (Zone III, 1030–1010°C) [Fig. 5(d)], badmintonlike (Zone IV, 1010–980°C) [Fig. 5(e)], and finally octopuslike (Zone V, 950–980°C) [Fig. 5(f)]. However, although the products formed in different temperature zones have different morphologies, they share some similar structural and compositional features. For example, as shown in Fig. 5, all the products are composed of a spherical liquid Ga ball and a silicon oxide nanowire bunch that is composed of numerous highly aligned amorphous silicon oxide nanowires with diameters of 40–80 nm and lengths of 10–20 μm. These complex morphologies can be well explained by two novel nanowire growth phenomena that were mainly observed in silicon oxide nanowire growth. One growth phenomenon is the branching growth, that is, one nanowire splits into two branches, and the newly formed branch also splits into two subbranches, and so on [Fig. 5(b)]. The other phenomenon is the batch growth; that is, the nanowires within a bunch tend to grow batch by batch so that the nanowire bunch can be regarded as a rick piled up with nanowires [see Fig. 5(a) as an example and refer to ref. 7 for detailed growth mechanism].

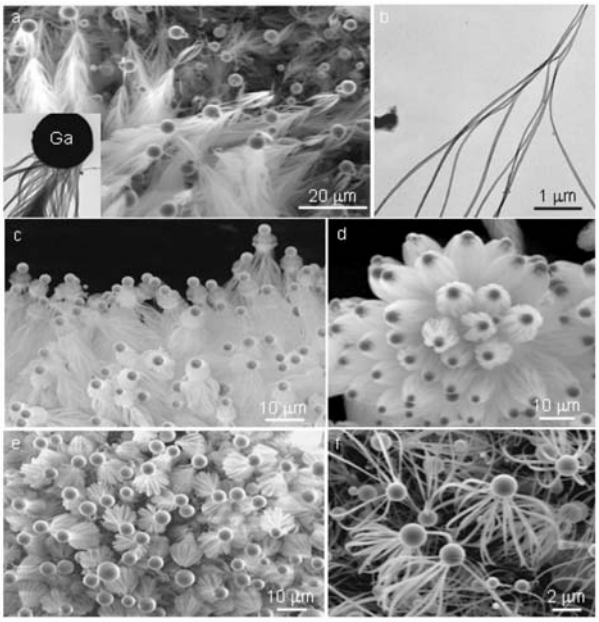


Fig. 5. SEM and TEM images of Ga-catalyzed SiO_x nanowire assemblies by using SiH_4 as the Si source. (a) Fishbonelike SiO_x nanowire assemblies formed in Zone I (1050–1100°C), showing typical batch-growth feature. The inset shows a TEM image of the tip part of a fishbonelike structure, showing many SiO_x nanowires growing out from the lower hemisphere surface of a Ga ball. (b) TEM image of a nanowire bundle displaying branching-growth phenomena. (c) Gourdlike SiO_x nanowire assemblies formed in Zone II (1030–1050°C). (d) Spindlelike SiO_x nanowire assemblies formed in Zone III (1010–1030°C). (e) Badmintonlike SiO_x nanowire assemblies formed in Zone IV (980–1010°C). (f) Octopuslike SiO_x nanowire assemblies formed in Zone V (950–980°C).

Figure 6 shows the SEM images of Ga-catalyzed SiO_x nanowire assemblies by using SiO powder as the Si source. The morphologies of the products formed by using SiO powder as the Si source are different from those formed from SiH_4. Figure 6(a) is a SEM image showing highly aligned and closely packed rodlike structures. Most of the rods have a carrotlike morphology with a sandwichlike topology that consists of three materials with different morphologies, states, and crystallographic structures [Figs. 6(b) and 6(c)]. The "carrot" top is a liquid Ga ball

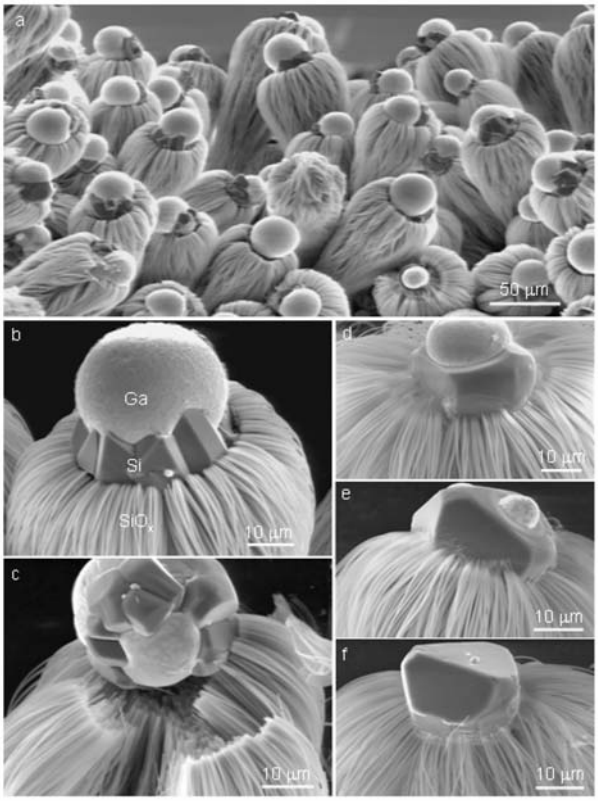


Fig. 6. SEM images of Ga-catalyzed SiO_x nanowire assemblies by using SiO powder as the Si source. (a) Low-magnification SEM image of the as-grown products, showing highly aligned and closely packed rodlike structures. Close-up view of the top part of an as-grown (b) and opened (c) sandwich-structured rod, revealing that the rod is composed of three materials with different morphologies, states, and crystallographic structures, as well as the relationship among these three parts. (d) A sandwichlike rod having a hemispherical Ga droplet on the top surface of a Si polyhedron. (e) A sandwichlike rod having two small hemispherical Ga droplets on the two top surfaces of a Si polyhedron. (f) A rod consisting of a polyhedral Si tip and a nanowire bunch.

with diameter of ~10 to 30 μm; the shoulder (middle part) is a Si ring usually composed of ~10 micrometer-sized, clearly faceted and crystalline Si polyhedrons that are arranged sequentially around the lower hemisphere surface of the Ga ball; the bottom part is a carrot-shaped bunch (~60 μm in maximum diameter and up to 200 μm in length) of highly aligned silicon oxide (SiO_x, $x = 1.5–2$) nanowires that grow perpendicularly from the downward facing facets of the Si polyhedrons. Besides the rods shown in Fig. 6(b), we also observed two other kinds of morphologies as shown in Figs. 6(d–f). Figure 6(d) and 6(e) show two typical derivatives of sandwichlike rods, respectively displaying 1 to 2 small hemispherical Ga droplets (5–15 μm in diameter) located on the top facets of a large Si polyhedron (~20 μm in width and 10–20 μm in height). In Fig. 6(f), however, the rod consists only of a polyhedral Si tip and a carrot-shaped nanowire bunch; no Ga droplet was observed on the Si crystal surface. Note that the batch-growth and branching-growth phenomena were also observed in the silicon oxide nanowire assemblies prepared from SiO powder.

Summary and Conclusions

As demonstrated in this project, well-aligned carbon nanotubes and quasi-aligned single-crystalline Ge nanowires were grown on mesoporous silica thin films doped with Fe and Au nanoparticles, respectively. Specially, the sol-gel-prepared silica film can be easily cast into patterned substrates by simply using TEM grids as the shadow masks so that aligned carbon nanotube patterned can be obtained. Because no sophisticated facilities such as lithography were involved in preparing the catalyst patterns, our technique represents a simple, low-cost approach to the micropatterning of aligned carbon nanotubes. These low-cost nanotube micropatterns might find application in the area of flat panel displays.

In this project, molten Ga was proved to be an effective catalyst for high-yield growth of carbon nanotubes and complex SiO_x nanowire assemblies. The Ga-catalyzed growth demonstrated several interesting new growth phenomena unlike any previously observed by using high-melting point Au and Fe as the catalysts, which greatly enrich the current VLS growth mechanism.

References

[1]J. Hu, W. Odom, and C. M. Lieber, "Chemistry and Physics in One Dimension: Synthesis and Properties of Nanowires and Nanotubes," *Acc. Chem. Res.* **32**, 435–445 (1999).

[2]R. S. Wagner and W. C. Ellis, "Vapor-Liquid-Solid Mechanism of Single Crystal Growth," *Appl. Phys. Lett*. **4**, 89–90 (1964).

[3]Z. W. Pan, H. G. Zhu, Z. T. Zhang, H. J. Im, S. Dai, D. B. Beach, and D. H. Lowndes, "Patterned Growth of Vertically Aligned Carbon Nanotubes on Pre-patterned Iron/Silica Substrates Prepared by Sol-Gel and Shadow Masking," *J. Phys. Chem. B* **107**, 1338–1344 (2003).

[4]Z. W. Pan, H. G. Zhu, Z. T. Zhang, H. J. Im, S. Dai, D. B. Beach, and D. H. Lowndes, "Hierarchically Ordered Carbon Tubes," *Chem. Phys. Lett.* **371**, 433–437 (2003).

[5]Z. W. Pan, S. Dai, and D. H. Lowndes, "Straight Single-Crystalline Germanium Nanowires and Their Patterns Grown on Sol-Gel-Prepared Gold/Silica Substrate," to be submitted.

[6]Z. W. Pan, S. Dai, D. B. Beach, N. D. Evans, and D. H. Lowndes, "Gallium-Mediated Growth of Multiwall Carbon Nanotubes," *Appl. Phys. Lett.* **82**, 1947–1949 (2003).

[7]Z. W. Pan, S. Dai, D. B. Beach, and D. H. Lowndes, "Temperature Dependence of Morphologies of Aligned Silicon Oxide Nanowire Assemblies Catalyzed by Molten Gallium," *Nano. Lett.* **3**, 1279–1284 (2003).

[8]Z. W. Pan, S. Dai, D. B. Beach, and D. H. Lowndes, "Liquid Gallium Ball/Crystalline Silicon Polyhedrons/Aligned Silicon Oxide Nanowires Sandwich Structure: An Interesting Nanowire Growth Route," *Appl. Phys. Lett.* **83**, 3159–3161 (2003).

Multifunctional Nanotube Composites

D. B. Geohegan,[1] P. F. Britt,[2] M. J. Lance,[3] A. A. Puretzky,[4] M. A. Guillorn,[5] M. Dadmun,[6] J. Devenyi,[7] A. Pedraza,[4] I. Ivanov,[1] H. Schittenhelm,[2] S. Viswanathan,[2] A. Rasheed,[4] S. N. Jesse,[6] B. White,[8] J. Y. Howe,[3] and K. Belay[9]

[1]*Condensed Matter Sciences Division*
[2]*Chemical Sciences Division*
[3]*Metals and Ceramics Division*
[4]*Materials Science and Engineering Department, University of Tennessee*
[5]*Engineering Science and Technology Division*
[6]*Department of Chemistry, University of Tennessee at Knoxville*
[7]*Department of Chemistry, University of Tennessee at Martin*
[8]*ERULF summer student, Saginaw Valley State University*
[9]*Visiting HBCU Faculty Program Participant, Florida A&M University*

This project focused on the development of new composite materials incorporating carbon nanotubes. The light weight, great strength, and high electrical and thermal conductivities of single-walled carbon nanotubes (SWNTs) are envisioned to lend new functionality to matrices of polymers, carbon, and metals. This project developed new capabilities in the synthesis and processing, characterization, and testing of multifunctional nanocomposites using nanotubes. First, coordinated laser-vaporization synthesis and chemical purification procedures were developed to produce the purest (>99.98 wt %) SWNTs known, an essential starting point to understand nanotube chemistry for strong bonding in polymer composites. Second, several novel processing techniques were developed to align the nanotubes and infiltrate them with matrix materials in order to form composites with directional electrical and thermal conductivity. Controlled aggregation techniques were developed to form pure SWNT papers and fibers. Electrospinning was developed to make polymer fibers containing aligned SWNTs. Pulsed laser deposition was used to coat and protect SWNT networks with amorphous diamond films. Infiltration techniques were used to encapsulate millimeters-long, vertically oriented arrays of carbon nanotubes (multiwall plus single wall). Finally, methods of imaging and testing the properties of these materials were developed, including (1) a new electron microscopy technique capable of imaging nanotubes within thin polymer films, (2) polarized Raman microscopy to measure nanotube alignment, and (3) Raman spectroscopy analysis of strain within thin composite films. The carbon nanotube chemistry, physics, and processing methods developed in this project resulted in nanoscience techniques, which greatly enhanced capabilities at ORNL, and prototype nanotube–based composites that have attracted interest from several federal agencies.

Introduction

Single-wall carbon nanotubes are exceedingly strong (>1 TPa axial Young's modulus) and have been theoretically predicted to form strong, lightweight composites far superior to those used today in aerospace, transportation, and commercial products. In addition to strength, their remarkable electronic and thermal properties have been envisioned to lend additional functionality to polymers, especially if the nanotubes can be aligned within the polymer matrix. However, due to a poor understanding of nanotube chemistry, good dispersion of nanotube bundles and strong bonding to polymer matrices have not been generally achieved and the predicted strength enhancements have not been realized. Moreover, reliable methods have not been developed to induce and measure nanotube alignment or methods to image and characterize electronic transport through nanotube networks in polymers. This project addressed these scientific needs and, through the fabrication of prototype nanotube-polymer composites, experimentally explored the improvements in polymer properties induced by carbon nanotubes.

Technical Approach

A major goal of this project was to understand how the synthesis conditions of single-wall carbon nanotubes (SWNTs) affect their chemistry during their purification. Irreproducible nanotube chemistry is largely attributable to chemical treatments employed to remove unwanted metal catalyst nanoparticles formed during their synthesis

and the resulting residual unconverted carbonaceous impurities which adhere to the walls of the nanotubes after these treatments. Laser vaporization was used to synthesize SWNTs at ORNL.

The technical approach for the chemical purification and functionalization of the nanotubes is outlined in Fig. 1. Raman spectroscopy of characteristic SWNT optical modes was used to gauge the yield of nanotubes in the raw material as well as the level of defects they contained, and all material was examined by scanning electron microscopy through each step of the purification procedure. Since SWNTs oxidize at higher temperatures than amorphous carbon (a-C), thermogravimetric analysis (TGA) was used to infer the fraction of SWNTs vs a-C after removal of the metal catalyst. Inadvertent variations in the synthesis conditions were found to affect the level of graphitic carbon–coated metal nanoparticles observed by TEM, so synthesis conditions were adjusted to minimize these while preserving high nanotube yields. The approach in this project of jointly synthesizing and purifying SWNTs differs from the general practice of purchasing SWNTs from vendors and attempting to perform reproducible chemistry; however, the successful reproducible achievement of ultrapure SWNTs (>99.98 wt %, as measured by ICP analysis) is a major accomplishment of this program which has not been achieved elsewhere.

Various approaches were derived to solubilize the nanotubes in various solvents and mix them with solubilized polymers to form stable nanotube-polymer solutions. To enhance the solubility of the SWNT in solvents, improve polymer-SWNT interactions, and enhance the dispersion of SWNT bundles, methods were explored to functionalize the SWNTs. Typically, the carboxyl groups (-COOH) produced from the oxidative purification of crude SWNTs are derivatized with long-chain alkyl groups to improve solubility. Unfortunately, FTIR analysis revealed that the purified SWNTs had no detectable carboxyl groups. Thus, various methods, including KNO_4, acids, and ozone, were attempted to introduce carboxyl groups into the pure SWNT without producing amorphous carbon.

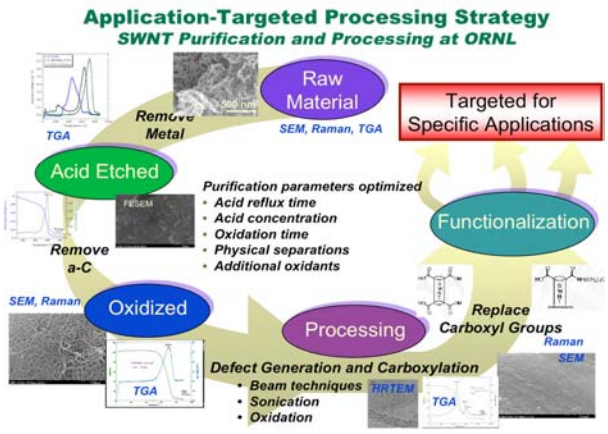


Fig.1. Chemical methods developed to purify ultrapure (>99.98 wt %) single-wall carbon nanotubes from "raw" 50 wt % material produced by laser vaporization, then functionalize them for use in polymer composites. Coordinated SEM, Raman, TGA, and HRTEM techniques are used to gauge effectiveness of acid-etching and selective oxidation steps to remove metal catalyst and amorphous carbon impurities. Synthesis conditions are readjusted to improve these steps. Once high-purity SWNTs are obtained, defects are reintroduced and assessed, enabling reproducible functionalization chemistry to be performed.

Nanotube-polymer composites were then formed from these solutions by spinning them onto silicon wafers with prepatterned electrodes, doctor-blading them into thin sheets, or electrospuning them into fibers. Alternatively, purified nanotubes in solvents were formed into aligned, pure nanotube papers and fibers through controllable aggregation using electric fields, drying techniques, shear flow, or combinations of these. Special consideration was given to methods of composite processing that simulataneously produced alignment of SWNTs in the finished composite. In addition, polystyrene composites containing 1 and 2 wt % octadecyl functionalized SWNTs were prepared by precipitation of a well-mixed solution of polystyrene and functionalized SWNTs followed by melt pressing the material into a bar.

To assess alignment of nanotubes within fibers, blocks, and thin films, several approaches were developed. Polarized resonant Raman scattering utilized the high anisotropy of the Raman scattering signal from SWNTs *along* vs *perpendicular* the nanotube length. In addition, frequency shifts of the characteristic nanotube breathing and tangential modes with strain were explored to understand load transfer of applied stress to nanotubes embedded in the composite.

Novel approaches were developed to characterize the composites and assess their mechanical, electrical, and thermal properties. Electrical conductivity of nanotube-polymer (PMMA) composites was explored in a specially fabricated scanning electron microscopy (SEM) stage which permitted the current-voltage characteristics of embedded nanotube networks to be recorded *while simultaneously imaging their morphology inside the polymer.*

Results and Accomplishments

Some of the highest-purity SWNTs produced worldwide were accomplished in this project, and are shown in Fig. 2. A wide variety of chemical oxidative treatments, including nitric acid, nitric/sulfuric acid, hydrogen peroxide, hydrogen peroxide/sulfuric acid, $KMnO_4$/sulfuric acid, as well as steam and thermal

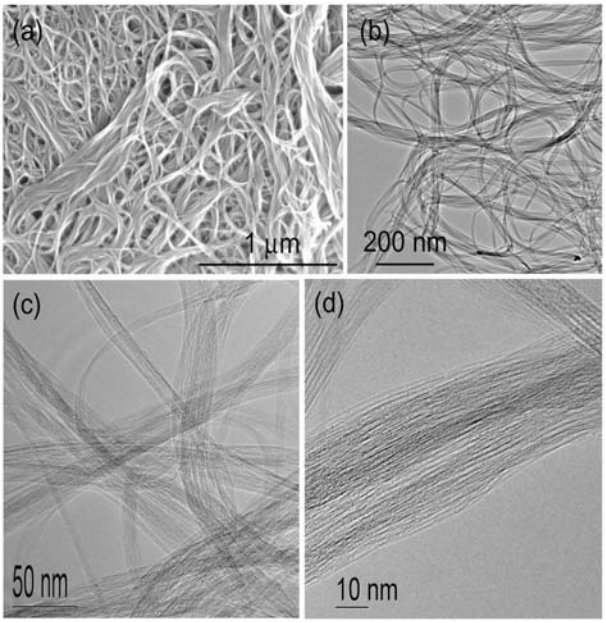


Fig. 2. Single-wall carbon nanotubes grown by laser vaporization and purified according to the procedure of Fig. 1 to remove metal catalyst nanoparticles (to <0.02 wt %) and amorphous carbon impurities. (a) Large-area scanning electron microscope image, (b) TEM image, and (c,d) HRTEM images of nanotube bundles. These SWNT are among the purest reported and serve as an important benchmark for purity and a starting point for reproducible nanotube chemistry.

oxidative treatments, were evaluated to remove the metal and carbon impurities while maintaining a reasonable yield of purified SWNTs. The carbon coating on the metal catalyst particles was notoriously difficult to remove without destroying the SWNTs. Moreover, the oxidative stability of the carbon impurities were found to be very sensitive to the laser ablation conditions so that each batch of crude SWNT had to be screened by TGA, FESEM with EDX, TEM, and Raman spectroscopy and the purification procedure optimized. Overall, the best purification method involved refluxing crude SWNT in 3 *M* nitric acid for 16 h, filtering, and optimizing the thermal oxidation treatment by TGA and FESEM (with EDX) assessment. Acid-refluxed samples were typically heated to 400 °C for 2 h in air, sonicated in concentrated HCl for 10 min, and washed with water. The samples were then subjected to a second thermal oxidation at 450 °C for 3 h, sonicated in concentrated HCl/HF, and washed with water. The two-stage thermal oxidation was necessary to enhance the yield of pure SWNTs. The final product was obtained in 8.0 ± 1.3% yield and contained <0.025 wt% Ni and Co (by ICP) and a very low content of amorphous carbon, by TEM and NIR measurements. These samples appear to be some of the highest purity SWNTs produced anywhere to date.[1]

A wide variety of novel processing techniques were developed to form nanotube-polymer and nanotube-carbon composites targeting several key application areas. Thin films, fibers, and blocks were obtained with greatly enhanced electrical and thermal conductivity. Several are shown in Fig. 3. The first SWNT-amorphous diamond composite films were formed by pulsed laser deposition of pure C onto dried, spray-deposited SWNT solutions while the electrical properties of the nanotube networks were monitored in situ. The conduciting nanotube networks were coated and protected with transparent, amorphous diamond [Fig. 3(a,b)] to form a novel all-carbon composite which were highly scratch resistant and of interest for applications as sensors and smart coatings.[2]

Thin films of nanotube-polymer composites were also formed [Fig. 3 (c, d) and Fig. 4]. Due to the poor contrast difference between carbon nanotubes and polymers in TEM, the morphology of embedded nanotube networks are virtually impossible to reveal nondestructively. In this project, it was discovered that nanotubes could be imaged *inside the polymer thin film* by scanning electron microscopy using certain bias conditions of the sample and accelerating voltages of the electron beam (Fig. 4).[3] These measurements provided *the first coordinated measurements of nanotube network morphology and conduction* inside polymers to understand how junctions

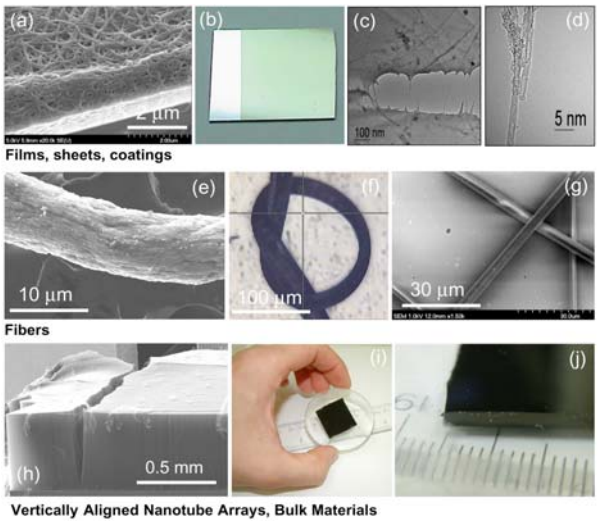


Fig. 3. Processing techniques were developed to form nanotubes into functional composites. (a, b) SWNT dried from solution after spray-deposition onto Si were coated with amorphous diamond by pulsed laser deposition to form highly scratch-resistant, transparent, electrically conductive coatings. (c,d) SWNTs encapsulated inside polymers (e.g., PMMA) by spin-deposition of nanotube-polymer solutions were investigated by HRTEM. (e) Neat fibers of SWNT formed by shear forces, solvent replacement, and flocculation of "nanotube ink." (f, g) Electrospun nanotube-polymer fibers (PMMA and PEO, respectively). (h) As-grown, vertically aligned array of multiwall carbon nanotubes. (i, j) Epoxy-infiltrated vertical MWNT arrays with preserved nanotube alignment.

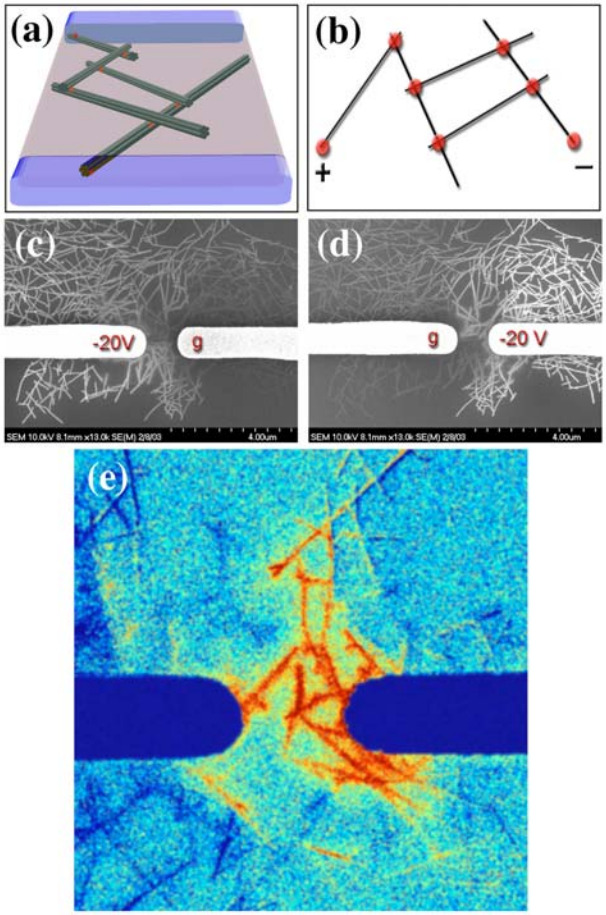


Fig. 4. Nanotubes lend orders-of-magnitude increase in electrical conductivity to polymers; however, nanotubes occur mainly in bundles as shown in (a), which shows a series of nanotube bundles inside a thin polymer film spanning two electrodes (blue). Connections between the nanotube bundles lead to an equivalent circuit as shown in (b). In this project, an electron microscopy technique was developed to image SWNT bundle networks inside *polymer films while they carried current between biased electrodes. As shown in (c) and (d), nanotube bundles (white) between grounded (g) or negatively biased (-20V) electrodes in PMMA are directly visible due to enhanced contrast induced by the electric field effect on secondary electron emission. The nanotubes participating in the measured current-voltage characteristic can be determined by image processing, as shown in (e).*

between nanotube bundles lead to overall conductivity of the composite. In addition, the conditions responsible for enhanced contrast between the nanotubes and the polymer were studied and analyzed and found to relate to the shift in secondary electron yield caused by electric fields in the test fixture. This technique appears to be a new, general one in electron microscopy we call *electric field–induced contrast* imaging. The electrical conductivity of SWNT networks in thin poly-methyl methacrylate (PMMA) films was measured vs weight loading, and over four-orders-of- magnitude increase was obtained with less than 1 wt % nanotubes.

For nanotube-polymer composites with directional electrical and thermal conductivity, nanotubes must be aligned. Fibers composed of aligned nanotubes are natural forms which satisfy this requirement. Graphitic carbon fibers are a mainstay of composite manufacturing, and SWNT fibers are predicted to have much higher strength and electrical/thermal conductivity. Neat nanotube fibers and aligned nanotube papers were formed by solubilizing SWNT in surfactants such as sodium dodecyl sullfate, then inducing aggregation during shear extrusion into another liquid or during drying on a plate at the meniscus. Long, hairlike prototype SWNT fibers were tested for mechanical strength and found to be remarkably strong (supporting over 30 g of weight for a 10-mm-diameter fiber). Electrospinning was developed in this project to synthesize nanotube-polymer fibers of over 15 polymers with diameters ranging from 100 µm down to 100 nm. Electrospinning involves high voltage to draw fibers from droplets of viscous (nanotube-polymer) solution. The nanotube alignment within the pulled fibers was determined by polarized Raman microscopy, with SWNT becoming highly aligned (18:1 polarization ratio) for fiber diameters less than 10 mm. In general, significant enhancements in the oxidative thermal stability (75°C for PMMA) were observed at low weight loadings (0.6 wt %) of SWNT—an important enhancement allowing polymers to be used in applications at elevated temperatures.

To enhance the solubility of the SWNT in solvents, improve polymer-SWNT interactions, and to enhance the dispersion of SWNT bundles, methods were explored to functionalize the SWNTs. Typically, the carboxyl groups (-COOH) produced from the oxidative purification of crude SWNTs are derivatized with long-chain alkyl groups to improve solubility. Unfortunately, FTIR analysis revealed that our pure SWNTs had no detectable carboxyl groups. Thus, various methods, including KNO_4, acids, and ozone, were attempted to introduce carboxyl groups into the pure SWNT without producing amorphous carbon. The best method involved treating SWNTs with piranha solution (a 4:1 mixture of 30% hydrogen peroxide and sulfuric acid) for 1 h at 70 °C. The carboxylated nanotubes were reacted with thionyl chloride ($SOCl_2$) followed by octadecylamine ($C_{18}H_{37}NH_2$) to produce octadecyl functionalized SWNT. This material could be suspended in organic solvents, such as ethylene dichloride, for >1 week while the unfunctionalized SWNTs would precipitate in a few hours. Attempts to produce a water-soluble SWNTs by functionalization with poly(ethylene glycol) monomethyl ether were unsuccessful.

Polystyrene composites containing 1 and 2 wt% octadecyl functionalized SWNTs were prepared by precipitation of a well-mixed solution of polystyrene and functionalized SWNTs followed by melt pressing the

material into a bar. The black composites containing functionalized and pure SWNTs were analyzed by dynamic mechanical analysis. Overall, the glass transition temperature, determined from the maximum of tan *d*, was not significantly affected (±5 °C) by the addition of the SWNTs to the polymer. However, the storage modulus (*E´*) at room temperature increased 57% for polystyrene plus 1wt % functionalized SWNTs, and a factor of 4.8 for polystyrene plus 2 wt % functionalized SWNTs compared to polystyrene alone. The storage modulus was also 60% larger for the functionalized SWNT than for the pure SWNTs. To investigate the dispersion of the functionalized SWNTs in the polystyrene, thin films were cast and examined by SEM analysis. Although large bundles of SWNTs were observed in the composite, the functionalized nanotubes appeared to be better dispersed than the non-functionalized SWNTs. Although functionalization did not significantly enhance the dispersion the bundles of SWNT, the mechanical properties of polystyrene were enhanced with low loading of SWNTs indicating load transfer between matrix and filler.

Finally, aligned multiwall carbon nanotubes grown by chemical vapor deposition into long arrays [Fig. 3 (h)] were infiltrated by epoxies to form composites which were electrically and thermally conductive [Fig. 3(i,j)]. Alignment of the nanotubes was preserved during polymer infiltration, resulting in an increase in thermal diffusivity of a factor of 20 compared to the epoxy alone, as measured by flash photolysis. Nanoindentation measurements of the composites both along and perpendicular to the nanotube axes revealed no change in hardness; however, a 47% increase in Young's modulus parallel to the nanotube axis was noted.

Summary and Conclusions

The chemistry of SWNTs and their processing into functional nanotube-polymer and nanotube-carbon composites were explored in this project. The *highest-purity SWNTs* reported were obtained in this project, distinguishing ORNL as a leader in combined SWNT synthesis and chemistry. Chemical functionalization of the nanotubes was studied and used to enhance the solubility and dispersion of the nanotubes in polymers, resulting in significantly higher strength (storage modulus, up to a factor of 5 higher for polystyrene). Electrical properties of nanotubes in polymers were studied through the development of an electric field–induced contrast-imaging electron microscopy technique, which permitted SWNT bundles and networks to be imaged during current flow in polymers for the first time. Several processing and characterization techniques were developed at ORNL to incorporate nanotubes into polymers, including electrospinning, shear, electrophoresis, fiber extrusion, and infiltration of vertically aligned nanotube arrays.

This project resulted in a wealth of new science involving the chemistry of carbon nanotubes which will enhance other DOE programs in both basic science (DOE-DMS/BES work at the Center for Nanophase Materials Sciences at ORNL and the other DOE nanoscience centers) and applied research (e.g., DOE-EERE Hydrogen and Fuel Cells programs). The protoype multifunctional nanotube composites developed in this project were prototypes designed not only for lightweight, fuel-efficient transportation vehicles with enhanced strength for DOE but for other government agencies (especially NASA, DARPA, and other DoD agencies).

Formal presentations have been made to secure follow-on funding at DARPA, NASA-Langley, NASA-Marshall, AMCOM, and TARDEC, with approximately five white papers prepared to date. DARPA is interested in thin multifunctional nanotube-polymer skins with optoelectronic, sensing, and energy storage capabilities. DARPA, NASA, and AFRL are all interested in the thermal management offered by embedded heat pipes formed by nanotubes in space age polymers. The Army is interested in nanotube-based materials for lighter-weight tanks and vehicles, as well as the optical limiting properties nanotubes offer embedded in transparent polymers. Most government agencies (especially NASA) have well-defined timelines for progress in nanotube-based composites. Our project—through the basic science and the resulting protoype composites it enabled—directly meshed with their needs, advancing their timelines and providing the scientific basis for continued efforts in this area.

References

[1]P. F. Britt. S. Viswanathan, D. B. Geohegan, I. Ivanov, A. A. Puretzky, J. Howe, H. Schittenhelm, B. White, "Preparation of High Purity Single-Wall Carbon Nanotubes Synthesized by Pulsed Laser Vaporization" (submitted).

[2]H. Schittenhelm, D. B. Geohegan, G. E. Jellison, A. A. Puretzky, M. J. Lance, P. F. Britt, "Synthesis and Characterization of Single-Wall Carbon Nanotube–Amorphous Diamond Thin-Film Composites," *Appl. Phys. Lett.* **81**, 2097 (2002).

[3]S. Jesse, M. A. Guillorn, I. N. Ivanov, A. A. Puretkzy, J. Y. Howe , P. F. Britt, and D. B. Geohegan, "Electron Microscopy Imaging of Single-Wall Carbon Nanotube Networks in Polymers" (submitted to *Nanoletters*).

Multiscale Modeling and Simulation of the Growth and Functionalization of Nanotube Crystals, Arrays, and Polymeric Composites

R. F. Wood,[1] D. W. Noid,[2] S. Pannala,[2] B. G. Sumpter,[2] J. C. Wells,[2] and Z. Zhang[1]
[1]*Condensed Matter Sciences Division*
[2]*Computer Sciences and Mathematics Division*

Calculations on both continuum and atomistic models were carried out as part of multiscale simulations of carbon nanotube nucleation and growth. The continuum model is used to demonstrate the key roles played by the size and shape of the catalytic particle in conjunction with the concentration and temperature gradients at the gas/solid interface and in the catalytic particle. First-principles electronic structure calculations relating to the formation of single-wall carbon nanotubes were carried out. Density-functional theory in various forms was used throughout. A 38-atom Ni cluster and several low-index Ni surfaces were investigated using pseudopotentials and plane-wave expansions. The energetic ordering of the sites for carbon atom adsorption was found to be the same, with the Ni(100) facet favored. In other calculations, Gaussian orbital basis sets were used to study a cluster or "flake" containing fourteen carbon atoms. The flake is a segment of three hexagons from an "unrolled" carbon nanotube, with an armchair termination. The binding energies of C, Ni, Co, Fe, Cu, and Au atoms to it were calculated to gain insight into the mechanism for the high catalytic activity of Ni, Co, and Fe and the lack of it in Cu and Au.

Introduction

The importance of nanoscience to advanced materials development and the potential opportunities associated with it have been repeatedly emphasized during the last decade. Within the nanoscience area, carbon nanotube (CNT) research is arguably the most likely to yield dramatic near-term results if certain issues can be resolved. Nanotubes are lightweight, exceptionally strong, and have excellent thermal properties. Their incorporation into composite materials could revolutionize a wide range of advanced structural materials. Likewise their electronic properties, particularly of single-wall carbon nanotubes (SWNT), have proved to be fascinating, exhibiting electrical conductivity that varies from metallic to semiconducting to insulating, depending on the growth conditions. It has also been claimed that carbon nanotubes can be a storage medium for atomic hydrogen. Unfortunately, there are daunting roadblocks to the development and commercialization of CNT technologies. As yet, there are no methods for growing nanotubes in the quantities and with the purity needed. And once grown, the separation of various types (single wall, multiwall, aligned, etc.) is time-consuming, unreliable, and costly. Also, there is no way at this time to control the electrical properties by growing nanotubes with specified diameter, length, and chirality.

Paramount among the issues impeding the development of CNT materials is the role of the metal catalyst, which is still completely unclear even for the simplest growth technique. It is widely agreed that a much better understanding of the nucleation and growth processes is needed. Since in situ techniques for directly observing the growth are not likely to be available anytime soon, simulation of the processes at some level is almost mandatory. But the time and space scales involved cover many orders of magnitude, and the mathematical and computational aspects of such "multiscale" problems are formidable. For example, first-principles, quantum-mechanical molecular dynamics (MD) calculations extensive enough for reliable parametric studies, even on a picosecond time scale, are well beyond the capabilities of current computers. Conversely, classical interaction potentials for dissimilar atoms embedded in a complex atomic environment, which might greatly speed up the MD calculations, have proved virtually impossible to construct. Consequently, the multiscaling used to date for the type of problems addressed here is at a relatively modest level. It consists mainly of using continuum and atomistic models in a complementary fashion to extract information of most interest to the experimentalists. Our efforts in this direction are described here.

Technical Approach

The approach taken in the research had two main focal points. In the first, CNT growth was studied using a continuum model based on coupled heat and mass

transport partial differential equations. The MATLAB program with the Partial Differential Equation (PDE) toolbox was used extensively. This provided a powerful and flexible framework in which to cast the continuum modeling. Unlike previous research on the growth of carbon nanotubes and nanofibers,[1] it was possible to use a finite element method (FEM) with unstructured grids, adaptive mesh refinement (AMR), etc., that greatly expanded the range of geometries that could be considered. A schematic of the continuum model for CNT growth in the so-called tip-growth mode is shown in Fig. 1.

The geometry in the figure was chosen based on a frequently observed shape of the catalyst. It was found that slight variations in the shape do not change the predictions drastically. An earlier two-dimension (2-D) simulation[2] for carbon nanofibers employed geometry similar to that of Fig.1 but with a simple, highly structured grid. Our grid takes advantage of the 2-D space-filling characteristics of triangles that lead to very efficient numerical integration algorithms. This unstructured, adaptive grid is optimized simply by changing the size and shape of the triangular elements over the computational domain, as needed for convergence. In fact, for steady-state calculations, the AMR feature of the computer program can be used to refine the grid based on a variety of criteria.

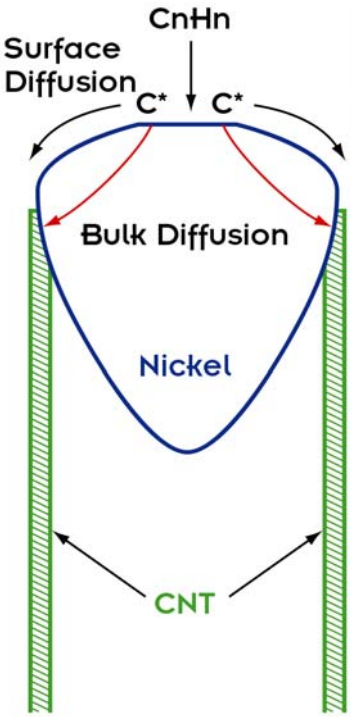


Fig. 1. Carbon nanotuibe growth via the tip growth mode. For a nanotube diameter of 1–2 nm, the catalytic particle is in the mesoscopic regime.

The other focal point included first-principles, atomistic calculations of several types of structures thought to be relevant to CNT formation. All of these calculations utilized density functional theory (DFT) as implemented in a number of readily available computer codes. In one set of calculations, pseudopotentials and plane wave expansions of the wave function were used, while in another, expansion of the wave function was carried out in Gaussian basis sets. In the latter approach, the NW Chem suite of programs developed at Pacific Northwest National Laboratory was employed.

Results and Accomplishments

Mesoscopic Continuum Model

Previous continuum modeling efforts in ref. 1 were directed to the growth of carbon filaments and fibers rather than single and multiwall nanotubes. In carbon filament growth, the carbon diffusion through the rather large catalytic particle is usually the rate-limiting step, and thus those studies cannot be extrapolated simply to the CNT growth from quite small particles. Also, using the latest advances in computational methods mentioned above, one can now simulate more complicated shapes and sizes of catalyst particles very efficiently. Later, the complexity of the calculations can readily be expanded to include interfacial dynamics.

Our starting point for these calculations was a set of coupled equations for heat and mass diffusion.[3] After considerable simplification of the diffusion equation, and with an assumption that steady-state growth has been reached, the mass diffusion equation can be reduced to

$$\frac{\partial}{\partial r}\left(rD(T)\frac{\partial Y_c}{\partial r}\right)+\frac{\partial}{\partial z}\left(rD(T)\frac{\partial Y_c}{\partial z}\right)=0\ ,$$

where Y_c is the carbon mass fraction, r the radial coordinate, z the axial coordinate, D the diffusion coefficient, and T the temperature. In principle, D is a function of both T and Y_c but since experimental data is not available for its dependence on the C concentration, only the T-dependence was considered. It was assumed to have the classical form $D(T) \sim \exp(-E/kT)$, in which E is an activation energy.

The above equation was solved with appropriate boundary conditions (BC). On the top surface (C_nH_n/Ni interface) in Fig. 1, a Dirichlet BC with a specified inlet carbon mass fraction was assigned. At the location where the nanotube is formed (CNT/Ni interface), another Dirichlet BC was specified, with Y_c corresponding to a carbon activity of unity. The outer radius of the nanotube

corresponds to that of the Ni particle, and an inner radius is determined by requiring the wall thickness to be 0.14 nm. On the rest of the Ni particle boundary, a Neumann BC with zero flux was specified.

The calculations show that as time proceeds carbon diffuses into the catalyst particle and the nanotube precipitates out when enough is supplied to the interface where the tube is forming. The concentration in the rest of the particle almost reaches that at the top surface. Thus, the entire particle acts like a reservoir for carbon, part of which diffuses to and is incorporated into the nanotube, while the deficit amount is replenished at the inlet source. This scenario, obtained from the steady-state solution, is also realized by running the transient simulation for extended times.

In the current simulations, three parameters were varied: inlet mass fraction, temperature of the catalyst particle, and the diameter of the Ni particle. The growth rate of the CNT was found, not unexpectedly, to have a linear dependence on the inlet C mass fraction, as shown in Fig. 2(a). This suggests that one way to increase the growth rates is to have surface conditions suitable to maintain higher concentrations.

The dependence of growth rate on particle diameter, Φ, is such that it decreases as the particle size increases. At steady state it is evident that the inlet carbon concentration drives the gradients that exist in the catalytic particle and thus controls the growth rate. The amount of carbon diffused through the top of the catalyst depends on the surface area. From this surface, the carbon diffuses to fill the entire volume of the particle. The surface area varies as the square of the particle diameter, whereas the volume varies as the cube. Consequently, the carbon concentration in the catalyst particle is inversely proportional to Φ. In addition, for the same carbon concentration at the inlet surface, the gradient varies inversely as Φ. This explains the inverse quadratic dependence of growth rate on particle diameter and also the fast growth rates of small diameter SWNTs observed in some experiments.

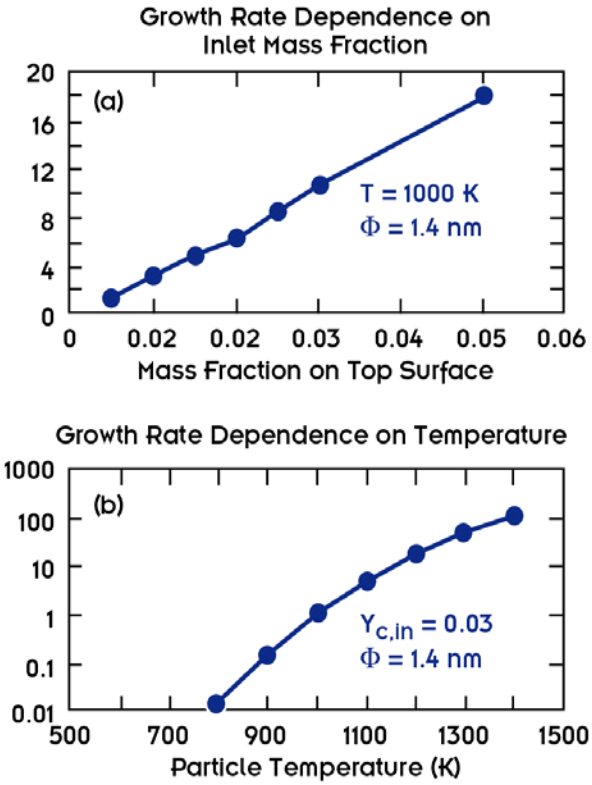


Fig. 2. Growth rate dependence on (a) inlet carbon mass fraction for T = 1000K and CNT diameter Φ = 1.4 nm; and (b) catalyst temperature for inlet carbon mass fraction of 0.03.

Because of the exponential dependence of $D(T)$ on the inverse temperature, the carbon diffusion in the Ni particle has an exponential dependence on temperature and thus the growth rate, which is directly proportional to diffusion, has the expected exponential behavior as shown in Fig. 2(b). The growth rates in these numerical simulations are at least an order of magnitude different from those reported in some experiments where the rate was measured indirectly. In very recent experiments, Geohegan et al.[4] at ORNL have measured growth rates directly by optical interference techniques. Their rates are in much better agreement with the ones shown here.

Perhaps the single most important result of the continuum calculation was one that demonstrated clearly how the number of walls in the CNT is related to the size of the catalytic particle and the concentration of carbon at the top surface. This is important because there is a concerted effort to grow CNTs with only predetermined radii, and it has been unclear how to do it experimentally. We are working closely with the experimentalists in the Condensed Matter Sciences Division on this problem.

Electronic Structure Calculations

An important common feature in the synthesis of single-walled carbon nanotubes is that small clusters of certain transition metals (TM) such as Fe, Co, and Ni are found to be essential to high-yield SWNT formation. The morphologies of SWNTs are directly related to the catalyst particle size. To understand the fundamentals of CNT growth, it is worthwhile to determine which facet is the preferred adsorption site for a carbon atom on TM clusters. Here adsorption energies for carbon nucleation on Ni surfaces were computed by first-principles DFT methods.[5] More specifically, the adsorption of one carbon atom on a 38-atom nickel cluster and on several low-index nickel surfaces was studied.

The results showed that the (100) site is favored for carbon adsorption. The energetic ordering of the sites for carbpm atom adsorption on both Ni_{38} cluster and the

extended nickel surfaces is exactly the same. These computational results are qualitatively consistent with experimental results showing that binding of carbon to nickel clusters saturates quickly with cluster size ($\sim n = 5$) to bulk values. Our DFT results "over predict" the binding energies by approximately 50%. Nevertheless, the experimental results do not yield the relative site energies and these are felt by some to be important in nanotube nucleation and growth.

In another series of calculations,[6] we considered small, flattened "flakes" that may be viewed as having been cut from an unrolled nanotube. The flakes are two-dimensional, and none of the strain energy due to rolling was included in the calculations. Initially, we followed Lee et al.[7] and chose a flake consisting of five pure carbon hexagons. After establishing that our results agreed well with those of ref. 7, in which similar computational approaches were employed, the flake size was cut down to three hexagons (^{14}C atoms) so that a large number of calculations could be carried out on an accelerated basis. The agreement between results for the five and three hexagon calculations was excellent. The main reason for this agreement is probably that the carbon bonds, as in many organic systems, are quite well localized and therefore additive to a good approximation.

The initial set of calculations consisted of adding C, Cu, Au, Ni, Fe, and Co to form a closed pentagon. The main purpose of these calculations was to determine if there are any clear differences between metals that are good catalysts and those that are not. The binding energies, E_b, of these elements to the carbon flake fell into two well-resolved energy ranges, with E_b for Cu and Au about an eV less than for the other elements. These differences are too large and consistent with each other to be the result of the computational approximations that were made. Consequently, we view them as a reliable guide to the relative interaction energies of the metal atoms with the carbon flakes.

To take these considerations a step further, calculations were carried out to determine if a carbon atom and a metal atom on neighboring pentagon sites preferred to remain on those sites, as in Fig. 3(a), or to combine to form a mixed hexagon site and an empty pentagon site, as illustrated in Fig. 3(b). It was found that Cu and Au atoms preferred to remain on the separated pentagon sites, while Ni, Fe, and C preferred to form hexagons. This result can be used to substantiate the elementary catalytic step proposed in ref. 7, but space precludes our discussion of this here.

Summary and Conclusions

As part of a focused computational effort on the simulation of CNT formation, we have developed and applied computer programs for coupled heat and mass flow in one and two dimensions. The results clearly illustrate the crucial role carbon diffusion plays in the growth of nanotubes, and they explain a number of the experimentally observed growth phenomena. They also demonstrate the key roles played by the size and shape of the catalytic particle in conjunction with the concentration gradients at the gas/solid interface. While it is apparent that an impressive amount of insight can be obtained from continuum calculations, the need to go beyond this simple approach is apparent. We want to incorporate more effectively the input from first-principles electronic structure calculations and, working with experimentalists, to focus on the most reliable data obtained under well-defined conditions.

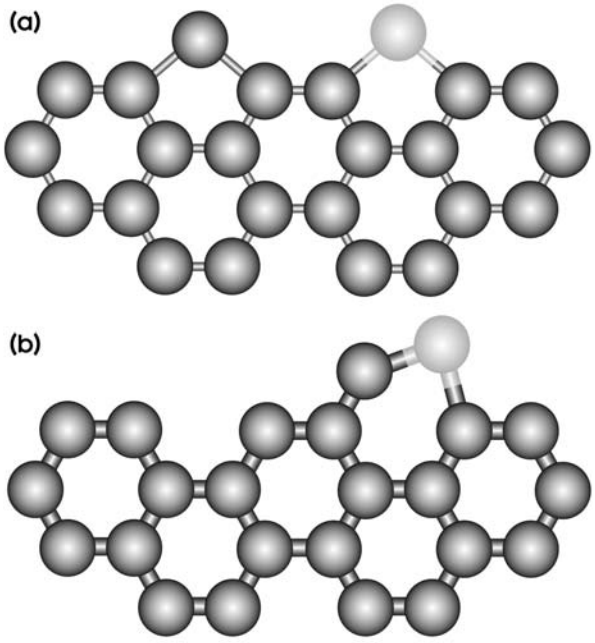


Fig. 3. Two separated pentagons, one with a carbon atom and one with a metal atom, as shown in panel (a), may combine to form the distorted hexagon in panel (b) if the total energy can be lowered in that way. Light circles are the metal atoms.

The binding of a carbon atom on various nickel surface and cluster facets was calculated. Small clusters yield the same results as extended surfaces. The Ni (100) facet is favored over the Ni (110) and Ni (111) facets. Electronic structure calculations of the binding energy of various metal atoms to a 14-atom carbon flake show that Au and Cu are less tightly bound than Ni, Co, and Fe by about 1 eV. Finally, comparison of the binding of a carbon and metal atom in a hexagon to the energy in separated pentagons indicates that for the catalytic metals the hexagon is stable against the breakup into two pentagons.

Preliminary results of the work described here were presented at the 2003 NASA/Rice Workshop on Single Wall Carbon Nanotube Growth Mechanisms. The

presentations were well received and have begun to establish ORNL as an important contributor to computational CNT research. The work on the continuum modeling has progressed to the point where significant concrete guidance can be given to the experimental efforts on CNT growth at ORNL. With continued development of the modeling and simulations, identification of the elementary catalytic step can be expected and with that a much deeper insight into how to control the growth conditions.

A white paper seeking follow-on funding was submitted to DARPA under one of their BAA solicitations. Submission of a full proposal was not encouraged, probably because of a weakness in the proposal not directly connected with the work reported here. The most promising avenue for follow-on funding is one that closely links the modeling to the experimental CNT effort at the CMSD at ORNL. Meanwhile, work on the modeling continues on a reduced level via funding from DOE's Office of Basic Energy Sciences.

References

[1]J. Alstrup, *Catal*. **109**, 241 (1988); R. T. K. Baker, *Carbon* **27**, 315 (1989); S. A. Safvi, E. C. Bianchini, and C. R. F. Lund, *Carbon* **29**, 1245–1250 (1991); R. T. K. Baker, *Encyclopedia of Materials: Science and Technology*, Elseiver Science Ltd., 932–941 (2001).

[2]P. Chitrapu, C. R. F. Lund, and J. A. Tsamopoulos, *Carbon* **30**, 285 (1992).

[3]S. Pannala and R. F. Wood, Proceedings of the NASA/Rice Workshop on Single Wall Carbon Nanotube Growth Mechanisms, *J. of Nanoscience and Nanotechnology* (in press).

[4]D. B. Geohegan, A. A. Puretzky, I. N. Ivanov, S. Jesse, G. Eres, and J. Y. Howe, *J. Appl. Phys*. (submitted).

[5]Q. M. Zhang, J. C. Wells, X.-G. Gong, and Z. Zhang, *Phys. Rev. B* (submitted 2003).

[6]J. C. Wells, D. W. Noid, B. G. Sumpter, R. F. Wood, and Q. Zhang, Proceedings of the NASA/Rice Workshop on Single Wall Carbon Nanotube Growth Mechanisms, *J. of Nanoscience and Nanotechnology* (in press).

[7]Y. H. Lee, S. G. Kim, and D. Tománek, *Phys, Rev. Lett.* **78**, 2393 (1997).

Development of Time-of-Flight Capabilities for Studies of Inelastic Neutron Scattering and the Dynamics of Soft Matter

P. D. Butler, W. A. Hamilton, and G. D. Wignall
Condensed Matter Sciences Division

Cold neutrons are a critical research tool for the investigation of scientific phenomena occurring on long (~ 5–1000 Å) length scales, and have been particularly successful for studies of "soft" matter such as polymers, complex fluids, and biological complexes. Over the past two decades, small-angle neutron scattering (SANS) has proven to be the most important technique for the study of such materials, revealing their structure on the nanoscale level, and it is well known that SANS instruments are the most over-subscribed facilities worldwide, attracting users from industry and government laboratories in addition to academic researchers. More recently, cold inelastic neutron scattering instruments have begun to come on line in this county to study the dynamics of these systems and are showing considerable promise. With the installation of a new cold source at the High Flux Isotope Reactor, this project's aim was to understand how to separate this dynamic information that must, in some cases, compromise regular SANS measurements, examining where that might be problematic, and perhaps more importantly to explore how SANS might be able to offer unique dynamical information, particularly for soft matter.

Introduction

To date, the treatment of SANS data has been based on the assumption that the scattering is predominantly elastic (i.e., the energies of the incident and scattered neutrons are the same) and the data may be integrated over all energies to give the time-averaged structure of the system. While this assumption is known to be true in the $Q \rightarrow 0$ limit, nearly all SANS experiments are conducted in the range $10^{-3} < Q < 1$ Å^{-1}, where the assumption of elastic scattering remains unverified. Moreover, those few studies in which energy analysis of the scattered beam has been performed indicate that this is not always a safe assumption. In 1999, for example, Rennie and coworkers found that in a number of common solvents (like water or toluene), inelastic SANS effects can be significant with *less than half* the 12-Å neutrons being scattered elastically as conventional SANS methodology assumes.[1] However, the motions on the nanoscale, which give rise to these inelastically scattered neutrons (i.e., scattered neutrons whose energy is different from that of the incident neutrons), hold important information about the dynamics (motions on the nanoscale) in the materials. Thus the goal of this project is to not only understand how inelastic "contaminations" may in some circumstances adversely affect our interpretation of SANS data and how to conduct experiments to avoid those pitfalls but also to open up new areas of physics by studying the inelastic component itself. In the process of working on developing the inelastic SANS technique, we identified several issues with the new class of detectors which ORNL has acquired for its SANS instruments. One of these has already led to the manufacturer developing an upgrade, while another is an ongoing concern both for the High Flux Isotope Reactor (HFIR) and spallation sources such as the Spallation Neutron Source (SNS). With ORNL building two best-in-class SANS machines, to rival the current best in Grenoble, France, the timely delineation of these critical issues will be invaluable to the success of the HFIR SANS program. As part of this project, we also developed a novel approach to dynamics using time-resolved SANS, which is discussed separately below.

Technical Approach

Although ORNL is building two state-of-the-art SANS machines, it does not currently have an operating SANS instrument. Thus all experiments were performed on the National Institute for Standards Technology (NIST) Center for Neutron Research (NCNR) SANS instrumentation in Gaithersburg, Maryland. The NCNR's SANS detectors are smaller than those which ORNL will use but are electronically identical. This new class of detectors provides an integrated data acquisition package. For historical reasons, NIST currently bypasses this, using instead its own acquisition electronics. Those electronics are inadequate for the purpose of inelastic measurements so that one of the requirements of this work was the implementation at NIST of this new integrated acquisition package, with its new time-stamping capabilities, to match

the anticipated ORNL configuration. The NCNR's aging electronics are obsolete, and in fact some critical components are no longer available. Thus the NCNR will eventually need to convert its electronics, making this a natural interagency collaboration.

Inelastic SANS

For purposes of this development program, we used simple optical choppers placed at the sample position to provide a time = 0 (T0) stamp on the neutrons. By measuring the arrival time of each neutron relative to that T0 and knowing the distance from the chopper to the detector, the energy of each neutron can be computed. The energy gain or loss can then be determined from the known incoming neutron energy provided by the upstream velocity selector. Only those neutrons arriving with zero energy change are elastically scattered, while all others are inelastically scattered. In order to provide the necessary pulse of neutrons, the optical chopper wheel was covered with cadmium (Fig. 1) containing two 2-mm-wide slits diametrically across from each other. In order to obtain reasonable time resolution, the 1.27-cm-diameter incoming beam was masked by cadmium containing a 2.5-mm-wide slit, making the incident beam 2.5 mm × 12.5 mm. The chopper was usually run at 40 Hz for a chopped beam of 80 Hz, providing a neutron-pulse width of roughly 0.3 ms FWHM. The chopper output timing signal was conditioned by feeding it through a pulse delay generator prior to sending it to the T0 input of the detector (Fig. 2). A routine implementation on a SANS instrument would provide a chopper package optimized for the geometry of the problem along with appropriate shielding. The instrument design implications are that sufficient space be available at the sample position for the insertion of such a device.

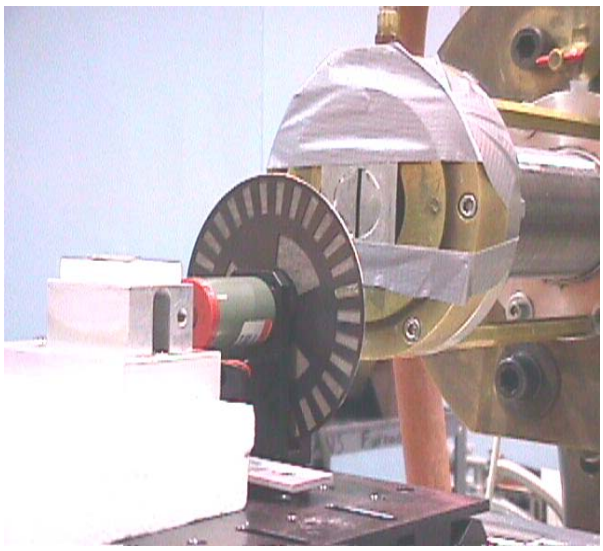


Fig. 1. Chopper and beam definition setup used for the inelastic measurements.

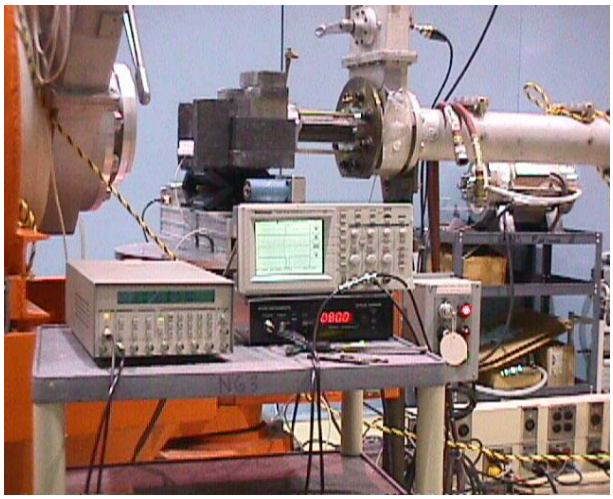


Fig. 2. Beam line operating with the chopper and signal conditioning electronics during the inelastic SANS run producing data such as shown in Fig. 3.

Time-Resolved SANS (T-SANS)

In order to do time-resolved SANS, the scattering signal must be time multiplexed into a series of histograms. The multi-time-binned SANS data collection is then cycled to build up statistics, synchronized at constant delay with respect to some experimental signal, in this case the stopping signal of our Couette shear cell motor. Given adequate electronics, the time resolution limit will be determined by the neutron time of flight and the wavelength spread in the incident beam and can extend down to hundredths of seconds, though duty cycles will usually be a more stringent limitation. It should be noted that our SANS direct neutron energy analysis experiments synchronize on a faster chopper phase signal, but in most respects the data collection issues of the two experiments are virtually identical so that lessons learned about data acquisition strategies from one set of measurements applies fairly directly to both techniques.

Time-resolved measurements have never been done with any significant resolution (i.e., more than 16 histograms) in the United States. Ideally, the new integrated acquisition system would provide nearly infinite histograms. However, as discussed below, implementation of the time-resolved modes of this acquisition package presented several challenges. Thus, in order to perform these measurements in a timely fashion, we opted for the creative approach of modifying the older, established electronics and acquisition software used at NIST. Specifically, the 128 × 128 detector pixel array of the two-

dimensional NIST SANS detector was rebinned in hardware into a 32×32 array, thus swapping spatial resolution for time resolution, giving us 48-constant-bin-width time channels. On the software side, we altered the indexing of the data arrays read into the live image display so as to be able to monitor our experiment in real time. In order to analyze the data, however, we opted to modify the ORNL SPECTOR software developed by John Hayter and Bill Hamilton for the old SANS at HFIR, as it is probably still one of the most versatile and well-designed acquisition/reduction packages available.

Results and Accomplishments

Acquisition electronics implementation

While this was simply a necessary by-product of the work, it produced several immediate benefits to the ORNL SANS program and so is described briefly here. The first obstacle encountered was the fact that the integrated acquisition system did not allow for external control (by, for example, an instrument computer), the programming assumption apparently being that the instrument would be controlled by the detector. Having identified the problem, we teamed with NIST to insist to the manufacturer that neither of us could operate in this mode. The availability of a live beamline at NIST to test the solutions allowed us to realize fairly quickly that the first simple solution was not robust enough for heavy user operation. While the final solution is still being tuned, the immediate benefit to ORNL and NIST is clear.

The discovery that these new detectors have a single input signal, programmed as a T0 reset signal was a second issue uncovered by this portion of the program. In this case, the NCNR's needs were more immediate and a solution was offered whereby this signal could be converted to an inhibit (or gate) input signal. This gate signal is normally used during routine SANS operation. However, time-of- flight applications, such as required for inelastic measurements, require the initial T0 signal, preferably with a gate signal also available. Converting the single input from gate to T0 and back is tedious and of course never provides both signals simultaneously. While we eventually took data with only a T0 input signal, future, more sophisticated implementations, both at HFIR and at NIST, will need both. Furthermore, spallation-source SANS instruments will probably want to use both signals routinely. Again working with NIST and the manufacturer, an upgrade board for these detectors has just recently been released which provides four inputs, two of which come factory programmed, one as a T0 and the other as a gate. ORNL has already purchased the upgrade for both its detectors, while NIST's most recently ordered detector was just delivered with the new board, making this another clear immediate benefit from this work.

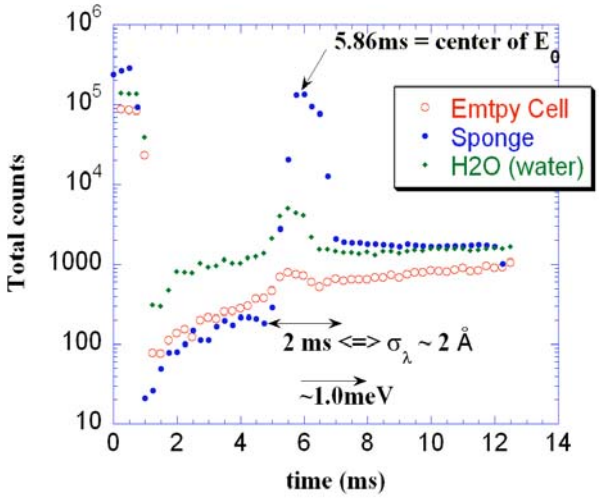


Fig. 3. Time-of-flight spectrum for three samples. The data represents the integrated counts overall the detector at each time. The filled circles are for a surfactant sponge sample, the filled diamonds a hydrogenous water sample, and the open circles an empty quartz cell. The incident wavelength used, λ, was 6 Å with a wavelength spread, λ/Δλ, of 10%. The effective time resolution is of the order of 2ms, making the wavelength resolution roughly 2 Å. The large central peak is the elastic signal. The energy resolution at the elastic peak is roufhly 1 meV. Note the abnormal behavior at short times.

Inelastic SANS

Figure 3 shows a time-of-flight spectrum of the whole detector integrated intensity (making this a low-Q resolution plot) for three different samples. The surfactant sponge sample was one of those used in the time-resolved SANS experiments to measure the energy of formation of passages and was hoped to show some evidence of the dynamic equilibrium of creation and disintegration of passages. The water sample is measured because it is known to have significant inelastic scattering and so serves as a test sample, while the empty quartz cell is expected to be a pure elastic scatterer to be used as a background. The incident wavelength used, λ, was 6 Å with a wavelength spread, λ/Δλ, of 10%.

The resolution of the measurement indicates roughly 1 meV. This can be improved with the use of a second counter-rotating chopper placed upstream to lower the effective wavelength spread. However, as can be seen, the data at higher energies than the elastic peak behave very oddly indeed. Unfortunately, due to the low energy of the cold neutrons, the energy gain side is the most likely place to see effects. It appears that the first 1 ms of data, regardless of configuration or time resolution used, contains a signal spatially identical to that in the elastic time bins, with a constant high intensity not related to the intensity of the elastic scattering. This seems to be somehow compensated for by having the time bins above 1 ms but below the elastic peak drop nearly to zero and

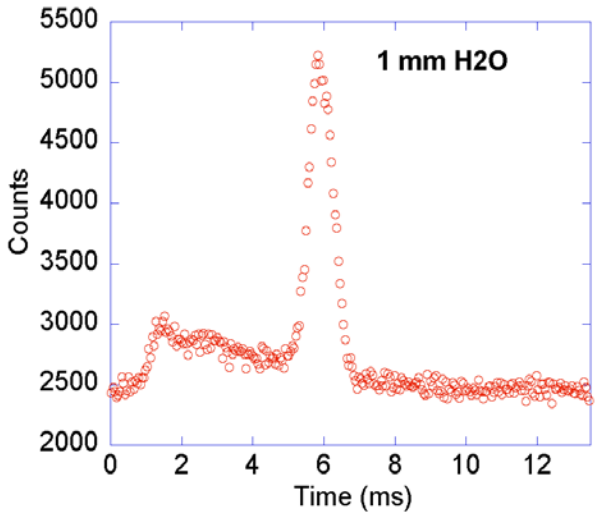


Fig. 4. Plot of the detector anode signal as a function of TOF for the same hydrogenous water sample as in Fig. 3. Note the difference in the low time behavior, with this data exhibiting the expected peak at short times.

come up only slowly. Figure 4 shows the water signal measured simultaneously by running the detector anode signal through a multichannel scalar. The resulting spectrum is exactly as expected with no intensity at 0 time (infinite energy) and significant inelastic scattering between 2 and 6 ms—with at least 2 peaks being easily discernable. This gives us reason to believe that the problem lies not in the detector itself but in the encoding done within the electronics. This issue has yet to be resolved. However, we note that this problem is not only important to ORNL SANS being able to provide inelastic capabilities, but it will also prevent us from doing T-SANS with this acquisition package, and of course precludes the use of the current integrated electronics with any spallation source such as the SNS SANS.

Time-Resolved SANS

Whereas normal SANS reveals the structures within materials, inelastic and quasi-elastic experiments reveal information about the dynamics within these materials. This is usually accomplished by measuring shifts in the scattered wavelength at each Q. The wavelength change reveals the energy involved, or equivalently the time scales, while the Q position speaks to the length scale of the motion. For much longer time scales, depending on the wavelength spread and distance from sample to detector (giving an effective TOF resolution), it is possible to measure some motions down to the hundreds of milliseconds resolution directly in the time domain as discussed previously. While a handful of such measurements of been reported on SANS machines worldwide, we are the first to relate the time sequence of structures back to an energy. Here we perform a time-resolved relaxation study of a shear-induced smectic surfactant membrane phase (L_α) transforming to its thermodynamic equilibrium sponge state (L_3), as shown in Fig. 5 with a time resolution of 0.2 s. Comparison of the T-SANS relaxation measurements with dynamic light scattering measurements of the pure diffusive membrane mode frequencies in this system (see figure) allowed us to determine the inelastic activation energy for the formation of the interconnecting "handles" which characterize the topological rearrangement of membranes required between these two states. This was in fact the first direct determination of the energy of this membrane fusion mode and the result, $E_A \approx 8k_BT$, was in excellent agreement with theoretical predictions for this system.[2,3]

Despite the rather rapid time sequencing used here, the motions in the normal sponge systems are still much too fast for these measurements. In order to perform this study, we therefore also employed the novel approach of using sugar as a viscosity enhancer, slowing the dynamics into an experimentally accessible region. This technique of adding sugar to slow the dynamics has now already been used by our Australian collaborator's Ph.D student working with gel systems while we are exploring it as an avenue to understand microemulsion behavior.

Summary and Conclusions

The modest progress in performing true inelastic measurements, with results showing better than 1 meV resolution, has generated significant interest both at ORNL and at NIST, and even other at facilities for that matter, and it is highly likely that this effort will be picked up by several groups, including our own. Meanwhile, the slow progress on getting the integrated acquisition package up

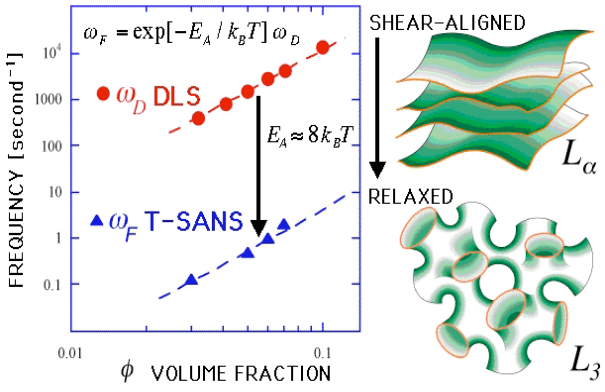


Fig. 5. Frequencies measured by both light scattering and T-SANS over a range of membrane concentration in a surfactant sponge phase and how the two are related through the energy of formation of handles which is thus determined.

and running on a NIST instrument provided both time to explore T-SANS as a new approach to SANS dynamics, which appears to hold tremendous promise, and vital information on the performance characteristics of the integrated acquisition system of the new class of detectors, which has already impacted the ORNL and NIST SANS programs as well as future SANS instrumentation worldwide, particularly at spallation sources.

The T-SANS work, along with its creative use of sugar to slow the dynamics of complex systems, is perhaps the most exciting aspect of this work, and again has also already had an impact on other programs. The results are clearly impressive. Further, the combination of T-SANS with TISANE (time-integrated small-angle neutron experiments, invented at ILL and being developed at NIST) and true inelastic SANS should allow a truly wide range of time and energy scales to be probed.

The main obstacle to general use of T-SANS now, and eventually of inelastic SANS, is the lack of analysis tools for handling such data. In order to make these techniques available and allow them to have the impact they should will require a significant software effort. One of us (P.D. Butler) is currently the principal investigator on an NSF subproject proposal (submitted) to produce state-of-the-art SANS analysis software.

References

[1] R. E. Ghosh and A. R. Rennie "Assessment of Detector Calibration Materials for SANS Experiments," *J. App Cryst* **32**, 1157 (1999)

[2] L. Porcar, W. A. Hamilton, P.D. Butler, and G.G. Warr, "Relaxation of a Shear-Induced Lamellar Phase Measured with Time Resolved Small Angle Neutron Scattering," *Physica B* (in press).

[3] L. Porcar, W. A. Hamilton, P. D. Butler, and G.G. Warr, "Topological Relaxation of a Shear-Induced Lamellar Phase to Sponge Equilibrium and the Energetics of Membrane Fusion," manuscript in preparation.

Selective Catalytic Oxidation of Hydrogen Sulfide

T. R. Armstrong,[1] N. C. Gallego,[1] S. H. Overbury,[2] W. A. Shelton,[3] X. Wu,[1] and A. K. Kercher[1]
[1]*Metals and Ceramics Division*
[2]*Chemical Sciences Division*
[3]*Computer Science and Mathematics Division*

Processing gas streams that produce hydrogen from fossil fuels commonly contain impurities, such as hydrogen sulfide (H_2S), which can poison shift catalysts and fuel cell electrodes if not removed. The removal of sulfur in these feedstocks by applying advanced desulfurization technologies is desired. The mission of this project was to evaluate one very promising technology, selective oxidation of H_2S to elemental sulfur on an activated carbon catalyst. The objective was to develop a fundamental understanding of the H_2S oxidation process on activated carbon catalysts, while further developing an understanding of the role of microstructure and surface chemistry on the catalytic process.

Introduction

Hydrogen is commonly produced by reforming of natural gas; however, reformate contains impurities, such as hydrogen sulfide (H_2S), which can poison shift catalysts and fuel cell electrodes. Even a low concentration (a few ppm) of hydrogen sulfide dramatically shortens the life of fuel cell catalysts. Major ongoing research efforts are seeking to develop fuel cell catalysts that are more sulfur tolerant. The mission of this research program was to develop an oxidative process to reduce sulfur levels to the parts-per-billion range using low-cost carbon-based catalysts.

Catalytic partial oxidation of H_2S using activated carbon as catalyst has been identified as a highly promising approach for reducing sulfur (H_2S) from well gas, fuel cell feedstocks, and hydrodesulfurizer streams down to ppb level. While activated carbons are commercially used for hydrogen sulfide removal from gas streams, the key mechanisms for H_2S removal and the critical features in an activated carbon catalyst are not well understood. Further, current removal strategies are strictly limited to adsorption of the sulfur species on the carbon followed by replacement and disposal of the spent carbon. Research efforts often have relied solely on commercial activated carbons, and none of the past development efforts have evaluated the use of carbon as a catalyst. In addition, the results reported by various investigators are often diverse and contradictory, especially concerning kinetics and reaction pathways. This project, therefore, aimed to develop optimized schemes for producing activated carbon catalysts by materials processing, molecular modeling, reaction mechanism studies, and engineering evaluation and to give new insights for the reaction mechanisms for the oxidation process in question.

Technical Approach and Results

The assembled research team allows for a processing-to-product approach to tackle the timely issue of hydrogen sulfide removal from reformate feedstocks of hydrogen. The feedback between materials processing, molecular modeling, reaction mechanism studies, and engineering evaluation provides a synergistic approach for the development of an activated carbon catalyst.

Different precursors and variations in activation conditions can produce dramatically different microstructures in carbon materials. A focused effort to identify their relationships and the influence of the resultant pore structure (pore volume, pore size, and distribution) on the performance of carbon catalyst was necessary. Surface functional groups play an important role in the catalytic activity of the catalyst; thus, it is important to identify the initial surface functional groups and then develop techniques to modify or alter surface functionality. To produce an optimal activated carbon catalyst, the functional group content should be tailored to optimize hydrogen sulfide adsorption and minimize side reactions, such as adsorption of CO and H_2O, that could result in the formation of release sequestered hydrogen sulfide products.

Inorganic impurities have also been shown to strongly catalyze hydrogen sulfide oxidation. In order to understand the influence of catalytic impurities on the reactions of interest, highly pure carbons will need to be synthesized to isolate their effects. Inorganic impurities must also be added to activated carbons of in order to elucidate their role on adsorption and catalysis. It is commonly accepted that metal impurities in activated carbons play important roles in catalytic performance. Key issues that must be clarified include (1) identification of the reaction mechanisms associated with specific impurities, (2) the

activity and selectivity of those impurities, and (3) the efficacy of these impurities after extensive sulfur deposition and catalyst regeneration.

Reaction mechanisms were determined using a combination of catalytic reactor experiments on simple gas mixtures, in situ analytical techniques (such as DRIFTS), and molecular modeling. Together, these techniques can isolate individual gas-carbon reactions, explore competition between gases for reaction sites, determine the mechanisms for deleterious side reactions, and predict ideal carbon structures for H_2S removal. This combination of analysis methods provided multiple perspectives of catalyst reactions, ranging from idealized thermodynamic reaction predictions to measured catalyst surface interactions to an overall perspective of downstream gas composition. Key reaction parameters included temperature, pressure, oxygen concentration, H_2S concentration, water content, and the concentration of other impurities.

Ultimately, an activated carbon catalyst is only successful for hydrogen sulfide removal if it can perform to required concentrations in an actual reformate gas stream for a reasonable time period using a practical reactor geometry. A catalyst must remove and retain hydrogen sulfide in the presence of the wide range of impurities found in reformate feedstocks. Optimized activated carbon catalysts need to be tested on simulated reformate feedstocks of hydrogen. The potential of promising catalyst candidates for regeneration also needs to be explored. Partnering with laboratories and industry will be explored as a means to evaluate down selected catalysts on actual reformed gas streams.

The selective catalytic oxidation process can occur both continuously (i.e., simultaneous adsorption and oxidation of H_2S to produce S_n) or in a batch mode whereby the reaction is separated into adsorption step and a separate oxidation step. Thermodynamic calculations using a Gibbs minimization approach have been carried out to determine the stability of possible side reactions to for SO_2 and COS. The thermodynamic analysis of the reaction between hydrogen sulfide–contaminated mixtures of air or CO is complicated by the possible formation of numerous species in the gas phase, as well as elemental sulfur. The gaseous species include S_n (n = 1,…8), SO_2, and COS, in addition to the sulfur-free oxidation production H_2O and CO_2. With the inclusion of reactants, a total of eighteen species can be assumed to be present. For the continuous process, reaction of a reformat stream with CO and 1 wt % H_2S with the stoichiometric amount of oxygen required by the reaction

$$H_2S + \frac{1}{2}O_2 \rightarrow H_2O + S_n$$

results in incomplete oxidation. The degree of oxidation obtained in practice could be higher if consumption of oxygen in the formation of CO_2 were kinetically inhibited. The predominant reaction products are COS and SO_2. For the batch process, these calculations have shown that no SO_2 or COS will form below 150°C when CO is not present. Calculations with CO present still need to be done.

The porosity and carbon structure of several commercial and laboratory-created activated carbons were determined. A catalytic oxidation system, which can be used to test the activity and selectivity of carbon catalysts at different experimental conditions by using a FTIR to monitor effluent gas concentrations, was constructed and tested. Ongoing tests on activated carbons focus on effect of temperature (100–200°C) and gas composition (CO, CO_2, H_2O). A Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS) reactor, which is capable of in situ measurement at reaction temperature, was used to identify the change of carbon surface chemistry in various gas atmospheres. Ongoing studies on the interaction of CO or H_2S with activated carbon have provided positive support for this approach. The results will be very helpful for interpreting the reaction mechanisms of H_2S oxidation and undesired side reactions.

The porosity and carbon structure of several commercial and laboratory-created activated carbons have been determined. Typical results are listed in Table 1 and Table 2. The activated carbons are derived from different precursors (pitch-derived carbon, coconut shell carbon,

Table 1. XRD parameters of assorted commercial and laboratory-activated carbons

	Precursor	Interplanar spacing (nm)	Crystallite height (nm)	Crystallite diameter (nm)
Anshan 1000 m^2/g	pitch fiber	0.3667	1.127	1.775
Anshan 1500 m^2/g	pitch fiber	0.3660	1.134	1.803
Calgon Centaur 4x6	coal	0.3574	1.157	1.813
Calgon OLC 12x40	coconut shell	0.3662	1.080	1.810
Calgon OVC 4x10	coconut shell	0.3764	1.107	1.847
ORNL SMM-19	carbon fiber monolith	0.3742	1.058	1.757
ORNL SMS-48	carbon fiber monolith	0.3743	1.081	1.736
PICA G55C-1	coconut shell	0.3737	1.037	1.821
PICA VA507	coconut shell	0.3767	1.053	1.886
Westvaco Ultramicroporous	wood	–	–	1.720

Table 2. Pore structure characterization of assorted commercial and laboratory-activated carbons from nitrogen adsorption analysis

	BET surface area (m^2/g)	D-R micropore volume (cc/g)	D-A average pore diameter (Å)
Anshan 1500 m^2/g	1393	0.756	22.3
Calgon Centaur 4x6	1144	0.506	21.3
Calgon OLC 12x40	1110	0.506	21.3
Calgon OVC 4x10	884	0.391	19.7
ORNL SMM-19	1730	0.895	22.8
ORNL SMS-48	1404	0.619	20.7

coal, and low-ash charcoal) in order to determine the effects of common inorganic impurities. Also, preliminary experiments sought to accurately determine their functional group content.

An in situ study has been initiated to understand the interaction of gas species such as H_2S, O_2, and CO with carbon surface and interpret the reaction mechanisms of H_2S oxidation and undesired side reactions for the formation of COS and SO_2. A DRIFTS reactor which is capable of in situ measurement up to 500°C was used to identify the change of carbon surface chemistry in various gas atmospheres. A preliminary study on the interaction of CO with activated carbon has provided positive support for this approach, and typical results are shown in Fig. 1

Reverse Monte Carlo modeling has been applied to diffraction data to produce a theoretical activated carbon microstructure. By introducing select modifications to theoretical microstructures, carbon microstructural variables can be kept constant except for a specific feature of interest (functional group content, pore size, carbon structural order, etc.), and the effect of specific feature changes on the thermodynamics and kinetics of H_2S removal can be predicted. Slight modifications of these theoretical microstructures will later be used to model the behavior of actual activated carbons, by incorporating the measured microstructural features of activated carbons.

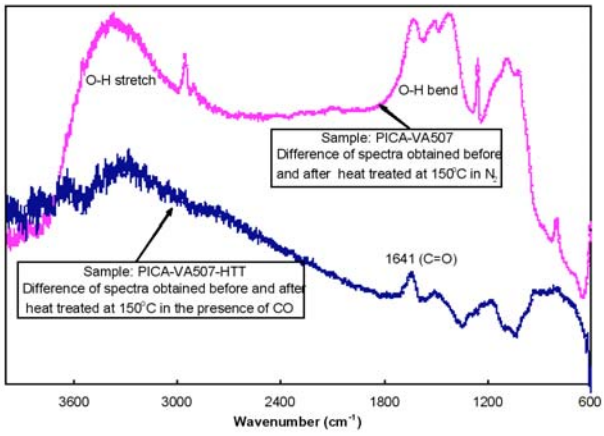


Figure 1. Preliminary results on the interaction of CO with activated carbon observed by DRIFTS.

Summary and Conclusions

This project demonstrated that H_2S can be oxidized and reduced by 99% in a single pass with little or no oxidation products being formed. Additional supporting theoretical and experimental data suggest that operation in a batch mode is preferred to ensure complete oxidation of H_2S and avoid formation of COS and SO_2.

Enhanced Performance and Energy Savings through Ultrahigh Magnetic Field Processing of Ferromagnetic Materials

G. M. Ludtka,[1] R. A. Kisner,[2] D. M. Nicholson,[3] J. B. Wilgen,[2] P. Kalu[4] G. Mackiewicz-Ludtka,[1]
G. M. Stocks,[1] and T. R. Watkins[1]
[1]*Metals and Ceramics Division,*
[2]*Engineering Science and Technology Division*
[3]*Computer Science and Mathematics Division*
[4]*National High Magnetic Field Laboratory, Florida State University*

This project is demonstrating and quantitatively documenting the influence of ultrahigh magnetic field processing on the phase equilibria and transformation kinetics for ferromagnetic materials. Novel microstructures with enhanced performance have resulted that inevitably will enable the implementation of new and lightweight materials that will lead to major energy savings and environmental benefits for the heat treatment and transportation sectors. A parallel simulation effort using modern first-principles electronic structure methods is being conducted to calculate the contribution to the free energy of the alloy phases that results from the application of these large magnetic fields. Based on the guiding principles established by this modeling effort, new compositions with their phase equilibria will be predicted and verified experimentally.

Objective

The broad goals for this research are to demonstrate and document the influence of ultrahigh magnetic field processing on the phase equilibria and kinetics for ferromagnetic materials and to develop predictive capability based on first-principles calculations. This ability to selectively control microstructural stability and alter transformation kinetics through appropriate selection of the magnetic field strength will be shown to provide a very robust mechanism to develop and tailor enhanced microstructures (even potential bulk nanostructures through accelerated product phase nucleation and transformation kinetics) with superior properties through a more efficient processing technology for a broad spectrum of material applications.

Technical Approach

The tasks and deliverables to successfully accomplish the objectives of this research include: (1) a predictive ultrahigh magnetic field processing simulation tool based on modern first-principles electronic structure methods and appended to a thermodynamic modeling software (ThermoCalc) with experimental validation; (2) quantitative experimental evidence of novel, ultrahigh magnetic field processed microstructures unattainable through conventional zero-field thermal processing methods; (3) ORNL capabilities for in situ characterization of magnetic field processing of materials; and (4) definition of future research thrust areas and funding resources for pursuing ultrahigh magnetic field processing as a major new ORNL research initiative. Of major importance to accomplish this research was the requirement to design and fabricate the first thermal-magnetic processing system for conducting ultrahigh magnetic field processing experiments in the U.S.A.

Results to Date

Significant progress has been made this past year in achieving these goals. These accomplishments will be highlighted under four subheadings defined as: Fe-15Ni Alloy, 52100 Alloy Steel, Bulk Amorphous Ferromagnetic Alloy, and Local Spin Density (LSD) Modeling.

Fe-15Ni Alloy: Experimental results on this INVAR-type alloy have quantitatively shown that magnetic field processing enables obtaining conventionally unachievable $\alpha + \gamma$ microstructures as exhibited by enhanced solid solution solubilities and phase distributions (volume fraction) compared with conventional equilibrium phase diagram data. Specifically, the γ phase Ni composition of a magnetically processed sample was 40 atomic percent for a 500°C heat treatment with an imposed field of 29 T compared to an equilibrium value of 30 atomic percent without a field. Similarly, the volume fraction of the α phase was magnetically enhanced by a factor of 2 to 28% compared to an equilibrium value of 14%.

52100 Alloy Steel: The application of a magnetic field at ambient temperature after an austenitization heat treatment and helium quench cycle resulted in an increase

in hardness of 14% when compared to a sample similarly processed but with no post quench magnetic field cycle. These results support the prediction that magnetic field processing raises the martensite start and finish temperatures of ferrous alloys and can enhance material properties. In addition, metallographic examination of other samples after various heat treatment cycles with a magnetic field imposed during continuous cooling from the paramagnetic austenite phase field indicates acceleration of the phase transformation kinetics, stabilization and enhancement of the ferromagnetic phases, and carbide precipitation normally only achievable with additional cryogenic processing.

Bulk Amorphous Ferromagnetic Alloy: Limited experiments this past year on a bulk amorphous ferromagnetic alloy annealed at 670ºC with and without an imposed 30-T magnetic field indicate an influence on the crystallization behavior of this material. These initial results suggest a potential impact of magnetic processing on the stabilization of precursors to ferromagnetic (or high susceptibility) phases which may allow for the formation of materials with unique magnetic properties.

Local Spin Density (LSD) Modeling: This task is enabling the prediction of phase diagrams accounting for the imposition of various magnetic fields while at elevated temperature. This past year the LSD model was modified to capture the coupling of volume and magnetic effects. This locally self-consistent multiple scattering, LSMS, code currently qualitatively captures the change in volume fraction of α and γ phases for the Fe-15Ni alloy system. The prediction of the shift in Ni concentration at which the α- and γ-phase free energy curves intersect indicates an increase in the martensite start temperature of this alloy and enhanced α phase stability as experimentally observed.

Benefits

Successfully demonstrating this novel materials and process development concept, while providing a fundamental understanding and predictive capability of the potential magnitude of the effects of this ultrahigh magnetic field processing technology, will result in significant future program and funding development interest from several major resources. The DOE OIT and OTT Programs definitely would fund additional technological development for this processing methodology. As an example, Dr. Peter Angelini, OIT Program Manager, indicated in a conversation that future funding under his Industrial Materials of the Future Program would be viable for further developing and implementing within industry this proposal's concepts once this initial research effort was completed. Similarly, the Department of Defense DARPA structural amorphous metals program would benefit from the phase destabilization and magnetic cooling aspects of this project to meet their goals of developing stabilized magnetic glasses. In addition, recent presentations to the U.S. Army Aviation & Missile Command in Huntsville, Alabama, have developed their interest in pursuing DARPA-funded magnetic processing research for ORNL for specific structural and aerospace applications.

Materials Science of Nanostructured Carbons and Graphites

T. D. Burchell,[1] J. W. Klett,[1] C. L. Fu,[1] N. Gallego,[1] J. Y. Howe,[1] M. K. Miller,[1]
D. B. Geohegan,[2] W. A. Shelton,[3] and J. Calo[4]
[1]*Metals and Ceramics Division*
[2]*Condensed Matter Science Division*
[3]*Computer Science and Mathematics Division*
[4]*Brown University*

The complex and disorganized structures of microporous carbons, in which the dimensions of the pore spaces are commensurate with the dimensions of the adsorbate molecule, make interpretation of adsorption very difficult. Indeed, our ability to advance the materials development is hindered by our lack of fundamental insight into the key structure-adsorption relationships. The objectives of this work are (i) the characterization of the nanostructure of a range of carbon materials, (ii) assessment of the H_2 storage capability of nanostructured carbons, (iii) the development of mathematical models that describe the "real" carbon structure and the interaction of H_2 with carbon and catalysts, and (iv) the application of those models to assess the nanostructure features that control the interactions of the carbon with the H_2.

A test matrix of ~20 different carbon samples, including commercial activated carbons, adsorbent carbon fibers and composites, graphite nanofibers, carbon nanowebs and single-walled carbon nanotubes (SWNTs), was assembled and characterized via N_2 adsorption at 77 K. Selected samples were examined for their ability to adsorb H_2 (at pressures up to 20 bar). The gravimetric uptake of H_2 adsorbed on the carbons tested increased with the surface area of the sample. At room temperature and 20 bar pressure, the maximum observed H_2 uptake was ~0.3% on a commercial activated carbon. This increased to ~6% at 77 K. Pure SWNTs exhibited an uptake of ~0.12% at room temperature and ~20 bar, increasing to ~6% at 77 K.

We used high-resolution transmission electron microscopy (HRTEM) to characterize SWNT purity and defect level before and after chemical purification. As-grown SWNTs may contain up to 50 wt % of impurities that mainly consist of the residual metal catalyst particles and amorphous carbon. HRTEM is an indispensable tool to evaluate the fine structures of the SWNTs at various stages, that is, as-grown and after purification. The availability of purified samples is key to determining whether H_2 adsorption on nanotubes samples is influenced by the presence of the metal catalyst. Contrast matching small-angle neutron scattering was conducted on microporous carbon monoliths to study the pore structure of the materials at nanoscale. Deuterated *n*-pentane was the matching agent as it is a chain-like molecule with similar neutron contrast to the carbon. It was found that the microporosity (<2nm) was all open porosity.

Using first-principles method, we have studied the binding interactions of H_2 molecules on the graphite surface with and without the presence of metal particles (e.g., Pd). First, without Pd, we find that H_2 molecules interact weakly with the graphite sheet (i.e., physisorption) with a low binding energy (0.08 eV); the H-H bond length is unaffected by the graphite surface. The adsorption of H_2 on the graphite surface is unlikely at room temperature (in agreement with our experimental evidence). Second, despite the strong in-plane C-C bond, Pd atoms can be adsorbed easily on the graphite surface. The H_2 molecule is found to bind preferentially to the Pd atom, with a binding energy as high as 1.36 eV (130 KJ/mole); the H-H bond length increases by 17% (with a sizeable reduction in the H_2 vibration frequency) with respect to that of the isolated H_2; that is, the presence of metal particles promotes the adsorption and dissociation of H_2. Third, while the electronic interaction between Pd and H_2 on the graphite surface is local, we find that a single Pd atom can promote the adsorption of up to six H_2 molecules on the graphite surface, indicating that the H_2 storage properties can be drastically improved even in the presence of a few metal particles.

We have additionally made significant progress toward devising a model for computing amorphous carbon structures from experimentally determined radial distribution data. To validate our approach, we have run a variety of test cases on other known amorphous carbon structures where we have successfully reproduced the structures. The simulations demonstrate that the structures are actually somewhat insensitive to the radial distributions

and that moderate broadening in the radial distribution peaks is accompanied by gross changes in the structure.

The ability to model the complex interactions between gases and nanoporous carbons will lead to an enhanced understanding of the physical adsorption process and guide the development of improved nanoporous carbon adsorbents. Such developments will support ongoing DOE programs, including research into the production and storage of hydrogen and methane, the separation and capture of carbon dioxide, the desulfurization of combustion and fuel gasses, and fuel cell research.

Moreover, improved nanoporous carbon adsorbents that can be made specific for certain chemical agents would be of great benefit in the individual/collective protection of personnel from chemical weapons threats.

Aberration-Corrected, Ultra-High-Resolution Electron Microscopy for Atomic-Level Characterization of the Structure and Chemistry of Nanophase Materials

S. J. Pennycook,[1] M. V. del Arco,[1] L. F. Allard,[2] I. M. Anderson,[2] and D. A. Blom[2]
[1]*Condensed Matter Sciences Division*
[2]*Metals and Ceramics Division*

The recent successful implementation of aberration correctors in two ORNL scanning transmission electron microscopes, coupled with the upcoming acquisition of a commercial version for the HTML user facility, puts ORNL in a strong position to develop advanced imaging techniques for the Transmission Electron Aberration-corrected Microscope (TEAM) programs. The goal of this project is to develop the necessary theoretical and computational groundwork for the simulation and interpretation of atomic-scale images and spectra enabled by these unprecedented instruments.

In the growing area of nanophase materials science, the ability to characterize structure and chemistry of individual nanostructures atom by atom has been a dream ever since the famous 1959 lecture of Nobel Laureate Richard Feynman, "There's Plenty of Room at the Bottom," which first predicted the importance of nanotechnology. In that lecture, Feynman explicitly called for improving the resolution of the electron microscopes of that day by two orders of magnitude, to 0.1 Å, through correcting the aberrations of electron lenses. We are now within striking distance of achieving Feynman's dream, and ORNL is strategically positioned among university, corporate, and government research laboratories in the United States and worldwide to capitalize on this potential due to recent and ongoing investments in sub-angstrom electron microscopy. ORNL will soon have three of the first aberration-corrected electron microscopes in the nation, and we expect to be the first in the world to achieve a 0.5-Å electron beam. However, the extremely rapid rate of instrumental development has far exceeded the pace of technique development. The necessary theoretical and computational groundwork for the simulation and interpretation of atomic-scale images and spectra enabled by these unprecedented instruments has not yet been developed. It is critical that work be supported now to achieve rapid application and maximum payoff of our new ultra-high-resolution characterization facilities for the new world of sub-angstrom imaging. To address these opportunities, this project is undertaking research goals that were unthinkable just a few years ago: simulation and interpretation of three-dimensional images of individual nanostructures, spectroscopic determination of the elemental identities and local electronic structure surrounding individual dopant atoms, and observation and catalytic clusters with atomic resolution and sensitivity. Furthermore, the necessary theoretical and computational tools are being developed for the simulation and interpretation of the atomic-scale images and spectra enabled by these unprecedented instruments. During FY 2003, progress was made in three areas as described below.

Development of Reliable Techniques for Simulation of Single-Atom Images in Nanostructured Materials

We have expanded our Bloch wave analysis of the imaging process to the larger apertures available with present and next-generation aberration-corrected scanning tunneling electron microscopes (STEMs). The image indeed becomes a direct image of the 1s Bloch states, as proposed by us previously based on analytical approximations to the image, and a paper has been submitted to the *Journal of Electron Microscopy*. In addition, it has become clear that the non-1s portion of the incident probe will become increasingly important with increasing aperture angle. We found that this part will allow three-dimensional (3-D) reconstruction by a new and direct method, depth sectioning, as described in the next section. We have also initiated the frozen phonon multislice algorithm, which is more efficient for some purposes, and have extended the code to sub-Ångstrom probes.

Development of 3-D Tomography at Atomic Resolution and Sensitivity

Initially, it was thought that 3-D information would need to be extracted from the dependence of the image intensity on thickness, based on accurate simulations. Although this is now feasible, based on the insights from the image simulations described above, we also investigated the possibility of selecting atoms at different depths simply by focusing on them directly. Simulations have confirmed that the depth resolution increases with increasing aperture angle but also that it is not negated by the tendency of electrons to channel along atomic columns. The simulations have led to the realization that both types of behavior are seen for any given probe, the ratio depending on aperture angle. Channeling dominates for small angles, the depth sectioning effect for large angles. Experimentally we have already achieved 4-nm depth resolution using a 0.7Å probe, and that result was submitted to *Nature*. Simulations show that our current system should be able to achieve 2-nm resolution. The next- generation STEM anticipated from the future TEAM program should improve this to 1 nm, and depth resolution at the atomic level would be possible in a future-generation instrument with an additional factor of two increase in aperture angle. This work led to three invited talks at this year's Microscopy Society conference and an invitation to speak on the depth sectioning concept at this year's TEAM workshop.

Development of Electron Energy Loss Spectroscopy with Ultimate Sensitivity and Resolution

For electron energy loss imaging, a *Physical Review Letter* has been published showing agreement between experiment and simulation at atomic resolution. Therefore reliable techniques are now available both for Z-contrast imaging and element-specific spectroscopic imaging. On the experimental side we have detected and identified individual atoms within a bulk material for the first time. The sample was specially grown and comprised lightly doped layers spaced 10 nm apart. Dynamical simulations have been performed and allow the atom depth to be determined by comparing the calculated and experimental EELS images. With increasing atom depth, as a result of beam broadening, signals from adjacent columns increase relative to that from the column that contains the dopant atom. These results are shown in Fig. 1 and have been submitted to *Science*.

Recently, DOE Secretary Spencer Abraham released the 20-year plan,"Facilities for the Future of Science: A Twenty-Year Outlook" (www.sc.doe.gov/Sub/Facilities_for_future/facilities_future.htm), which included as a high priority aberration-corrected electron microscopy. To quote from his remarks to the National

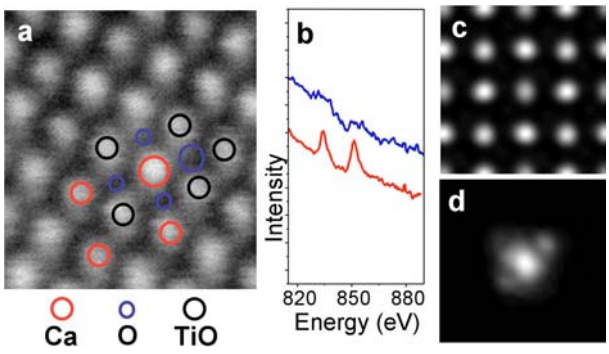


Fig. 1. (a) Z-contrast image of $CaTiO_3$ doped with a very low concentration of La. The column with the large red circle is on the Ca sublattice but is brighter than the other Ca columns, indicating the presence of a single La atom located within the specimen thickness. (b) EELS data from specific columns in the image identifying the La $M_{4,5}$ edge. A strong signal is obtained with the beam over bright Ca column, but a weak signal is also seen with the beam over the adjacent O column (large blue circle). (c) Simulated Z-contrast image with an aberration-balanced probe and 0.4 Å Gaussian broadening. (d) Simulated EELS image showing that significant La signal is anticipated from the adjacent O column due to beam spreading.

Press Club on November 10, 2003, "A new generation electron microscope can help us study how atoms combine to form materials, and how materials respond to external factors such as electric fields. This new instrument, the Transmission Electron Achromatic Microscope or TEAM, will help us design lighter, more efficient materials for everything from automobiles to advanced fuel cells." ORNL is an active participant inTEAM and already has two aberration-corrected instruments that represent prototypes for this next-generation instrument. This LDRD involves developing the methods of image simulation and analysis to support the capabilities of advanced hardware of this nature. We are developing advanced simulation tools to (1) quantify images at the ultimate resolution, (2) to reconstruct three-dimensional atomic structure from such images, and (3) to apply these advanced methods to energy problems of national interest. Such problems include materials science, catalysis, and nanoscience, which are high priorities for atomic-scale research with major potential for technological payoff to improve the nations economic well-being. Therefore this LDRD directly benefits many science and technology programs of DOE and other agencies.

Biologically Driven Controlled Synthesis and Directed Assembly of Nanophase Inorganic Materials

M. L. Simpson,[1] M. J. Doktycz,[2] T. E. McKnight,[1] and N. Samatova[3]
[1]*Engineering Science and Technology Division*
[2]*Life Sciences Division*
[3]*Computer Science and Mathematics Division*
University Collaborators
D. E. Morse,[4] G. S. Sayler,[5] J. J. Collins,[6] J. Turner[7]
[4]*University of California, Santa Barbara*
[5]*University of Tennessee Center for Environmental Biotechnology*
[6]*Boston University*
[7]*University of Tennessee Chemistry Department*

Our goal is to "learn the rules" of biologically driven controlled synthesis and directed assembly and eventually to apply these rules to the creation of engineered nanophase material systems. We are focusing ORNL's and our university collaborators' unique capabilities in biology, computation, and the nanosciences on mechanisms of biology that manipulate energy, information, and materials with spatial and temporal specificity on the nanoscale. As there are more than 60 different inorganic minerals produced by organisms from bacteria to humans, we could look at a wide variety of biological systems that synthesize inorganic nanophase materials. However, it was recently discovered in the laboratory of one of our collaborators that 'silicateins' (silica proteins) can be harnessed to direct the nanofabrication of other technologically important materials such as titanium dioxide, zinc oxide, and gallium oxide. This provides an appropriate model system to focus the science of this project while providing the potential of highly significant technology results. This work includes the following components: identifying/describing molecular processes useful for controlled synthesis and directed assembly of inorganic materials to create inorganic nanostructures abiotically; analysis, modeling, and simulation of natural gene/biochemical circuit and network topologies that control synthesis and direct assembly at the nanoscale in biological systems; the development of synthetic gene/biochemical networks for the engineered control/mimicking of biologically driven controlled synthesis and directed assembly; and the creation of nanoscale synthetic elements to deliver and control the expression of genetic materials within viable cells.

With a precision of nanostructural control that far exceeds present human capabilities, biological systems fabricate a remarkable diversity of three-dimensionally organized materials. Proceeding from instruction encoded within the genome to the completed nanostructures requires a set of complex processes that provides for the transport of material, energy, and information at the nanoscale that leads to the controlled synthesis and directed assembly of nanophase material systems. It is these processes, pathways, and complex network topologies that we are investigating. Stated in broad terms, our goal is to "learn the rules" of biologically driven controlled synthesis and directed assembly and eventually to apply these rules to the creation of engineered nanophase material systems. In FY 2003, we have made progress along several fronts as described below.

Molecular Processes Useful for the Controlled Synthesis and Directed Assembly of Inorganic Materials

Our principal objective has been to extend present expression and cellular display studies with the silicatein molecule on the surfaces of microbial cells to new and collaborative work on biofilms, in which the expressed silicatein molecules in microbial biofilms are used as structure-directing catalysts for the formation of oriented silica and siloxane-based nanostructured composite materials. Morse's lab succeeded in expressing silicatein alpha (the principal silicatein occluded in the biosilica needles produced by the marine sponge, *Tethya aurantia*) as a fusion protein with the bacterial surface protein, *OmpA*. This protein codes for a transmembrane protein that is produced in many copies and projects from the

surface of the bacterium, *E. coli,* into the surrounding medium. Morse's lab constructed the genetic fusion of the silicatein and *OmpA* genes in such a way to ensure that the silicatein sequence would be displayed in high copy number on the surfaces of the bacterial cells. When these bacteria were prepared as a film, and subsequently exposed to the silica precursor, TEOS (tetra-ethoxyorthosilicate) at neutral pH, the silicatein catalyzed the hydrolysis and subsequent polycondensation of TEOS to form silica. Evidence for silicatein/biofilm–directed templating of the silica synthesis was suggested by the close association of this silica with the microbial surfaces, but disappointingly, no higher-order regularity has yet been seen. Our further efforts to enhance nanoscale and higher-order regularity of the templated synthesis will be focused on: (1) optimization of conditions of synthesis; (2) use of organic silicon trialkoxides, to impose steric hindrance and other (e.g., pi-pi stacking) constraints on the degrees of freedom of synthesis; and/or (3) use of precursors of crystalline semiconductor metal oxides in place of the silicon alkoxides, to thus augment the structure-directing properties of the silicatein-containing biofilms.

Although limited by budget constraints, we were able to make some progress exploring the use of carbon nanofibers as templates for biologically driven synthesis. As a precursor to the use of silicatein enzyme systems, we have covalently attached soybean peroxidase (SBP) onto nanofiber scaffolding and validated the reactivity of the attached enzyme via colorimetric assay and precipitation visualization techniques. The homogeneity of the precipitant appears to increase if the immobilization reaction of SBP onto fibers is conducted following oxygen plasma etching of the fibers, apparently due to improved surface coverage of carboxylic acids and therefore SBP upon the fiber scaffold. FY 2004 for this thrust area is obvious: extend the vertically aligned carbon nanofiber derivatization work to the silicateins.

Analysis and Simulation of Gene/Biochemical Systems

We have developed an Exact Stochastic Simulator (ESS) which reads systems biology mark-up language (SBML) level 2 models and simulates biological models using the Gillespie algorithm with the Gibson and Bruck enhancements. ESS allows reactions to be specified using elementary rate constants or Hill Repression Kinetics. ESS was submitted and subsequently incorporated into BioSPICE version 2.0 as the standard stochastic simulator. Although the ESS tool is quite fast (with over 420,000 reactions per second achieved), computational demands for full cellular modeling vastly exceed this speed. To address this need, we developed a hardware accelerator we named FASST (FPGA Accelerated Stochastic Simulation Tool) which allows simulation of 4.5 million reactions per second—an increase in simulation speed of more than 10. To provide deeper insights into the structure/function relationship within gene/biochemical circuits and networks, we have developed a novel analytical approach that we refer to as the frequency domain chemical Langevin approach. In particular, compared to other noted gene circuit analysis, we find our approach leads to a deeper understanding of the architecture of gene circuits and networks and the emergences of robust behavior from systems composed of highly imperfect components.

This research directly addresses needs of DOE-BES programs in the design and synthesis of nanomaterials that focus on efficiently designing and fabricating materials "from the bottom up." Such challenging tasks are the essence of nanotechnology and materials science for the new century and therefore are a central focus of the DOE Nanoscale Science Research Centers. Additionally, the biological approaches pursued in this project accomplish the synthesis of nanostructured materials at relatively low temperatures, ambient pressure, at near-neutral pH, and in the absence of caustic chemicals, in marked contrast to present human manufacturing technologies. This provides a possible enabling technology for the controlled synthesis of nanoscale materials with a considerable energy savings and low environmental impact. This project aims to harness this capability for the abiotic production of nanoscale materials in direct support of DOE,s energy resources, and environmental quality missions.

Nanoscale Control of Collective Phenomena Using Artificially Structured Materials

H. N. Lee, H. M. Christen, and D. H. Lowndes
Condensed Matter Sciences Division

Advances in modern epitaxial growth techniques have enabled the synthesis of nearly perfect thin films, superlattices, etc., by the atomic-scale control of surfaces and interfaces. This enables the formation of artificial oxide superlattices that do not exist in bulk forms. During FY 2003, we have refined the growth techniques for metal-oxide heterostructures. Atomic-scale control of surfaces and interfaces led to the growth of epitaxial superlattices containing thousands of individual layers (at a total thickness of 1 micron) with extraordinary crystalline quality. Therefore, this capability now allows us to synthesize entirely artificial layered materials, combining dissimilar characteristics and leading to unprecedented physical properties.

Finding new physical phenomena at the nanoscale in complex metal oxides, including ferroelectric, ferromagnetic, and multiferroic materials, and manipulating the resulting properties in specifically designed epitaxial heterostructures, is a key challenge in ORNL's Condensed Matter Sciences Division. Epitaxial, artificial superlattice materials constitute a significant element of this research; however, many technical and scientific challenges related to their synthesis remain to be addressed. During FY 2003, most of the effort of this project was thus focused on the preparation of single-stepped substrates and the growth of conducting ruthenates as well as dielectric heterostructures by pulsed laser deposition (PLD).

As a first step, atomically flat, single-stepped surfaces were prepared on commercial strontium titanate ($SrTiO_3$) substrates using combined chemical (buffered hydrogen fluoride etch) and thermal treatments (at 1100–1200°C in air). Thin films of particle-free, atomically flat, electrically conducting $SrRuO_3$ were subsequently grown by a novel quasi-eclipse technique. In addition, the thermal stability of single-stepped epitaxial $SrRuO_3$ thin films (as required for subsequent growth of dielectric heterostructures) was explored as a function of temperature (25–800°C) and oxygen pressure (10^{-7}–10^{-2} Torr) by ex situ reflection high-energy electron diffraction (RHEED), low-energy electron diffraction (LEED), atomic force microscopy (AFM), scanning tunneling microscopy (STM), and X-ray photoelectron spectroscopy (XPS). We found that the decomposition mechanism of $SrRuO_3$ (yielding metallic Ru at the surface) is thermally activated with the activation energy E_a = 63 kJ/mol. A stability region was thus identified for the growth of dielectric materials on top of these electrode layers, and AFM, TEM, and X-ray investigations demonstrated the near-perfect transfer of crystalline structure from substrate to dielectric material through these $SrRuO_3$ films.

For artificial heterostructures, asymmetric tri-color superlattices (TCSs) with a combination of $BaTiO_3$/$SrTiO_3$/$CaTiO_3$, in which the inversion symmetry can be broken by a composition gradient, have been grown on $SrRuO_3$/$SrTiO_3$ substrates at 720°C in 10 mTorr O_2 (i.e., within the above-determined stability region). The controlled growth of artificial crystals in the form of epitaxial superlattices containing thousands of individual layers led to atomically flat surfaces even at a total thickness of 1 micron. This capability, therefore, gives us the freedom of materials choice in the design and synthesis of a multitude of multilayers containing materials with dissimilar properties.

Additional specific scientific observations during FY 2003 include (1) the first observation of spiral growth in a single-perovskite and the unambiguous correlation of these features with merging terraces on the $SrTiO_3$ substrates (bulk screw dislocations), and (2) the growth of submicron-sized oxide islands of an insulating perovskite on a conducting layer, by careful control of the growth mode.

During FY 2004, magnetic and multiferroic materials consisting of ferroelectric and ferromagnetic layers will be grown and tested by employing the growth technique established during FY 2003, in particular to investigate the charge and orbital ordering, ferromagnetism, ferroelectricity, and magnetic control of ferroelectric properties (or vice versa).

Such materials being investigated are highly relevant to DOE missions and will impact many federal agencies because of their potential usefulness for sensors, actuators and related devices. Thus, they may contribute to the entire range of security, energy, environmental, and science applications.

Materials Needs for Successful Implementation of Lean NO_x Treatment Technology

C. K. Narula,[1] L. F. Allard,[1] C. Goralski,[2] and C. S. Daw[3]
[1]*Metals and Ceramics Division*
[2]*Ford Motor Company*
[3]*Engineering Science and Technology Division*

The regulatory requirements for diesel emissions (excluding particulates) can be met by deploying NO_x trap systems (LNTs) for exhaust treatment; however, the performance of LNTs deteriorates rapidly. There is experimental evidence that repeated regeneration cycles (especially sulfation-desulfation) and associated high-temperature exposure of trap materials leads to the deterioration of LNT performance. Therefore, it is essential to understand the microstructural changes in trap materials associated with LNT performance deterioration. Such information is necessary for systematic research of new materials that can withstand the harsh conditions of regeneration cycles and, therefore, could lead to enhanced durability of LNTs. In this report, we describe our collaborative programs initiated to address the LNT durability issues and a summary of technical accomplishment of these programs.

The introduction of diesel engine or lean-burn-engine-based passenger vehicles depends on a successful development of a strategy to treat nitrogen oxides (NO_x) emissions, since a catalyst (or a combination of catalysts) that can convert NO_x into inert gases under oxidizing conditions over a complete range of exhaust temperature does not exist. Among NO_x treatment strategies, LNT are the most likely candidates for early deployment because they are consumer transparent and can be easily system integrated into current vehicle control strategies. However, several materials issues related to LNT aging need to be addressed before LNTs can be deployed in mass-produced diesel /lean-burn gasoline-powered vehicles. This research program is focused towards identification and alleviation of aging-related materials issues.

During the first year of this project, we investigated two issues: (1) microstructural changes in LNT materials under lean and rich conditions at high temperatures and (2) improving high-temperature durability of LNT materials.

Microstructural Changes in LNT Materials

A complete electron microscopic analysis of supplier catalyst aged under lean, rich, and stoichiometric conditions on a pulsator for lean gasoline conditions shows some interesting features:

- The fresh sample is a two-layer system on a honeycomb cordierite substrate. The inner layer is Pt/BaO-Al_2O_3, and the outer layer is CeO_2-ZrO_2. There are alumina islands in CeO_2-ZrO_2 layer that contain some barium. A dark field image of the CeO_2-ZrO_2 layer shows bright areas for the alumina islands and dark spots for segregated zirconia. This segregated zirconia is present only in fresh sample and is not seen in any of the aged samples. This suggests that segregated zirconia reacts with the CeO_2-ZrO_2 during aging. Since CeO_2-ZrO_2 forms under moderate conditions and crystallizes in a cubic fluorite phase regardless of ratio, the reaction of segregated ZrO_2 with CeO_2-ZrO_2 is not surprising. The dark field image of BaO-Al_2O_3 layer shows a uniform distribution of platinum particles. In general, platinum particles are about 5 nm. There is no indication of platinum or barium in CeO_2-ZrO_2 layer.
- The two layers in lean-aged samples remain intact. The segregated zirconia in CeO_2-ZrO_2 layer, seen in the fresh sample, is no longer segregated. The representative platinum particles, seen in the dark field image, have considerably increased in size as compared with fresh samples. The partial migration of barium from BaO-Al_2O_3 layer to CeO_2-ZrO_2 layer can be inferred from the EDS of an area near the outer edge of the CeO_2-ZrO_2 layer.
- The microstructure of a rich-aged sample is identical to that of the lean-aged sample. The aging under stoichiometric condition leads to partial migration of barium from BaO-Al_2O_3 layer to CeO_2-ZrO_2 layer. The sintering of platinum particles also occurs, but the extent of sintering is smaller than that under lean or rich conditions. (Note: the stoichiometric conditions refer to condition under which three-way catalyst operates in current gasoline-powered passenger vehicles.)

Thus, there are two pathways that can lead to reduced platinum–barium oxide contact. The increase in platinum particle size due to sintering reduces the number of catalyst sites that are available and in contact with barium oxide. The partial migration of barium into CeO_2-ZrO_2 leads to reduced barium in alumina layer that can be in contact with or in close proximity of platinum. Both pathways occur during lean or rich aging of LNT catalysts.

Model Catalyst Materials with Improved Thermal Durability

The complexity of the LNT described above warrants a complex model system that can be analyzed and evaluated in laboratory. Improvements in the model system then can be implemented in the full-size catalyst that can be tested on a vehicle. While at Ford, CKN started addressing the thermal durability of LNT materials. He found that a model system [10%CeO_2-ZrO_2-90%(2%La_2O_3-98%BaO.6Al_2O_3)] retains its beneficial surface properties even after thermal aging at 1050°C. This material can be prepared as follows:

1. Impregnation of $Ba(NO_3)_3$ on commercial high-surface-area alumina, pyrolysis, and sintering at 750°C.
2. Impregnation of step 1 material with lanthanum nitrate, pyrolysis, and sintering at 750°C.
3. Ball milling step 2 material with commercial high-surface-area CeO_2-ZrO_2.

As a next step, we impregnated this powder with 1% platinum and investigated platinum distribution on the powder. The electron micrograph shows that platinum particles are about 1 nm in size.

Next Steps

We have recently completed the complete redesign of an ex situ reactor for rapid screening of LNT materials for their durability under operating conditions by monitoring microstructural changes using TEM. We have procured funding and purchased components to set up a bench-top flow reactor at NTRC to evaluate LNTs using simulated exhaust. The model materials and new materials that address aging issues will be evaluated on these two systems.

Catalysis is a major thrust area for DOE-EERE FreedomCAR and Vehicle Technologies (OFCVT). The ability to prepare, characterize, and evaluate new LNT materials under this program complements the DOE-EERE-funded diesel programs at NTRC on evaluation of LNT and HTML involvement in several catalysis-related efforts, generally addressing the changes that occur at the microstructural level in collaboration with Ford, Cummins Engines, Caterpillar, etc.

Production of Hydrogen Using Nuclear Energy and Inorganic Membranes

B. L. Bischoff,[1] C. W. Forsberg,[2] L. D. Trowbridge,[2] and L. K. Mansur[1]
[1]*Metals and Ceramics Division*
[2]*Nuclear Science and Technology Division*

The Iodine-Sulfur (IS) thermochemical process is the leading candidate worldwide for production of hydrogen (H_2) using nuclear energy. This process thermocatalytically cracks water yielding hydrogen and oxygen. The IS process consists of a series of chemical reactions where all the chemicals are recycled in the process except for water. The high-temperature chemical reaction in this cycle is the thermal decomposition of sulfuric acid. The IS process is efficient, scalable to large sizes, and uses no expensive chemical reagents. However, it has one major disadvantage: high operating temperatures (800 to 900°C). This is not a show stopper. There is potential to lower the peak temperature by more than 200°C if the high-temperature decomposition products of sulfuric acid can be separated using an inorganic membrane. Lowering the peak process temperature may (1) allow existing designs of high-temperature reactors to be used for H_2 production, (2) reduce development times for a nuclear-H_2 system by many years with the potential that nuclear H_2 is a relatively near-term option, (3) improve the economics, and (4) improve the safety of the process.

The goal of this project is to conduct proof-of-principle experiments and associated analysis to demonstrate the potential for inorganic membranes to dramatically improve the IS process. The challenges that must be overcome are the development of membranes that promote the separation of O_2, H_2O, and SO_2 from SO_3, testing under simulated conditions of the IS process, and development of a model that describes the separation of multiple species from a multicomponent gas stream. ORNL leads the world in development of inorganic separation membranes. The technology is based on the technology used in the gaseous diffusion uranium enrichment plants. If this technology can be shown to potentially lower the IS peak temperatures, ORNL would become a major player in development of technologies for generation of H_2 using nuclear energy and a major development effort supporting GEN IV reactors would follow.

The overarching objective of this project is to demonstrate that inorganic membranes can dramatically improve the IS process by lowering the peak temperature. Technically, this requires demonstration that inorganic membranes can be employed to separate O_2, H_2O, and SO_2 from SO_3 and that this separation will enhance the conversion of SO_3 to SO_2 and lower the temperature required for almost complete conversion by 100 to 200°C. Separation will be determined by measuring and comparing single-gas permeances. The measurements will be conducted at temperatures up to 600°C, and these results will be extrapolated to the expected operating temperatures of 650 to 750°C. These results will be incorporated into a model that is being developed by Sandia National Laboratory and General Atomics to show what effect inorganic membrane separations will have on the efficiency of the IS process.

Measuring the permeance of the gases, SO_2 and SO_3, involves some unique challenges. First these gases have a low vapor pressure. In order to operate at pressures up to 100 psig, the gas cylinders, piping and valves, pressure transducers, and flow meters all must be heated to over 100°C for SO_3 and to about 60°C for SO_2. Second, these gases are both corrosive and toxic. Selection of wetted materials is very important, as is minimizing the concentration of water vapor in the gas stream in order to maintain the integrity of the system piping and its components. However, safety was the most important factor that went into the design of the membrane test system. The system design was completed in July 2003 and is shown in Fig. 1. The complete system was designed to fit into an existing fume hood located in Building 4501. This lab had the necessary security systems to handle the testing of these sensitive membranes. The system design includes a chemical trap to minimize the possibility of any of these corrosive gases entering the fume hood ductwork and also includes a diluent gas stream to minimize their corrosivity. All the components compatible with both the gases and the high temperatures were ordered in July and August, the system fabrication was completed in October 2003, and system checkout is currently taking place.

Concurrently with the above work, a series of membranes was fabricated and characterized and are

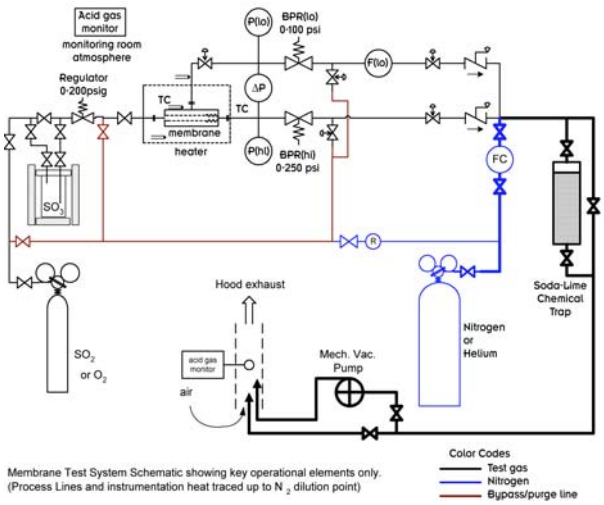


Fig. 1. Membrane test system schematic showing key operational elements only. Process lines and instrumentation heat traced up to N_2 dilution point.

candidates for initial testing in the system. The membranes were estimated to have pore sizes ranging from 0.62 nm to almost 0.9 nm.

A Research Safety Summary was completed and approved in August 2003, and the operational checkout of the test system should be completed in November 2003. Initial evaluation of the first-generation membranes by measuring the pure gas permeances at several temperatures will be completed by January 2004. Second-generation membranes will be fabricated and evaluated during the remaining months of FY 2004 using both pure gases and gas mixtures. These results will demonstrate that the ORNL membranes can have a major impact on the success of using nuclear energy to produce hydrogen.

MATERIALS SCIENCE AND ENGINEERING

Seed Money Fund

A New Microwave-Driven Pack Cementation Coating Process

J. Kiggans,[1] T. Tiegs,[1] M. Losego,[2] S. Nunn,[1] L. Riester,[1] M. Williams,[1] and R. Lowden[1]
[1]*Metals and Ceramics Division*
[2]*Office of Science, Energy Research Undergraduate Laboratory Fellowship*

Experiments were performed to determine if microwave heating can be used to apply diffusion coatings to metal alloy parts. In order to maintain a broad scope for this project, three metal alloys were selected which represent materials of interest for future research areas. These alloys included a tool steel (A2), a low-carbon steel (1018), and a nickel-chrome super alloy (625). Three types of diffusion coating processes were chosen: aluminization, chromization, and boronization. Control samples were processed by conventional heating. Aluminum-, chrome-, and boron-based diffusion coatings were successfully created using microwave heating. These microwave coatings show some similarities to the conventionally produced coatings, although significant differences also existed in some cases. Very encouraging results were obtained with the A2 tool steel.

Introduction

The use of microwave (MW) energy or processing as a tool for materials processing emerged during the late 1950s, but only within the past two decades has MW research really gained a foothold in the scientific community. Researchers in both the scientific and industrial communities have begun to explore MW technology as a way to improve even well-established processing practices.[1]

Pack cementation is a surface modification technique primarily used for metallic components to increase oxidation/corrosion resistance, wear resistance, and/or surface hardness. Pack cementation is a diffusion-based technique in which the desired species diffuses into the material through the surface. This diffusion creates a solid solution region near the surface with enhanced properties. Common diffusion-driven surface treatments include aluminizing, chromizing, and borodizing. The actual pack cementation process involves packing the untreated metal parts in a powder bed or pack and then heat-treating at temperatures between 800–1000°C. The powder pack consists of three general components: the desired diffusion source material, a halide activator (such as NaCl or NH_4Cl), and an inert filler (typically alumina). At elevated temperatures, the source material reacts with the halide to form a gas, which permeates the porous pack. Upon reaching the surface of the sample, another reaction occurs, depositing the source material on the substrate. Due to the high processing temperatures, the deposited element typically diffuses into the surface, leading to the desired surface treatment. The filler material is present simply to deter sintering.[2] Due to its simplicity and ability to coat even complicated parts, pack cementation continues to be a popular surface treatment method.

Current knowledge surrounding MW processing practices seems to indicate that pack cementation could be improved with the use of MW heating. Exposure of the pack cementation materials to MW energy should exploit the unique characteristics of MW heating behavior, as compared to conventional heating, which should provide several benefits. Depending on the dielectric properties of each component, they will absorb energy differently and therefore heat at different rates. It is well known that dispersed metal particulates heat extremely well in a MW furnace and they should do so in the pack cementation bed. In the above example, the aluminum powder would heat preferentially. The halide activator may or may not heat well by microwaves, depending on its dielectric properties.[3] For instance, the NH_4Cl in the above example would heat better than the NaCl (based on dielectric loss tangents), but neither would heat up as well as the aluminum. However, if one uses $CuCl_2$ as the activator, it heats just as well as the aluminum. Thus, by choosing appropriate halides, that heat or do not heat, further control of the cementation process can be obtained. In general, the inert filler powder would not heat well in a microwave furnace if an oxide, such as Al_2O_3, were employed. However, if fillers such as SiC are used, it could also be modified to tailor the heating response of the cementation pack.

Typically MW energy is confined within the furnace without much energy loss until there is an interaction with an absorber or the targeted sample (coupling). Hence, the energy transfer has much greater efficiency than most conventional heating methods, which employ radiant heating. Ordinarily, a dense metal part would not be expected to heat with microwaves. Thus, MW exposure

of the pack cementation process would most probably preferentially heat the bed containing the precursor source material and the halide activator. However, even in the metal parts there is some microwave absorption in the wear surface region. An estimate of the MW penetration (or skin depth) into a metal part during pack cementation can be made using the following equation:

$$\text{Skin Depth} = 5030\sqrt{\text{Resistivity}/(C \bullet \text{Frequency})}$$

As shown, the skin depth is dependent on the resistivity of the part and the MW frequency (C is a constant and is normally equal to 1). Thus, the approximate penetration depth at 2.45 GHz and a resistivity of 1×10^{-4} W·cm is on the order of 10 μm. Because the resistivity of most metals increases with temperature, the penetration depth should also increase as the temperature increases.

Microwave heating offers several advantages over conventional heating, mainly because it can selectively heat materials in a multicomponent system, creating a unique temperature profile. The potential advantages would include the following.

1. Because the temperature gradients will be higher than those in a conventional heating situation, the concentration gradient of the diffusing species into the component will also be steeper.[4] In most cases, this should increase the enhancement of surface properties and decrease the effects on bulk properties of the part being treated.
2. The process will be faster. Selective absorption of the MW energy will heat the packed bed rapidly to temperatures sufficient for the halide activator to operate. Conversely, because the surrounding insulation is cooler, the part should also cool faster. In addition, MW enhancement of diffusional processes in the near-surface regions, if any, would further improve the kinetics during the treatment.[5,6]
3. The bulk properties of the treated part will undergo only minimal alteration by thermal exposure. Since the process is faster, the parts are not subjected to the same high temperatures for extended periods of time as compared to the conventional heating situation. Consequently, less grain growth and other microstructural changes will occur that would affect the bulk properties.[7]
4. Microwave pack cementation should be safer and more environmentally friendly. With the heating confined to the volume directly surrounding the part to be treated, there should be less halide volatility and therefore less chance for halides being released from the reaction crucible. Also, because of the concentrated heat, it is possible that less halide activator will be required for the process. In addition, since the metal MW cavity is relatively cold, it acts as a getter for fumes, which may evolve during the reaction. Thus, noxious by-products will not be exhausted from the furnace.
5. Because the MW process is faster, safer, and less apt to change bulk properties, it will also be more efficient and economical on a total life cycle basis.

Some challenges do remain, though, including dealing with materials that poorly absorb MW power and controlling thermal runaway in samples that couple too well. Although the metal substrates used for pack cementation are expected to be MW reflectors,[8] the metal powders used as the source material are known to be excellent couplers of MW radiation.[8] Therefore, such heating problems are not expected for this project. However, due to possible shielding effects of the metal samples, it is difficult to predict the compositional and thermal uniformity. Furthermore, accurate measurement of temperatures in MW experiments has proven difficult historically due to temperature gradients that occur naturally in such processes.[9]

Technical Approach

For these experiments, we prepared cylindrical (1.25-cm diam × 1.25 long) and rectangular (1.25 cm × 1.25 cm × 1.25 cm). Three types of diffusion coating processes were chosen: aluminization, chromization, and boronization. For all three processs, the pack consisted of NH_4Cl as the activator and –325 mesh tabular aluminum oxide (Al_2O_3) as the filler. The depositing source materials (15 wt %) were aluminum metal powder (Al) for aluminization; chrome oxide (Cr_2O_3) for chromization; and boron carbide (B_4C) powder for boronization. After placing samples in the packing powder in an alumina crucible, the crucible was sealed with a ceramic paste and an alumina lid, heated to 100°C to set the glue, and evacuated and backfilled twice with argon. Loaded crucibles were heated at either 850 or 925°C for 4 h in flowing argon in either a 2.45-GHz MW furnace or in a conventional tube furnace. Control experiments were performed in which samples were heated in the alumina powder without the NH_4Cl activator and without the source powder. Processed and control samples were weighed and measured both before and after surface treatment. Control and processed samples were analyzed using optical microscopy, scanning electron microscopy, and energy dispersive X-ray spectroscopy (EDS). The hardness values of control and boronized samples were measured using both a nano-indentor and a micro-hardness testing device.

Results and Accomplishments

Aluminization Results

Aluminum-based coatings were successfully deposited on all three alloys using MW heating. Approximately 800 W of power was needed to maintain the temperature at 850 and 925°C for the 40 g of samples and the ~425 g of packing powder. Coatings were also deposited on all alloys using conventional heating. Figure 1 shows weight gain values obtained for the three metal alloys aluminized at 850 and 925°C using either MW or conventional heating. The results show that there was more aluminum deposited on the MW-treated samples at both temperatures. Scanning electron microscopic (SEM) examination of the samples was used to determine the extent of the aluminum diffusion into the samples. SEM examination (photos not shown) show that the aluminum coatings exist as both deposited and diffusion layers. Figures 2 and 3 show data generated from the SEM examination of both the A2 tool steel and the 1018 steel processed by either microwave or conventional heating at 850 or 925°C. The aluminized coatings were much thicker for the A2 and 1018 steel alloys processed to 925°C by MW heating. The data also shows that most of this coating

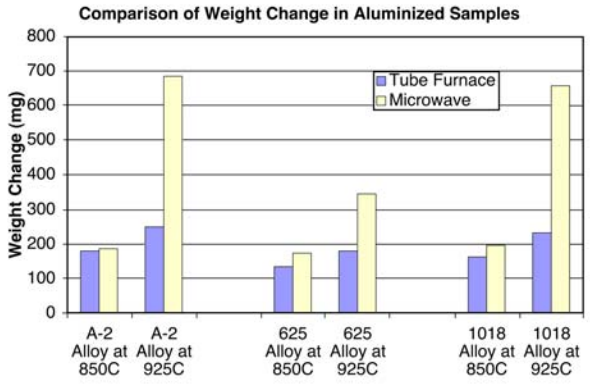


Fig. 1. Observed weight changes during aluminization.

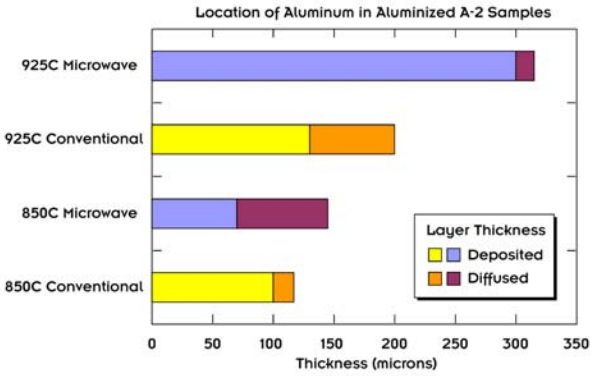


Fig. 2. Thickness of deposited and diffusion layers in aluminized A2 samples.

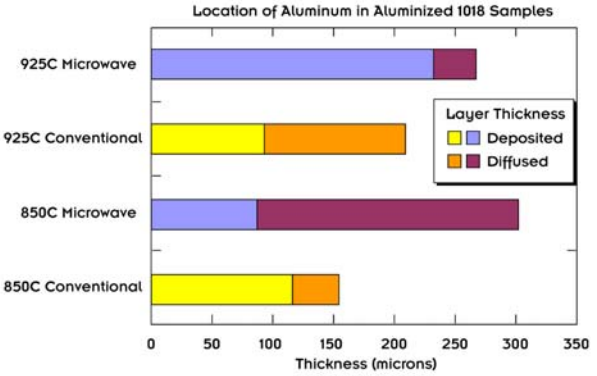


Fig. 3. Thickness of deposited and diffusion layers in aluminized 1018 samples.

was deposited, rather than diffusion, coating. Dimensional increases of up to 300 µm were observed. Aluminization of both alloys at 850°C yielded slightly thinner coatings; however, a larger percentage of the coatings exist as diffusion coatings, especially for A2 and 1018 samples processed using MW heating. This suggests that the deposition occurred rapidly at the higher temperature and effectively shielded the surface from MW interaction. On the other hand, at the lower temperature, the deposition was slower, so increased diffusion due the MW interactions of the aluminum could take place. Figure 4 shows a SEM micrograph for an A2 sample conventionally aluminized 925°C for 4 h. Figure 5 shows a SEM micrograph of a much thicker coating produced in an A2 sample aluminized using microwave heating at 925°C for 4 h. Note there was minor cracking in both the conventional and microwave coatings, possibly due to the large thickness of the coatings.

According to collected X-ray diffraction (XRD) data, the A2 and 1018 alloys showed the formation of an Al_5Fe_2 phase in the deposited aluminum layer under all processing conditions except at 850°C in the MW furnace. However, since iron is an excellent absorber of copper K_a radiation, this data is only applicable to about the first 20 to 40 µm. Therefore, this conclusion may be slightly oversimplified. Although most of the backscatter electron (BSE) images reveal a compositionally homogenous layer, the 1018 sample conventionally aluminized at 925°C clearly has multiple phases. Hence, the Al_5Fe_2 phase seems to be favored when the pack is fully activated at these processing temperatures, although the formation of other phases is possible due to the temperature of concentration gradients. Structurally, this deposited layer seems highly prone to crack formation in both MW and conventional samples at 925°C. This probability for cracking decreases somewhat at 850°C but is not totally eliminated. According to Koo and Yu's work,[10] this cracking is the result of thermal expansion differences (not volumetric increases) and will lead to local corrosion attacks. Therefore, a shorter process

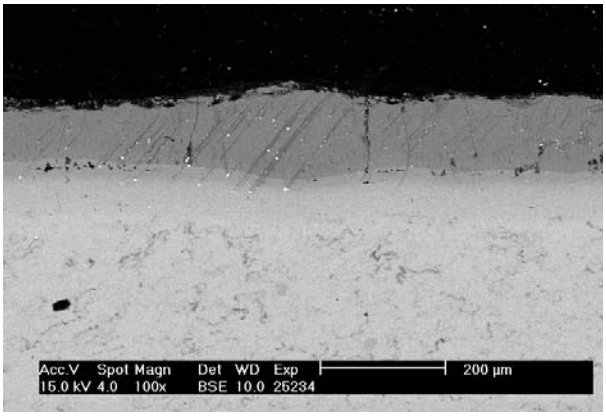


Fig. 4. SEM micrograph of A2 sample conventionally aluminized at 925°C for 4 h.

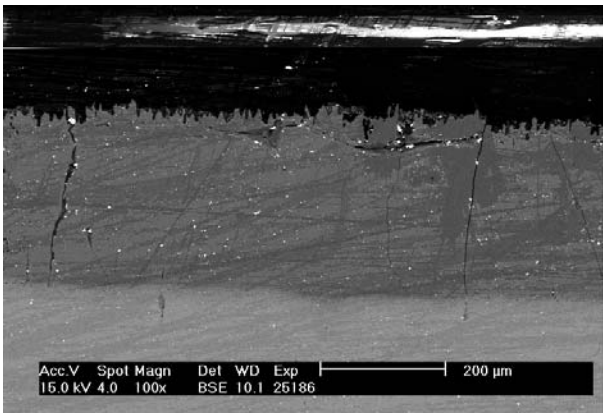


Fig. 5. SEM micrograph of A2 sample aluminized at 925°C with microwaves.

time is advisable for future aluminizing experiments. In comparing MW to conventionally heated samples, the largest structural difference is the tendency for porosity to form near the original surface of MW-processed samples. Unlike the 1018 samples, aluminized A2 samples tend to form chrome-vanadium carbide precipitates in the diffused region. Such precipitates were not observed in any of the control samples. Thus, this precipitation suggests that the diffused aluminum is forming a solid solution with the iron and making the formation of carbide thermodynamically favorable. These precipitates were observed in all aluminized A2 samples.

Chromization Results

Chrome-based coatings were successfully deposited on all three alloys using MW heating. Approximately, 600 W of power was needed to maintain the temperature at 850 and 925°C for the 40 g of samples and the ~450 g of packing powder. Coatings were also deposited on all alloys using conventional heating. Figure 6 shows the weight changes for the three alloys heated to either 850 or

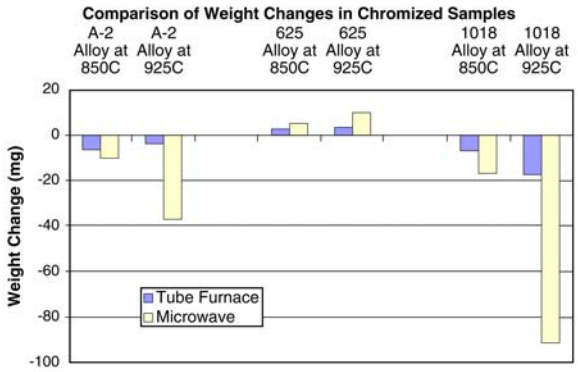


Fig. 6. Observed weight changes during chromization.

925°C using conventional and MW heating. The 625 nickel alloy gained a small amount of weight for both conventional and microwave processing. Weight losses were observed in the A2 and the 1018 alloys for samples processed by both conventional and microwave heating, with higher weight losses in microwave-heated parts. SEM analysis showed that the chromization of MW-treated pieces was better at the higher processing temperature; however, even at 925°C for 4 h, the diffusion layer was not as continuous as that observed in the conventionally heated samples. SEM analysis showed that the chrome appeared to diffuse into the first 2 to 10 µm at the samples surfaces. The MW samples appear to be less uniformly chromized. Unlike conventional samples that display a nearly continuous diffused chrome region near the surface, MW samples seem much more prone to show patches of chrome diffusion. A possible solution is substituting metallic chrome powder for the Cr_2O_3 source used in these experiments. The weight losses shown in Fig. 6 suggest iron volatilization from the surfaces. Energy dispersive X-ray (EDX) analysis indicated higher-than-normal iron concentrations near the surface of the chromized 625 nickel samples provide further evidence that iron volatilization from the other samples occurred during chromization of all three alloys. Chromized A2 and 1018 alloys display many similarities by XRD analysis. Typical phases found in these samples are Cr_7C_3 and Cr_2O_3. Once more, though, the MW samples treated at 850°C deviate significantly from this trend. Carbide formation was not unexpected, since it is a very stable phase, and often steels are aluminized first before chromization to prevent such a layer.[11]

Boronization Results

Boron-based coatings were successfully deposited on all three alloys using MW heating. Approximately, 350 W of power was needed to maintain the temperature at 850 and 925°C for the 40 g of samples and the ~350 g of packing powder. For control experiments, approximately

1050 W of microwave power was needed to maintain the temperature at 850 and 925°C with the alumina powder pack. Coatings were also deposited on all alloys using conventional heating. Very little changes in hardness changes were observed for the hardness at the surface of the 625 and 1018 alloys processed by microwave and conventional heating (data not shown). Since borides of iron or nickel are not very stable, this is not surprising. However, very interesting results were obtained for the A2 tool steel, where the alloying additive readily form borides. Figure 7 shows a graph of the nanohardness values for the A2 as-received material, A2 material processed in the tube furnace (TF), and A2 material processed in the MW furnace with alumina powder or with B_4C present. Both the MW controls heated in alumina powder alone and the A2 material heated in the B_4C pack using MW heating showed higher hardness values than the control untreated sample. Figure 8 shows a finer analysis of the near surface of the A2 samples processed in a tube furnace (top) or a microwave furnace (bottom). Due the size of the nano-indentor tip, one sees large hardness spikes as the tip hits hard and softer grains. Figure 9 shows micro-hardness hardness data for A2 alloy processed under several conditions. The microhardness tester was chosen to obtain a larger indent size, and thus a more average number is obtained than with the nano-indentor. The data shows the highest hardness for MW-treated samples packed in B_4C powder. The next two higher hardness levels were for A2 samples packed in alumina powder and processed conventionally and for A2 samples packed in alumina powder and processed using MW heating. The hardness values for the as-received A2 sample and the A2 samples processed in the tube furnace with B_4C powder were much lower. Figure 9 data suggests that there is a surface hardening effect with not only B_4C powder, but

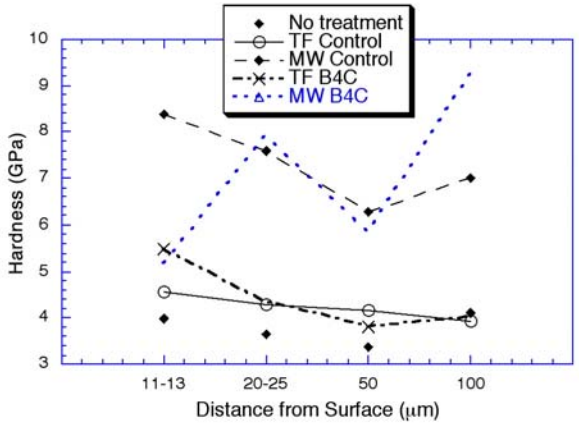


Fig. 7. Hardness as measured with a nano-indentor for as-received A2 steel, control samples heated in Al_2O_3, and samples treated in a B_4C pack at 925°C for 4 h using conventional (TF) and MW heating.

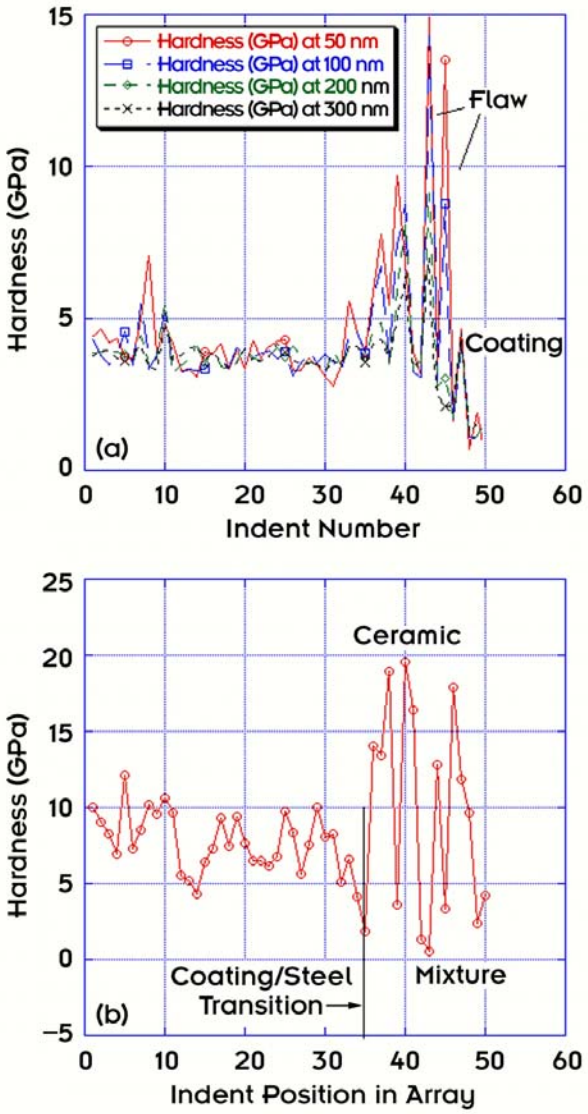


Fig. 8. Nano-indent hardness data for A2 alloy heated in a B4C pack at 925°C for 4 h by conventional (top) or MW heating (bottom).

also the alumina powder with the A2 steel alloy. The fact that the A2 alloy heated in the MW transparent alumina powder bed was an unexpected result.

Summary and Conclusions

Results of this broad, scoping project indicate that the MW heating shows strong potential as a method for surface treating alloys to obtain deeper diffusion of alloy elements into metal parts. Very low MW power is required to heat pack cement parts to 925°C. In addition, surprising results were obtained in control samples processed using MW heating in aluminum oxide, with no boron-hardening additive. These results indicate that microwave heating may be an additional method for surface hardening of metal

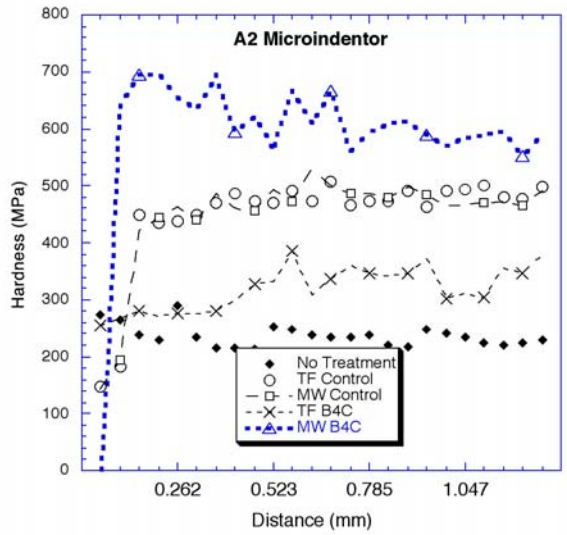


Fig. 9. Hardness as measured with a nano-indentor for an A2 alloy control sample and an A2 alloy sample heated using MW processing with B_4C packing at 925°C for 4 h.

parts. The work described has also led to an interest and FY 2003 funding provided by the Office of Energy Efficiency and Renewable Energy for pack cementation of ceramic parts for improved corrosion resistance.

References

[1]W. H. Sutton, "Microwave Processing of Ceramics—An Overview," pp. 33–38 in *Microwave Processing of Materials III.* MRS Symposium Proceedings Volume 269, Pittsburgh, 1992.

[2]R. Bianco, R. A. Rapp, and J. L. Smialek, "Chromium and Reactive Element Modified Aluminide Diffusion Coatings on Superalloys: Environmental Testing," *Journal of the Electrochemistry Society* **140**, 1191–1203 (1993).

[3]S. L. McGill et al., "Effects of Power Level on the Microwave Heating of Selected Chemicals and Minerals," pp. 247–252 in *Mater. Res. Soc. Symp. Proc.*, Vol. 124, Mater. Res. Soc., Pittsburgh, Pa., 1988.

[4]D. F. Stein et al., "Microwave Processing—An Emerging Industrial Technology," pp. 3–7 in *Mater. Res. Soc. Symp. Proc.*, Vol. 347, Mater. Res. Soc., Pittsburgh, Pa., 1994.

[5]S. J. Rothman, "Critical Assessment of Microwave-Enhanced Diffusion," pp. 9–18 in *Mater. Res. Soc. Symp. Proc.*, Vol. 347, Mater. Res. Soc., Pittsburgh, Pa., 1994.

[6]M. A. Janney et al., "Enhanced Diffusion in Sapphire During Microwave Heating," *J. Mat. Sci.* **32**, 1347–1355 (1997).

[7]Z. Xie et al., "Microwave Processing and Properties of Ceramics with Different Dielectric Loss," *J. Eur. Ceram. Soc.*, **19**, 381–387 (1999).

[8]D. E. Clark, W. H. Sutton, and D. A. Lewis, "Microwave Processing of Materials," pp. 61–96 in *Microwaves: Theory and Application in Materials Processing IV,* American Ceramic Society, Westerville, 1997.

[9]J. G. P. Binner and T. E. Cross, "Microwave Processing in the UK," pp. 33–28 in *Microwave Processing of Materials III.* MRS Symposium Proceedings Volume 269, Pittsburgh, Pa., 1992.

[10]C. H. Koo and T. H. Yu, "Pack Cementation Coatings on Ti_3Al-Nb Alloys to Modify the High-Temperature Oxidation Properties," *Surface and Coatings Technology* **126**, 171–180 (2000).

[11]Z. Minhui and R. A. Rapp, "Simultaneous Aluminizing and Chromizing of Steels to Form $(Fe, Cr)_3Al$ Coatings," *Oxidation of Materials* **49**, 19–31 (1997).

Ultrasonic Processing of Ultrafine Materials

Q. Han,[1] C. L. Xu,[2] G. R. Romanoski,[1] D. T. Hoelzer,[1] M. M. Menon,[3] and R. P. Taleyarkhan[4]
[1]*Metals and Ceramics Division*
[2]*University of Tennessee*
[3]*Nuclear Science and Technology Division*
[4]*Engineering Science and Technology Division*

High-intensity ultrasonic energy can be used to produce oscillating strain/stress fields in solid materials. The purpose of this seed money project was to explore novel use of the strain/stress induced by ultrasonic vibration for the production of ultrafine materials. Two new approaches have been developed conceptually and tested experimentally. The first approach was to use ultrasonically induced strain/stress to induce plastic deformation in materials. The second approach was to use ultrasonically induced oscillating stress fields to induce repeated phase transformation in the solid materials. Initial experimental results indicate that, for the first time, nanostructures are formed in the ultrasonically treated metal. The size of the nanostructures obtained using the first approach is smaller than 200 nm and possibly in the range of 100 nm.

Introduction

Nanostructured materials offer unique and entirely different mechanical, electronical, optical, and magnetic properties compared with conventional micro- or millimeter-sized materials. In metals, for example, the hardness of nanocrystalline copper increases with decreasing grain size, and 6-nm copper grains show five times the hardness of conventional copper.[1] Nanosized Al-Ni-In alloys exhibit a tensile strength larger than 1200 MPa, greater than conventional high-strength aluminum alloys.[2] Nanostructured M50 steel is more fatigue and fracture resistant than conventional M50 steel that is widely used in the aircraft industry as the main-shaft bearings in gas turbine engines.[3] In semiconductors, nanoclusters are often referred to as quantum dots, nanocrystals, and Q-particles. As an example, quantum dots can be developed to emit and absorb a desired wavelength of light by changing the particle diameter. This feature allows the construction of a finely tunable and efficient semiconductor laser.

Traditional methods for producing nanostructured materials include gas atomization and ball milling, followed by consolidation and rapid solidification.[4–6] These processes tend to be expensive and capable of producing only small amounts of material. Recently approaches for producing nanostructured materials include severe plastic deformation.[7–8] equal channel angular extrusion (ECAE) is one of the methods that uses severe plastic deformation to produce nanostructured materials, but the size of the sample is limited due to the high friction forces that exist between the material and the wall.[9] These developments suggest that methods that can produce severe plastic deformation have the potential in the production of nanostructured materials.

Ultrasonic vibrations produce oscillating strain and stress fields in solid materials.[10] Experiments have shown that ultrasonic vibrations above a critical intensity increase the concentration of dislocations and point defects in the solids. The density of the defects increases with increasing ultrasonic intensity. When the density of the dislocations is high enough, an alignment of dislocations occurs. It has been observed that a distinct cellular structure, about 2 μm in size, is formed in ultrasonically treated aluminum monocrystals after it was exposed to ultrasonic vibration for only 8 s.[11] The experimental evidence suggests that ultrasound is a powerful tool in producing severe plastic deformation in solid materials. However, the grains size was still a few orders of magnitude larger than the size of nanostructures. Also the oscillating nature of the strain and stress induced in solid metal by ultrasonic vibrations is not explored for grain refinement.

The purpose of this research was to investigate the possibility of producing nanostructures in metals using ultrasonic vibrations. Two new approaches have been developed conceptually and tested experimentally. Initial experimental results indicate that, for the first time, nanostructures are formed in the ultrasonically treated metal. The size of the nanostructures obtained using the first approach is smaller than 200 nm and possibly in the range of 100 nm.

Technical Approach

The First Approach

The idea in the first approach was to use the ultrasonically induced strain/stress to enhance dislocation and vacancy densities in metal in the hope that the line/point defects can align themselves to form ultra-fine structures. Extruded 1010 steel was used, and the initial grain size in the steel was about 10 μm. A 20-kHz, 1.5-kW ultrasonic unit was used for the experiments. Initially ultrasonic energy was directly injected into steel bars. No substantial grain refinement was observed, and the bar fractured due to fatigue. This was because in an oscillating stress field, the stresses are periodic (positive and negative). Usually materials fracture more easily under tensile stress.

A novel approach was then developed to offset the instantaneous tensile stress induced by ultrasonic vibration. High-intensity ultrasonic energy was then focused on specific locations in the specimen for 60 s, resulting in a severe plastic deformation in those locations. The deformed region was polished, etched, and characterized using scanning electron microscopy (SEM) and transmission electron microscopy (TEM).

The Second Approach

The idea in the second approach was to inject acoustic energy into metal at its solid/solid phase transformation temperature such that the oscillating stress generated by ultrasonic vibration may lead to a repeated phase transformation at high frequency. The solid/solid phase transformation temperature varies with external stress. Under the influence of oscillating stress, the phase transformation temperature will also oscillate around its equilibrium temperature. When the specimen under acoustic vibration is held at the equilibrium solid/solid phase transformation temperature, oscillating undercooling occurs for both phases, which may result in repeated nucleation for both of the phases involved. Since the diffusion coefficient in the solid state is very small, the growth rate of the newly formed phase will be small, and the grain size of the newly formed phase will be mainly determined by the nucleation stage of the phase transformation.

In the second approach, bar specimens were made out of 1010 steel and placed in a furnace with a known temperature gradient. The temperature was 1000°C at one end of the specimen and was room temperature at the other end to ensure that the α to γ phase transformation occurred at a certain location in the specimen. Ultrasonic energy was then injected into the specimen for 5 to 20 s until the specimen was quenched in water to preserve the grain size at high temperatures.

Results and Accomplishments

Ultrasonically Induced Plastic Deformation

Sever plastic deformation occurred in the specific location of specimen injected with high-intensity ultrasonic vibration under compressive stress. Optical microscopy revealed that the original α grains were severely deformed into thin strips. The SEM image shown in Fig. 1 indicates that the strips are narrower than 200 nm. Within each nanosized strip, a large number of small features/grains in the neighborhood of 200 nm can be barely seen.

Figure 2 is a TEM image showing the size of the nano particles/grains. A few particles less than 200 nm can be clearly seen. On the top side of the image, dislocation densities are extremely high, indicating that severe deformation occurred in those locations.

Results shown in Figs. 1 and 2 indicate that grain size smaller than 200 nm have been obtained by injecting ultrasonic energy into 1010 steel specimens. By using ultrasonic vibrations, the smallest grain size never before obtained is 200 nm. The grain sizes are in the nanosize scale, and it was one of the objectives of this project.

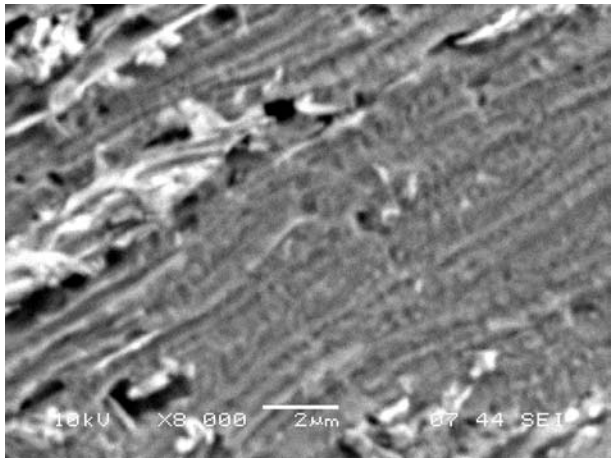


Fig. 1. An SEM image of the severely deformed region in the 1010 steel specimen. The fine strips are the deformed α phase, and the coarse structure are the pearlite. Most of the fine strips are narrower than 200 nm and contain a large number of particles smaller than 200 nm.

Repeatedly Induced Phase Transformation

Results obtained in the second approach were not as anticipated. When ultrasonic energy was injected into the test bar (~1.9 cm diam), the specimen was quickly broken at the location corresponding to the δ to γ phase transformation (A3 temperature). This phenomenon, which was not reported before, might be related to the volume change during repeated δ–to-γ phase transformation at high frequencies.

In order to observe the microstructure at the δ–to-γ phase transformation, the specimen was quenched before

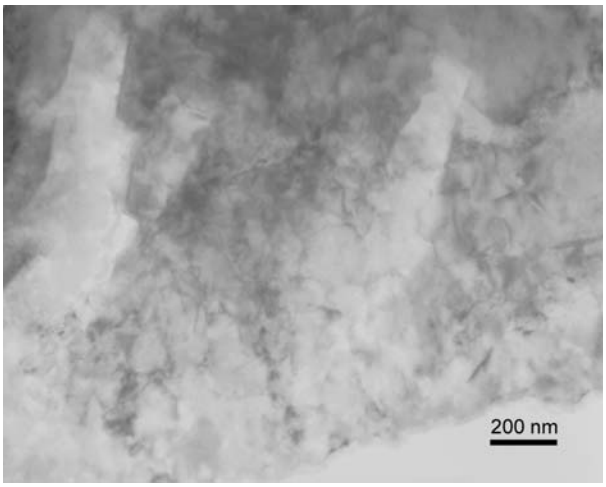


Fig. 2. A TEM image showing particles smaller than 200 nm in the severely deformed region. The dark areas on the image contain a large number of dislocations.

being broken. Figure 3 shows the microstructure of the quenched specimen. Parallel plates like those of martensite structure occurred in the specimen. Fractures can be observed at the grain boundaries. Adjacent to the cracks, a large number of nanosized features/grains exist, which has also not been reported before. It looks as if the δ–to-γ phase transformation occurred only at the grains boundaries, and the reactions didn't reach to the center of the previous α grains before cracking occurred.

Although bulk nanostructures were not really obtained in this approach, two nanosized features were observed: one occurred at the grain boundaries adjacent to the cracks, and the other was the two-dimensional parallel plates, which might be martensite.

Fig. 3. A TEM image showing the microstructure in the quenched specimen. The specimen was subject to ultrasonic vibrations for 15 s at its α–to-γ phase transformation temperature before quenching. Cracking occurred along the grain boundaries.

Summary and Conclusions

Nanostructures have been obtained in 1010 steel using ultrasonically induced severe plastic deformation. The size of the grains in the severely deformed regions is smaller than 200 nm.

The experiments on repeated phase transformation induced by ultrasonic vibration revealed some new phenomena not reported in the open literature. Fine, nanosized grains occurred near the grain boundaries. Parallel nanoplates form during quenching of 1010 steel injected with ultrasonic energy.

The project results have the potential for scaling up for the production of critical metal components for transportation and defense applications, leading to energy savings. Contact will be made with DOE and DOD program managers for future funding to continue this work.

References

[1]H. S. Nalwa, *Handbook of Nanostructured Materials and Technology*, ed. H. S. Nalwa, Academic Press, 1996, p. ix.

[2]K. Kita, K. Saito, A. Inoue, and T. Matsumoto, *Mater. Sci. Eng.* **A226–228**, 1004 (1997).

[3]K. E. Gonsalves, S. P. Rangara, and J. Wang, *Handbook of Nanostructured Materials and Technology*, ed. H. S. Nalwa, Academic Press, 1996, p. 27.

[4]R. W. Seigel, *MRS Bulletin* **15**, 60 (1990).

[5]M. Umemoto, Z. G. Liu, and K. Masuyama, *Scripta Mater.* **44**(8–9), 1741–45 (2001).

[6]I. G. Brodova, D. V. Bashlykov, and A. B. Manukhin, *Scripta. Mater.* **44**(8–9), 1761–64 (2001).

[7]I. V. Alexandrov and R. Z. Valiev, *Scripta Mater.* **44**(8–9), 1605–08 (2001).

[8]I. Altenburger, B. Scholtes, and V. Martin, *Mat. Sci. Eng.* **A264** (1–2), 1–6 (1991).

[9]K. Zhang, I. V. Alexandrov, and K. Lu, *Nanostruct. Mater.* **9**(1–8, 347–350 (1997).

[10]O. V. Abramov, High-Intensity Ultrasonics Theory and Industrial Applications, Gorden and Breach Science Publishers, Singapore, 1998.

[11]B. Langenecker, *Proc. High Power Ultrasonics*, 1971, p. 32.

Compound Semiconductors on Silicon: Shedding Some Light on the Matter

O. W. Holland,[1] B. C. Larson,[1] D. B. Beach,[2] and M. F. Chisholm[1]
[1]*Condensed Matter Sciences Division*
[2]*Chemical and Analytical Sciences Division*

The integration of lattice-mismatched material onto Si wafers is key to increasing the functionality of integrated circuits and, in particular, to integrating both optical and digital processing on a single chip. A method is proposed which is capable of providing a growth template on Si for films with a lattice mismatch of ≤4%. It involves implantation of Ge into Si to form a dilute solid solution. Enrichment is achieved by thermal oxidation of the implanted layer to snowplow the Ge ahead of the growing oxide interface to form a uniform pseudomorphic SiGe film. The composition of the film (and thus the lattice parameter) can be varied from pure Ge to Si rich by adjusting the oxidation conditions. Once formed and relaxed to its natural lattice spacing, the SiGe film is an ideal template for growth of defect-free GaAs and/or strained-layer SiGe layers. Methods of achieving this relaxation without plastic deformation are discussed.

Introduction

Silicon, an indirect bandgap semiconductor, is well suited for digital electronics but is an inefficient emitter of light. Many III-V compound semiconductors such as GaAs have a direct bandgap and, as such, are ideally suited for optical applications. However, the lattice mismatch (4.1%) between silicon and GaAs is too large for growth of defect-free, oriented, single-crystal films. Ge is lattice matched to GaAs and can be grown on Si by a number of deposition techniques, but the large misfit results in the creation of misfit dislocations, which partially relieve the strain. Dislocations that thread through the film have the potential to severely degrade the optical and electrical integrity of any material grown on this layer.

Technical Approach

The proposed process involves sequential Ge^+-implantation and thermal oxidation of the top Si layer in silicon-on-insulator (SOI) material. Implantation initially yields a dilute SiGe mixture in the near surface that is progressively Ge enriched during oxidation until it forms a distinct pseudomorphic GeSi-layer on the Si substrate. The method relies on the complete rejection of the implanted Ge from the oxide during growth. This segregation results in the formation of a distinct Ge-rich layer at the interface that is epitaxially oriented on the underlying Si. A method is demonstrated for relaxing the strain in the Ge layer without plastic deformation so that it can be used as an intermediate layer to facilitate growth of lattice mismatched films on Si. The process involves decoupling the Ge layer from the Si substrate by using SOI material. The material consists of a buried oxide layer (BOX) beneath the surface that isolates a top Si layer from the bulk Si. In general the Ge layer formed by this process is biaxially strained near the lattice parameter of Si. The temperatures involved in the oxidation of Si may allow the germanium layer to partially relax to its natural lattice spacing (i.e., as a result of viscoelastic flow of the SiO_2 or buckling of the thin Ge film). Further relaxation of the film has was achieved by ion irradiation without misfit dislocation formation.

This process is substantially different from other growth techniques that generally achieve layer-by-layer growth by atomic attachment at a free surface. The proposed two-step process is more akin to a bulk process in that it does not involve a free surface (although it does depend on a interfacial boundary), and it achieves layer formation by enrichment rather than by an adatom process. These differences offer substantial benefits in forming strained layers that yield a greater thermal stability (against relaxation) and, thus, different pathways or mechanisms for relaxation. The thickness of the pseudomorphic Ge-rich layers formed by this technique has been shown to far exceed the critical value established by Matthew[1] (i.e., the thermodynamic limit). The kinetic barrier for forming misfit dislocations is sufficiently great that relaxation generally occurs only when the elastic strain within the pseudomorphic film far exceeds its critical value (i.e., where dislocation formation is energetically favored). Thus, studies have shown that the onset of misfit formation can only occur as a result of nonhomogeneous strain with the layer. Such is the cause when growth involves a free surface that can develop morphological instabilities. It has been shown that such instabilities (i.e., roughening) redistribute the strain within the film to produce localized

strain that is capable of nucleating misfit growth. Thus, the absence of a free surface and the inherent stability against roughening in the layers formed by the proposed technique is considered to be responsible for its substantial benefits over other techniques.

Results and Accomplishments

Cross-section, transmission electron micrographs (XTEM) in Fig. 1 show a SiGe film formed by this technique. The sample shown is an SOI wafer implanted with Ge at 80 keV at a dose of $5 \times 10^{16} cm^{-2}$. It was oxidized in dry O_2 in several stages. An initial cycle of 1150°C/3 h was used followed by a 1050°C cycle and a final cycle at 930°C. The oxidation temperature was decreased as the Ge content of the film increases to avoid melting (which occurs at 934°C for pure Ge). These oxidation cycles removed most of the top Si layer and left only a well-defined SiGe layer encapsulated between the thermal oxide (on the top) and the buried oxide (BOX) of the SOI. Both the bright-field image in Fig. 1(a) and the dark-field image in Fig.1(b) demonstrate that partial relaxation of the film (~50%) has occurred by dislocation formation. Interestingly, both micrographs indicate that the segregated film has bifurcated into a Ge-rich layer at the top and a Si-rich layer at the bottom. Threading dislocation segments can be seen originating from the bilayer interface. The results demonstrate that the viscoelastic response of the BOX is insufficient to accommodate the strain in the SiGe film. These results are consistent with those by Hobart,[2] who reported relaxation in a compressively strained SiGe film bonded to a borophosphorosilicate glass by buckling. Buckling was avoided by patterning the film into small areas prior to annealing (i.e., relaxation).

It is clear that another technique is required to achieve defect-free relaxation of the segregated film. XTEM of a sample that was processed similarly to the one previously discussed is shown in Fig. 2. It is clear from the figure that the SiGe film (estimated to be 70% Ge) is quite uniform in thickness with planar interfaces at both oxide layers. There is no indication of bifurcation within the film. An inspection across the entire field of view (~several microns) within the microscope failed to find any dislocations within the film. This yields a rough estimate for the upper limit for threading dislocations density of $\sim 10^7 cm^{-2}$. The main difference in processing of this sample compared to the one shown in Fig. 1 was that it was irradiated (with 200 keV Si^+ ions) prior to the final oxidation cycle at 900°C in an attempt to promote relaxation within the Ge layer. While no additional relaxation was observed compared to a non-irradiated control, the ion mixed sample achieved a 50% relaxation of the in-plane spacing without plastic deformation (i.e., no misfit dislocations), while a high dislocation density was observed in the control sample.

The use of energetic ions to promote relaxation of highly strained, pseudomorphic films offers an innovative method to form dislocation-free, relaxed films. The effect derives from the ability of the energetic ions to transfer sufficient energy to the Ge atoms in the film to physically remove them from the film. This ballistic effect is shown pictorially in Fig. 3. The biaxial compressed film is relaxed by the physical removal of excess Ge atoms in the film, which allows the film to relax to its natural lattice spacing.

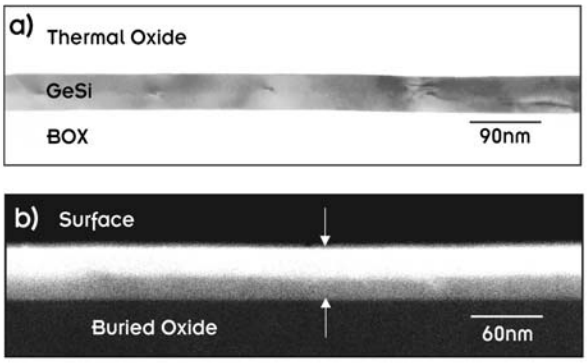


Fig. 1. XTEM micrograph of a Ge-rich film encapsulated within a SOI material by a buried oxide (BOX) and a superficial thermal oxide at the surface. The SOI was implanted with 80 keV Ge^+-ions at a fluence of $5 \times 10^{16} cm^{-2}$. (a) Bright field image revealing a dislocation array near the centerline of the film with threading segments spanning to the top and bottom interface. (b) z-contrast image showing bilayer: top 24 nm of film is Ge rich compared to bottom 16 nm. Viscoelastic response of the oxide substrate (BOX) is insufficient for large area films to achieve defect-free relaxation of the encapsulated film. Strain relaxation is observed, but it appears to be predominantly due to misfit dislocation formation.

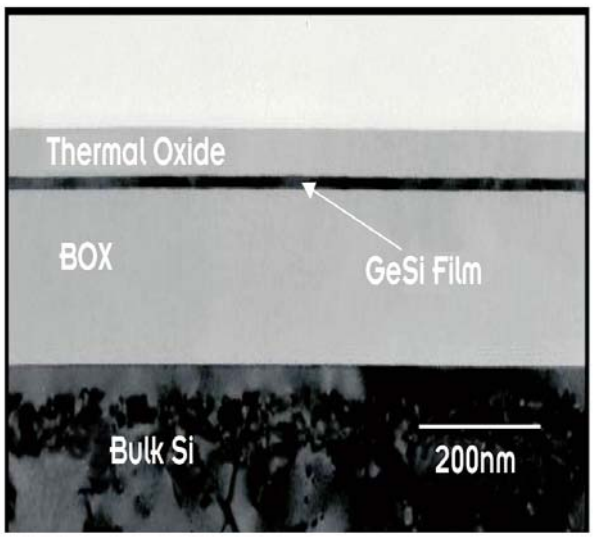


Fig. 2. XTEM of a continuous 15 nm $Si_{0.3}Ge_{0.7}$ film with no dislocations by inspection over a wide area.

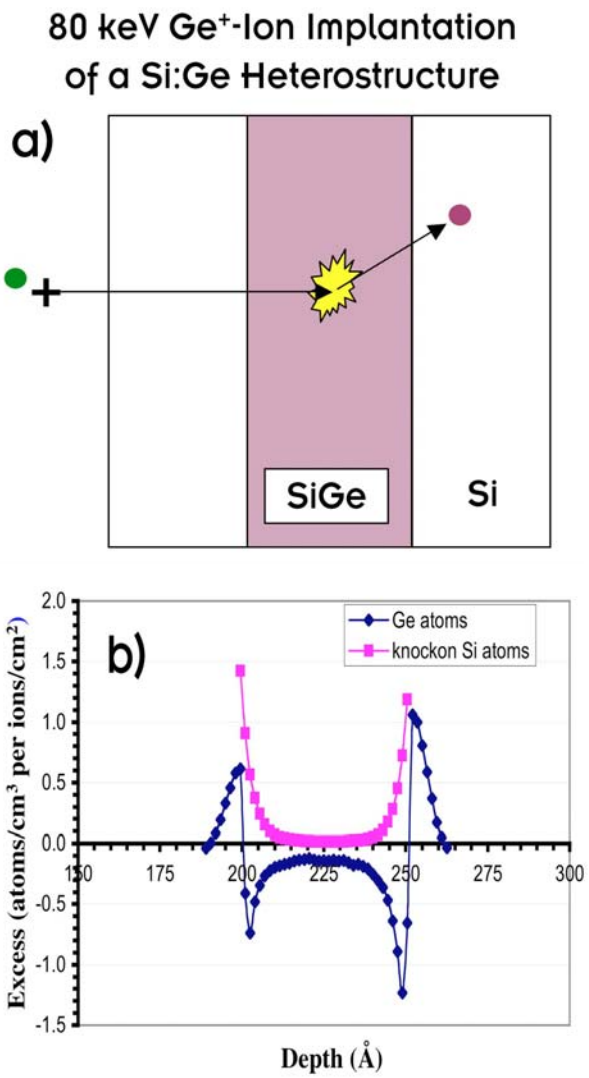


Fig. 3. (a) Ballistic rarefaction of Ge film by ion implantation, and (b) TRIM simulation of rarefaction process demonstrating the ability of energetic ions to ballistically remove atoms from the film to allow it to relax without dislocation formation.

Results of a computer simulation of this process are also shown in Fig. 3. The simulation shows that, although Ge atoms are removed from the film, Si atoms are knocked into the film. However, the process strongly favors removal of Ge atoms due to the larger scattering cross section of the ions with the higher atomic number. Simulation results show that complete relaxation can be achieved at a dose of ~10^{16}cm^{-2}.

Summary and Conclusions

A method of forming a dislocation-free, relaxed Ge layer on SOI was demonstrated. The method involved Ge^+-implantation and thermal oxidation to form a thin, pseudomorphic Ge layer in SOI. Relaxation of the segregated Ge film was achieved using a ballistic effect due to the interaction of energetic ions with the Ge atoms. The partial relaxation of the film was achieved without any plastic deformation in the film. Fully optimized, this technique should be capable of achieving almost full relaxation of the Ge films, which will permit integration of defect-free GaAs and other III-V films onto Si substrates.

This project explored methods of relaxing strained thin films without plastic deformation (dislocations). An innovative technique involving the use of energetic ions to promote relaxation was discovered to produce 50% relaxation of the strain without the introduction of dislocations. While the method is rather specialized, the principal benefit will be increased understanding of strain relaxation in thin films. This improves the basic science foundation of the DOE and other programs that rely on the integration of lattice mismatched film/substrate systems.

References

[1]J. W. Matthews and A. E. Blakeslee, *J. Cryst. Growth (Netherlands)* **27**, 118 (1974).

[2]K. D. Hobart, F. J. Kub, M. Fatemi, M. E. Twigg, P. E. Thompson, T. S. Kuan, and C. K. Inoki, *Journal of Electronic Materials* **29**, 897 (2000).

Ferromagnetism in Dilute Magnetic Semiconductors: Getting to the Science with Neutron Scattering

B. C. Sales,[1] D. Mandrus,[1] H. A. Mook, Jr.,[1] L. A. Boatner,[1] and T. C. Schulthess[2]
[1]*Condensed Matter Sciences Division*
[2]*Computer Science and Mathematics Division*

This project focused on the growth and characterization of single crystals of dilute magnetic semiconductors (DMS). DMS are one of the key components in the next generation of electronic devices that make use of the electron spin as well as its charge. The origin of the ferromagnetism in DMS is not well understood, in part because virtually all of the DMS have been prepared in thin film form using nonequilibrium synthesis methods. Single crystals can be studied with a larger variety of experimental methods, including neutron scattering. A large number of magnetically doped single crystals were prepared during the early stages of this project (chalcopyrites, ZnO, $KTaO_3$, Ge, and TiO_2—doped with Mn, Co, and Fe). Initial studies on these crystals indicated that much of the room-temperature ferromagnetism reported in the literature was due to clustering and/or an impurity phase. To avoid clustering, we have grown and started to characterize large crystals (≈0.7 g) of a stochiometric DMS: $Yb_{14}MnSb_{11}$. In this ferromagnetic compound, all of the Mn atoms are located at a well-defined crystallographic site, and hence there is no problem with clustering. $Yb_{14}MnSb_{11}$ and related compounds appear to be ideal materials on which to unravel the magnetism in DMS. Because of the extra work involved in showing that earlier studies were incorrect, neutron scattering experiments were not conducted as part of this seed money. However, the $Yb_{14}MnSb_{11}$ materials were shown to be extremely interesting and DOE Basic Energy Sciences funding has been redirected to fund neutron scattering experiments that are currently in progress.

Introduction

"Until recently, the spin of the electron was ignored in mainstream charge-based electronics. A technology has emerged called spintronics (spin transport electronics or spin based electronics), where it is not the electron charge but the electron spin that carries information, and this offers opportunities for a new generation of devices combining standard microelectronics with spin-dependent effects that arise from the interaction between the spin of the carrier and the magnetic properties of the material."[1] This quote from a recent review article in *Science* summarizes the emerging field of spintronics that includes everything from spin valves to quantum computing. This is a vast new area of science and technology that has already resulted in new read heads for computer hard drives (based on a spin valve)[1] and the first demonstration of a primitive quantum computer.[2]

One of the key components in the next generation of proposed spintronic devices is a ferromagnetic semiconductor that can be used as a source of current in which the spins of the carriers are highly polarized.[1,3] The recent observation of ferromagnetism near room temperature in semiconductors containing a few percent of magnetic atoms was a surprise and is not well understood.[4,5] To date, all of the DMS have been prepared in thin-film form using nonequilibrium synthesis methods. Neutron scattering is the most powerful technique for probing and understanding magnetism in a solid, but it does not have enough sensitivity to study a DMS thin film. The goal of this project is to identify and prepare one or more DMS as a large single crystal (>1 mm^3) suitable for neutron scattering investigations.

Technical Approach

A large number of different materials were grown as single crystals based on various reports in the literature. These reports usually indicated the observation of ferromagnetism in a thin-film semiconductor doped with a few percent of a magnetic element such as Co, Mn, or Fe via a nonequilibrium method. Based on these reports, we prepared crystals of the following materials:

1. ZnO doped with Mn or Co (Crystals were grown via vapor transport in a sealed silica tube at temperatures between 600–800°C using NH_4Cl as a transport agent.)
2. Ge doped with Mn (Crystals grown by slow cooling of melt in a silica crucible from 1100°C or by quenching the melt to room temperature followed by annealing.)

3. $CdSiP_2$ doped with Mn (Phosphides with the chalcopyrite structure were grown from a molten tin flux. The excess tin was removed from the crystals using a centrifuge.)
4. $ZnSnAs_2$ doped with Mn (Arsenides with the chalcopyrite structure were grown by pre-reacting the elements in a sealed silica tube at 600°C followed by melting at 1150°C and slow cooling.)
5. TiO_2 crystals doped with Co (Small doped crystals of TiO_2—anatase structure—were grown via vapor transport.)
6. $KTaO_3$ doped with Mn, Fe, Co, and Ni (Large crystals of doped $KTaO_3$ were grown using a self-flux method.)

Preliminary structural, magnetic, and electrical measurements were made on all of the crystals listed above. We found no evidence of ferromagnetism in any of the magnetically doped semiconducting crystals. In some of the materials (such as $KTaO_3$), the equilibrium solubility of magnetic elements in the structure was too low (0.1–0.3%); in others (ZnO), the coupling between the magnetic elements was too weak to produce ferromagnetism above 4 K. In several of the crystals (Ge: Mn and $ZnSnAs_2$: Mn), ferromagnetism could be traced to a metallic impurity phase.

The disappointing results discussed in the previous paragraph indicated a need to change our technical approach. The problem of clustering and/or impurity phases in a semiconducting host could be eliminated if we could find a ferromagnetic semiconducting compound where all of the magnetic ions are at well-defined but widely separated crystallographic sites. The physics of the magnetism in such a compound would also be much easier to study using neutron scattering. David Mandrus pointed out that there is indeed such a class of materials, the Zintl compounds such as $Ca_{14}MnSb_{11}$. These semiconductors are ferromagnetic at about 60 K,[6] contain only 3.8 at. % of the magnetic element Mn, and the Mn ions are at well-defined crystallographic sites that are separated by at least 1 nm.

Results

We have been able to grow large crystals (0.7 g) of the Yb member of this family of materials ($Yb_{14}MnSb_{11}$) from a Sn flux.[7] This compound crystallizes in the tetragonal $Ca_{14}AlSb_{11}$ structure-type, $I4_1/acd$ space group (Z = 8) with a = 16.61 Å and c = 21.95 Å. $Yb_{14}MnSb_{11}$ has 208 atoms in the unit cell and contains only 3.8 at. % Mn. The Yb is divalent in this compound and hence is not magnetic. A photograph of one of the flux-grown crystals is shown in Fig. 1. Magnetic susceptibility and magnetization data from one crystal (Fig. 2) clearly demonstrate that this compound is ferromagnetic below

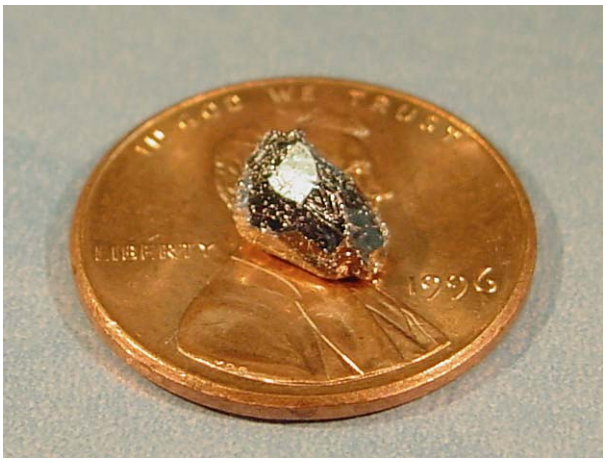


Fig. 1. Flux grown crystal of $Yb_{14}MnSb_{11}$.

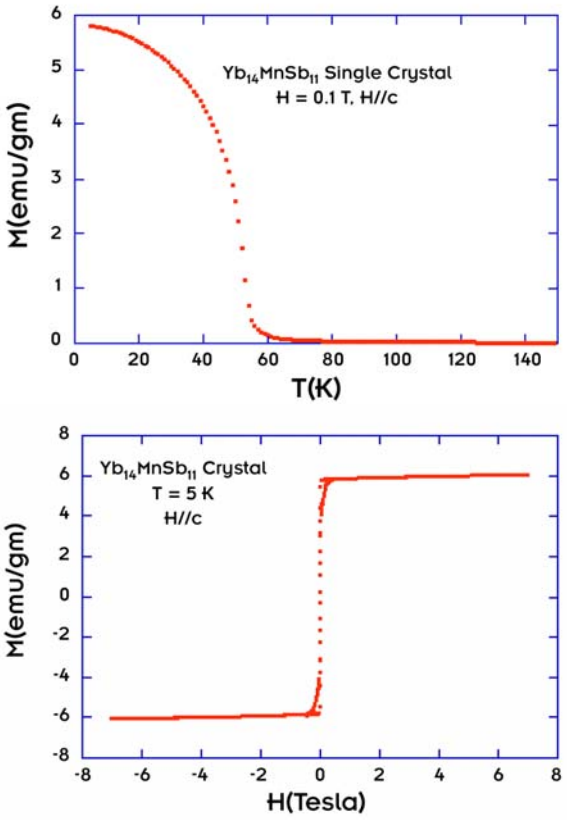


Fig. 2. Magnetic susceptibility (top) and magnetization data (bottom) from a $Yb_{14}MnSb_{11}$ crystal. The ferromagnetic transition at $T_c \approx 55$ K is clearly evident.

about 55 K. The magnitude of the saturation magnetization corresponds to a magnetic moment of about 4 μ_B per Mn. The electrical resistivity of undoped $Yb_{14}MnSb_{11}$ crystals is about 1 mΩ-cm at room temperature, which is indicative of a heavily doped semiconductor. Although we have still not unambiguously measured the carrier concentration in this material, we have determined from doping studies that

the dominant carriers are holes. We have grown single crystals of La-doped $Yb_{14}MnSb_{11}$ and Te-doped $Yb_{14}MnSb_{11}$. Both of these dopants should increase the concentration of electrons. Since the Curie temperature,

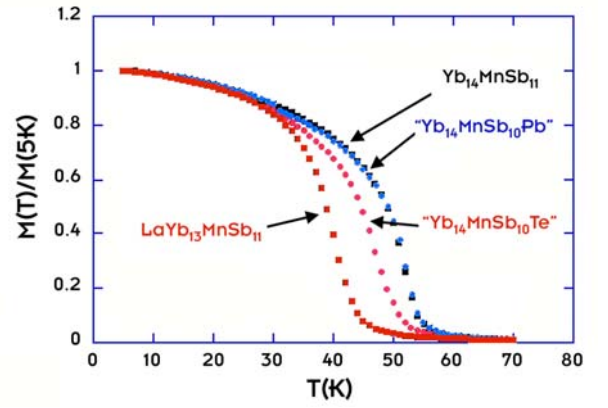


Fig. 3. Normalized magnetization data from undoped and doped crystals of $Yb_{14}MnSb_{11}$. Electron doping with La or Te lowers T_c and suggests that holes are the dominant carriers in $Yb_{14}MnSb_{11}$.

T_c, decreases with La or Te additions, it is likely that the overall carrier concentration has decreased, suggesting that holes are the dominant carriers. The decrease of T_c with electron doping is shown in Fig 3.

Summary and Conclusions

A large number of crystals were grown in an attempt to duplicate many of the claims in the literature concerning the existence of DMS. Several claims of DMS in bulk materials (Ge:Mn for example) were found to be due to impurity phases and poor science. For other material systems, we found that equilibrium crystal growth was not able to duplicate results found in thin films prepared using nonequilibrium methods. Either the solubility limit of various magnetic elements in the semiconducting host was too low to induce ferromagnetism ($KaTaO_3$: Mn for example) or the carrier concentration could not be increased enough to couple the magnetic moments to each other (ZnO: Mn for example).

We found that a more successful approach was to prepare single crystals of an unusual class of ferromagnetic semiconducting compounds where all of the magnetic ions are at well-defined but widely separated crystallographic sites: $Yb_{14}MnSb_{11}$ or $Ca_{14}MnSb_{11}$ for example. We found we could grow relatively large crystals (0.7 g) of $Yb_{14}MnSb_{11}$ and could change the carrier concentration through doping on one of the non-magnetic sites (Yb or Sb sites) with La or Te. This system appears to be ideal to investigate the physics of carrier-mediated ferromagnetism, a phenomenon that is at the heart of all DMS materials envisioned for spintronic applications. Because these new materials were discovered near the end of the seed money project, there was no time for neutron scattering experiments. However, the initial results from the $Yb_{14}MnSb_{11}$ were so interesting that DOE Basic Energy Science funding has been redirected to study these materials and neutron scattering experiments are currently in progress.

References

[1]S. A. Wolf, D. D. Awschalom, R. A. Buhrman, J. M. Daughton, S. von Molnar, M. L. Roukes, A. Y. Chtchelkanova, and D. M. Treger, *Science* **294**, 1488 (2001).

[2]L. M. K. Vandersypen, M. Steffen, G. Breyta, C. S. Yannoni, M. H. Sherwood, and I. L. Chung, *Nature* **414**, 883 (2001).

[3]S. Das Sarma, J. Fabian, X. Hu, and I. Zutic, *Solid State Commun.* **119**, 207 (2001).

[4]H. Ohno, A. Shen, F. Matsukura, A. Oiwa, A. Endo, S. Katsumoto, and Y. Iye, *Appl. Phys. Lett.* **69**, 363 (1996).

[5]T. Dietl, H. Ohno, F. Matsukara, J. Cibert, and D. Ferrand, *Science* **287**, 1019 (2000).

[6]J. Y. Chan, M. N. Olmstead, and S. M. Kauzlarich, *Chem. Mat.* **10**, 3583 (1998).

[7]I. R. Fisher et al., *Phys. Rev. B.* **59**, 13829 (1999).

Growth of MgB_2 Thin Films for Superconductor Applications

T. M. Besmann, A. Goyal, and D. M. Kroeger
Metals and Ceramics Division

The recent discovery of the superconducting properties of MgB_2 has stirred significant scientific and technological interest. Films of MgB_2 have potential advantages over oxide superconductors in their applicability to Josephson Junctions and as high-power, medium-temperature superconductor wires and tapes. A major issue in the use of the material is its fabrication in a cost-effective manner. While a number of attempts to produce films have been successful, they all use very expensive, low-throughput techniques such as laser ablation. We proposed to use a unique chemical vapor deposition (CVD) precursor system that would allow, for the first time, the deposition of the MgB_2 phase using this technique. Resources were sufficient to prepare the CVD for the deposition of MgB_2 using two unique precursors, and for a limited number (5) of runs. The films were characterized with regard to their crystallinity and composition. It was determined that MgB_2 was not formed under the conditions utilized, although elemental magnesium was deposited. This was encouraging for future attempts, as well as valuable in itself.

Introduction

The recent discovery of superconductivity in magnesium diboride, MgB_2, by Akimitsu et al. caused a flurry of activity among researchers all over the world.[1,2] MgB_2 is a simple binary intermetallic compound with a transition temperature, T_c, of 39 K. This value is nearly twice that of any previously known metallic superconductor. Bud'ko et al.[3] demonstrated a boron isotope effect in $Mg^{10}B_2$ with an increase of T_c to 40.2 K and concluded that the compound behaves as a phonon-mediated BCS superconductor. The growth of thin films of MgB_2 with T_c near 39 K and with critical current densities greater than 10^5 A/cm^2 has been reported.

The problem in growing high-quality MgB_2 films in situ using physical vapor deposition techniques is that the very high vapor pressure and volatility of magnesium in the temperature range suitable for formation of MgB_2 (850–900°C) results in magnesium-deficient films. Hence, investigations have focused on using an ex situ approach so far. Such investigations use a two-step approach wherein a boron precursor film is first deposited on the substrate and then the sample is annealed in a magnesium-containing environment. Zeng et al.[4] have very recently reported a high-pressure chemical vapor deposition (CVD) in which they bulk heat magnesium to provide the vapor source and react it with diborane. Besides the tremendous toxicity of diborane, the use of high pressures is problematic in that deposition on desired surfaces is difficult to control. Here, we propose to use a new method to fabricate high-quality MgB_2 films—metalorganic CVD (MOCVD), which may be useful for both epitaxial and nonepitaxial potential applications. Very recent reports from the Pennsylvania State University confirm that a quasi-CVD approach can produce high-quality MgB_2 films, although the boron source and volatility of the magnesium are still problematic.[3]

Technical Approach

This project used the organometallics bis-(cyclopentadienyl)-magnesium (Cp_2Mg) as the magnesium source and *o*-carborane, *o*-$C_2H_2B_{10}H_{10}$, as the boron source and were commercially available from Aldrich, Inc. The required CVD equipment was available in our laboratory.

The efforts focused on the conditions needed to produce dense, uniform coatings of ~0.2–1 μm in thickness on various single-crystal substrates of interest such as C-plane sapphire. Variables that were explored centered on the evaporation temperatures of the organometallic sources. Because of the differing thermochemical stability of the organometallics employed, different evaporation temperatures at the same pressure are possible. Argon carrier gas was used to deliver precursor molecules into the reaction zone. As Cp_2Mg is an air-sensitive compound, it required specialized handling, including the use of glove bags.

A set of evaporation temperatures for Cp_2Mg and *o*-carborane temperatures were explored to deposit MgB_2 films. The films were characterized with respect to crystallinity, composition, and microstructure using electron microscopy/scanning Auger spectroscopy and X-ray diffraction.

Results and Accomplishments

A CVD system was modified to accept the organometallic precursor source vessels. The precursors were acquired and were installed in the deposition system. The entire system was tested for leak-tightness and thermal stability. Since the vapor pressures of the precursors were unknown, the range of vapor source temperatures had to be estimated. These resulted in an apparent high rate of evaporation of the *o*-carborane and a more modest evaporation rate of the magnesium precursor.

Five runs at various temperatures and flows were performed and are summarized in Table 1.

Table 1. Summary of CVD run conditions

Run No.	*o*-carborane					Bis-Mg					Furnace temp. (°C)	Run time (min)
	Temp (°C)	H_2 flow (sccm)	Weight (g)			Temp (°C)	H_2 flow (sccm)	Weight (g)				
			Initial	Final	Change			Initial	Final	Change		
1	105	100	1178.7	1176	2.7	140	100	1198.8	1195.6	3.2	700	15[a]
2	50	250	1176	1176.1	–0.1	100	250	1195.6	1189.3	6.3	700	120
3	100	250	1181.3	1180.1	1.2	50	250	1189.3	1188.8	0.5	700	37[a]
4	75	250	1180.6	1178.6	2	50	250	1188.8	1187.8	1	700	120
5	75	250	n/a	n/a	n/a	75	250	n/a	n/a	n/a	700	120

[a]Clogging in *o*-carborane lines stopped run. *o*-carborane pot was refilled after second run.

The weight gains for the coatings on sapphire substrates are listed in Table 2. Coatings were optically observed on the substrates (Fig. 1), although the weight gain on the ~6-mm-square samples was low. Low-angle and normal-incidence-angle X-ray diffraction did not indicate the presence of a detectable coating. However, scanning Auger analysis did reveal an elemental magnesium coating on specimens of run 2 (Fig. 2), as well as oxygen and carbon. The oxygen is likely a contaminant from the air, as magnesium will form a native oxide. The carbon can be either an environmental contaminant or actually deposited from the carbon in the precursor. There was no observation of boron in the coatings or of MgB_2. It was observed that the cooled exhaust end of the reactor condensed crystalline material during the run that had the same appearance as the *o*-carborane. It was thus assumed that the *o*-carborane was transported through the CVD furnace unreacted.

Table 2. Results of CVD runs

Run	Specimen	Initial (g)	Final (g)	Gain (g)
1	1	0.3179	0.3183	0.0004
	2	0.3198	0.3203	0.0005
	3	0.3170	0.3175	0.0005
2	1	0.3190	0.3397	0.0207
	2	0.3196	0.3274	0.0078
	3	0.3182	0.3244	0.0062
3	1	0.1470	0.1483	0.0013
	2	0.3176	0.3202	0.0026
	3	0.3170	0.3196	0.0026
4	1	0.1473	0. 1474	0.0001
	2	0.1473	0. 1475	0.0002
	3	0.1469	0. 1471	0.0002
5	1	0.1470	0. 1476	0.0006
	2	0.1469	0. 1468	–0.0001
	3	0.1467	0. 1474	0.0007

Summary and Conclusions

Given the uncertainties entering into the project and the limited resources, it was not possible to accomplish the deposition of MgB_2. Encouraging results were seen in the deposition of elemental magnesium, which in itself is difficult to accomplish. The bis-(cyclopentadienyl)-magnesium does indeed reduce to magnesium under appropriate CVD reactor conditions. These were high hydrogen-to-magnesium precursor gas ratios, approximately 700°C ambient, and reduced total pressure.

The stability of *o*-carborane was too great to allow deposition of the diboride. The precursor appeared to transport through the reactor unaffected. Subsequent considerations resulted in the conclusion that less-stable boron sources, such as diborane, would have had a substantially greater probability of success. Unfortunately diborane was not considered in the project due to its both high toxicity and pyrophoric/explosive nature. Although the use of diborane is possible, it would have required a very significant upgrade in the safety configuration of the CVD apparatus.

A major benefit of this research has been the demonstrated ability to deposit elemental magnesium. This led directly to the development of a preliminary proposal

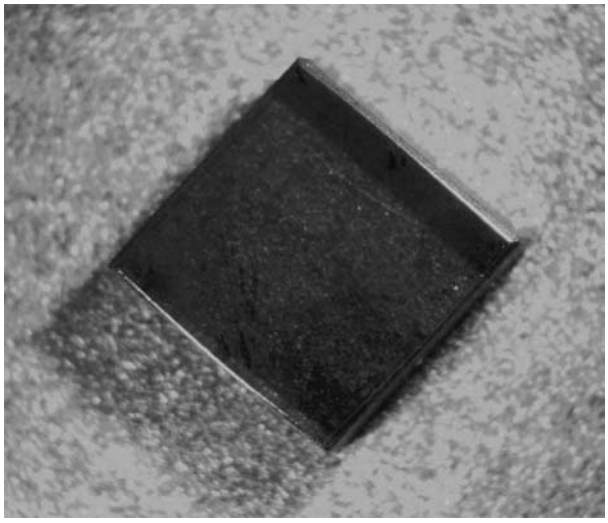


Fig. 1. Optical image of a 6-mm-square coated sapphire substrate.

to participate in a DOE virtual center for research on the storage of hydrogen. Although this proposal was not successful, efforts are continuing to develop magnesium-based hydrogen storage projects based on these results.

References

[1]J. Akimitsu, Symposium on transition metal oxides, Sendai, January 10, 2001.

[2]J. Nagamatsu, N. Nakagawa, T. Muranaka, Y. Zenitani, and J.Akimitsu, *Nature* **410**, 63 (2001).

[3]A. V. Pogrebnyakov, J. M. Redwing, J. E. Jones, X. X. Xi, S. Y. Xu, Q. Li, V. Vaithyanathan, and D. G. Schlom, *Appl. Phys.Lett.* **82**(24), 4319 (2003).

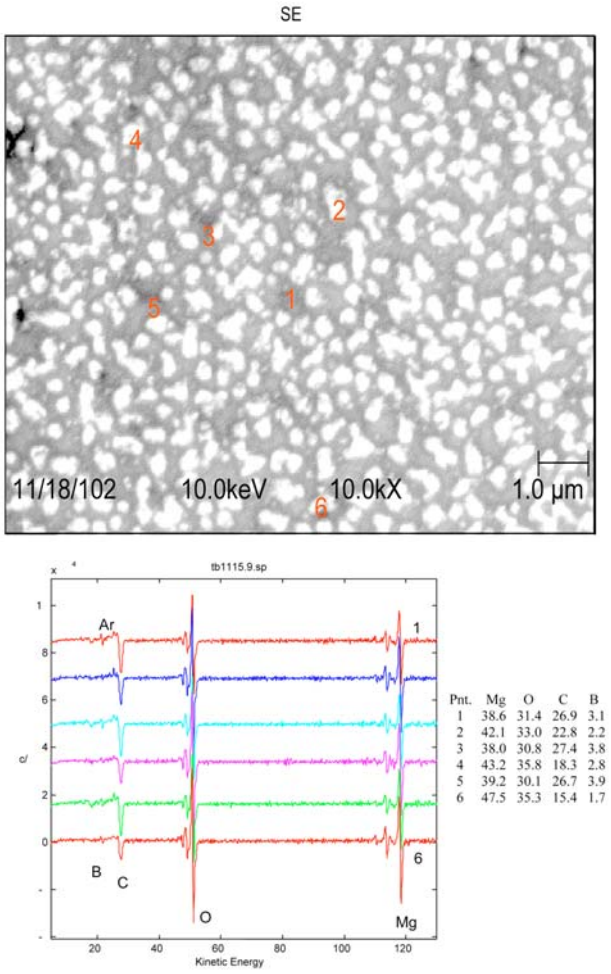


Fig. 2. Scanning Auger results for specimen 3 of run 2 (Tables 1 and 2). The upper image is a secondary electron image indicating positions where the Auger signals were acquired. The Auger energy spectra are shown in the lower image with indicators for the elemental identification. The table is an integrated numerical determination of the relative percentage of the elements observed.

Light-Emitting Nanoscale Tunnel Junctions

T. L. Ferrell
Life Sciences Division

This project was initiated to fabricate and study light-emitting, metal-oxide-metal, electron tunneling junctions on the nanoscale. It is well known that macroscopic metal-oxide-metal (MOM) tunnel junctions radiate with low efficiency. On the nanoscale, the junctions, in which the tunneling electrons engender localized radiative surface plasmons, can radiate with an efficiency 137 (reciprocal of the fine-structure constant) times greater than that of previously investigated millimeter-scale junctions. With each junction of a size much less than the wavelength of the photons emitted, an array of the junctions would radiate in dipolar mode with sufficient intensity to provide an extremely high-resolution, three-dimensional, thin, flat-screen, color display. The frequency of emission would be eV/h for a bias voltage V so that tunable operation in the visible would require 1.5–3 volts. Additionally, operation in reverse would provide a new type of photovoltaic device with stacked transparent layers illuminated laterally by p-polarized light. Finally, due to the sensitivity of the surface plasmon resonance to the optical index of an adsorbate, a novel sensor array could be realized that presents high spatial resolution. Previous studies on the millimeter scale produced low emission intensities due to the multiple couplings needed for radiative emission, but this work established methods of obtaining practicable operation lifetimes.

Introduction

Metal-oxide-metal (MOM) tunnel junctions have been explored for many years motivated by interest in the associated fundamental physics and by the possibility of obtaining thin-film, tunable, light emitters in the visible portion of the electromagnetic spectrum. A considerable body of literature was amassed due to a variety of potential applications. Unfortunately, in practice, the millimeter-scale junctions have poor emission intensities and tend to suffer breakdown due to oxide surface irregularities, impurities, or diffusion of metal ions. MOM tunnel junctions have in the past invariably been constructed using flat wire electrodes of millimeter size. The poor emission intensities are due to the fact that the tunneling first engenders nonradiative surface plasmons that must thereafter couple to dipolar surface plasmons on surface roughness features on the upper electrode. In this project, a high-density nanoscale alternative was explored with far superior characteristics and performance with the ideas of enabling high-resolution, three-dimensional displays, sensors, and photovoltaic cells, and potentially a new type of tunable laser.

Technical Approach

Fundamental theory predicts that by eliminating the need for an intermediate coupling stage involving the nonradiative mode, a much higher emission intensity would be evinced. As a result, the objective was to produce an array of nanoscale tunnel junctions using microlithographic masks prepared by electron-beam lithography and applied using the standard lift-off process for metal deposition. Such an array consists of aluminum wires of 120-nm width separated by a slightly larger amount, oxidized, and orthogonally crossed with similarly sized and arranged gold wires. Elementary electrodynamics provides that a radiative dipole is induced on any surface that is bounded in one or more dimensions on the nanoscale (dimensions smaller than the wavelength of the emission) when the electron distribution is locally disturbed from equilibrium. In the present geometry, each junction is bounded normal to the surface and diagonally at the inside "corners." Hence, direct dipole radiation is emitted normal to the surface and tangential to the surface by the respective surface plasmon modes engendered by the tunneling electrons. This eliminates the intermediate-stage coupling to the nonradiative surface plasmons and thus is of much higher intensity since losses occur at each necessary stage of coupling.

Fabrication by electron-beam lithography was used in this project in order to obtain the nanometer-scale linewidths required. The fabrication was carried out via a subcontract with Dr. Greg Norden of the University of Alabama in Huntsville (UAH). The procedure used was to first layout the design (Fig. 1) and then carry out the following lithography process:

1. Sputter 500-nm SiO_2 on silicon wafer for insulation purposes.

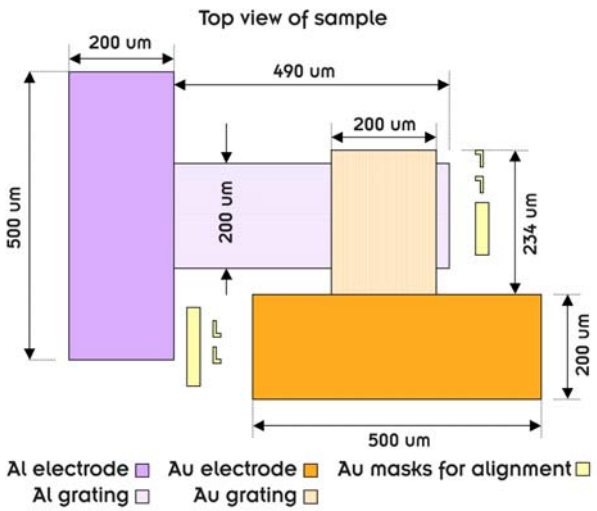


Fig. 1. Layout drawing for tunnel-junction array.

2. Spin coat 150 nm PMMA A4 on the surface and bake the resist at 400°C.
3. Fabricate two sets of "L"-shaped Au masks by e-beam exposure followed by Au metallization and lift-off.
4. Repeat spin and bake of PMMA on the sample.
5. Expose the Al pad and grating pattern by using one of the two alignment mask sets followed by Al metallization and lift-off.
6. Grow Al_2O_3 on the top of the Al grating (200° and 15 min in oven).
7. Repeat spin and bake of PMMA on the sample.
8. Transfer the Au electrode pattern by using the same alignment mask set as used in step 5 followed by a Cr-Au metallization and lift-off.
9. Repeat spin and bake of PMMA onto the sample.
10. Transfer the Au grating pattern by using the second alignment mask set followed by purely Au metallization and liftoff.

A complete theory of the light emission wavelength, intensity, angular distribution, and polarization in the present case was developed at ORNL. The parameters of the device were thus known to UAH for purposes of fabrication. Testing at ORNL was conducted for similar mesoscopic samples in order to prepare the substantial portion of the test procedures for the nanoscale devices.

Results and Accomplishments

The first devices using the above procedure were successfully fabricated in the first six months of the project. The contacts to the device were then attempted using a commercial wirebond instrument. In this step the Au-Al_2O_3 contacts peeled and the device was not then viable. In the ensuing attempts at fabrication, the scanning electron microscope began displaying anomalies that grew worse over time. Further samples thus displayed irregularities that prevented their use in testing. Service on the scanning-electron microscope was requested from the vendor but was unavailable for the remaining three months of the project due to corporate-level changes by the manufacturer. Device images are shown in Figs. 2 and 3.

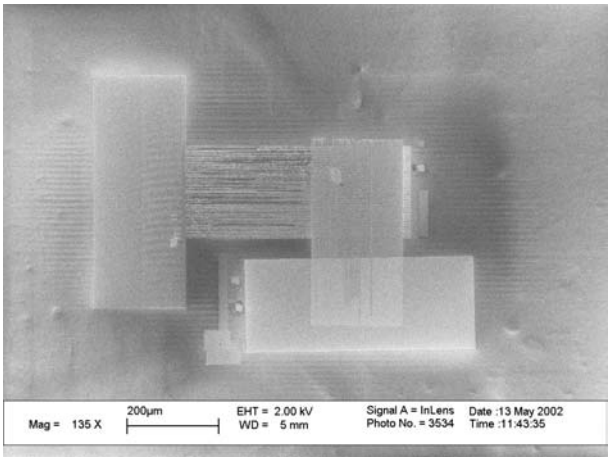


Fig. 2. First tunnel-junction array fabricated.

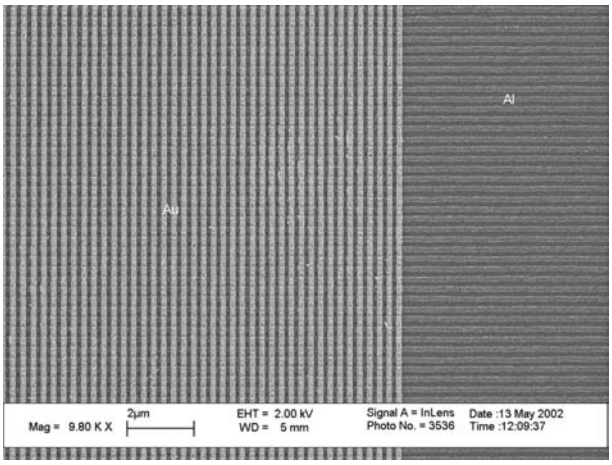


Fig. 3. Scanning electron microscope image of first array.

Images of all samples prepared are shown in the figures. It was thus established that the desired tunneling junction array can be fabricated by electron-beam lithography, but samples could not be further studied.

Summary and Conclusions

Although the technical difficulties prevented our obtaining samples that could be tested for light emission, the fact that samples were fabricated led us to apply for capital funding for an ORNL electron-beam lithography system that would allow further work. We were able to obtain $125,000 in capital funds from the DOE for this purpose and are currently preparing proposals using the preliminary sample images and results.

Dynamic Transport in Nanostructures

Z. Zhang,[1] A. Eguiluz,[1,2] S. T. Pantelides,[1,3] and K. Varga[1]
[1]*Condensed Matter Science Division*
[2]*University of Tennessee, Knoxville*
[3]*Vanderbilt University , Nashville*

Conventional microelectronics is already confronting the nanoscale, where quantum phenomena are inevitable. The time is approaching fast when nanoscale devices will have to exploit quantum phenomena instead of avoiding them. The theoretical understanding of the transport in nanoscale devices is very important. We have developed new approaches to transport problems which are simpler and more efficient than previous methods. Our research focused on two important areas: the development of a novel "source and sink" approach to transport and the development and implementation of the time-dependent density functional theory (TDDFT) for large-scale simulations of the dynamical aspects of quantum transport. We have implemented a Lagrange interpolants approach into the TDDFT, which combines the rapid convergence of spectral methods with the matrix sparsity of the local (real space) representations. This new technique speeds up the calculations by an order of magnitude, facilitating direct-time-step integration in TDDFT.

Introduction

It is now clear that silion-based technologies are expected to reach insurmountable roadblocks beyond the year 2020. One of the most promising directions for overcoming the envisioned obstacles lie in the development of a robust nanoscale electronic systems, which promises to open a whole new world of novel devices and sensors with outstanding capabilities. Indeed, the past decade has seen considerable progress in producing experimental prototypes of such devices. Theoretical progress in this field is considerably slower, and it has so far been limited primarily to DC transport. There is, however, a substantial interest in time-dependent transport in nanostructures; dynamic conductance, photon-assisted tunneling, AC characteristics, nanostructures in time-varying electromagnetic fields, electron pump, electron turnstiles, and frequency-dependent localization, are just a few examples where time-dependent electron dynamics plays an extremely important role. In the last few decades, the emphasis of first-principles quantum mechanical calculations for molecules and solids has been the electronic structure of the ground state and the exploration of the available unoccupied states (excited states). It is important to note that the transport problem is far more complicated than the usual ground-state problem or the problem of excited states of molecules or crystalline solids, which represent closed systems (the Schrödinger equation amounts to solving an eigenvalue problem). The transport problem has distinctly different boundary conditions, namely links to reservoirs of electrons with a steady-state current. We have developed two new and novel approaches to solving the time-dependent or dynamic quantum transport problem. In the first approach we evaluate an approach to quantum transport by directly building in "sources and sinks" for the current into the ab initio problem. In the second approach, we combine the Schrödinger and Maxwell equations and explicitly integrate them in time.

Technical Approach

In the approach we have developed, the circuit (i.e., the open system consisting of the device, the leads and the battery) is replaced by a closed finite system that with a source and a sink which are connected to the device as shown in Fig. 1. The source and the sink are modeled by imaginary Dirac delta-function potentials. A positive imaginary potential injects electrons, while the negative imaginary potentials absorb the wave function at the boundaries. The novelty of this method is that the source and sink potentials may be determined self-consistently, and that they therefore drive the same current through the molecular device with the same voltage drops.

To see this, we consider the continuity equation for the current density j of a device subject to a Hamiltonian that contains an imaginary potential $W(r)$:

$$\nabla j = 2W(r)\rho(r) .$$

Depending on its sign, W describes a source or sink for the current flowing through the system. When a Dirac delta-like source is added to the right, along with a

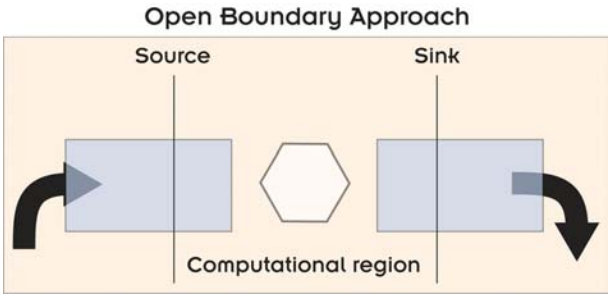


Fig. 1. Schematic of a two-probe structure with "source/sink" boundaries.

corresponding "sink" term (same potential with the opposite sign) to the right, the corresponding imaginary potential takes on the form:

$$iW(r) = \frac{I}{2\rho(r)}[\delta(z-L) - \delta(z-R)] \quad .$$

To impose these injecting and absorbing boundary conditions, the Schrodinger equation with wavefunction

$\phi(r) = \exp[iG(r)]\ \varphi(r)$

with real functions $G(r)$ and $\varphi(r)$ may be decomposed into its real and imaginary parts:

$$H + \frac{1}{2m}[\nabla G(r)^2]\varphi(r) = E\varphi(r)$$

and

$$\nabla^2 G(r)\ \varphi(r) + 2\nabla\ G(r)\ \nabla\ \varphi(r) = 2W(r)\ \rho(r) \ .$$

The first equation is a Schrödinger equation with an extra potential term for the real wave function $\varphi(r)$, while the second equation is continuity equation of the probability current $j(r) = \nabla\ G(r)\ \rho(r)$,

$$\nabla j(r) = 2W(r)\ \rho(r) \quad .$$

The Schrödinger equation and the continuity equation for ∇G need to be solved self-consistently.

Results and Accomplishments

Our initial numerical calculations for simple systems, for example, for an analytically solvable square-well potential barrier (see Fig. 2) or for a three-atom sodium nanowire connected to jellium electrodes (see Fig. 3), show that this approach can work very efficiently and accurately.

We believe that this approach has the following potentially unique features:

- The transport problem for open systems is transformed into an eigenvalue problem. The eigenvalue problem is very similar to the usual Kohn-Sham equations, and therefore the existing advanced techniques developed for the solution of the Kohn-Sham equation can be utilized.
- The current is calculated explicitly using current-carrying complex wave functions. The method therefore can be used in current density functional theory approaches as well.
- Multiterminal devices can be described by introducing several sources and sinks. This makes possible the direct investigation of FET transistors and other devices where the current is controlled by a gate voltage.
- A further advantage may be that this approach may be extended to time-dependent transport problems. In this case, the boundary conditions depend on time and time-dependent "source and sink" terms are needed.

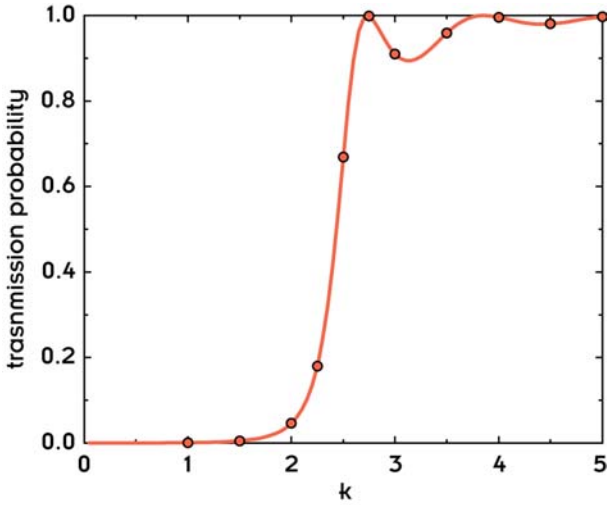


Fig. 2. Comparison of the exact and numerically calculated transmission probability for a square-well potential barrier.

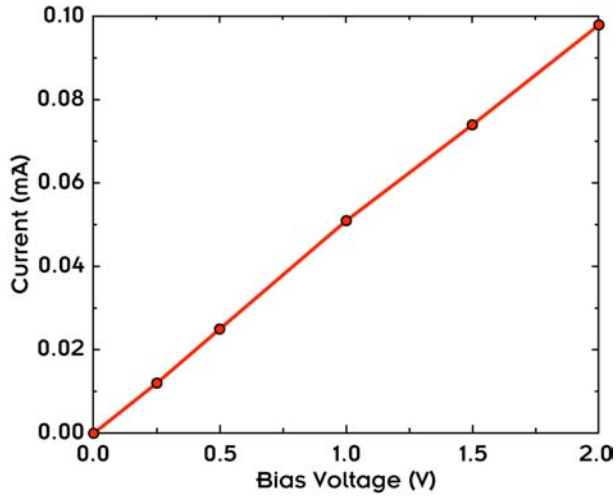


Fig. 3. I-V characteristic of a three-atom sodium nanowire connected to tellium electrodes.

Our aim was to develop tools for large-scale time dependent density functional theory (TDDFT), which directly couple the Schrödinger and Maxwell equations. This approach, while computationally demanding, is important because it provides a natural scheme for a first-principles treatment of the electronic conduction in such "difficult" cases as nonequilibrium and strongly inhomogeneous systems. Indeed, TDDFT enables one to meaningfully discuss the dynamical effects of quantum confinement in the presence of important correlations between the electrons, which are treated on the same footing as the electronic structure and bonding characteristics of the molecular device. To this aim we have developed a new very powerful approach for the direct time integration of the TDDFT equations. In this method the real space part of the wave function is represented by a Lagrange mesh basis, which dramatically reduces the basis dimensions to facilitate the fast direct solution of the time-dependent problem. The Lagrange basis functions are defined by the property that they are equal to unity in a single grid point but vanish on all other grid points.[1,2] It is a variational, spectral method with exponential convergence; the basis functions are very flexible and Hamiltonian is sparse. This basis preserves the most advantagous property of the real space mesh methods, but the basis functions are defined everywhere, not only on the mesh points. Due to the Lagrange interpolation condition, the potential matrix elements are diagonal and the matrix elements of the kinetic energy operator can be easily calculated by numerical integration.

The implementation of the Lagrange mesh technique dramatically speeds up the convergence and reduces the size of required basis dimensions. We have compared the convergence of the energy of the C_6H_6 molecule using the Lagrange mesh and a traditional finite-difference scheme in Table 1. This shows that the number of basis functions is about an order of magnitude smaller then in the previously applied methods. This results in a significant speed up in time-dependent simulations, and it also gives us a better control of accuracy.

Table 1. Convergence of the energy (in eV) of C_6H_6 as a function of number of basis states N using Lagrange mesh (LM) and finite difference (FD) approaches

N	LM	FD
8^3	–1014.40	
12^3	–1016.85	
16^3	–1016.72	–1014.67
20^3	–1016.73	–1015.82
24^3	–1016.73	–1016.51
28^3	–1016.73	–1016.72
32^3	–1016.73	–1016.73

Summary and Conclusions

We have developed a very efficient new approach to transport problems[3] in which using a sink and a source potential maps the open boundary problem to a closed eigenvalue problem. Its accuracy and efficiency have been illustrated through several benchmark calculations. To facilitate fast and reliable direct-time integration of TDDFT equations, we have developed a new Lagrange mesh–based approach[4] which speeds up the convergence and reduces the basis dimensions considerably. These new techniques allow us faster and more efficient calculations of molecular electronic devices.

Follow-on funding has been received from DOE for "Integrated Multiscale Modeling of Molecular Computing Devices."

References

[1]D. Baye and P.-H. Heenen, *J. Phys.* **A19**, 2041 (1986).

[2]B. I. Schneider and N. Nygaard, *J. Chem. Phys.* **A106**, 10773 (2002).

[3]K. Varga, Z. Zhang, and S. T. Pantelides, "Open Boundary Approach to Transport Problems," in preparation.

[4]K. Varga, Z. Zhang, and S. T. Pantelides, "Lagrange Mesh Approach to the Solution of the Density Functional Equations," in preparation.

Nanocrystalline Giant Magnetostrictive Materials for Microactuator Applications

C. T. Liu,[1] J. A. Horton,[1] C. L. Fu,[1] E. P. George,[1] and J. H. Zhu[2]
[1]*Metals and Ceramics Division*
[2]*Tennessee Technological University*

This project focused on the development of nanocrystalline giant magnetostrictive (GMS) materials for next-generation microactuators which will exhibit large-displacement, high-actuating force/energy density, low-actuating and saturation magnetic fields, rapid response, and compatibility with Si microfabrication. The project demonstrated that the amorphous state in the $(Tb,Dy)Fe_2$–based system could be stabilized by combining alloy design and melt spinning. Initial work showed that by controlling the annealing condition (i.e., temperature), the crystallization process could be controlled and nanocrystalline GMS materials could be synthesized. It was further indicated that the nanocrystalline GMS materials exhibited higher magnetostriction and/or magnetostrictive strain coefficient, in comparison to both amorphous and fully crystalline GMS materials.

Introduction

In the past several years, microelectromechanical systems (MEMS) have evolved rapidly from an academic curiosity into a relatively new technology of considerable commercial and military interest. A crucial element of MEMS technology is the development of suitable microactuators. Thin-film/thick-film magnetostrictive materials have been recognized recently as a promising material for making MEMS microsensors and microactuators,[1–5] due to the low actuating voltage, high actuating force, and fast response. Microactuator devices utilizing magnetostriction as the actuation mechanism have been designed and demonstrated, including micropumps,[3–6] linear ultrasonic motor,[3] small flying machine,[7] etc.

Terfenol-D, based on a ternary Laves phase system with composition of stoichiometric $(Tb_xDy_{1-x})Fe_2$, is the preferred magnetostrictive material, with strains about 0.2% in a field of 10,000 Oe at room temperature.[8] One important issue related to MEMS application of magnetostrictive materials is to increase the magnetostriction at low magnetic field and to lower the saturation field. Recent studies indicated that stabilization of the amorphous state in $(Tb,Dy)Fe_2$ via alloying additions (such as boron) and rapid cooling could significantly reduce the saturation field.[9–11]

Even though amorphization has been found to reduce the saturation field of $(Tb,Dy)Fe_2$ significantly, it has an adverse effect in decreasing the maximum magnetostrictive strain. What is really desirable for microsystem application is to increase the magnetostrictive strain coefficient and maximum magnetostrictive strain while maintaining the low saturation field through some innovative approach. In this project, nanocrystalline magnetostrictive materials based on $(Tb,Dy)Fe_2$ are proposed to achieve the desired performance for MEMS applications.

Technical Approach

To achieve the desirable performance for the microactuators, it is proposed to form aligned nanocrystalline $(Tb,Dy)Fe_2$ with a size of 1–10 nm in an amorphous matrix by crystallizing amorphous $(Tb,Dy)Fe_2$ alloys in controlled magnetic fields. The rationale for such an approach is as follows: the amorphous matrix and the nanoscale crystalline phase will reduce the magnetoelastic anisotropy of the material, thus retaining the low saturation magnetic field of the amorphous alloy; under the influence of magnetic field during crystallization and cooling, the nanocrystalline phase will align themselves in such a way that the magnetostrictive strain will be maximized, up to or even surpassing the limit of the single-crystal material (0.2%); the reduction of grain size to nanoscale range will improve magnetic softness of the alloys; for example, coercivity and hysteresis will decrease dramatically.

Since the magnetostrictive properties correlate with the crystal size, it is important to develop scientific bases to control the grain size of these materials. One very attractive way to control the grain size of the crystalline materials is to first stabilize the amorphous state of the material, followed by controlled crystallization to get the desirable nanoscale grain size via thermal treatments. This approach requires a fast cooling rate or enhanced glass formability to obtain a supercooled liquid state.

Recent studies indicate that three metallurgical factors strongly affect the glass forming ability of bulk amorphous alloys:[12–14]

1. multicomponent systems consisting of more than three elements;

2. significant difference in atomic size ratios above 12% among the three main constituent elements; and
3. negative heats of mixing among the three main constituent elements.

According to the physical metallurgy principles required for forming amorphous structures, additional elements with tiny atomic size such as B and C, small atom size such as Co and Ni, and large atom size such as Zr and Al have been added to (Tb,Dy)Fe_2 system to achieve the large difference in atomic sizes among the main constituent elements. With careful control of alloying elements, it is expected that the glass-forming ability of the (Tb,Dy)Fe_2 phase will be enhanced, and amorphous (Tb,Dy)Fe_2-based ribbons can be synthesized by melt-spinning.

Once the formation of amorphous structure has been accomplished, controlled heat treatments will be conducted to crystallize the materials to obtain nanocrystalline structures. By controlling the annealing temperature/time, it is possible to get extremely fine, nanoscale structures with the crystalline phase distributed in an amorphous matrix, or a mixture of several crystalline phases. Ultimately, magnetic field will be utilized during heat treatment to self-align the magnetic domains and thus further improve the magnetic and magnetostrictive properties of the nanocrystalline GMS materials.

Results and Accomplishments

We first characterized the magnetostrictive properties of commercial GMS Terfenol-D bar processed by directional solidification (DS) (supplied by ETREMA Products, Inc.) and melt-spun ribbon fabricated in our laboratory using the ETREMA material. It was found that the Terfenol-D ribbons synthesized by melt-spinning were crystalline, and the amorphous state could not be stabilized by modifying the melt-spinning parameters alone. Furthermore, the magnetostriction of the Terfenol-D ribbon was lower than the Terfenol-D bar and annealing of the ribbon at 500°C for 1 h did not improve the magnetostriction of the ribbon. This can be explained by the fact that the Terfenol-D bar was directionally solidified, and its grains have been preferentially oriented to optimize its magnetostrictive properties.

To develop amorphous (Tb,Dy)Fe_2-based alloys with enhanced glass formability, a large number of alloys with various alloying additions were arc melted using high-purity starting materials; furthermore, ribbons were prepared by melt-spinning, followed by characterization with X-ray diffraction and other techniques to assess if the amorphous state was stabilized in the ribbon. A magnetostriction measurement rig was constructed which could be used to measure the magnetostriction of the ribbons by the strain-gauge method, as shown in Fig. 1.

A series of alloy compositions have been identified, which retained amorphous structure after melt-spinning, as indicated by X-ray diffraction (Fig. 2). The alloy compositions are based on the multicomponent system of $Tb_{30}Fe_{(70-a-b-c)}X_aY_bZ_c$, where X and Y are small atoms compared to Tb or Fe, and Z is a reactive element added to scavenge oxygen and nitrogen in the alloys. The optimal contents of X, Y, and Z have been determined. Transition-metals such as Co and Ni were also added, which did not noticeably affect the glass-forming ability of the alloys. Preliminary results on heat treatment of the amorphous ribbons indicated that after crystallization at 500°C for 1 h a crystalline Laves phase is formed from the amorphous matrix (see Fig. 2 for X-ray diffraction results and Fig. 3(b) for a scanning electron micrograph). The precipitates were very fine, and the sizes were in the range of 5–50 nm. As a comparison, the as-spun ribbon shows a featureless, complete amorphous structure, see Fig. 3(a). On the other hand, heat treatment at 600°C for 1 h fully crystallized the material and more than two crystalline phases could be identified (see Fig. 2 for X-ray diffraction results). Obviously, TEM observation will be needed for further revealing the fine structure in these materials.

Preliminary magnetostriction measurements of the ribbons after different heat treatments are summarized in

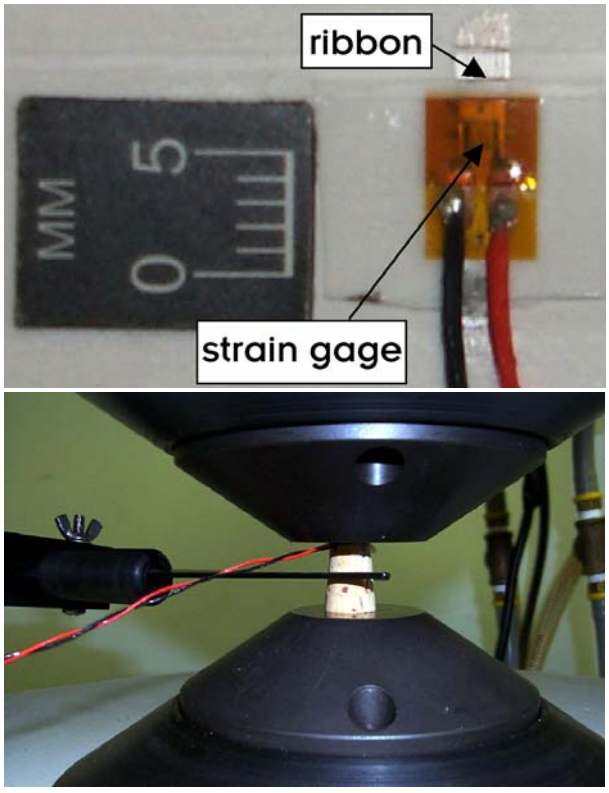


Fig.1. Test setup for measuring the magnetostriction of the ribbon:(a) ribbon with strain gauge on it; and (b) setup of the measurement rig.

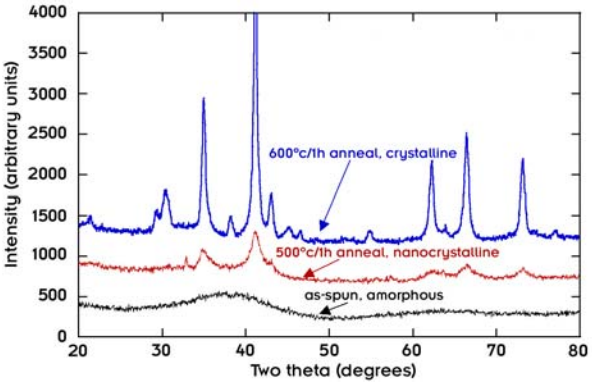


Fig. 2. X-ray diffraction results on melt-spun ribbons based on $Tb_{30}Fe_{(70-a-b-c)}X_aY_bZ_c$, indicating amorphous state for as-spun ribbon (black line), nanocrystalline state after 500°C/1 h anneal (red line), and fully crystallized state after 600°C/1 h anneal (blue line).

Fig. 4. Here, it was observed that the saturation magnetostriction was 280 μm/m for the amorphous ribbon, while heat treatment of 400°C for 1 h decreased the saturation magnetostriction slightly, yet increased the initial magnetostrictive strain coefficient. Heat treatment of 500°C for 1 h increased the saturation magnetostriction significantly, up to the level of 500 μm/m, possibly due to the formation of nanocrystalline microstructure. Further increase in annealing temperature decreased, instead of increased, the saturation magnetostriction.

Compared to the same alloy after arc-melting/casting, the nanostructured ribbon obtained by 500°C for 1-h heat treatment showed significantly higher magnetostrictive strain coefficient and required much lower actuating magnetic field to achieve saturation magnetostriction, as shown in Fig. 5. This clearly indicated that the amorphous/nanocrystalline structure was advantageous over the large-grained crystalline structure in the GMS materials. Figure 6 is a back-scattered electron micrograph showing the large, random crystalline structure in the as-cast $Tb_{30}Fe_{(70-a-b-c)}X_aY_bZ_c$ alloys. Corresponding electron probe microanalysis indicated that three phases were co-existing in the material, with the composition of each phase included in Table 1.

Summary and Conclusions

A series of alloy compositions with enhanced glass formability that retained the amorphous structure after melt-spinning have been identified. Initial heat treatment experiments have demonstrated that the crystallization process could be manipulated to achieve nanocrystalline microstructure in the developed alloys. The new alloy, after a 500°C for 1 h nanocrystallization treatment, possessed higher saturation magnetostriction and initial magnetostrictive strain coefficient and required smaller magnetic field for actuation, compared to the same alloys in the amorphous, as-cast, or fully crystallized conditions.

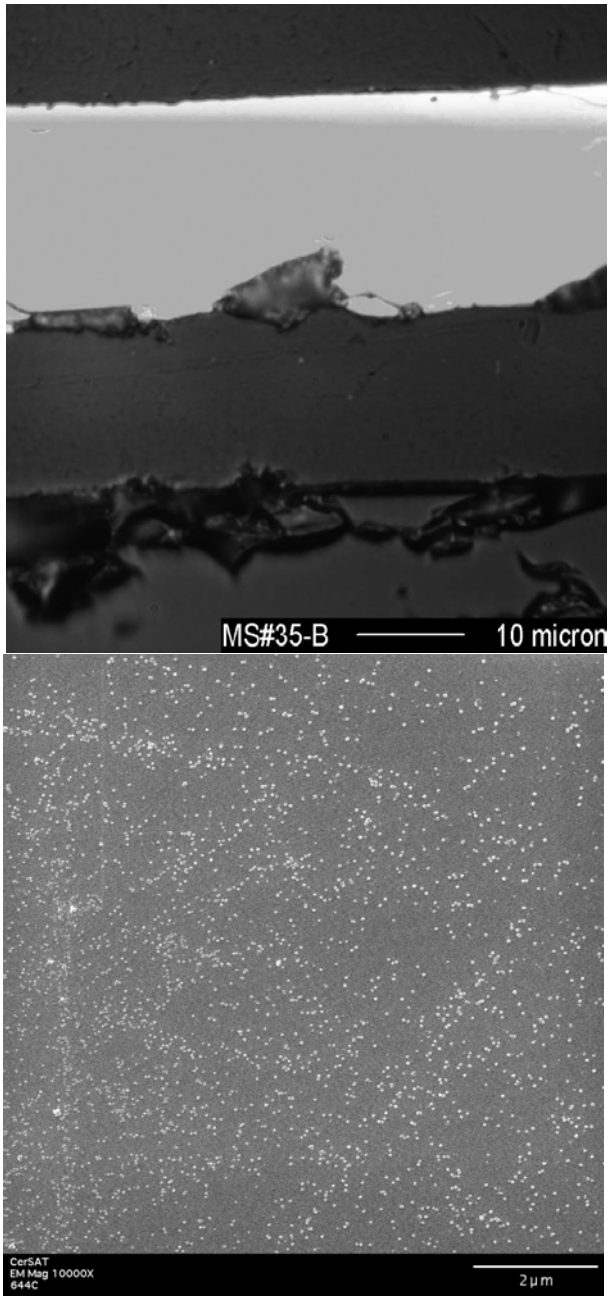


Fig. 3. Microstructure of amorphous and nanocrystalline ribbons ($Tb_{30}Fe_{(70-a-b-c)}X_aY_bZ_c$): (a) in as-spun condition, showing featureless amorphous structure; (b) upon annealing of 500°C for 1 h, showing fine precipitates of the Laves phase in an amorphous matrix.

References

[1]E. Quandt, H. Holleck, "Materials Development for Thin Film Actuators," *Microsystems Technologies* **1**, 178–184, (1995).

[2]G. Flik, M. Schnell, F. Schatz, and M. Hirscher, "Giant Magnetostrictive Thin Film Transducers for Microsystems," *Proc. Actuator* **94**, Bremen, 232 (1994).

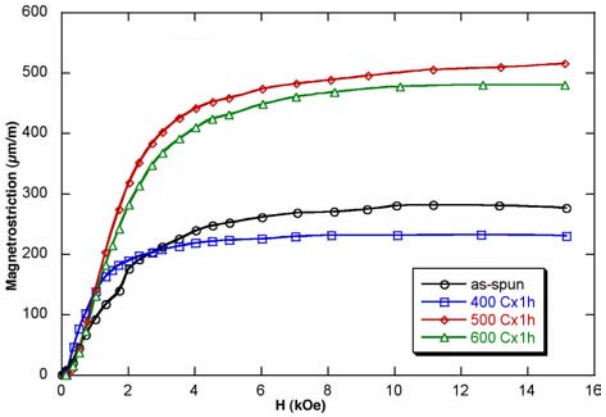


Fig. 4. Magnetostriction of the $Tb_{30}Fe_{(70-a-b-c)}X_aY_bZ_c$ ribbons after different heat treatments.

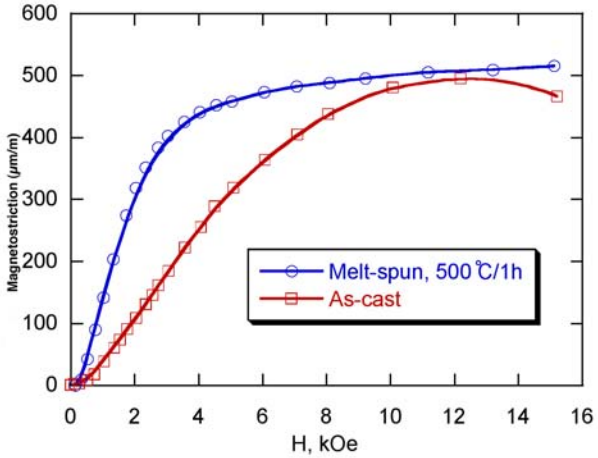


Fig. 5. Comparison of magnetostriction of nanostructured $Tb_{30}Fe_{(70-a-b-c)}X_aY_bZ_c$ ribbon (after 500°C for 1h annealing) and crystalline button (after arc-melting/casting). Note that both of these curves are for no crystallographic alignment while the maximum for DS Terfenol material is a strain of 2000 μm/m.

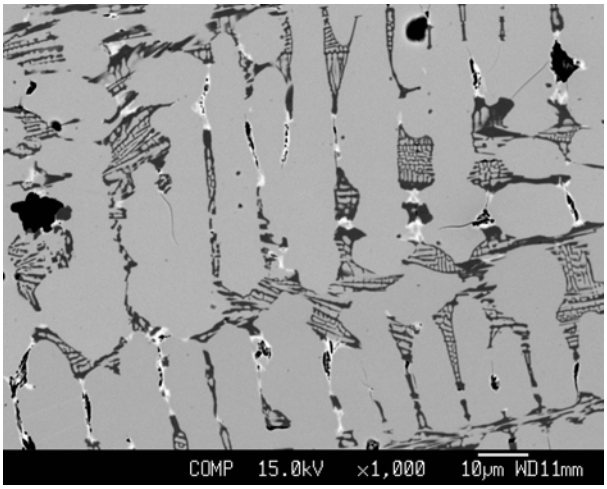


Fig. 6. Back-scatter electron image of bulk $Tb_{30}Fe_{(70-a-b-c}X_aY_bZ_c$ alloy after casting.

Table 1. EPMA results of the compositions of the phases coexisting in the $Tb_{30}Fe_{(70-a-b-c)}X_aY_bZ_c$ alloy after casting

Phase	Fe	Tb	X
Matrix	59.1	31.4	9.3
Dark Phase	43.2	14.1	42.4
White Phase	27.2	36.1	36.3

[3]E. Quandt, "Giant Magnetostrictive Thin Film Materials and Applications," *J. Alloys & Compounds* **258**, 126–132 (1997).

[4]J. Betz, K. Mackay, J.-C. Peuzin, B. Halstrup, and N. Lhermet, "Torsion Based Drift-Free Magnetostrictive Microactuators," *Proc. Actuators 96,* Bremen, 283 (1996).

[5]E. Quandt and K. Seemann, "Fabrication and Simulation of Magnetostrictive Thin-Film Actuators," *Sensors and Actuator A* **50**, 105 (1995).

[6]E. Quandt and K. Seemann, "Magnetostrictive Thin-Film Microflow Devices," in: H. Reichl, A. Heuberger (Eds.), *Micro Systems Technologies* **96**, VDE-Verlag, Berlin, 451 (1996).

[7]K. I. Arai, W. Sugawara, K. Ishiyama, T. Honda, and M. Yamaguchi, "Fabrication of Small Flying Machines Using Magnetic Thin Films," *IEEE Trans. Magn.* **31**, 3758 (1995).

[8]A. E. Clark, "Magnetostrictive Rare Earth-Fe_2 Compounds," *Ferromagnetic Materials,* Vol. 1, ed. by E. P. Wohlfarth, North-Holland, (1980).

[9]W. J. Ren, Z. D. Zhang, A. S. Markosyan, X. G. Zhao, X. M. Jin, and X. P. Song, "The Beneficial Effect of the Boron Substitution on the Magnetostrictive Compounds $Tb_{0.7}Pr_{0.3}Fe_2$," *J. Phys. D: Applied Phys.* **34**, 3024–3027, (2001).

[10]N. H. Duc, K. Mackay, J. Betz, Z. Sárközi, and D. Givord, "Magnetic and Magnetostrictive Properties in Amorphous $(Tb_{0.27}Dy_{0.73})(Fe_{1-x}Co_x)_2$ Films," *J. Phys.: Condens. Matter* **12**, 7957–7968, (2000).

[11]H. Fujimori, J. Y. Kim, S. Suzuki, H. Morita, and N. Kataoka, "Huge Magnetostriction of Amorphous Bulk (Sm,Tb)Fe_2-B with Low Exciting Fields," *J. of Magnetism and Magnetic Materials* **124**, 115–118 (1993).

[12]A. Inoue, "Stabilization of Metallic Supercooled Liquid and Bulk Amorphous Alloys," *Acta Mater.* **48**, 279–306 (2000).

[13]A. Inoue, A. Takeuchi, and T. Zhang, *Metall. Mater. Trans.* **29A**, 1779 (1998).

[14]A. Inoue, *Bulk Amorphous Alloys,* Trans. Tech. Publications, Zurich (1998).

Probing Charge Transport in Oriented Conducting Polymer Nanostructures: Toward Integrated Molecular Optoelectronics

M. D. Barnes
Chemical Sciences Division

This project focused on measurement and modeling of charge transport and optically gated conductivity in oriented polymer nanorods. This idea was based on newly discovered techniques for producing oriented semiconducting polymer nanoparticles with structural and photophysical properties that are quite different from bulk material. The specific goals of this project were to probe charge transport behavior of isolated polymer nanoparticles and electronic coupling between polymer chains and molecular wire/logic structures as first steps toward enabling an optical interface to molecular-level electronic circuitry, and use of polymeric materials in the newly emerging field of nanoscale optoelectronics.

Introduction

Nanophotonics—or molecular-scale optoelectronics—is a rapidly emerging area of research that combines nanoscale optical elements (e.g., single-molecules, or quantum dots) and molecular-scale electronics. Such "photonic elements" could ultimately provide a route for connecting the macroscopic and nanoscopic worlds through absorption or emission of photons coupled with electron (or hole) motion to or from, say, a molecular wire or gate structure. Here the materials challenge in design and fabrication of such elements is daunting: For practical applications, one needs nanoscale components with well-defined transition energies and dipole moment orientation for spatially selective addressability, as well as robust photostability and facile charge transport properties. Because of the inherent size scalability, responsiveness to optical excitation, and long history in mesoscale photonics, semiconducting polymers make attractive candidates for nanoscale optoelectronics. However, in conventional thin film formats, these species are formed with random in-plane orientation with a broad distribution of chain morphologies and tend to be quite fragile with respect to repeated optical excitation.

The inherent size scalability and large parameter space associated with structural and electronic properties of semiconducting polymers[1] make them interesting candidates for the purpose of interfacing and interconnecting nanoscale logic units. The work of James Tour,[2] for example, has focused on synthetic techniques to fabricate molecular "wires" from conducting polymers with specifically designed branch architectures and end-group functionalities. However, in addition to difficulties in controlling the orientation and chain alignment of semiconducting polymers such as polyphenylene vinylene (PPV) derivatives, these species tend to be quite fragile with respect to repeated electronic excitation. Furthermore, most strategies for organizing conducting chain segments are based on molecular self-assembly which restricts the range of device geometries and position of the desired species. While gel-processing[3] or templating techniques[4] have been used to produce axially aligned conducting polymers, these strategies lack the molecular-level addressability and connectivity needed for multiplexed optoelectronic applications.

Our research program was designed to exploit recent experimental results in our laboratory that have demonstrated the ability to "print" single molecules of a conducting polymer with dramatically altered structural and photophysical properties relative to conventional thin-film preparations.[5,6] We observed luminescence characteristics strikingly similar to those recently observed for cadmium-selenide quantum rods[7] suggesting a highly ordered chain configuration with exciton delocalization within the particle. In addition to linear polarization in emission and modification of spectral characteristics, we found uniform particle orientation and a two-order-of magnitude enhancement in photostability relative to similar molecules deposited (from the same solution) by spin-casting. This modification of structural and photophysical properties derives in part from the effect of three-dimensional confinement by the liquid droplet of solution as it evaporates en route to the substrate.[8] As a result, this is a highly general method of preparing different kinds of nanostructured polymer systems for molecular scale optoelectronics with tailorable properties (absorption/emission frequencies, chain morphologies etc.). We find that many important luminescence characteristics are preserved—and in some cases enhanced further—by doping the polymer solution with other materials.

The many consistent experimental observations we have made on these species lead to a picture of a crystalline-like ordering of conjugated segments within an individual nanostructure (single chain). As such, many of the static electronic structure properties are modified – as manifested in the spectral narrowing and discrete emission frequencies seen in fluorescence. It was therefore natural to expect that the dynamic electronic properties, such as ballistic charge transport, would be greatly enhanced by virtue of the structural order as well. It was the primary goal of this project to make experimental measurements of "molecular wire" behavior in these oriented polymer nanostructures. In order to achieve this goal, it was necessary to develop experimental methodologies to orient these species on conducting substrates such as graphite, gold or indium-tin-oxide. In early experiments, it became clear that control over the (static) charge state of these materials, as well as the experimental parameters for making photochemically robust polymer nanostructures, was crucial to the ultimate program goals. We therefore concentrated our efforts here, and one of the most important results of this program was the development of a hybrid Kelvin-force scanning probe microscopy that allowed us to quantitate the magnitude and sign of static charge on individual polymer nanoparticles.

Technical Approach

Experiments were performed on a Dimension 3100 atomic force microscope (Digital Instruments) using silicon cantilevers coated with Cobalt and Chromium. A "Lift-Mode" scan was used in all these experiments, so that the topographic features could be decoupled from the image features generated by the long-range electrostatic forces. The surface potential was measured by nulling the electric field between the tip and the sample at each point in the image scan, by applying a bias to the tip. To quantify the charge on the sample, we used electric field gradient microscopy (EFM). In this technique the field gradient generated by the long-range electrostatic forces (between the charged sample and the biased tip) was kept at a constant level using a feedback loop which varied the drive frequency of the cantilever as it scanned the sample surface. The charge image was generated by plotting the frequency shift (measured using a lock-in amplifier) for each point in the scan. The shift in the drive frequency was used to calculate the force experienced by the cantilever. These forces arise from Coulomb interactions of the sample charge, its image charges in the AFM tip and substrate, and the induced charges due to the bias voltage applied to the cantilever. From an electrostatic analysis of the tip-sample system modeled using a parallel-plate geometry, the sample charge was quantified. An ultradilute (10^{-12} M) solution of cyano-polyphenylene vinylene (CN-PPV) in toluene was nebulized in a stream of dry nitrogen gas using a drawn glass capillary, and the samples were collected on a cleaned glass substrate and used for the EFM experiments.

Results and Accomplishments

We tested this technique by using a sample of polystyrene nanoparticles (20-nm diameter) which were deposited on a cleaned glass substrate using an electrospray device. By controlling the polarity of the voltage applied to the electrospray needle, we could control the charge on the polystyrene beads. We found excellent correlation between surface potential images and topographic scans, where each polystyrene bead spot shows a negative surface potential. The average surface potential per particle was 30 mV, corresponding to a charge state of about 15 individual carriers. For CN-PPV nanoparticles, topographic scans show heights ranging from 8–15 nm, values that are in good agreement with the persistence length of PPV molecules measured from light scattering experiments. Figure 1 shows the field gradient scan on a CN-PPV nanoparticle sample that provides quantitative information on the charge state. Frequency shifts in the range of 4–8 Hz were measured, corresponding to an excess surface charge of 2–5 electrons. EFM measurements on a freshly prepared CN-PPV sample showed frequency shift of 55 Hz, which corresponds to approximately 10 excess electrons on the CN-PPV nanoparticles. The change in the number of excess charges could be due to charge migration or dissipation phenomena, which is more pronounced in samples that

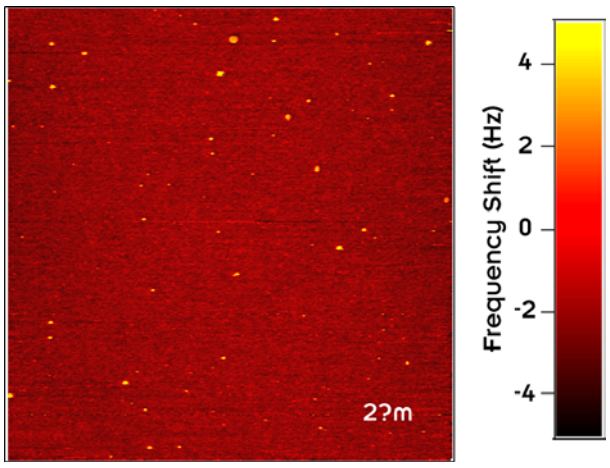


Fig. 1. EFM scan of CN-PPV nanoparticles showing the charge image. Frequency shift ranged from 4–8 Hz, which corresponds to an excess surface charge of 2–5 electrons. (A. Mehta, et al., to be published in Applied Physics Letters.*)*

are exposed to ambient conditions for long periods of time. These results confirm that the CN-PPV nanoparticles produced by the nebulization technique are negatively charged with an excess charge ranging from two to ten electrons. This suggests that these particles could be attached to a conducting surface with an appropriate cationic coating. Using a metal-coated biased cantilever, individual nanoparticles could be addressed to study the I-V characteristics of these nanostructures.

Summary and Conclusions

Our results provided definitive proof that the semiconducting nanoparticles produced by the nebulization technique are negatively charged with an excess charge ranging from two to ten electrons.[9] This value is in good agreement with values estimated from Rayleigh stability limits for a particle of this size. These results suggests that these particles could be attached to a conducting surface with an appropriate cationic coating. Using a metal-coated biased cantilever, individual nanoparticles could be addressed to study the I-V characteristics of these nanostructures. Many questions remain: Are specific charge states (higher or lower) correlated with photochemical properties? How does the chemical potential (electron affinity) depend on charge state? This question is of particular importance in relation to nanoscale optoelectronic device functionality. While our experimental path diverged somewhat from our original proposal, we have made significant progress in our microscopic understanding of the nature of these highly novel polymeric species and the results described here provide a solid scientific foundation for further investigation.

References

[1]A. J. Heeger, "Semiconducting and Metallic Polymers: The Fourth Generation of Polymeric Materials," *J. Phys. Chem. B* **105**, 8475 (2001).

[2]J. M. Tour, "Molecular Electronics: Synthesis and Testing of Components," *Acc. Chem. Res.* **33**, 791 (2000).

[3]T. W. Hagler, K. Pakbaz, and A. J. Heeger, "Polarized-Electroabsorption Spectroscopy of a Soluble Derivative of Poly(P-Phenylenevinylene) Oriented by Gel Processing In Polyethylene- Polarization Anisotropy, the Off-Axis Dipole Moment, and Excited State Delocalization," *Phys Rev. B* **49**, 10968 (1994).

[4]T. Q. Nguyen, J. J. Wu, V. Doan, B. J. Schwartz, and S. Tolbert, "Control Of Energy Transfer In Oriented Polymer-Mesoporous Silica Composites," *Science* **288**, 652 (2000).

[5]P. Kumar, A. Mehta, M. Dadmun, J. Zheng, L. Peyser, R. M. Dickson, T. Thundat, B. G. Sumpter, and M. D. Barnes, "Observation of narrow-bandwith spontaneous luminescence from oriented conjugated polymer nanostructures," *J. Phys. Chem. B* **107**, 6252–6257 (2003). [Letter to the Editor]

[6]A. Mehta, P. Kumar, M. Dadmun, J. Zheng, R. M. Dickson, T. Thundat, B. G. Sumpter, and M. D. Barnes, "Oriented cylindrical nanostructures of a semiconducting polymer: Polarization evidence for highly ordered individual macromolecular structures," *Nanoletters* **3**(5), 603–607 (2003).

[7]J. Hu, L-S. Li, W. Yang, L. Manna, L-W. Wang, and A. P. Alivisatos, "Linearly polarized emission from colloidal semiconductor quantum rods," *Science* **292**, 2060 (2001).

[8]M. D. Barnes, B. G. Sumpter, and D. W. Noid, "Polymer-blend microparticles: A new approach to nanoscale composites with tunable properties," *Materials Today* (invited) **2**(3), 25–27 (1999).

[9]A. Mehta, P. Kumar, T. Thundat, and M. D. Barnes, "Investigation of static charge states of oriented semiconducting polymer nanostructures," to be published in *Applied Physics Letters.*

Synthesis of Highly Textured Ternary Carbide Compounds for Power Generation and Other Industrial Applications

E. Lara-Curzio, T. R. Watkins, C. A. Blue, M. Radovic, and S. Waters
Metals and Ceramics Division

This project focused on the synthesis of highly textured ternary compounds of Ti, Si, and C. These compounds, such as Ti_3SiC_2, have been shown to have excellent physical and mechanical properties for applications at elevated temperatures. Furthermore, surfaces containing the basal planes of the crystalline structure of this compound have exhibited a very low coefficient of friction. Such materials would be ideal to reduce parasitic losses in reciprocating engines that would operate at higher temperatures.

Attempts to extrude cylindrical bodies of Ti_3SiC_2 using ORNL's 1200-ton press were not successful because of the temperature limitations of the system. However, subsequent efforts to extrude cylindrical objects using a press with higher temperature but lower force capabilities were partially successful and provided indications of the possibility of achieving preferential crystallographic orientation of the basal planes Ti_3SiC_2, through hot extrusion. These subsequent experiments also allowed the identification of processing parameters for subsequent research.

Introduction

During the 1980s and 1990s, significant efforts were focused on developing monolithic ceramics to manufacture components for more efficient reciprocating engines (e.g., cam roller followers, intake and exhaust valves, turbocharger rotors, exhaust port liners, and pistons). These materials were chosen because of their excellent wear and corrosion resistance and their ability to retain their mechanical properties at elevated temperatures.[1] Yet, the widespread use of monolithic ceramics in reciprocating engines has not been realized because of high manufacturing costs and concerns about their reliability.

Ti_3SiC_2, which belongs to a class of materials known as MAX phases, might possess the necessary combination of properties to overcome the limitations exhibited by monolithic ceramics and thus enable the development of more efficient reciprocating engines and industrial applications. MAX phases are made up of an early transition metal (M), an element from an A group in the periodic table (A), and either carbon or nitrogen in the composition $M_{n+1}AX_n$ where n is 1, 2, or 3. These materials are layered with a hexagonal crystallographic structure in which $M_{n+1}X_n$ layers are interleaved with layers of pure A.[2–3]

It has been reported that the coefficient of friction of Ti_3SiC_2 on surfaces containing its basal planes is less than 5×10^{-3} and that this ultra-low value persisted even after exposure to ambient air for several months.[4–5] These results are comparable to, and potentially lower than, the lowest coefficient of friction measured on any solid lubricant surface. This unique property of Ti_3SiC_2, along with its high specific stiffness and strength, good thermal conductivity, and machinability make it an ideal candidate material for the manufacture of intake and exhaust valves and valve guides for reciprocating engines, for example.

This project focused on demonstrating the feasibility of synthesizing highly textured Ti_3SiC_2 by hot extrusion using ORNL's 1200-ton press. The objective of this project was to demonstrate the feasibility of forcing the alignment of the basal planes of the crystalline structure of this compound, parallel to the outer surface of extruded bodies, in order to achieve surfaces with low coefficient of friction. The demonstration of such technology could enable the manufacture of intake/exhaust valves and other reciprocating engine components, and in turn, contribute towards the development of higher efficiency engines and industrial processes.

Technical Approach

Dense, polycrystalline bodies of Ti_3SiC_2 have been synthesized by milling Ti, C, and SiC powders followed by cold-pressing and reactive hot-pressing.[2–3] By varying the hot-pressing temperature and time, the average grain size can be controlled. However, the grains in the resulting material are oriented randomly. The objective of this project was to demonstrate the feasibility of synthesizing highly textured polycrystalline bodies of Ti_3SiC_2 that had the crystalline basal planes oriented parallel to their outer surface.

Fig. 1. Photograph of hot-extruded Ti_3SiC_2 in a tantalum-lined 304 stainless steel can. The extrusion products had tears and broke in several pieces.

It has been documented that during extrusion, or during other processes that involve shearing of viscous fluids, anisometric particles align so that their long axis is parallel to the extrusion direction due to shear developed between the extrusion nozzle and the material being extruded.[6] In crystalline materials, slip planes are usually the planes with the highest density of atoms. In the case of hexagonal crystals slip will occur on the basal planes. Therefore, during hot extrusion, when the material can deform plastically, deformation will be easiest along the basal planes, which will align parallel to the extrusion direction. For example, pronounced textures have been achieved through the hot extrusion of superconducting compounds of $YBa_2Cu_3O_x$ and $Bi_2Sr_2CaCu_2O_{8+X}$ where the basal planes became aligned along the direction of extrusion.[7–8]

To achieve the objective of this project, a series of hot extrusion runs were carried out using ORNL's 1200-ton extrusion press. Starting powders of Ti_3SiC_2 with particle size of 1–5 μm were canned under vacuum (>4 × 10^{-4} torr) inside tantalum-lined 304 stainless steel or molybdenum cans, which were subsequently sealed by electron-beam welding. Extrusion runs were carried out after the cans had been heated inside a furnace at temperatures between 1100°C and 1300°C for 1 h. Additional attempts to extrude Ti_3SiC_2 were carried out using a hot-press with higher temperature capabilities, but lower force capacity. In this case, conical dies were manufactured with three different extrusion ratios between 1:2.6 and 1:4 and an angle of 60° using high-strength graphite. The dies were 127 mm in diameter and 171 mm in length. To minimize friction, the die was lined with a 50-μm thick layer of Graphoil®. Hot extrusion runs at temperatures as high as 1700°C were carried out and the microstructure of the resulting materials was characterized using conventional metallographic techniques and X-ray diffraction methods for phase identification.

Results and Accomplishments

Attempts to extrude Ti_3SiC_2 at temperatures between 1100°C and 1300°C and with forces as high as 500 tons were unsuccessful using ORNL's 1200-ton press. One-half of the cans that were prepared failed to extrude through dies with diameters of either 12.5 mm or 15.8 mm. The remaining extrusions had tears or broke in several pieces after exiting the die and revealed poor consolidation of the Ti_3SiC_2 powders (Fig. 1). These results suggested that the temperatures used were not sufficiently high to consolidate the powders and to allow for creep/plastic flow. Although the furnace used to heat up the cans before extrusion can operate in excess of 1300°C, there aren't many materials to fabricate cans capable of withstanding temperatures above 1300°C in air.

The failure to hot-extrude Ti_3SiC_2 using ORNL's 1200-ton extrusion press prompted the search for alternative strategies to accomplish the objectives of this project. Additional hot-extrusions were carried out using a hydraulic press equipped with a graphite furnace capable of 2000°C in vacuum or inert environment. Extrusion experiments were carried out in two steps using this system. In the first step powders were densified at 1400°C in vacuum under an axial force of 10 kN inside a conical graphite mold lined with Graphoil®. This resulted in very dense products. In the second step, the densified bodies were extruded at temperatures between 1400°C and 1700°C and forces as high as 17.5 kN using high-strength conical graphite dies with extrusion ratios between 1:2.6 and 1:4 and an angle of 60° using. The dies were 127 mm in diameter and 171 mm in length. To minimize friction, the die was lined with a 50-μm-thick layer of Graphoil®. Although it was possible to extrude the material (Fig. 2), the amount of extrusion was very small. There was evidence of reaction between the layer of Graphoil® and Ti_3SiC_2, which resulted in the formation of titanium carbide. An X-ray diffraction analysis also revealed that the intensity of the peak associated with the basal plane on a surface of the extruded body parallel to the extrusion direction was 3.5 times larger than the intensity of the peak associated with this plane on a surface perpendicular to the extrusion direction (Fig. 3).

While the results from this investigation have demonstrated the possibility of extruding Ti_3SiC_2, albeit in a small amount, and the preferential alignment of its crystalline planes, they also showed the good resistance of Ti_3SiC_2 to creep deformation, which should be considered a good trait. The ability to hot-extrude Ti_3SiC_2 could be improved by using an alternative lubricant (e.g.,

Fig. 2. Photograph of dense Ti_3SiC_2 body after hot-extrusion at 1400°C in a graphite die using a hydraulic press and a graphite furnace.

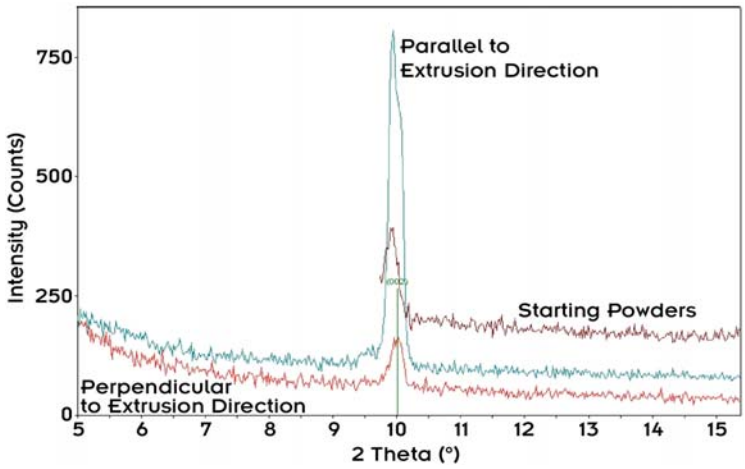


Fig. 3. X-ray diffraction patterns obtained from surfaces parallel and perpendicular to the extrusion direction. The intensity of the basal plane peaks for surfaces parallel to the extrusion direction were 3.5 times larger than that associated with this plane on a surface perpendicular to the extrusion direction.

BN) to reduce reactivity of the material with graphite, and by hot-extruding the material in several steps with a small extrusion ratio in each step. While the creep deformation of materials is a temperature-activated process, higher temperatures also promote grain growth in Ti_3SiC_2, making it more creep resistant and more difficult to extrude. This study has identified the range of temperature and stresses needed to facilitate the hot-extrusion of Ti_3SiC_2.

Summary and Conclusions

Attempts to extrude Ti_3SiC_2 using ORNL's 1200-ton press at temperatures between 1100°C and 1300°C were unsuccessful. Subsequent attempts to hot-extrude sintered powders of Ti_3SiC_2 using a hydraulic press equipped with a high-temperature furnace resulted in bodies with crystalline texture, but the amount of extrusion was limited. This is attributed to the good creep resistance of the material and the lack of force capacity of the press used for these experiments. However, the range of operating conditions to facilitate the hot-extrusion of Ti_3SiC_2 have been identified.

References

[1]U.S. Department of Energy, Energy Information Administration, *Annual Energy Review 1988* (1988)

[2]M. W. Barsoum and T. El-Raghy, "The MAX Phases: Unique New carbide and Nitride Materials," *American Scientist* **89**, 334–343 (2001).

[3]T. El-Raghy and M. W. Barsoum, "Processing and Mechanical Properties of Ti_3SiC_2:I, Reaction Path and Microstructure Evolution," *J. Am. Ceram. Soc*. **82**, 2849–54 (1999).

[4]A. Crossley, E. H. Kisi, J. W. B. Summers, and S. Myhra, "Ultra-low friction for a layered carbide-derived ceramic, Ti_3SiC_2, investigated by lateral force microscopy (LFM)," *J. Phys. D: Appl. Phys*. **32**, 632–638 (1999).

[5]S. Myhra, J. W. B. Summers, and E. H. Kisi, "Ti_3SiC_2—A layered ceramic exhibiting ultra-low friction," *Materials Letters* **39**, 6–11 (1999).

[6]S. Y. Lienard et al., "Texture Development in Si_3N_4/BN Fibrous Monolithic Ceramics," *J. Mater. Sci*. **35**, 3365–3371 (2000).

[7]I-W. Chen et al., "Texture Development in $YBa_2Cu_3O_x$ by Hot Extrusion and Hot Pressing," *J. Am. Ceram. Soc*. **70**, C388–C390 (1987).

[8]X. Wu and I-W. Chen., "Hot Extrusion of Ceramics," *J. Am. Ceram. Soc*. **75**, 1846–53 (1992).

Microscale and Mesoscale Strain Measurements in Cement-Based Materials

C. R. Hubbard,[1] J. J. Biernacki,[1,2] and J. Bai[1,3]
[1]*Metals and Ceramics Division*
[2]*Tennessee Technological University*
[3]*University of Tennessee*

Understanding the microscale and mesoscale load bearing and load transfer characteristics of concrete could revolutionize the way that this complex heterogeneous material is engineered. Synchrotron X-rays were used to make measurements of microscale strains under mechanically generated stresses. Stressors were applied by in situ loading of specimens, and diffraction measurements were made to establish the stress states of various crystalline phases including calcium hydroxide, unreacted clinker phases, and aggregate phases.

Introduction

Concrete is the ubiquitous building material used in virtually every inhabited environment on earth. Despite concrete's benefits of low initial cost, formability, and field fabricability, there are numerous durability-related problems that limit the practical lifespan and life-cycle cost- effectiveness. The four primary mechanisms of degradation are (1) mechanical, (2) thermal, (3) chemical, and (4) drying in origin. While fundamentally different, these four families of *stressors* manifest in the generation of loads (stresses) that result in strains (deformations) that lead to formation of cracks and ultimately to failure.

The goal of this research is to develop and demonstrate experimental techniques that utilize ORNL's strength in synchrotron facilities to study mechanically induced stresses in cement, mortar, and concrete on a mesoscale and microscale. The diffraction experiment provides a direct measurement of the lattice spacings of each phase of a material. These spacings change as a function of mechanical, thermal, or chemical effects. Diffraction techniques provide direct measures of strains in individual phases and can also discriminate between tensile and compressive deformation.

Portland cement concrete is comprised of aggregate, cement paste, and water-filled or partially filled porosity. Both the aggregate and cement paste are themselves heterogeneous in nature, aggregate typically being sand (fine) and whatever rock (coarse) is locally available, and the paste being comprised of hydrated and unhydrated cement. The unhydrated cement fraction includes tricalcium silicate, dicalcium silicate, tricalcium aluminate, tetracalcium aluminoferrite, and gypsum.The hydrated cement fraction includes calcium silicate hydrate, calcium hydroxide, monosulfoaluminate, and ettringite, the respective hydration products of the anhydrous phases.

Research on fracture processes in concrete and other quasi-brittle materials shows that cracks initiate as microscale features.[1] As damage progresses, the scale and number of defects increases. Eventually the number density and size of defects grows such that the microscale and now mesoscale cracks merge to form macro flaws leading to complete failure. Prior research has focused on the measurement of macroscopic stress and strain measurements to interpret the cumulative effect of microscale degradation mechanisms. For example, macroscopic shrinkage measurements due to drying of concrete have been combined with macroscale constitutive relationships to predict drying shrinkage cracking in constrained concrete. Until now, virtually no direct measurements of the microscale strain states and strain distributions have been made for cement, mortar, or concrete.

Schulson et al.[2] recently demonstrated that stress levels within hardened cement paste subject to thermal loads could be quantified by measuring strains in the calcium hydroxide (CH) phase using neutron diffraction. Furthermore, their work demonstrated that stress distributions on a microscale are anisotropic. Although they were unable to measure strain levels and stresses developed upon freezing due to lack of instrumental resolution, Schulson et al.[3] also demonstrated that neutron diffraction can be used to study ice formation in cement paste.

Technical Approach

ORNL facilities at the National Synchrotron Light Source (NSLS) are suited for this research. The high flux and highly parallel X-ray beam at X14A makes it far superior to conventional laboratory X rays and results in extremely high resolution and flux needed for measuring very small strains.

Sample Preparation

Type I ordinary Portland cement meeting ASTM C150 specifications and Ohio River sand were blended in the appropriate proportions with tap water to form samples of cement paste, mortar, and concrete. Since the hardened test specimens were to be cubes no larger than 7.6 mm on a side, the particle size of the sand was separated into suitable fractions to produce mortar and concrete samples with workable aggregate distributions. Cement paste samples were prepared using a water/cement (w/c) ratio of 0.34. The Portland cement was mixed by hand with the water until a uniform paste was formed. The paste was then cast into a 1.9-cm-diameter glass jar to a level of about 5 cm. The jar was sealed and placed into a constant temperature water bath at 35°C.

Sand with particle size <425 μm was used to produce mortar samples (primarily quartz or Q). An aggregate (sand)-to-cement ratio (a/c) of 2.0 was used with a w/c of 0.45. The higher w/c ratio was necessary to provide a workable mortar mixture. Mortar mixtures were prepared by hand, sealed in glass jars, and cured at 35°C.

Concrete specimens were made by separating the river sand into two fractions and reconstituting it in proportions that would be more representative of a simulated concrete. Since the test specimen size is smaller than most single coarse aggregate particles, the concrete had to be prepared by blending the coarse fraction of the sand with the fine fraction of the sand to make a reasonable simulation. The sand was separated into <425-μm and >425-μm fractions. A simulated aggregate blend was made by reconstituting these fractions in an x/y ratio of coarse to fine fractions. The aggregate blend was then mixed with cement in an a/c ratio of 2.0 and with water at a w/c of 0.45. The components were mixed by hand, cast into glass jars, and cured at 35°C.

The specimens were cured in the sealed environment for at least 28 days, after which time they were removed from their jars and cut into 7.6-mm cubes using a diamond saw.

X-Ray Experiments

X-ray strain experiments were performed using beamline X14A at NSLS at a wavelength of 1.5425 Å (energy of 8.0377 keV, nominally equivalent to CuK_α) with a Ge analyzer crystal. The X-ray goniometer was fitted with a small load frame capable of applying up to 500 lb_f. Strain data was collected for one mortar and two cement samples. Θ-2Θ scans were made at various applied loads between zero and 500 lb_f for $2Q$ between 124 and 134°. This range was used because it contains two calcium hydroxide (CH) peaks, two quartz (Q) peaks, and one calcium carbonate (C) peak at angles 2Θ = 126.8660°, 130.430°, 127.537°, 131.516°, and 128.262°, respectively (d-spacings of 0.8623, 0.8496, 0.8598, 0.8458, and 0.8571 Å). These diffraction peaks are associated with hkl planes (3 1 0)-CH, (2 1 4)-Q, (3 0 5)-C, (1 1 6)-CH and (2 3-10)-Q. This rather high 2Θ range was selected since peak shift increases with increasing 2Θ for a constant applied load. X-ray scans were performed in sequences of load, no load, load or no load, load, no load to replicate conditions and observed changes in strain states due to sequential stressing.

A procedure called Y tilting was used to obtain the in-plane stress. Since a free surface is required for Θ-2Θ scanning, the load was applied in the sample plane (perpendicular to the diffraction vector when the tilt angle is zero). Θ-2Θ scans were performed at Ψ angles of zero, 25.66, 37.76, 48.59, and 60°. This creates a series of five equally spaced $sin^2\Psi$ values for the given five Ψ angles. In some cases additional Ψ angles selected from the following series; 22.52°, 32.03°, 32.8°, 41.56°, 43.21°, 50°, and 54.11° were also scanned.

Data Analysis

The strain measured in the laboratory reference frame ε'_{33} is related to the strain in the lattice reference frame ε_{33} according to the following equation:

$$(\varepsilon'_{33}) = \frac{d_{\phi\Psi} - d_0}{d_0} = \varepsilon_{11}\cos^2\phi\sin^2\Psi + \varepsilon_{12}\sin 2\phi\sin^2\Psi + \varepsilon_{22}\sin^2\phi\sin^2\Psi + \varepsilon_{33}\cos^2\Psi + \varepsilon_{13}\cos\phi\sin 2\Psi + \varepsilon_{23}\sin\phi\sin 2\Psi \quad (1)$$

where $d_{\phi\Psi}$ = the measured strained d spacing, d_0 = the measured unstrained d spacing, ε_{ij} = the lattice strains in directions 1, 2, and 3, ϕ = the angle of rotation of the sample about laboratory direction 3, and Ψ = the laboratory tilt angle. As a result, the observed strain ε'_{33} can be a complex function of $sin^2\Psi$.[4] Three functional behaviors are possible: (1) linear, (2) nonlinear split, and (3) nonlinear oscillating.

Diffraction d-spacings ($d_{\phi\Psi}$) were determined using profile fitting and Bragg's Law. Each 2Θ range was fit assuming a pseudo-Voigt peak form with either constant or variable peak width. The apparent strain was computed using the left-hand side of Eq. (1). All strains were referenced to the Ψ = 0.0°; no load condition (taken as d_o). The strains, so determined, were then plotted as a function of the $sin^2\Psi$ for conditions of load and no load.

Results and Accomplishments

Strain in Cement Specimens

Plots of strain as a function of $sin^2\Psi$ were made for each of the samples tested—two cement cubes and one mortar cube. Cement sample 1 was tested in an initial no load, load, no load, load sequence with loads of roughly 10 lb_f, 250 lb_f, 20 lb_f and 300 lb_f, respectively. The initial no load data, however, was limited to a single Ψ angle of 0.0°. Figures 1 and 2 summarize the results. Figures 1(a) through 1(e) include $sin^2\Psi$ plots for two peaks associated with CH (d = .8623 and .8496, 2Θ = 126.866 and 130.430°), two peaks associated with quartz (d = .8598 and .8458, 2Θ = 127.537 and 131.516°), and one peak

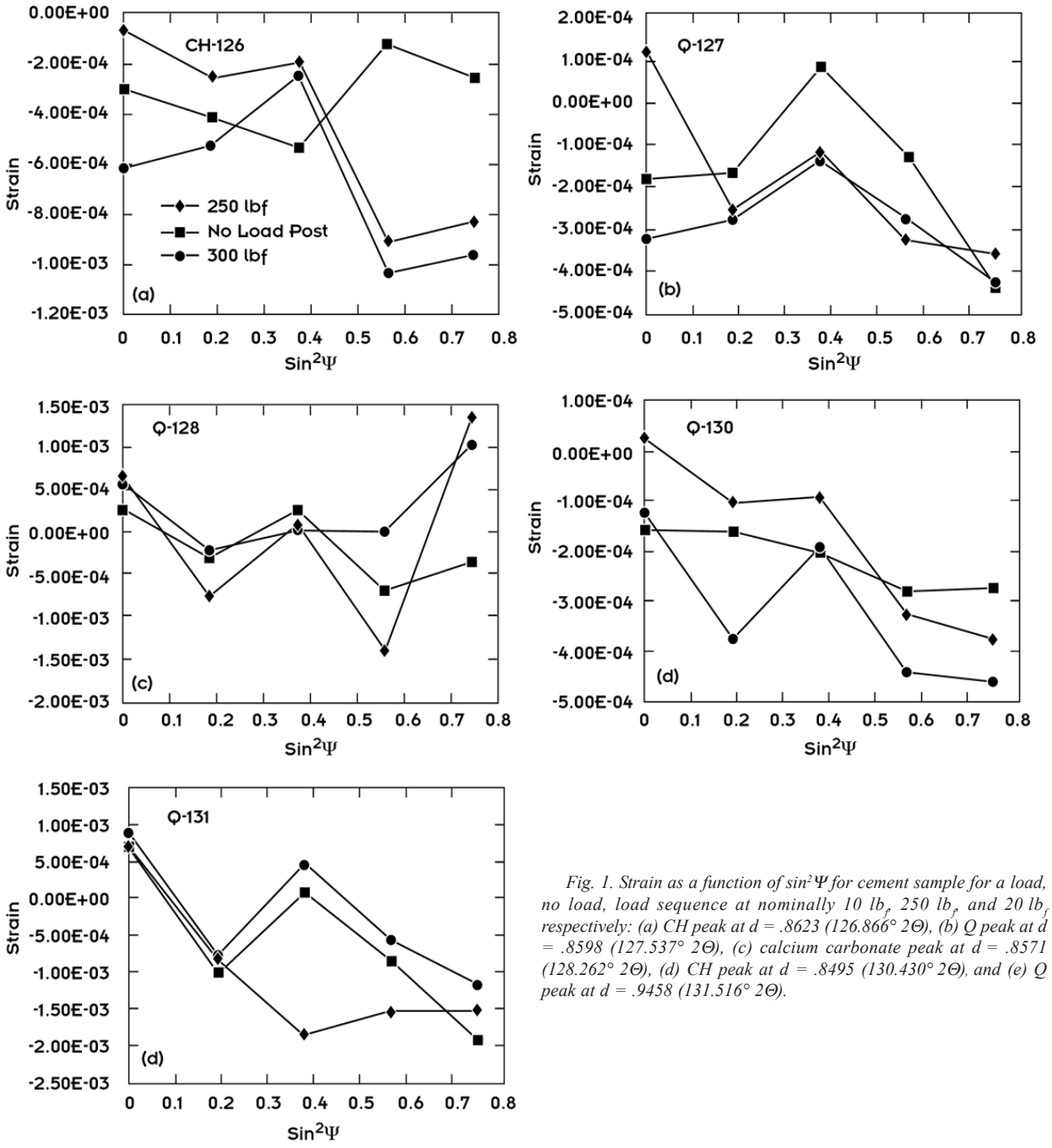


Fig. 1. Strain as a function of $sin^2\Psi$ for cement sample for a load, no load, load sequence at nominally 10 lb_f, 250 lb_f, and 20 lb_f respectively: (a) CH peak at d = .8623 (126.866° 2Θ), (b) Q peak at d = .8598 (127.537° 2Θ), (c) calcium carbonate peak at d = .8571 (128.262° 2Θ), (d) CH peak at d = .8495 (130.430° 2Θ), and (e) Q peak at d = .9458 (131.516° 2Θ).

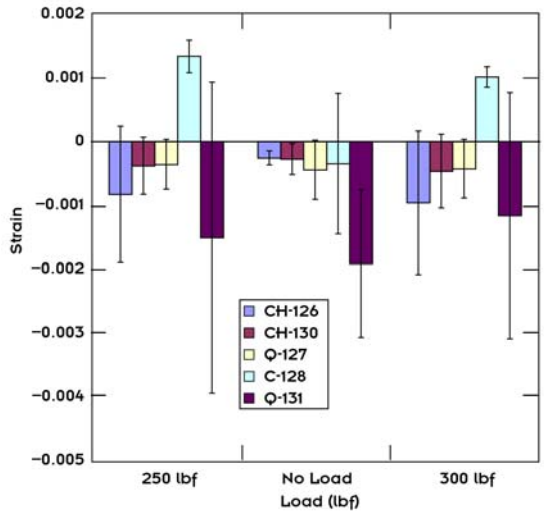


Fig. 2. Strain at $\Psi = 60°$ as a function of load for cement Sample 1 for various CH, Q, and calcium carbonate peaks.

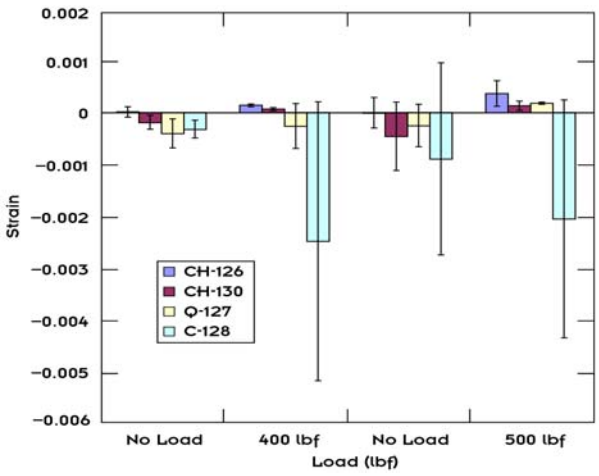


Fig. 3. Strain at $\Psi = 60°$ as a function of load for cement sample 2 for various CH, Q and calcium carbonate peaks.

associated with calcium carbonate (d = .8571, 2Θ=128.262°). Figure 2 is a bar chart comparing the strain at Ψ = 60° for the load, no load, load sequence all referenced to the initial no load d-spacings at Ψ = 0.0°.

Figure 1 suggests that the strain behavior for hardened cement paste is of a complex nonlinear form. Although the diffraction peaks used are from different phases and different hkl planes, it is expected that the strain behavior would be qualitatively similar, albeit the absolute strains may differ due to crystal anisotropy. While complex and somewhat variable, the general trend in the $sin^2\Psi$ plots is the same for the CH and Q peak pairs, Figures 1(a) and 1(d), and 1(b) and 1(e). When loaded, CH exhibits a larger oscillatory strain behavior, as a function of $sin^2\Psi$, than when unloaded. When the load is removed, the behavior suggests some level of compressive residual loading (relative to the no load Ψ = 0.0° d-spacing) which may be somewhat oscillatory in nature. Most distinct is the repeatability of the compressive strain state noted for Ψ = 60° (Fig. 2). Finally, when loaded, the strain behavior, while similar on repeated cycles, appears to vary some from cycle to cycle, suggestive of plastic deformation in some of the phases in the concrete. This, however, is difficult to establish at this time since, in this case, only two no-load load cycles were recorded.

The quartz peak pair illustrates a similar oscillatory $sin^2\Psi$ behavior. Again, similarities are shared between the peak pairs. Consistent with CH, the strain behavior appears to change with load cycle. The strain state at Ψ = 60° is also consistent and appears repeatable for both peaks (Fig. 2).

A single carbonate peak was observed. Again the strain behavior appears to change from load to no load and with loading cycle. As for CH and Q, calcium carbonate displays a distinct and repeatable change in strain state most evident at Ψ = 60° (Fig. 2).

A second series of experiments was also conducted on a different cement sample. Figure 3 summarizes the strain response for the same series of CH, Q, and calcium carbonate peaks for a no-load/load/no-load/load sequence. Again the results suggest a somewhat variable yet repeatable pattern of changing strain states with changing stress states. Detailed plots of strain as a function of $sin^2\Psi$ are similar to those for cement Sample 1 shown in Fig. 1.

Strain in Mortar Specimens

A single mortar specimen was tested in a no-load, load, no-load cycle. While the results are similar to those found for the cement sample, they are also somewhat more ambiguous. Figure 4 summarizes the strain response for Ψ = 60°. Again, the responses appear systematic except

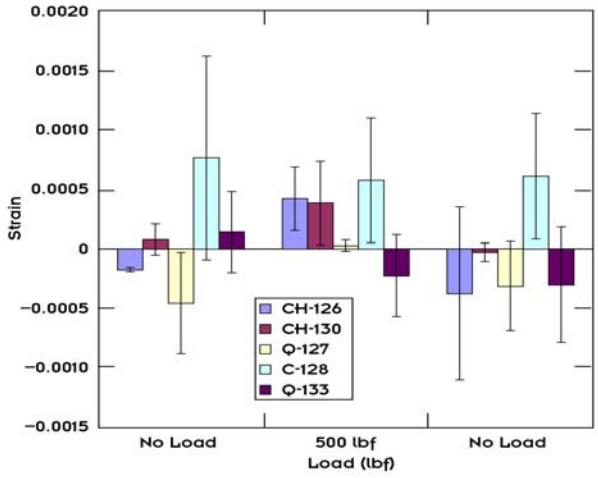


Fig. 4. Strain at $\Psi = 60°$ as a function of load for mortar Sample 1 for various CH, Q and calcium carbonate peaks.

for the calcium carbonate peak that was difficult to resolve in the presence of the strong Q peaks in mortar. Although systematic with no-load, load, no-load cycle, both the Q peak pair and CH peak pair demonstrate compressive behavior for one hkl plane and tensile behavior in the other. At this time it is unclear if there is a physical interpretation for this response. Furthermore, since standard construction grade sand was used, the sand was not *pure* quartz and appears to contain a variety of other compounds that interfere in the 124° to 134° 2Θ range scanned.

Summary and Conclusions

These results, while based on limited experiments, support the proposed hypothesis that strain states in several phases in Portland cement–based materials can be measured using synchrotron X-ray diffraction methods. Further, it clearly demonstrated that the response of different phases are different and that understanding stressors in cements requires such a microscale probe. Based on this research, current efforts to develop models that aim to predict crack initiation due to applied loads or other stressors must include the response of each phase instead of assuming a bulk average.

The variability is possibly due to the low depth of penetration of ~1.54-Å X rays and the highly parallel beam resulting in a consequential limited number of grains sampled. Further, the limited penetration depth results in the X-ray probe sampling just near surface grains. Possible solutions could be improved methods to involve more grains contributing to the X-ray results via oscillation of the sample and/or use of higher-energy X rays. Another alternative is to use neutron diffraction methods due to the far greater depth of penetration of neutrons. However, for neutron diffraction the hydrogen should be replaced with deuterium to minimize incoherent background scattering. These approaches will be included in follow-up proposals.

The development and demonstration of synchrotron-based diffraction techniques to quantify mesoscale and microscale stress and strain behavior in concrete has broad reaching application-based implications. The concrete industry has recently prepared a vision document and a draft technology roadmap in which they identify specific research needs.[5,6]

Among the needs identified in Roadmap 2030 as "high priority items" are development of "new materials to reduce shrinkage and cracking," "reduction of alkali-silica reactions in concrete," "prediction methods and models for . . . cracking, durability and performance (including environmental interaction)," "tools and data for quantifying benefits of using alternative materials," and "multiscale modeling to connect microstructure with engineering properties." Each of these areas and others can benefit from the synchrotron method and the proposed neutron diffraction method. Once fully developed, the phase-sensitive nondestructive analytical methods would enable a researcher to test the microscale impact of prospective shrinkage-reducing admixtures on stress and strain development. For the first time, researchers could study the impact that chemical modification of cement has on alkali-silica interaction-induced stresses. Further, the experimental data could be used to validate and evolve multiscale model development for prediction of cracking, durability, and performance as they relate to interactions such as sulfate, alkali-silica, freeze-thaw, and other environmental stressors.

The joint TTU/ORNL team lead by PI Biernacki (TTU) has secured follow-on funding through the National Science Foundation (NSF) for a three-year $235,000 effort to extend this line of research—specifically to further explore the use of synchrotron XRD and to initiate efforts with neutron diffraction, Raman spectroscopy, and direct microscopic measurement of stressed samples. This project will be conducted in part at ORNL facilities.

References

[1]S. P. Shah, S. E. Swartz, and C. Ouyang, *Fracture Mechanics of Concrete : Applications of Fracture Mechanics to Concrete, Rock and Other Quasi-Brittle Materials*, Wiley, New York (1995).

[2]E. M. Schulson, I. P. Swainson, and T. M. Holden, "Internal Stress Within Hardened Cement Paste Induced Through Thermal Mismatch Calcium Hydroxide versus Calcium Silicate Hydrate," *Cem. Concr. Res.* **31**, 1785–1791 (2001).

[3]E. M. Schulson, I. P. Swainson, T. M. Holden, and C. J. Korhonen, "Hexagonal Ice in Hardened Cement," *Cem. Concr. Res.* **30**, 191–196 (2000).

[4]I. C. Noyan and J. B. Cohen, *Residual Stress Measurement by Diffraction and Interpretation*, Springer-Verlag, New York, pp. 276 (1987).

[5]*Vision 2030: A Vision for the U.S. Concrete Industry*, Strategic, Development Council, American Concrete Institute (2001).

[6]*Roadmap 2030: The U.S. Concrete Industry Technology Roadmap*, Strategic Development Council, American Concrete Institute (Draft, May 10, 2002).

Investigation of Tribological Properties of Graphitic Foam Reinforced Carbon-Carbon Composites

P. J. Blau,[1] J. Qu,[2] J. Klett,[1] and B. Jolly[3]
[1]*Metals and Ceramics Division, ORNL*
[2]*Post-doctoral Research Associate, ORISE, Oak Ridge, TN*
[3]*Pellissippi State Technical College, Knoxville, TN*

Research at ORNL has led to the development of a new class of high-thermal-conductivity carbon foam materials. While the applications of these materials are primarily aimed at rapid heat removal, recent experiments have revealed their potential as bearing surfaces as well. The three primary advantages are (1) they can efficiently remove frictional heat, (2) their natural porosity can trap debris, and (3) the porosity can serve as a lubricant reservoir. A series of pin-on-disk experiments was conducted at room temperature and 400°C to compare the sliding friction and wear characteristics of a densified form of the carbon foam material, mated against M-50 tool steel or alumina, to those of conventional bearing materials such as graphite, bearing bronze, poly-tetrafluoroethylene (PTFE), bearing steel, and a cobalt-based superalloy. At room temperature and under low contact pressure, the tribological behavior of the densified carbon foam material was comparable to that of graphite and better than that of other bearing materials. At 400°C, traditional graphite exhibited a 'dusting' wear regime and experienced a very high friction coefficient. In contrast, the carbon foam demonstrated the ability to maintain low friction and wear at elevated temperature. Further work is needed to elucidate the sliding mechanisms and optimize material properties for high-temperature tribological applications.

Introduction

High-thermal-conductivity graphite foam, developed by researchers at ORNL, has shown considerable promise for use in electronics cooling systems and vehicle radiators.[1] With a thermal conductivity comparable to that for aluminum and one-fifth of that metal's density, this material has the potential to solve thermal management problems in a variety of industrial and military applications. However, in many instances the low density and low strength of the porous foam limit its applicability. In order to overcome the limitations of the as-formed carbon foam, a novel densification process has been developed to improve the density, thermal conductivity, and compressive strength.[2–3]

Certain characteristics of carbon foam, such as its low density, inherent porosity, and attractive thermal properties, suggest the possibility of using it in tribological (sliding bearing) applications. For example, surface-breaking voids have been reported to reduce the friction coefficient by serving as a repository for wear debris[4] or to hold lubricants. The porosity of carbon foam may provide a built-in ability to do these things. Secondly, the maximum operating temperature for the carbon foam is 500°C higher than that for most commercial bearing materials. Thirdly, the high thermal conductivity of the foam could help dissipate frictional heat. This could allow bearings to run at higher speeds without experiencing thermal failure. Finally, the low density of the foam is attractive for weight-sensitive applications and, hence, energy efficiency improvements.

Technical Approach

The friction and wear characteristics of densified carbon foam were measured at room temperature at 400°C for comparison to traditional bearing materials tested under the same conditions. Table 1 lists the composition, density, and Vickers microindentation hardness of the materials used in this research. Friction and wear tests were conducted using a flat-ended pin-on-disc geometry. A custom-built, pin-on-disk testing machine was used. The pin specimen was a bearing ball whose tip was ground and polished to create a flat spot of either 2 or 4 mm in diameter. The use of a flat-ended pin avoided the concentrated contact stresses associated with a spherical tip. Initial tests with rounded pins crushed the ligaments between pores to create a deep wear groove. Tests were conducted at a normal load of 10 N and a sliding speed of 1.0 m/s for a total pin sliding distance of 5000 m. The disk wear tracks were all of the same diameter (21 mm).

It was originally proposed to use a Sub-Scale Brake Tester (SSBT) for high-speed pin-on-disk testing (up to 10 m/s). This required the use of 230-mm-diameter disks, but the large, densified carbon foam disks were not ready for testing by the end of FY 2003. However, large disks

Table 1. Test materials

Specimen	Material designation	Composition/description	Density (g/cm^3)	HV, 100 gf load (GPa)
Pin	M-50 tool steel	Nominal composition (wt%): 0.85 C, 0.10 W, 4.25 Mo, 4.00 Cr, 1.00 V, bal. Fe (Winsted Precision Ball Co., CT)	7.97	9.08
	Polycrystalline alumina	99.5 wt% dense aluminum oxide, ABMA grade 25 ball (McMaster-Carr)	3.90	24.75
Disk	Densified carbon foam	Graphitic foam reinforced carbon-carbon composites, 46% relative density (ORNL—raw foam, SMJ Carbon Technology—densification)	1.26	*N/M*[a]
	Graphite ACF-10Q	Commercial grade, fine-grained graphite, 5 µm particle size, 0.8 µm pore size, 80% relative density (POCO Graphite, TX)	1.80	*N/M*[a]
	Alloy 932 bronze	Nominal composition (wt%): 83 Cu, 7 Sn, 7 Pb, 3 Zn (McMaster-Carr)	8.91	1.34
	PTFE (Teflon)	Virgin electrical grade Teflon™ sheet (McMaster-Carr)	2.14	*N/M*[a]
	AISI 52100 steel	Nominal composition (wt%): 0.95–1.1 C, 1.3–1.6 Cr, 0.2–0.5 Mn, 0.35 max. Si, 0.025 max P, 0.025 max. S (McMaster-Carr)	7.81	8.32
	Stellite 6B	Heat composition (wt%): 1.09 C, 29.58 Cr, 2.75 Fe, 0.06 Mo, 1.48 Mn, 2.66 Ni, 0.01 N, <0.01 P, 0.58 Si, 3.82 W, bal. Co (High Performance Alloys, IN)	8.39	5.26

[a]N/M = not measurable with a Vickers indentation test.

specimens have been obtained and are nearly finished at the time of this report. High-speed tests of the materials are planned for a new follow-on effort which has recently been funded by DARPA.

Results and Accomplishments

Friction and Wear Behavior at Room Temperature

The friction and wear results of the densified carbon foam and other bearing material disks against M-50 tool steel and alumina pins at room temperature are summarized in Tables 2 and 3. The wear factor, as used here, is defined as the wear volume normalized by the applied load and the sliding distance of the pin.

The carbon foam had relatively low friction coefficients, 0.31 and 0.23, for M-50 steel and alumina pins, respectively. The friction coefficient trace was fairly stable and free from significant fluctuations. More promisingly, this material showed fairly high wear resistance. Fracture and a high wear rate, which might have been intuitively expected for foam-structured materials, did not occur. The wear factor of the densified carbon foam disk was relatively low, 5.6×10^{-6} and 1.3×10^{-5} mm^3/N-m for the steel and alumina sliders, respectively. No debris was found on the wear track (Fig. 1), possibly having been trapped by the pores. The worn surface was flat and smooth without evidence of surface damage. The porous macrostructure survived. The swirled microstructure between pores disappeared after the unidirectional sliding test—probably due to the surface crystallites being oriented with their basal planes roughly parallel to the surface. The steel and alumina pins had no measurable wear. A thin film of carbon was formed on the sliders, however, and this helped to stabilize the sliding conditions after a brief running-in period.

Graphite performed similarly to the carbon foam material with even lower friction coefficient and wear rate. The friction coefficient was very stable. The corresponding wear track is shown in Fig. 2. Sliding action smeared the surface and wiped off the porous microstructure, leaving little wear debris. A thin transfer film of graphite on the pin specimens protected the sliders from damage.

Bronze and Teflon disks had relatively low friction coefficients, but high wear rates, as shown in Tables 2 and 3. The friction coefficient traces for bronze and Teflon show higher variations compared to carbon foam or graphite. Large amounts of flake shape debris were collected on the bronze and Teflon disks, evidence of low shear strength which contributed to the low friction and high wear. The tests on 52100 steel and Stellite 6B alloy had high average friction coefficients ($0.5 \sim 0.6$) with large variations ($>\pm 0.1$), as shown in Table 2. The contact surfaces on both the pins and disks were damaged, and the wear factors were on the order of $10^{-6} \sim 10^{-5}$ mm^3/N-m. Material removal and deformation were observed on the wear tracks of the 52100 steel and Stellite 6B disks. The pin surfaces were scratched and covered with transferred material.

Table 2. Kinetic friction coefficient

Friction coefficient	Room temperature		400°C
	M-50 steel	Alumina	M-50 steel (test #1,#2)
Carbon foam	0.31 ± 0.005	0.23 ± 0.005	0.15 ± 0.003, 0.57 ± 0.01
Graphite	0.21 ± 0.002	0.24 ± 0.002	1.2 ± 0.2, 0.8 ± 0.1
Bronze	0.19 ± 0.02	0.20 ± 0.03	–
PTFE	0.21 ± 0.01	0.19 ± 0.02	–
52100 steel	0.60 ± 0.08	0.61 ± 0.12	–
Stellite 6B	0.50 ± 0.13	0.56 ± 0.16	0.55 ± 0.06, 0.54 ± 0.04

Table 3. Wear rates of disk specimens[a]

Wear rate	Room temperature		400°C
(mm^3/N-m)	M-50 steel	Alumina	M-50 steel (test #1,#2)
Carbon foam	5.6×10^{-6}	1.3×10^{-5}	1.8×10^{-5}, 1.7×10^{-5}
Graphite	1.0×10^{-6}	2.5×10^{-6}	1.6×10^{-5}, 2.2×10^{-5}
Bronze	2.8×10^{-4}	2.9×10^{-4}	–
PTFE	1.1×10^{-3}	1.0×10^{-3}	–
52100 steel	7.5×10^{-6}	1.6×10^{-5}	–
Stellite 6B	5.8×10^{-5}	7.6×10^{-5}	3.5×10^{-7}, $<1.0 \times 10^{-7}$

[a]Typical wear rates of bearing materials for unlubricated sliding should be less than ~10^{-5} mm^3/N-m. Values for well-lubricated sliding should be less than ~10^{-7} mm^3/N-m.

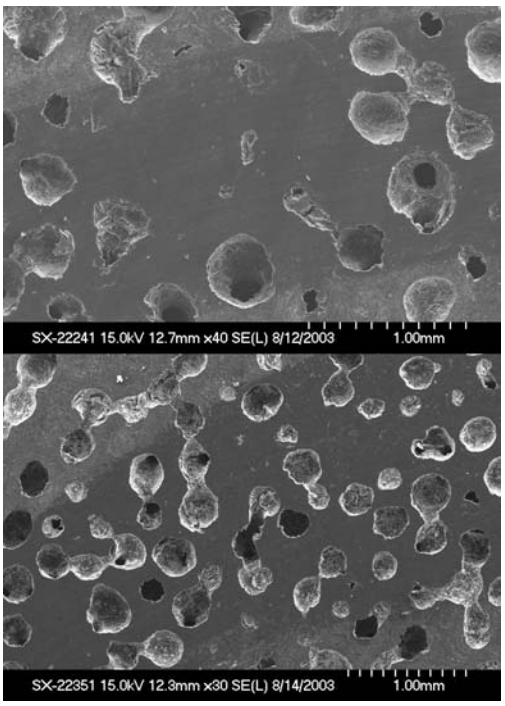


Fig. 1. SEM images of wear tracks on densified carbon foam disks at room temperature (top) and 400°C (bottom).

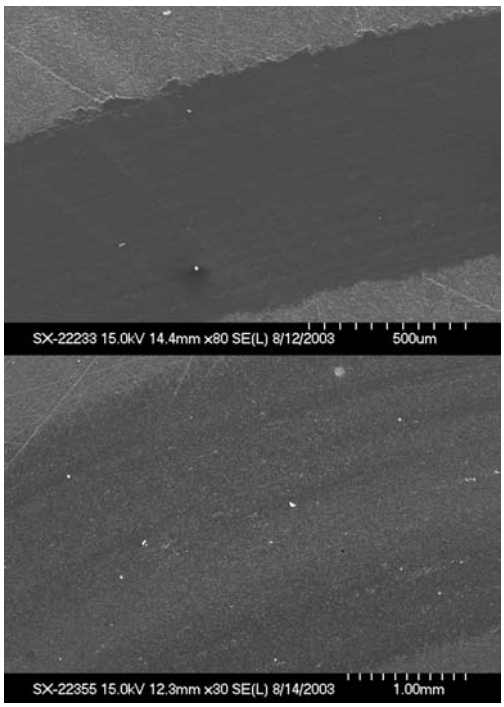


Fig. 2. SEM images of wear tracks on graphite disks at room temperature (top) and 400°C (bottom).

Notably, there was a lower friction coefficient for the alumina pin than for the M-50 steel pin against the carbon foam disk. In contrast, the friction coefficient for the alumina pin against other disk materials was comparable or slightly higher than that for the M-50 steel.

Friction and Wear Behavior at Elevated Temperature

Friction and wear tests were also conducted at 400°C in ambient air. Due to softening and oxidation problems, bronze, Teflon, and 52100 steel cannot be used at temperatures higher than about 300°C. Consequently, graphite and Stellite 6B were used as a basis for comparison with the densified foam. The slider was an M-50 tool steel pin with a 4 mm-diameter flat contact surface.

Graphite and carbon-carbon composite materials may significantly deteriorate their tribological properties in vacuum,[5] inert gases,[6] or elevated temperature,[7] by showing high friction coefficients (0.4 ~ 1.0) and high wear rates. The propensity of carbon materials to expel clouds of fine wear particles is known as "dusting." It is now widely accepted that the self-lubricating properties of graphite are environmentally dependent, and the shear strength along the basal planes is lowered only in the presence of moisture, oxygen, or certain other gases that can be adsorbed into the graphite structure.[8–9]

A critical temperature for graphite and carbon-carbon composites to transition to dusting has been reported in the range of 150 ~ 200°C,[7] above which high friction and wear occur due to the desorption of water. In this study, however, high wear rates were not produced in 400°C tests on either the graphite or the densified carbon foam. Further study of the surface chemistry of sliding films is needed and will be pursued in follow-on efforts. The densified carbon foam exhibited a very low friction coefficient in just one of the two high-temperature tests, and while results demonstrate that low friction is possible, it still remains for us to learn how to achieve such low friction repeatedly.

As shown in Tables 2 and 3, graphite had very high friction coefficient, $\mu = 1.2$ and 0.8, but relatively low wear factors, 1.6×10^{-5} and 2.2×10^{-5} mm^3/N-m, in the two tests at 400°C. The friction coefficient trace of test #2 is shown in Fig. 3. There was a low-friction running-in stage ($\mu = 0.04 \sim 0.16$), apparently from residual water vapor in the surface porosity, then μ increased quickly as the water vapor was removed by heating and sliding. The friction coefficient eventually reached steady state (0.8 ± 0.1). Test #1 had similar friction behavior, but the friction coefficient climbed to a surprisingly high level, 1.2 ± 0.2. This indicated strong adhesion in the contact area. The worn surface (Fig. 2) was smooth, with the porous microstructure remaining intact. Smears and deformation were significantly less than that at room temperature, indicating higher resistance to shear at elevated

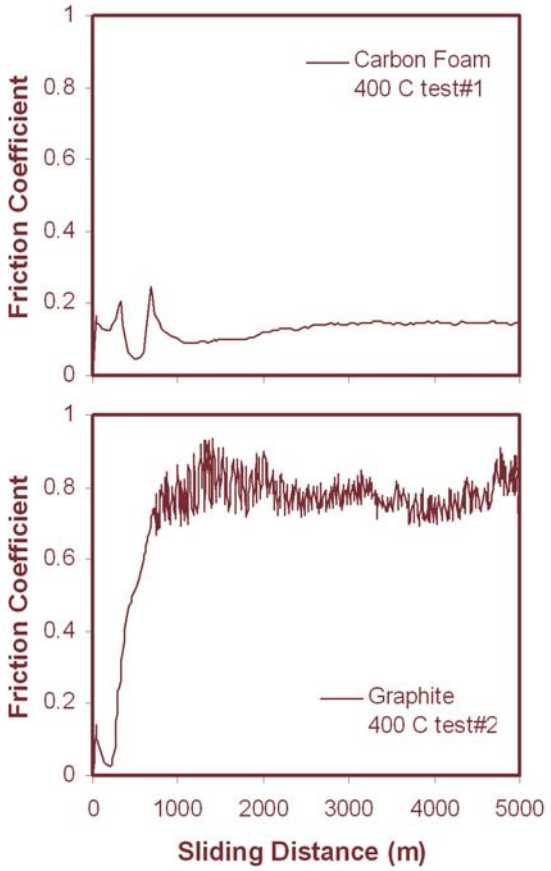


Fig. 3. Friction coefficient traces of densified carbon foam (top) and graphite (bottom) at 400°C.

temperature. Two types of debris were observed: very fine black dust and larger-sized shiny particles. According to EDS X-ray analysis, the black dust was carbon generated during dusting and the shiny particles were iron, chromium, and their oxides, worn off the steel slider. Loud noises, which sounded like metal scraping against metal, was generated in the tests of steel against graphite at 400°C.

Similar to the tests on graphite, the tests on the densified carbon foam material also had a low-friction running-in stage. The friction coefficient traces of the two replicates had a similar initial trend, showing double peaks in the range of $\mu = 0.2 \sim 0.25$ and a single valley around $\mu = 0.05$. After the second peak, the two traces headed in opposite directions. The first one (Fig. 3) dipped and stabilized at $\mu = 0.15 \pm 0.003$. The second one (not shown here) rose after 1000-m sliding and eventually transitioned into the dusting regime with $\mu = 0.57 \pm 0.01$ by the end of the test. Despite different frictional behavior, the wear factors for the two tests were low and similar, 1.8×10^{-5} and 1.7×10^{-5} mm^3/N-m, respectively. Results indicate that unlike graphite, the densified carbon foam material does not necessarily transition into a high friction regime at elevated temperature in ambient air. This suggests the

possibility for using carbon foam as a high-temperature bearing material, but the favorable behavior must be made repeatable.

Stellite 6B showed comparable friction and wear behavior at room temperature and 400°C. Energy-dispersive X-ray analysis indicated that the wear track on the Stellite 6B disk was covered by a transfer layer of iron and iron oxides.

Summary and Conclusions

The friction and wear characteristics of a novel, high thermal conductivity, densified carbon foam material were compared with those for several conventional bearing materials at room temperature and 400°C without additional lubrication. High contact stresses can crush the foams, but if the load is distributed over a planar contact region, low friction and wear can be obtained without fracture or chipping problems. At room temperature, the friction and wear rates of carbon foams were comparable to those of graphite and better than those of the other bearing materials we tested. At the elevated temperature, Stellite 6B behaved similarly to room temperature, but graphite exhibited dusting and experienced high kinetic friction coefficients, up to $\mu = 1.2$. In one test against M-50 tool steel at 400°C, the carbon foam material had a low friction coefficient ($\mu = 0.15$), but in a repeated test, it had a high friction coefficient ($\mu = 0.57$). Despite this, the wear factors for both tests were similar. We have shown that it is possible to obtain low friction and wear on the densified carbon foam at elevated temperature. As the sliding mechanisms are clarified, it should be possible to optimize the material properties to obtain low friction and low wear behavior repeatedly. Support from the Defense Advanced Research Projects Agency that resulted from this project will enable such research to be started during FY 2004.

References

[1]J. Klett, R. Hardy, E. Romine, C. Walls, and T. Burchell, "High-Thermal-Conductivity, Mesophase-Pitch-Derived Carbon Foams: Effect of Precursor on Structure and Properties," *Carbon* **38**, 953–973 (2000).

[2]P. G. Wapner, W. P. Hoffman, and S. Jones, *Carbon and Ceramic Matrix Composites Fabricated by a rapid Low-Cost Process Incorporating in-situ Polymerization of Wetting Monomers,* U.S. Patent 6,309,703 (2001).

[3]J. Klett, S. Jones, L. Klett, and C. Walls, "High Thermal Conductivity Graphitic Foam Reinforced Carbon-Carbon Composites," *SAMPE 2003*, May 11–15 2003, Long Beach, California (2003).

[4]A. P. Semenov, "Tribology at High Temperature," *Tribology International* **28**(1), 45–50 (1995).

[5]R. I. Longley, J. W. Midgley, A. Strang, and D. G. Teer, "Mechanism of the Frictional Behavior of High, Low, and Non-Graphitic Carbon," *Lubrication and Wear Convention*, 198–209 (1963).

[6]F. Robert, D. Paulmier, H. Zaidi, and E. Schouller, "Combined Influence of an Inert Gas Environment and a Mechanical Action on a Graphite Surface," *Wear* **181–183**, 670–690 (1995).

[7]B. K. Yen, T. Ishihara, and I. Yamamoto, "Influence of Environment and Temperature on 'Dusting' Wear Transitions of Carbon-Carbon Composites," *Journal of Materials Science* **32**, 681–686 (1997).

[8]J. K. Lancaster and J. R. Pritchard, "The Influence of Environment and Pressure on the Transition to Dusting Wear of Graphite," *Journal of Physics D: Applied Physics* **14**, 747–762 (1981).

[9]M. Brendle and P. Stempe, "Triboreactions of Graphite with Moisture—a New Model of Triboreactor for integrating Friction and Wear," *Wear* **254**, 818–826 (2003).

Tailoring the Properties of Crystalline Solid Solutions by Magnetic and Stress Annealing

E. D. Specht,[1] C. J Sparks,[1] R. K. Williams,[1] and A. Gali[2]
[1]*Metals & Ceramics Division*
[2]*Department of Materials Science and Engineering, University of Tennessee*

It is demonstrated that magnetic annealing affects the atomic short-range order of $FeNi_3$. While this in accord with the commonly accepted model for magnetic annealing of substitutional alloys, annealing is shown to produce order which is *strong* and *isotropic*, while the model predicts a *weak* and *anistropic* affect on short-range order. The Seebeck coefficient is shown to provide a rapid measure of short-range order.

Introduction

The beneficial effects of annealing magnetic materials in magnetic field have been known since 1913.[1] Similarly, magnetic anisotropy can be induced by annealing in a stress field.[2] The commonly accepted model for both these effects is that both these effects are due to local redistribution of atoms creating a *small* preference for atomic pairs oriented to minimize their energy in the magnetic or stress field.[3] In an effort to make the first direct test of this hypothesis, we discovered magnetic annealing conditions which create a *large* increase in atomic short-range order in $FeNi_3$ alloys. This project confirms this increase and measures the Seebeck effect to assess its potential as a rapid screening tool to determine which annealing conditions give rise to enhanced order.

Technical Approach

Samples were annealed using the apparatus shown schematically in Fig. 1. Sealed in a quartz capsule, $FeNi_3$ samples were first heated to high temperature to create local atomic disorder, then heated in vacuum with an applied magnetic field under conditions which lead to large increases in short-range order. One set of control samples was annealed in magnetic field under conditions leading to small increases in short-range order, and another set of control samples was annealed in no magnetic field.

Diffuse X-ray scattering is used as a direct measure of atomic short-range order. At high temperatures, $FeNi_3$ is disordered, with Fe and Ni atoms distributed randomly on an fcc lattice. At low temperatures, $FeNi_3$ forms an ordered L12 crystal with *fundamental* Bragg peaks on the fcc reciprocal lattice points and *superstructure* Bragg peaks on the other cubic reciprocal lattice points. Both these structures have cubic symmetry; the applied field leads to no anisotropy. Samples are annealed so as to produce local order only; the fundamental Bragg peaks then remain sharp, while the superstructure Bragg peaks are broadened into short-range order peaks. The width of the short-range order peaks provides a measure of the length scale of the local order. Any anisotropy due to the applied field is seen as a difference in the intensities of the various superstructure peaks. For example, if the sample orders along the [100] direction, the (100) short-range order peak will be stronger than the (010) or (001) peaks.

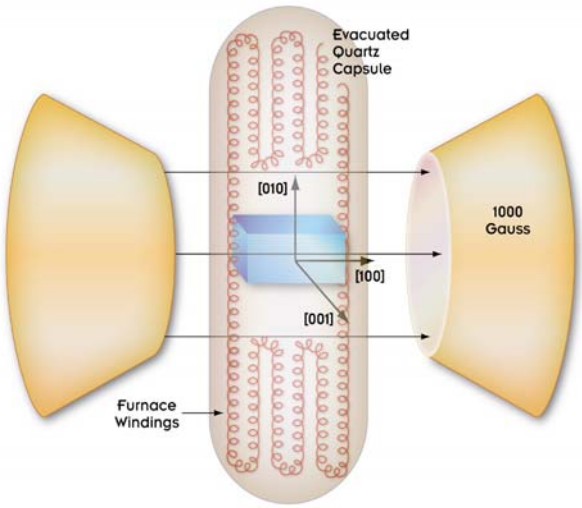


Fig. 1. Apparatus used to anneal $FeNi_3$ in an applied magnetic field.

While diffuse X-ray scattering provides the most direct measure of short-range order, it is a slow measurement which requires the use of synchrotron radiation sources which are very limited in availability.

Seebeck coefficients were measured to assess their value as a rapid probe of short-range order. The Seebeck apparatus is shown schematically in Fig. 2; we followed the procedure described elsewhere in detail.[4] One end of the sample is heated to create a thermal difference of ~1°C. Chromel-alumel thermocouples are attached to the hotter and cooler end of the sample. The voltages of all four leads are measured. The temperature of each end of the sample is calculated from the thermocouple voltages V_{1c}- V_{1a} and V_{2c}- V_{2a} (voltages are defined in Fig. 2). The differences in voltage for each lead material, V_{1a}- V_{2a} and V_{1c}-V_{2c}, are due to the combined Seebeck effect of the thermocouple material and of $FeNi_3$. Subtracting the effect of the thermocouple material leaves the Seebeck voltage of $FeNi_3$; dividing by the temperature difference T_1- T_2 gives two independent measurements of the Seebeck coefficient of $FeNi_3$. Each sample is measured at two temperatures.

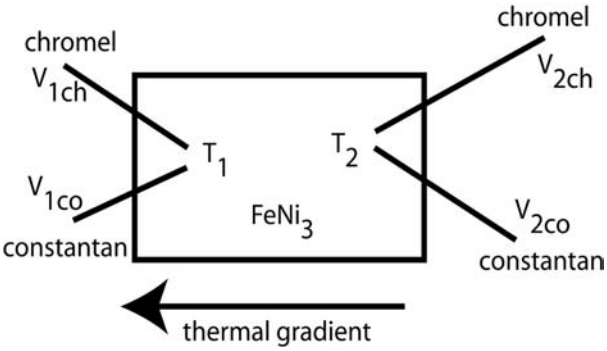


Fig. 2. Apparatus used to measure Seebeck coefficients.

Results and Accomplishments

Diffuse X-ray scattering confirmed our preliminary results. Samples annealed in zero magnetic field show very weak short-range order, barely extending beyond near-neighbor atoms. Annealing conditions which give weak short-range order produce ordering which extends ~0.5 nm. Annealing conditions which give the strongest short-range order produces ordered domains of up to 5 nm.

Repeating these measurements at various superstructure reciprocal lattice vectors, however, yield data which contradicts the commonly-accepted model that magnetic annealing leads to magnetic anisotropy through the local rearrangements of atoms. All the short-range order peaks have the same intensity, within experimental uncertainty of ~5%. Consequently, none of the magnetic anisotropy can be attributed to anisotropic local atomic ordering.

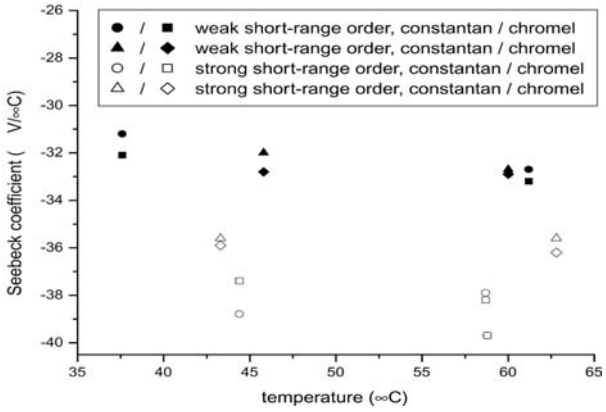


Fig. 3. Seebeck coefficients for $FeNi_3$, measured with both constantan and chromel leads. Two samples are annealed in magnetic fields to give weak short-range order, two to give strong short-range order.

The Seebeck coefficients (Fig. 3) for the samples annealed for weak short-range order are consistent with the results for highly disordered $FeNi_3$, while those for the samples annealed for strong short-range order are between those for disordered and fully ordered $FeNi_3$.[4]

Summary and Conclusions

The effects of magnetic annealing on short-range atomic order in $FeNi_3$ have been confirmed; under the proper conditions, short-range order forms to a far greater extent than in the absence of a field. The isotropic nature of the short-range order, however, is at odds with the prevailing model for magnetic annealing. These studies can form the basis for reassessment of the fundamental mechanisms by which magnetic and stress annealing affect the properties of solid solution alloys. We will pursue funding for this effort from Basic Energy Sciences; successful explanation of the magnetic and stress annealing effects will lead to funding in applied areas.

We have demonstrated that measurement of Seebeck coefficients provides a rapid measure of short-range order, which will greatly facilitate future studies.

References

[1]H. Pender and R.L. Jones, *Phys. Rev.* **1**, 259 (1913).

[2]R. M. Bozorth, *Ferromagnetism*, D. Von Nostrand, New York, 1951.

[3]J. C. Slonezewski, p. 205 in *Magnetism*, Vol. 1, edited by G. T. Rado and H. Suhl, Academic Press, New York, 1963.

[4]J. P. Moore, T. G. Kollie, R. S. Graves, and D.L. McElroy, *J. Appl. Phys.* **42**, 3114 (1971).

Nanoelectronic Devices Made from Doped Nanofibers

J. B. O. Caughman,[1] M. L. Guillorn,[2] D. B. Beach,[3] V. I. Merkulov,[2] L. R. Baylor,[1] and L. F. Allard[4]
[1]*Fusion Energy Division*
[2]*Engineering Science and Technology Division*
[3]*Chemical Sciences Division*
[4]*Metals and Ceramics Division*

The purpose of this project is to understand and develop a process for depositing doped carbon and boron carbonitride nanofibers that will be used to make nanoelectronic devices in a way that will be useful for large-scale manufacturing. Vertically aligned nanofibers are being grown using a high-density plasma-enhanced chemical vapor deposition technique that allows control of gas-phase and surface chemistry. In contrast to current growth methods, the electrical properties of the nanofibers will be controlled by substitutional doping of boron and/or nitrogen to produce metallic or semiconducting doped carbon or boron carbonitride nanofibers. The electrical characteristics of individual vertical nanofibers will be determined by measuring the current flowing through the fiber as a function of applied voltage across the fiber. Nanofiber rectifying devices containing metal-semiconductor (Schottky) junctions will be grown and measured. Our approach for growing nanofibers has advantages over current methods in terms of controlling electrical properties of individual nanofibers, controlling the geometry of the growth, and the potential use as a practical method for realizing vertical integration of dense arrays of nanoelectronics.

One of the main objectives of our project is to be able to control the electrical characteristics of nanofibers by controlling their composition. The other objective is to be able to grow individual, isolated nanofibers in a controlled way. Such an approach has applications for massively parallel vertical nanofiber devices for nanoelectronics. Our approach for growth of the fibers uses a plasma-enhanced chemical vapor deposition technique using an inductively coupled radio frequency plasma source operated at low pressure (<100 mTorr). The gases used during the growth process include hydrogen and a carbon-containing gas, which is either acetylene or methane. Details of the system and initial growth results can be found in a recently published paper (Caughman et al., *J. Appl. Phys.*, Vol. 83, 2003, p. 1207). Our research to date has focused on understanding of the growth of vertically aligned carbon-based nanofibers and how to change their composition. We have conducted experiments to determine the role of various parameters on the growth results, started the electrical characterization of the nanofibers, and are starting to explore changes to their composition with the addition of boron.

Many parameters influence the growth and structure of vertically aligned carbon nanofibers (VACNFs), including the gas composition, the substrate bias, and the growth temperature. As the gas composition changes from being hydrogen rich to a carbon rich (i.e., by increasing the acetylene flow relative to the hydrogen flow), the structure of the fibers changes from being tall and thin cylinders to broad-based cones. The carbon-rich plasma condition creates excess carbon species that condense on the sides of the fibers to create the cone-like structure. The role of substrate bias is related to ion energy, which controls the physical etching that takes place during VACNF growth. We have discovered that a minimum amount of substrate bias is required to grow fibers. If the ion energies are too low, then a layer of carbon builds up on the surface and VACNF growth stops. We have also discovered that growth temperature is important. If the growth temperature gets too high, then the carbon-containing gas (acetylene of methane) will substantially decompose on the substrate surface and inhibit VACNF growth. There is a balance between growth temperature and gas flow. We have added an optical pyrometer to the system to better control the temperature during growth.

The electrical characterization of the VACNFs has started. We have measured the VACNF conductivity by using a four-point probe technique. We have found that the VACNFs are conducting, with a conductivity similar to graphite at 1 mΩ-cm, which is similar to that measured for VACNFs grown with a DC PECVD technique by Mike Simpson's group at ORNL.

The changes to the fibers as a function of composition are also being explored. For growing boron-containing

VACNFs, we have added dilute diborane to the system (4% in hydrogen). Tests with the mass spectrometer have shown that the plasma is very efficient at breaking up the diborane in the gas-phase, which is probably advantageous for the boron to become bonded in the nanofiber structure (less surface temperature needed to break the chemical bonds because the plasma already did it). High diborane flows have been shown to inhibit nanofiber growth, while low flows do not. The change in nanofiber composition as a function of diborane flow is currently being determined.

There should be numerous benefits for DOE and potential for follow-on funding. The area of nanoscience is of great interest to DOE, and the knowledge that we gain from this work will help us to understand the growth mechanisms of nanofibers. Potential applications of interest include nanoelectronics, field-emission devices, hydrogen storage, and super-hydrophobic materials. We are currently pursuing follow-on funding as part of a white-paper to DARPA.

Development of a New High-Temperature Proton-Electron Mixed Conductor for Hydrogen Separation

E. A. Payzant, R. D. Carneim, S. A. Speakman, and T. R. Armstrong
Metals and Ceramics Division

High-temperature ion transport membranes (proton conductors) are presently limited in availability and performance. Current high-temperature proton conductors (HTPCs) either have low conductivities (e.g., $SrZrO_3$-based materials) or are highly susceptible to chemical attack by contaminants such as sulfur (H_2S) and CO_2 (e.g., $BaCeO_3$-based materials). *The purpose of the proposed research is to develop a practical HTPC, where high conductivity is the primary requirement, based on a novel approach investigating proton conductivity near phase boundaries.* Durability is addressed by avoiding the use of polyvalent lanthanides and transition metals, which bind readily with sulfur—especially under the reducing conditions found in reformate or syngas streams. The perovskites $LaYO_3$ and $SrZrO_3$, and the pyrochlore $La_2Zr_2O_7$, meet the stability requirement and have demonstrated proton conductivity. These binary oxides have low proton conductivities; however, by studying pyrochlore–perovskite binary systems, it was hoped a ternary composition will be produced that meets the primary requirement for a practical HTPC—high conductivity. This novel approach relies on there being at least one phase boundary between the two end-point compositions. It is expected that various properties, including protonic conduction, of each end-point phase can be increased since transport mechanisms are often enhanced near phase boundaries. Rapid synthesis and characterization techniques are used to screen a large number of initial compositions, which was followed by conductivity measurements, structural characterization, and modeling.

X-ray diffraction and dc conductivity measurements were used to survey the functionality of two pyrochlore-perovskite binary systems as high-temperature proton conductors (HTPCs). An efficient HTPC could be employed in a number of electrochemical devices, such as gas separation membranes, fuel cells, or direct chemical conversion stacks. The primary driver for HTPC development is the need for hydrogen separation membranes that can provide an unadulterated hydrogen stream from a reformate feed for petrochemical processing. Equally important, though, are the advantages an efficient HTPC could provide to a future generation of fuel cell technology.

A solid ceramic ion transport membrane has the greatest potential for producing pure hydrogen at high temperatures. Nafion® membranes produced by DuPont are currently the most effective low-temperature proton conductors; however, their maximum operating temperature (~160°C for hybrid membranes) is below the desired operating temperature (500 to 600°C) for petrochemical processes.[1] Microporous membranes operate in this temperature range with high fluxes; however, they are at best 99% selective to hydrogen.[2] Ceramic membranes operate well at high temperatures and can be very selective; however, they are intolerant to H_2S and CO_2 contaminants and their flux is too low for practical application.[2]

Ionic conductivity in ceramics can be enhanced by disorder. Typically, disorder is related to structural defects such as oxygen vacancies or to configurational features such as the rotation and tilting of oxygen polyhedra. However, disorder can also be created by fluctuations in crystal structure. Displacive properties, such as piezoelectricity, are enhanced at compositions near a phase boundary, where little separates one phase from another and localized regions can dynamically fluctuate between different crystal structures.[3] The inherent chaos of these structural fluctuations increases the kinetics of the system. It has recently been shown that transport properties can also be enhanced at a phase boundary, providing a new mechanism for developing fast ion conduction in ceramic oxides.[4]

The crystal structure for pyrochlore, $A_2B_2O_7$, has octahedrally coordinated B-site cations, 8 coordinated A-site cations, and two distinct oxygen sites. The crystal structure of perovskite, ABO_3, also has octahedrally coordinated B-site cations, as well as 12 coordinated A-site cations. The pyrochlore-perovskite binary system was chosen for study because the crystal structures share several features, particularly the BO_6 octohedra. Additionally, small, localized regions of pyrochlore-type ordering were recently observed in perovskite materials.[5]

Therefore, it was anticipated that compositions near the phase boundaries in these systems might exhibit enhanced conductivity due to structural fluctuations. The pyrochlore $La_2Zr_2O_7$ and the perovskites $LaYO_{3-}$ and $SrZrO_3$ were selected for investigation because all have demonstrated proton conduction.[6–9] Furthermore, these systems are stable in the presence of both CO_2 and H_2S. Samples were tested across a full compositional range, thereby including doped-pyrochlore regions, doped-perovskite regions, near-phase boundary regions, and pyrochlore-perovskite heterophase mixtures.

The glycine-nitrate method was used to produce precursor powders for the test materials.[10] Nitrate solutions of the appropriate cations were mixed with glycine fuel and then combusted. The resulting precursor was calcined at 900°C for 1 h. The calcined powders were pressed into pellets and sintered at either 1000°C for 96 h, 1500°C for 2 h, or 1500°C for 15 h. X-ray diffraction (XRD) was used to analyze the precursor, calcined, and sintered powders. Peak positions were determined by profile fitting individual diffraction peaks with pseudo-Voigt functions. The peak positions of standard reference materials were used for internal 2θ calibration of the diffraction data.[11] The calibrated peak positions of the sample were analyzed with a cell refinement algorithm to determine the unit cell lattice parameters.[12]

Discs were sintered at 1500°C, had Pt lead wires attached with silver paint, and then were annealed at 900°C. The electrical conductivities were then measured in air at temperatures between 300 and 900°C using a four-point van der Pauw technique.[13] Electronic measurements were made with two Keithley 6517A electrometers and an Agilent 34970A switch unit. One electrometer simulated a current source, applying voltages (V_{appl}) of ±10, 5, 2.5, 1.25, and 0.675 V and measuring the corresponding current (I_{source}) across two electrodes, while the second electrometer measured the resulting potential (V_{obs}) across the two parallel electrodes. Eight measurements made at different V_{appl} were analyzed for linearity and consistency. The linear portion of the $I_{source}V_{obs}$ curve was then used to determine the resistance. These measurements were repeated for four configurations of the electrodes M, N, O, and P: $I_{MN}V_{OP}$; $I_{NO}V_{PM}$; $I_{OP}V_{MN}$; and $I_{PM}V_{NO}$. The measurements of $I_{MN}V_{OP}$ and $I_{OP}V_{MN}$ were averaged together to produce R_A; likewise, $I_{NO}V_{PM}$ and $I_{PM}V_{NO}$ were averaged together to produce R_B. The resistivity (ρ) was then solved from the equation:

$$\exp\left(-\frac{\pi d}{\rho}R_A\right)+\exp\left(-\frac{\pi d}{\rho}R_B\right)=1 \quad , \tag{1}$$

where d is the thickness of the pellet. These measurements were performed in a Thermolyne F21130-33 tube furnace at different temperatures. The sample temperature was measured with a thermocouple that was independent from the one used to operate the furnace. The resulting temperature and conductivity data were used in an Arrhenius analysis to determine the activation energy (E_A) and pre-exponential geometric factor (σ_o) according to the equation:

$$\sigma_{ion}=\frac{\sigma_o}{T}\exp\left(-\frac{E_A}{kT}\right) \quad . \tag{2}$$

Activation energies and geometric factors (E_A and σ_o) at 600°C were determined by linear regression analysis of the observed conductivity data. Figure 1 shows the electrical conductivity for the various compositions in air at 600°C, while Fig. 2 shows the variation of E_A and σ_o with composition.

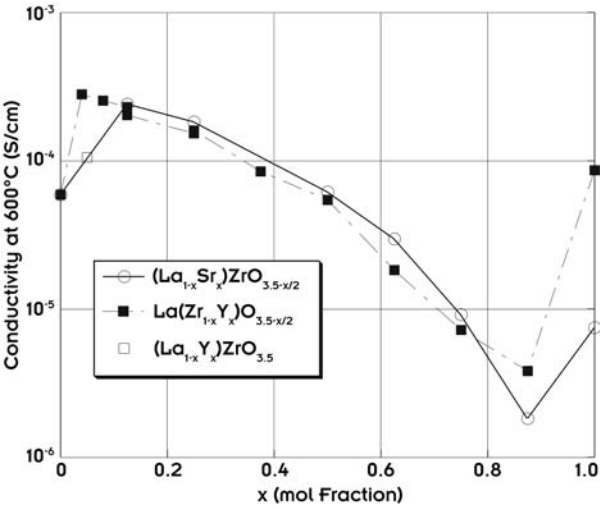


Fig. 1. The electrical conductivity at 600°C in air.

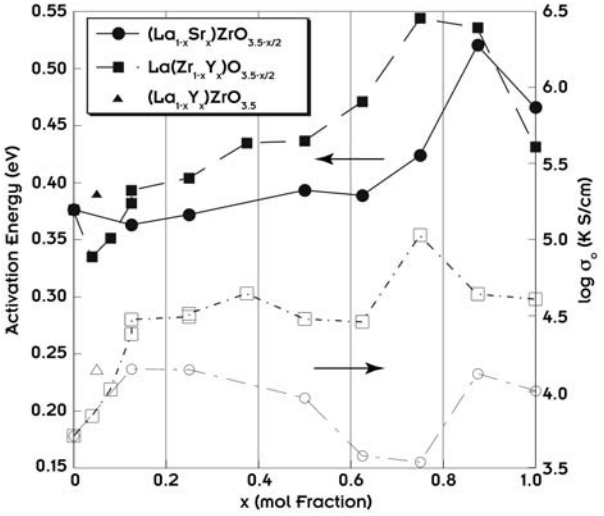


Fig. 2. The activation energies (solid data points) and pre-exponential factors (hollow data points) for electrical conductivity at 600°C in air.

The conductivities of both $(La_{1-x}Sr_x)ZrO_{3.5-x/2}$ and $La(Zr_{1-x}Y_x)O_{3.5-x/2}$ varied similarly with composition. Low levels of doping in $La_2Zr_2O_7$, whether by the substitution of Sr for La or the substitution of Y for Zr, produced a modest increase in the high-temperature conductivity. Likewise, both $(La_{1-x}Sr_x)ZrO_{3.5-x/2}$ and $La(Zr_{1-x}Y_x)O_{3.5-x/2}$ exhibited a depression in conductivity at low levels of doping in the perovskite phase, whether La-doped $SrZrO_3$ or Zr-doped $LaYO_3$. The conductivities of compositions $0.125 < x < 0.875$ varied almost linearly and empirically appeared to be a mixing function of the conductivity at $x = 0.125$ and $x = 0.875$.

Though the conductivities of $(La_{1-x}Sr_x)ZrO_{3.5-x/2}$ and $La(Zr_{1-x}Y_x)O_{3.5-x/2}$ behaved similarly, their dopant solubilities did not. In $(La_{1-x}Sr_x)ZrO_{3.5-x/2}$, a phase pure product was only formed for $x = 0$ and 1. All other compositions produced heterophase mixtures of perovskite and pyrochlore phases. The lattice parameters of $(La_{1-x}Sr_x)ZrO_{3.5-x/2}$ did not vary much with composition for either the pyrochlore or perovskite phases, confirming that little or no doping occurred.

The $La(Zr_{1-x}Y_x)O_{3.5-x/2}$ phase diagram was more complicated than that of $(La_{1-x}Sr_x)ZrO_{3.5-x/2}$, despite the similarities in their conductivities. A pyrochlore phase was observed for $0 \le x \le 0.25$ in $La(Zr_{1-x}Y_x)O_{3.5-x/2}$, though a small amount of La_2O_3 impurity was observed at $x = 0.125$ and 0.19 and an unidentified phase was observed at $x = 0.25$. Examination of lattice parameters, shown in Fig. 3, revealed that Y did not substitute strictly for Zr as anticipated but rather substituted for both Zr and La at higher doping levels. According to Vegard's Law, if the substitution of Y into $La_2Zr_2O_7$ followed a consistent pattern, then the variation of the lattice parameter with composition would be linear. The lattice parameters for $0 \le x \le 0.25$ were not co-linear but rather demonstrated a discontinuity between $x = 0.08$ and 0.125. In Fig. 3, the lattice parameter of $(La_{0.95}Y_{0.05})ZrO_{3.5}$ is also shown; it was significantly smaller than that of $La_2Zr_2O_7$. These data support the hypothesis that Y substituted on both the A and B sites. The substitution of La by the smaller Y cation on the A site produced a smaller unit cell in $(La_{0.95}Y_{0.05})ZrO_{3.5}$, while the substitution of Zr by the larger Y cation on the B site produced a larger unit cell in $La(Zr_{0.96}Y_{0.04})O_{3.48}$. For $0.125 \le x \le 0.25$, Y substituted onto both the A and B sites, producing a mixed effect on the size of the unit cell. This hypothesis was further supported by the observation of La_2O_3 in the XRD patterns of $x = 0.125$ and 0.19. Since the composition was formulated for pure B site substitution, the occupancy of Y on the A site produced excess La_2O_3. The lack of La_2O_3 in the X-ray diffraction (XRD) pattern of $x = 0.25$ deviated from this hypothesis; however, there was a second (presently unidentified) phase observed. It was suggested that this phase was a La-rich phase that formed instead of La_2O_3, though this was not proven. The exact composition at which Y began to substitute onto the A site has not yet been clearly identified. The lattice parameter of $x = 0.08$ lies below a line defined by the lattice parameters at $x = 0$ and $x = 0.04$, suggesting that the lattice parameter at $x = 0.08$ was not as large as it would be if Y substituted only for Zr on the B site and therefore that a small amount of A site substitution may have occurred at $x = 0.08$.

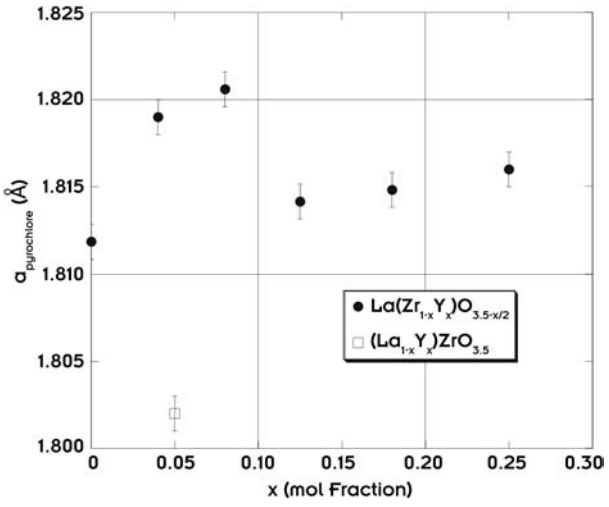


Fig. 3. The variation of the lattice parameter of the pyrochlore phase $La(Zr_{1-x}Y_x)O_{3.5-x/2}$.

As a consequence of the A-site substitution by Y, the electrical conductivity did not increase above its level at $x = 0.04$ and 0.08. The substitution of La with Y did not produce oxygen vacancies and therefore had little effect on the conductivity. The substitution of Y for Zr in $La_2Zr_2O_7$ at 0.04 and 0.08 mole fraction increased σ_o and decreased E_A, thereby increasing the overall conductivity; however, the Y-doping at greater levels of 0.125, 0.19, and 0.25 mole fraction consisted mostly of La substitution and consequently increased E_A, leading to a decrease in conductivity.

The conductivity behavior of $La(Zr_{1-x}Y_x)O_{3.5-x/2}$ at $x \ge 0.875$ paralleled that of $(La_{1-x}Sr_x)ZrO_{3.5-x/2}$, but the phase composition was markedly different, showing a dependence on thermal history. Samples sintered at 1000°C were an orthorhombic polymorph of $LaYO_3$.[14] Pellets sintered at 1500°C had different phases on the surface and within the interior of the pellet. The surface phase was a metastable cubic polymorph of $LaYO_3$,[15] while the interior of the pellet was a presently unindexed polymorph of $LaYO_3$, best described with an orthorhombic unit cell, similar to the orthorhombic polymorph formed at 1000°C but with an a/b ratio closer to 1 (i.e., more nearly tetragonal) and a different symmetry. A definitive indexing was not achieved. In order to test the effect of these polymorphs on conductivity, two pellets of $La(Zr_{0.125}Y_{0.875})O_{3.0625}$ were sintered at 1500°C. One was

ground to remove the surface layer, then electroded; the other was not ground and was electroded. Conductivity data empirically indicated that the surface cubic phase surface layer of $LaYO_3$ acted as a barrier to conduction.

Doping of $La_2Zr_2O_7$, $LaYO_3$, and $SrZrO_3$ was shown to affect the electrical conductivity. Substitutional doping of Sr for La and Y for Zr in $La_2Zr_2O_7$ increased the conductivity, with the greatest gains achieved by 4 and 8 mol % substitution of Y for Zr. Yttrium was more soluble in $La_2Zr_2O_7$ than Sr; however, its effectiveness as a dopant was hindered by its tendency to substitute for both A and B site cations at higher doping levels. Enhancement of conductivity due to phase boundaries has not yet been demonstrated, primarily owing to the lack of miscibility between the systems studied to date, and perhaps also influenced by the occupancy of Y on both A- and B-sites in $La_2Zr_2O_7$, which precludes a single, well-defined phase boundary between pyrochlore and pyrochlore-perovskite. The complex system of polymorphs of $LaYO_3$-based perovskites may also have hindered the dynamics at the boundary between pyrochlore-perovskite and perovskite.

Computer simulations of empirical potential models are currently being used to identify pyrochlore-perovskite binary systems with greater miscibility and therefore a greater chance of producing enhancement of the conductivity near the phase boundaries. Conductivity data is scheduled to be collected on this system in a hydrogen atmosphere in order to separate the proton contribution to the overall conductivity. Neutron powder diffraction data will be collected in early October 2003 for Rietveld refinement of these systems to facilitate more exact identification of site occupancies in these systems and a greater understanding of the relationship between structure and conductivity. Finally, additional compositions with Ce, Sc, and Ga dopants are being synthesized in an effort to improve the overall conductivity.

References

[1]G. Alberti and M. Casciola, "Solid State Protonic Conductors, Present Main Applications and Future Prospects," *Solid State Ionics* ***145***(1–4), 3–16 (2001).

[2]Y. S. Lin, "Microporous and Dense Inorganic Membranes: Current Status and Prospective," *Separation and Purification Technology* **25**(1–3), 39–55 (2001).

[3]R. Newnham, "Phase Transformations in Smart Materials," *Acta Cryst*. **A54**(6.1), 729–37 (1998).

[4]H. Yamamura, H. Nishino, K. Kakinuma, and K. Nomura, "Electrical Conductivity Anomaly around Fluorite-Pyrochlore Phase Boundary," *Solid State Ionics* **158**(3–4), 359–65 (2003).

[5]J. T. S. Irvine, "High Temperature Fuel Cell Materials-Structure at Ion's Length," presented at Materials for Energy Production and Storage Conference, Rutherford Appleton Laboratory, March 6–7, 2003.

[6]J. A. Labrincha, J. R. Frade, and F. M. B. Marques, "Protonic Conduction in $La_2Zr_2O_7$-based Pyrochlore Materials," *Solid State Ionics* **99**(1–2), 33–40 (1997).

[7]E. Ruiz-Trejo and J. A. Kilner, "Oxygen Diffusion and Proton Conduction in $La_{1-x}Sr_xYO_{3-d}$," *Solid State Ionics* **97**(1–4), 529–34 (1997).

[8]J. A. Labrincha, F. M. B. Marques, and J. R. Frade, "Protonic and Oxygen-Ion Conduction in $SrZrO_3$-based Materials," *J. of Materials Science* **30**(11), 2785–92 (1995).

[9]T. Schober, F. Krug, and W. Schilling, "Criteria for the Application of High Temperature Proton Conductors in SOFCs," *Solid State Ionics* **97**(1–4), 369–73 (1997).

[10]L. A. Chick, L. R. Pederson, G. D. Maupin, J. L. Bates, L. E. Thomas, and G. J. Exarhos, "Glycine-Nitrate Combustion Synthesis of Oxide Ceramic Powders," *Materials Letters* **10**(1–2), 6–12 (1990).

[11]R. Jenkins and R. L. Snyder, *Introduction to X-Ray Powder Diffractometry*, pp. 281–5, John Wiley & Sons, Inc., New York, 1996. Materials Data, Inc., *Jade v. 6.5.5*, Livermore, California, 2003.

[12]L. J. van der Pauw, "A Method of Measuring the Resistivity and Hall Coefficient on Lamellae of Arbitrary Shape," *Philips Technical Review* **20**(8), 220–4 (1958).

[13]J. Coutures and M. Foex, "Etude à haute température du diagramme d'equilibre du système formé par le sesquioxyde de lanthane avec le sesquioxyde d'yttrium, " *J. Solid State Chem.* **11**(4), 294–300 (1974).

[14]O. Yamaguchi, H. Kawabata, H. Hashimoto, and K. Shimizu, "New Modification of $LaYO_3$," *J. Am. Ceram. Soc.* **70**(6), C131–2 (1987).

Selective Area Chemical Vapor Deposition of Carbon Nanotube Films Using Seeded Molecular Beams

G. Eres
Condensed Matter Sciences Division

Because carbon nanotubes (CNTs) are not a stable form of carbon, synthetic routes for formation of CNTs must be kinetically controlled. The objective of this project is to explore beams of small carbon-containing molecules for catalytic growth of CNTs. The molecular beam environment allows reaction control at the single collision level, revealing the molecular parameters that influence the catalytically induced conversion of hydrocarbons into CNTs. The key benefits of this approach include reducing the growth temperature of CNTs, controlling the structure (single wall vs multiwall) of CNTs, increasing the yield, and circumventing formation of undesirable side products such as nanoparticles and amorphous carbon.

The main obstacle to widespread practical application of CNTs is the lack of controllable synthesis methods. Currently single-wall CNT (SW-CNT) synthesis is performed either by evaporation and spontaneous condensation of carbon or by high-temperature chemical vapor deposition of hydrocarbons using transition metal catalysts. The extreme reaction conditions in these growth techniques induce a large number of secondary reactions that, in addition to SW-CNTs, produce numerous undesirable side products such as encapsulated nanoparticles and amorphous carbon. These side products interfere with applications, and must be removed before the CNTs can be used.

The molecular beams are generated by mixing small amounts of hydrocarbon source gases (2%) with an inert carrier gas such as He. The beam of carbon molecules is directed at a heated substrate containing a thin film of transition metal catalyst. Upon collision with the substrate, the carbon molecules either stick to the surface or bounce off unchanged and are pumped away. For a fixed catalyst composition, the outcome of the collision depends on the nature of the carbon-containing molecule, the incident velocity of the molecule, the incident angle of the molecule, and the substrate temperature. Each of these reaction parameters can be independently controlled to form the desired reaction products, namely, pure SW-CNTs. The following results have been achieved during the first half of the funding period in this project.

The key accomplishment of this project is the growth of SW-CNTs from molecular beams. There are no previous reports in the literature of SW-CNT growth from molecular beams. The substrate temperature for SW-CNT growth was in the range from 530°C to 680°C. Interestingly, CNT growth ceases above 680°C, suggesting that the mechanism for CNT growth is governed by a surface-limited reaction. Compare these growth temperatures with typical CNT growth temperatures by hot-wall CVD of 900°C or higher.

The molecular structure was found to play an important role in determining the CNT yield from hydrocarbons. No SW-CNT growth was observed from CH_4, and only a small amount of SW-CNTs was found with C_2H_4. The best precursor for SW-CNT growth was C_2H_2. The order CH_4, C_2H_4, C_2H_2 represents a chain of stable intermediates in carbon formation by thermal decomposition of methane. This trend implies that the molecule that is the closest stable intermediate to carbon formation should be chosen to obtain the highest CNT yield.

Several transition metal catalyst were explored in the form of thin films, and chemically synthesized catalyst nanoparticles, deposited on various substrates including Si, sapphire, and quartz. The most effective catalyst was Fe, followed by Co. Surprisingly no CNT growth was observed with Ni. The composition and the thickness of the catalyst films also affected SW-CNT growth. No CNT growth was observed using chemically synthesized nanoparticles of Fe, Co, FeMo, and FePt.

In the second half of this project, work will be focused on the chemical reaction dynamics of the most efficient catalyst-carbon source molecule system. The goal is to use the molecular reaction parameters to tailor the properties of as-grown SW-CNTs in application-relevant configurations.

High-T_c Silicon-Compatible Ferromagnetic Semiconductors

H. H. Weitering, J. Shen, and Z. Zhang
Condensed Matter Sciences Division

Semiconductor "spintronics" holds great promise for novel, faster devices that consume much less power than conventional electronics and that may even facilitate quantum bit or "qubit" operations for quantum computing. The control of the electron spin in semiconductor devices is still at a conceptual stage, awaiting the science and engineering breakthroughs to create new materials and technologies. The objective of this project is to create silicon-compatible semiconductor materials with high ferromagnetic ordering temperatures which could serve as the spin injector or spin detector in silicon-based spintronic devices.

The revolution of spin-based electronics ("spintronics") in materials physics is likely to impact our lives in ways reminiscent of the early days of the transistor and microelectronics industry. Unlike current microelectronic devices, spintronic devices utilize both carrier spin and charge to carry or store information. Spin is a purely quantum phenomenon which lends itself elegantly to the logic of "ones" and "zeroes." One example of a spintronic device is the "spin valve," a layered structure of magnetic and nonmagnetic metal films, which is now widely used in read heads for magnetic hard disk drives. The discovery of high-temperature ferromagnetism in diluted magnetic semiconductors in the late 1990s has defined a radically new avenue in spintronics research.[1,2] If spintronic devices can be made from semiconductors (i.e., if semiconductors could be made magnetic), then in principle we could build spin amplifiers and integrate electronic, opto-electronic, and magneto-electronic functionality on a single device. We may even think of injecting spin-polarized currents into semiconductors and controlling the spin state of the charge carriers, which may allow quantum bit operations for quantum computing.

There are two ways to realize spin injection. One of these is to fabricate a ferromagnetic-metal/semiconductor (MS) heterostructure; the other is to use a dilute magnetic semiconductor (DMS) as the spin aligner. The former method is hampered by chemical intermixing or lattice mismatch at the MS interface, which usually results in significant loss of spin polarization. The large "conductivity mismatch" between the metal and semiconductor furthermore limits the spin injection efficiency for "transparent" MS contacts. Low Curie temperatures (T_c) limit the latter method using a DMS. The objective of this program is to create semiconductors with ferromagnetic ordering temperatures (i.e., Curie temperatures) approaching room temperature so as to create atomically abrupt interfaces between a high-T_c DMS and a nonmagnetic semiconductor for spin-injection purposes. These novel materials are grown using molecular beam epitaxy, an advanced growth technique that allows researchers to create artificially structured materials that cannot be grown under thermodynamic equilibrium conditions. The synthesis efforts are guided by theory efforts, aimed at predicting the proper growth conditions and magnetic properties.

We have succeeded in making germanium semiconductor ferromagnetic by doping it with a few percent of manganese (Mn). The Mn atoms are distributed randomly in the crystal lattice, and the highest Curie temperature that can be achieved by random doping is about 110 K. Although similar results had been reported in 2002 by researchers at the Naval Research Laboratory in Washington, D.C.,[3] we have firmly established the parameters for optimum growth and significantly increased the ferromagnetic ordering temperature by about 40 K through "digital doping" (Fig. 1). In digital doping experiments, Mn atoms are located in two-dimensional sheets. Parallel theoretical studies by Fishman and collaborators, using dynamical mean field theory, have indicated that these materials are so-called "frustrated" magnets and that the ferromagnetic transition temperature can be substantially increased by alleviating magnetic frustration. In a digitally doped heterostructure, magnetic frustration can be removed from each plane of Mn atoms when the Mn magnetic moments are aligned normal to the plane. The consistency between theory and experiment show that it is indeed possible to boost the magnetic ordering temperature in artificially structured materials, as we originally envisioned.

Spintronics is a major research area for the DOE and other federal agencies. Within the first year, this project produced a novel heterostructure for spin injection in a

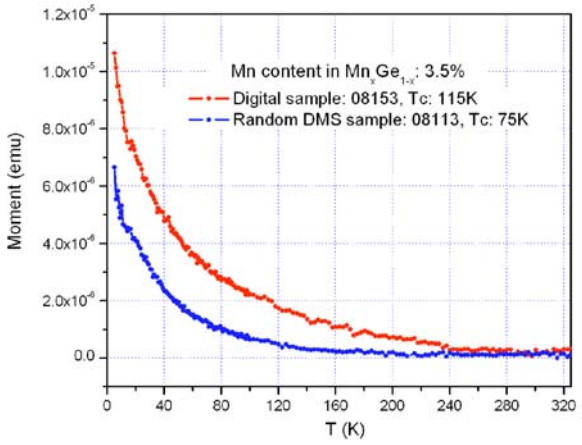


Fig. 1. Magnetization of a "random" and a "digital" manganese-germanium heterostructure as a function of temperature. Digital doping significantly enhances ferromagnetism.

silicon compatible geometry and produced a powerful strategy toward enhancing the ferromagnetic ordering temperature in a dilute magnetic semiconductor. These accomplishments represent major contributions toward solving the bottleneck issues for spintronics research and technology.

References

[1]S. A. Wolf, D. D. Awschalom, R. A. Buhrman, J. M. Daughton, S. von Molnár, M. L. Roukes, A. Y. Chtchelkanova, and D. M. Treger, "Spintronics, A Spin-Based Electronics Vision for the Future," *Science* **294**, 1488 (2001).

[2]T. Dietl, "Ferromagnetic Semiconductors," *Semicond. Sci. Technol*. **17**, 377 (2002)

[3]Y. D. Park, A. T. Hanbicki, S. C. Erwin, C. S. Hellberg, J. M. Sullivan, J. E. Mattson, T. F. Ambrose, A. Wilson, G. Spanos, and B. T. Jonker, "A Group-IV Ferromagnetic Semiconductor: Mn_xGe_{1-x}," *Science* **295**, 651 (2002).

An Innovative Technique for Bimaterial Interface Toughness Research

J. A. Wang,[1] K. C. Liu,[2] I. G. Wright,[2] and L. R. Xu[3]
[1]*Nuclear Science and Technology Division*
[2]*Metals and Ceramics Division*
[3]*Vanderbilt University*

In general, the weakest link in bimaterials occurs at the interface between dissimilar materials, such as the interface between a thin film and its substrate. In order to make multilayered electronic devices or structural composites with a long-term reliability, the fracture behavior of these interfaces must be known. Unfortunately, none of the state-of-the-art testing methods for evaluating interface fracture toughness fully conform to fracture mechanics theory, as is evident in existing data that show severe scatter and procedure dependence in thin-film evaluation methods. This project addresses the problems associated with this deficiency and offers an innovative testing procedure for determination of interface fracture toughness applicable to thin-coating materials in general.

The drive toward increased performance, efficiency, and reduced environmental pollution in heat engines, energy conversion processes, and many chemical processes involves operation of equipment at higher temperatures, often in increasingly corrosive environments. All materials of construction require environmental protection, and in high-temperature environments, the ability of the protective barrier to remain adherent to the surfaces is critical. As a result, measurement of the toughness of the substrate-barrier interface addressed in this project is expected to enhance the ability to improve the performance of many of the components needed in, for example, the DOE's Vision 21 and distributed generation programs, as well as in DOD programs to improve liners for combustors and thermal barrier coatings for gas turbine engines, and to develop better thermal protection systems for space vehicles.

An innovative technology for measuring interface toughness will be demonstrated for oxide scales formed on high-temperature alloys. This new approach includes the development of analytical and experimental procedures and associated bimaterial fracture mechanics theory for evaluating interface fracture toughness. The expected outcomes are (1) improved life prediction capability for high-temperature environmental and thermal barrier coatings, (2) understanding of the mechanism of the interface crack initiation, and (3) determination of unambiguous values for interface fracture toughness. This innovative technique is expected to greatly assist the development of coating materials with improved protective capabilities and provide a reliable method for use in assessing material performance.

In the Phase I of the proposed project, the following progress was made as scheduled:

- Upgraded biaxial test machine with a new load cell capable of high resolution required for torsional testing on thin film materials.
- Selected a Ni-based super alloy (MA956) currently used in high-temperature heat exchangers as the base material on which a thin coating will be formed by oxidation. Two types of notch configurations, a V-shape and a U-shape with different aspect ratios, were used in the new spiral notch torsion test (SNTT) specimens. The specimens are currently being fabricated at a local machine shop.
- To enhance the latitude of this project, a collaboration was discussed with Dr. Matt Ferber of the ceramic structure group at the High Temperature Materials Laboratory (HTML) for using his thermal-barrier-coating (TBC) material in the proposed study. The specimen design appropriate for this material was also completed.
- Developed a fatigue precrack procedure, which utilizes underpeak detector to detect and control fatigue crack growth.
- Developed the finite element models for the SNTT specimens with specified notch configurations and the boundary conditions.

The great technical challenges are expected to be met both in developing experimental procedures and testing apparatus, and theoretical and analytical evaluation of interface toughness for the proposed configurations, which will be carried out in the Phase II research in FY 2004.

The development of this innovative testing procedure will support innovation in the development and use of materials and coatings by providing reliable methods for industry to use in assessing their performance and provide industry with the means to develop and use better (more advanced) materials and coatings through a better understanding of how they behave in practical situations. Knowledge of how the toughness of coating-substrate interfaces is influenced by alloy and coating parameters will greatly aid improved environmental/thermal barrier design.

Nanoporous Inorganic Membranes for High Selectivity Hydrogen Separations

L. K. Mansur and B. L. Bischoff
Metals and Ceramics Division

The United States is committed to a future hydrogen economy. Hydrogen is not generally available as an elemental resource. It can be dissociated from compounds, such as from hydrocarbons or water, by high-temperature chemical reactions. Subsequent to dissociation, effective recovery of hydrogen from gas mixtures becomes a central issue. We are addressing this issue by research on the nanoscience of inorganic gas separation membranes. Pore sizes in commercially available membranes range from about 4 nm to many micrometers. Knudsen diffusion, surface transport, and other reasonably well-understood permeation processes provide limited gas selectivity in such commercial membranes. Recently, we have fabricated experimental inorganic membranes with pore sizes of 1 nm or less. Proof-of-principle experiments are now being carried out on these membranes with the goal of demonstrating exceptionally high selectivity for hydrogen with respect to a number of other gases including a hydrocarbon and carbon dioxide. Guided by theoretical models, we are analyzing these data for evidence of the dominant separation mechanisms in operation and, in particular, for thermally activated atomic-level separation mechanisms.

Approximately 40 nanoporous membranes have been fabricated and characterized over the past 2 years in earlier research. Much of the permeation work to characterize the selectivity of these membranes employed helium and sulfur hexafluoride (SF_6) up to temperatures of 275°C. This base of historical data was examined at the outset of the present research effort.

Since hydrogen separation from hydrocarbon mixtures is the focus of our research, H_2 was added to the membrane evaluation protocol and propane (C_3H_8) was chosen to replace SF_6. A modification to the test system was completed for these flammable gases. The required safety documentation was completed and approved. The system is now fully operational for analysis of the permeation of individual gases, including H_2, He, CO_2, and C_3H_8 up to 275°C. Measurements with SF_6 will also be carried out for comparison with earlier work. In order to help determine the fundamental operating mechanisms of gas separation, it is essential that data be obtained over a wider range of temperatures. Therefore, more recently, a second system has been modified to measure the permeation of individual gases up to approximately 400°C. In addition, this second system gives us the capability to measure the separation of hydrogen from mixtures of gases. During the past month, permeation measurements were initiated. Nanoporous membranes are currently being evaluated with the series of gases mentioned. In several cases the preliminary data show selectivities higher than that for Knudsen diffusion, indicating that the membrane pore size is indeed in the desired nanometer-size regime.

In order to most effectively analyze the experimental data being accumulated, the dominant operating mechanisms of membrane transport need to be known. To this end we have distilled parametric dependencies of key mechanisms of gas transport into simple and readily applicable expressions. Key parameters include temperature T, pressure P, molecular mass m, kinetic molecular diameter d_m, pore diameter d_p, molecular collision mean free path λ, and several thermal activation energies. The latter are H_a, the heat of gas adsorption, E_s, the activation energy for surface diffusion, and E_d, the activation energy for nanopore diffusion.

The most important characteristic of membranes that dictates the dominant transport mechanism is the pore diameter or, more precisely, the ratio of the pore diameter to two important physical characteristics of the gas, λ and d_m. Most gases of interest have kinetic diameters between one-fourth and two-thirds of a nanometer. Generally, for pore diameters >2 nm, the important permeation mechanisms in operation are Knudsen diffusion and surface transport, together with molecular diffusion and viscous flow (Poiseuille or laminar flow) at larger pore diameters. For pore diameters of about 1 nm and smaller, other mechanisms, which are covered by the term "nanopore diffusion," come into operation. Several mechanisms of gas transport through a membrane are summarized in Table 1. Expressions for selectivity and permeance also are given in the table for those cases where the dependencies on important variables can be expressed in simple form without extensive qualifying statements or limiting conditions.

Table 1. Characteristic expressions for key membrane gas-transport mechanisms

Mechanism	Pore Diameter	Selectivity	Permeance
Viscous Flow	$\lambda < d_p$	None	$d_p^2 P T^{-1}$
Molecular Diffusion	$\lambda < d_p$	None	
Knudsen diffusion	$\lambda > d_p$	$m^{-1/2}$	$m^{-1/2} d_p T^{-1/2}$
Surface Transport	All d_p	Variable	$d_p^{-1} P \exp[(H_a - E_s)/RT]$
Capillary Condensation	$f(P)$	Variable	
Nanopore Diffusion	$3 d_m > d_p$	Highest	$m^{-1/2} d_p T^{-1/2} \exp[-E_d/RT]$

This research project is expected to help DOE to accomplish its goals set forth in the "National Hydrogen Energy Roadmap" released by Secretary of Energy Spencer Abraham on November 12, 2002. The work can benefit programs within the Office of Fossil Energy, including the separation of hydrogen during syngas production and the Oil Processing Program, where hydrogen is used to upgrade low-grade heavy petroleum in a refinery. Government agencies, such as NASA, that heavily utilize hydrogen as an energy source can benefit from the capability to separate hydrogen more efficiently.

High-Aspect-Ratio Carbon Nanofiber Probes for Scanning Probe Microscopy

M. L. Simpson,[1] M. A. Guillorn,[1] C. Rouleau,[2] D. Hensley,[1,2]
R. J. Kasica,[1] M. J. Doktycz,[3] and A. V. Melechko[1,4]
[1]*Engineering Science and Technology Division*
[2]*Condensed Matter Sciences Division*
[3]*Life Sciences Division*
[4]*University of Tennessee*

The objective of this project is to develop the first technique for the synthesis of high-resolution, high-aspect-ratio scanning-probe-microscope (SPM) tips that can be implemented in a large-scale fabrication process. This technique will be based on the ORNL-developed vertically aligned carbon nanofibers (VACNF) technology. While other methods currently exist for the production of this type of probe tip, the use of the VACNF will surpass these techniques since VACNFs can be synthesized completely deterministically in a wafer-scale fabrication process. This allows for thousands of identical tips to be fabricated simultaneously in precise locations with total control over their morphology.

Scanning probe microscopy (SPM) has become an indispensable tool for the analysis of surfaces and the exploration of general substrate morphology at the nanoscale and microscale. Moreover, SPM techniques have been used to investigate a wide variety of material properties beyond substrate topography. The range of applications spans a vast array of scientific disciplines ranging from solid state physics to biology, environmental science to polymer engineering. The currently used methods for producing high-resolution, high-aspect-ratio tips for SPM applications are based on processes serial in nature; that is, only a single tip can be fabricated at a time. The goal of this project is to integrate synthesis of vertically aligned carbon nanofibers (VACNFs) with standard cantilever fabrication processes and to characterize the high-aspect-ratio VACNF SPM tips obtained in such method.

We have demonstrated that VACNFs can be synthesized in a completely deterministic manner using a catalytic dc, plasma-enhanced, chemical vapor deposition process. Features such as the location, length, tip diameter, shape, and chemical composition of VACNFs can be precisely controlled during the synthesis process. Functional microfabricated device structures have been fabricated that exploit both the electrical and mechanical properties of single and multiple VACNF. These structures have been fabricated on a variety of substrates including whole 10-cm-diameter Si, Si-on-insulator (SOI) and quartz wafers. This material meets the required specifications for high-aspect-ratio SPM tips. In addition, it offers a number of other features that may allow it to surpass all other materials presently used in SPM tip manufacturing.

The fabrication process for microcantilevers with VACNF tips have been developed during FY 2003. Individual processing steps compatible with handling microsctructures with high-aspect-ratio features such as VACNFs have been developed and tested. The first batch of cantilevers with VACNF tips has been obtained, thus proving the feasibility of the idea. Figure 1 shows an SEM image of a finished device. The sharpness of the VACNF tips has been improved via changes in the catalyst preparation process. The diameter of the nanofiber coincides with the diameter of the catalyst particle. We have found the techniques to prepare smaller nanoparticles that can be prepared using standard photolithography, thus making the complete fabrication an inexpensive and high-throughput method.

The devices obtained after optimization of all processing steps will be characterized, and their performance will be compared to currently available atomic force microscope tips in application areas that are critical to DOE and NIH. Micro- and nano-fabricated structures and complex biological systems, two major research thrust areas at ORNL, possess a varied surface topography that is often best characterized by SPM techniques. However, they can only be imaged accurately with high-resolution high-aspect-ratio tips. We will explore the application of the VACNF probe tips in these two target areas as they show significant promise to benefit programs at ORNL.

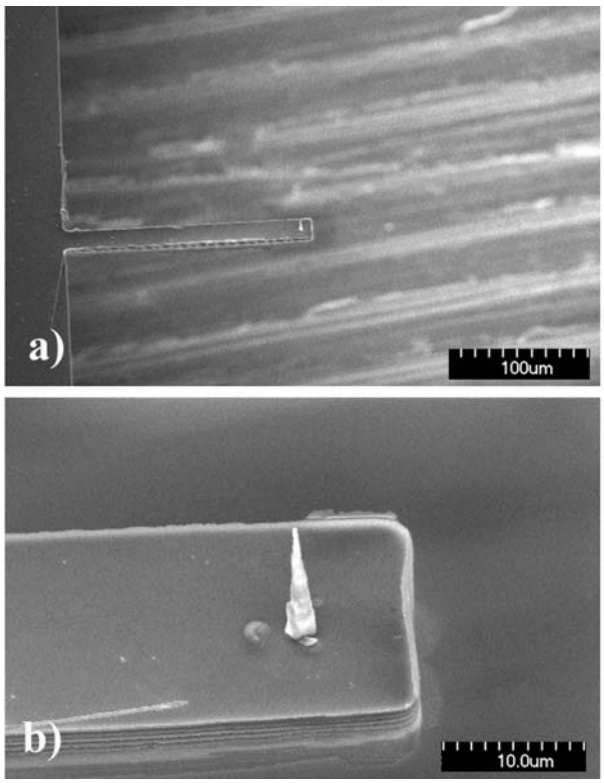


Fig. 1. Scanning electron micrographs of a cantilever with a VACNF tips at 45° viewing angle at two different magnifications: (a) 300× and (b) 3000×.

In Situ Studies of Hydrogen Storage Materials Using Neutron Scattering

C. J. Rawn,[1,2] J. Y. Howe,[1] B. C. Chakoumakos,[3] J. L. Robertson,[3] and F. C. Montgomery[1]
[1]*Metals and Ceramics Division, Oak Ridge National Laboratory*
[2]*Materials Science and Engineering Department, University of Tennessee*
[3]*Condensed Matter Sciences Division, Oak Ridge National Laboratory*

This project concentrates on using in situ neutron diffraction studies to answer structural and thermodynamic issues about the role of hydrogen in hydrogen containing compounds. The first tasks underway are collaborating with personnel from the neutron scattering facility at Chalk River to collect neutron diffraction data at ambient conditions on candidate hydrogen storage materials and to design and construct a new pressure cell for use at different temperature and pressure ranges of interest for the investigation of suitable hydrogen storage materials. Future research includes proof-of-principle experiments using the new pressure cell to demonstrate its utility for implementation of neutron scattering experiments on hydrogen-containing compounds.

Progress Report

To obtain a pressure cell for use at different temperature/pressure ranges of interest for the investigation of suitable hydrogen storage materials. To verify the efficacy of the cell by studying $NaAlD_4$ and demonstrate that we can generate details surrounding the role of H atoms in the atomic structure.

To collect neutron powder diffraction data on the candidate hydrogen storage materials $NaAlD_4$ doped with the addition of $TiCl_3$ as a catalyst. With high-quality neutron powder diffraction data we can use the Rietveld method to determine the positions of Ti in the crystal structure. Additional in situ experiments have been proposed to observe hydrogen absorption and desorption under service conditions of hydrogen storage materials. From these experiments we hope to understand the path that the hydrogen atoms take during absorption and desorption and to evaluate a high-temperature/high-pressure cell, designed by the group at Chalk River, for possible procurement and use at the High Flux Isotope Reactor.

Our two-phase proposal to obtain beamtime on the powder diffractometer at the Chalk River neutron scattering facility has been review and approved, and we are hoping to start data collection in early March. The first experiments we wish to conduct are at ambient temperature and ambient pressure on $NaAlD_4$ with and without the addition of $TiCl_3$ as a catalyst. There are several explanations in the literature for where the Ti must be located within in the crystal structure; however, to date this has not been shown experimentally using neutron scattering techniques. Our intent is to collect a high-quality neutron powder diffraction pattern and examine the data using the Rietveld method to determine the sites occupied by Ti in the $NaAlD_4$ unit cell. A third data set is to be collected on $NaAlD_4$ + $TiCl_3$ desorbed to Na_3AlD_6 + products. We are collaborating with Joachim Schneibel (Metals and Ceramics Division) to synthesize $NaAlD_4$ with $TiCl_3$ from as-received $NaAlH_4$ by desorbing the material and recharging it with D_2 gas.

One of the main goals of this proposal is obtain a high-pressure/high-temperature cell to study hydrogen storage materials with in situ neutron powder diffraction techniques. To this end we are collaborating with Lachlan M.D. Cranswich, a Research Council Officer from the Neutron Program for Materials Research at Chalk River Laboratories (Chalk River, Ontario, Canada). In July we learned from Mr. Cranswick that the Chalk River group was in the first stages of designing such a cell, and in October Mr. Cranswich visited ORNL and discussed the early stages of a high-pressure/high-temperature cell design. Figure 1 shows the mass flow control layout designed to minimize the amount of tubing. The design of this pressure cell is such that it will work in conjunction with a commercial Cryofurnace, which we currently have available at the High Flux Isotope Reactor (HFIR), for changing the temperature. For the current cell design, the highest obtainable pressure would be 10 bar, and we have proposed modifications to the design to obtain pressures of 50 bar. For this increased pressure range, several factors such as the choice of container material and the addition of an external gas booster are being considered. We have received two of mock-up sample holders that are the sizes and geometry used on the powder diffraction beamline at the Chalk River facilities. These cells contain porous metal filters that will allow the gas into the sample holders

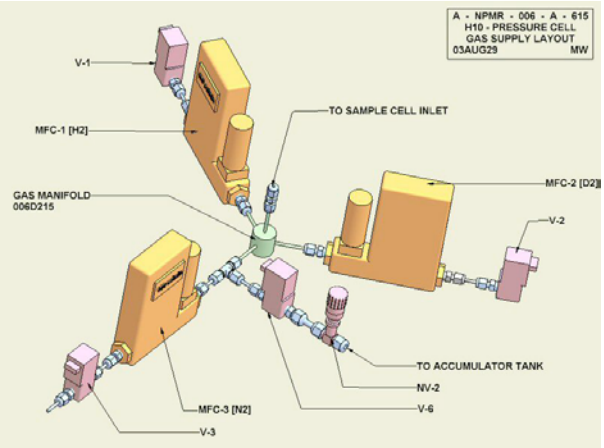


Fig. 1. The Chalk River design for the mass flow controllers for the 10 bar hydrogen pressure cell. The design attempts to minimize the amount of tubing used.

without blowing the powder samples out. These cells are important to our understanding of how much powder will be needed per experiment.

This research will result in expertise in the characterization of hydrogen-storage material systems at the atomic level and position ORNL to compete for future funding from the Hydrogen, Fuel Cells, and Infrastructure Technology Program and other areas where hydrogen fuel is the focus.

Alanates for High-Capacity Hydrogen Storage

J. H. Schneibel,[1] D. S. Easton,[2] S. Babu,[1] and S. A. Speakman[1]
[1]*Metals and Ceramics Division*
[2]*Consultant, Metals and Ceramics Division*

We propose to investigate the hydrogen absorption/desorption properties of alanate compounds [$NaAlH_4$, $LiAlH_4$, AlH_3, $Mg(AlH_4)_2$] capable of storing up to approximately 10 wt % hydrogen. The synthesis of these materials will be accompanied and guided by thermodynamic modeling and crystal structure analysis. Alloying will be carried out by high-energy milling. Elements or compounds will be added to improve the absorption/desorption kinetics and to shift the absorption/desorption conditions closer to ambient conditions.

To reduce the country's dependence on foreign oil supplies and to promote a cleaner environment, the President has proposed the FreedomCar (Freedom Cooperative Automotive Research) initiative. Its long-term goal is to develop technologies for hydrogen-powered fuel cell cars and trucks. This project focuses on a particularly important issue of the initiative, namely, hydrogen storage. The objective is to develop materials which can absorb and desorb large amounts of hydrogen (e.g., 10 wt %) at near-ambient conditions.

Alanate materials (i.e., materials containing AlH4$^-$ ions, such as for example $LiAlH_4$) will be prepared by high-energy ball milling of powder precursor materials. Small quantities of elements or compounds with catalytic activity will be added to improve the absorption/desorption kinetics ("micro-alloying"). Macro-alloying (i.e., ≈1 wt % and above) with the aim of shifting the absorption/desorption conditions closer to ambient will be guided by thermodynamic calculations as well as electronegativity and atomic size considerations.

During FY 2003, this project focused on experiments designed to assure that we are able to reproduce existing results for high-performance hydrogen-storage materials and that our equipment for measuring hydrogen storage capacity values is performing properly. Considerable effort was spent on determining the precision and reproducibility of a recently acquired piece of equipment for characterizing hydrogen-storage materials. This equipment measures the hydrogen pressure as a function of the hydrogen concentration in the material, for absorption and desorption at preselected temperatures (the so-called pressure-composition-isotherm, PCI). The equipment also measures, as a function of temperature and pressure, the rates with which hydrogen is absorbed and desorbed. Following published information, commercially available sodium alanate, $NaAlH_4$, was high-energy milled with 2 wt % of $TiCl_3$. Its hydrogen absorption and desorption behavior was measured as a function of pressure. The curves obtained demonstrated the two-stage decomposition reaction:

$$3NaAlH_4 \rightarrow Na3AlH6 + 2Al + 3H_2 \rightarrow 3NaH + 3Al + 5/2\ H_2 \quad (1)$$

According to a desorption PCI curve measured by us, the hydrogen storage capacity was 4.95 wt %. The hydrogen storage capacity was independently measured by weighing the fully charged as well as the material discharged according to the right-hand side of Eq. (1). The storage capacity measured in this way was 4.66 wt %, which is in reasonable agreement with the PCI value of 4.95 wt %. These values are not as high as the theoretical capacity, 5.6 wt %, but are in agreement with typical values found in the open literature. We have now reached the point at which we will be able to perform experiments with novel compositions.

If improved hydrogen storage capacities can be obtained, and if absorption and desorption occur at close-to-ambient conditions, this project will be very valuable for the Energy Efficiency and Renewable Energy (EERE) program of DOE.

Other federal agencies involved in supporting cutting edge research in carbon nanotube synthesis and electronic application such as DARPA and NASA will directly benefit from increased control over the carbon nanotube film growth process. A specific example is the application of carbon nanotubes as field emitters. The ability to control the growth rate and the growth temperature is also relevant to mass production of carbon nanotubes. Agencies that are involved in supporting carbon nanotube composite research such as NASA would directly benefit from a high-rate carbon nanotube synthesis process.

Enhancing Performance of Hydrogen Storage Materials through Nanoscale Design

V. K. Sikka,[1] G. Muralidharan,[1] T. J. Huxford,[1] T. M. Besmann,[1] N. C. Gallego,[1] and M. P. Paranthaman[2]
[1]*Metals and Ceramics Division*
[2]*Chemical Sciences Division*

Insufficient gravimetric capacities and slow hydriding/dehydriding kinetics have been recognized as hurdles to be overcome in the quest for future hydrogen storage materials. The objective of this project is to examine a novel approach for enhancing the hydriding/dehydriding kinetics based on the use of nanometer-thick films. The feasibility of this approach will be studied by depositing films of magnesium on the interior surfaces of a microporous silica gel with a large surface area (up to 1000 m^2/g) and examining the hydriding/dehydriding characteristics of the resulting material. The use of silica gel as the template material is a key to obtaining improved kinetics without sacrificing storage capacity. Improved kinetics and lower operating temperatures are benefits anticipated as a result of using this approach.

Hydrogen-storage materials are of great interest for the new national and international interest in the hydrogen economy. Based on the potential use of hydrogen storage materials in automotive applications, a storage criteria of 9 wt % of hydrogen has been set as a national goal and there are many efforts under way to develop such materials. In addition to meeting the gravimetric capacity, such materials should also be capable of adsorbing and desorbing the hydrogen at temperatures less than 100°C at pressures of 1–10 atm. A number of metal hydrides have been studied for their potential use as hydrogen-storage materials. Such studies have shown that slow kinetics of the hydriding/dehydriding reaction, high temperatures required to absorb/release the hydrogen, or insufficient storage capacities limit the use of metal hydrides.

A number of approaches, such as physical and chemical treatments, have been carried out to improve the hydriding and dehydriding kinetics of metal hydrides with limited success. Our approach to addressing this problem is based on developing a general scheme for improving the hydriding/dehydriding kinetics of all hydrogen storage materials of interest. We plan to use magnesium as a model material for developing and demonstrating our concept. In addition to improving the hydriding/dehydriding kinetics using this methodology, our objective is to also achieve reduced operating temperatures in the range of 100–150°C in such films.

The core theme of our approach is the use of thin films of hydrogen-storage materials with a thickness of less than 50 nm. A significant benefit in using such films is that diffusion is required only over very short distances when the surfaces of the thin films are exposed to the gas. In addition, to achieve sufficient gravimetric storage capacity, the internal surfaces of a foam-like material with a large surface area and an interconnected pore structure that is permeable to the surrounding gas medium have to be coated with the thin film that can store hydrogen. An example of this kind of a material is silica gel. These gels are commonly used for humidity control, and catalyst support. The combined use of the silica gel as a template and a thin-film hydrogen-storage material is unique and novel.

Significant progress has already been made over the first three months of the project. Several candidate silica gel materials have been identified and are currently being evaluated for use in this work. The pore surface area has been evaluated for one particular batch of materials using Brunauer-Emmett-Teller gas adsorption measurements. The results show a typical total accessible pore surface area of about 430 m^2/g. Experiments are under way to evaluate deposition techniques that would be suitable for coating the internal surfaces of these gels with Mg/Mg-alloy composite thin films. Initial experiments have explored the use of physical vapor deposition methods to coat the silica gels with magnesium. These techniques exploit the fact that the vapor pressure of magnesium is high at relatively low temperatures due to its low melting point. Silica gels were exposed to an atmosphere consisting of a high partial vapor pressure of magnesium through heat treatment in a furnace either in a sealed quartz tube or in a flowing Ar/H_2 atmosphere. The resulting material was crushed and the phases present were analyzed through X-ray diffraction. MgO peaks were observed in the diffraction pattern, indicating that magnesium was indeed deposited but was subsequently oxidized during exposure to air. Future experiments will focus on protecting the magnesium films through handling in an inert atmosphere and subsequently depositing palladium to prevent

oxidation. An alternate and novel technique using rapid sequential heating of magnesium and palladium with a high-power IR lamp in a flowing Ar/H_2 atmosphere has also been attempted to minimize the problems due to oxidation of magnesium. Further characterization of this material is in progress.

Chemical vapor infiltration will also be pursued as an alternative technique for depositing composite thin films. Experiments to evaluate the hydriding/dehydriding properties of the material produced have also been initiated.

Hydrogen storage and delivery materials are of significant interest to DOE and other federal agencies such as DARPA. If this scheme is successful, engineering hydrogen-storage materials based on magnesium with gravimetric capacities up to 9.3 wt % may be feasible, thus satisfying both the existing and future DOE criteria. There is also the additional advantage that, if necessary, this methodology can be applied to other hydrogen-storage materials.

COMPUTER AND COMPUTATIONAL SCIENCES

Director's R&D Fund

Synthesis of High-Performance Algorithms for Electronic and Nuclear Structure Calculations

D. E. Bernholdt,[1] R. J. Harrison,[1] J. B. White III,[1] D. J. Dean,[2] M. R. Strayer,[2] G. Baumgartner,[3] S. Hirata,[4] M. Nooijen,[5] R. M. Pitzer,[6] J. Ramanujam,[7] and P. Sadayappan[3]

[1]*Computer Science and Mathematics Division*
[2]*Physics Division*
[3]*Department. of Computer and Information Sciences, Ohio State University*
[4]*Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory*
[5]*Department of Chemistry, University of Waterloo*
[6]*Department of Chemistry, Ohio State University*
[7]*Department of Computer Science, Louisiana State University*

In order to increase both the productivity of scientific programmers, and the performance of the software they create, we are developing tools and techniques to translate a high-level expression of a computational problem into highly optimized source code, which can then be compiled and run as usual. This approach allows significant optimization of the algebraic expressions, trade-offs between storage space and computation time, and distribution of the data across the parallel computer during code generation. We are applying these techniques to produce a *Tensor Contraction Engine* targeted at a class of problems in computational quantum chemistry and computational nuclear physics. To be able to validate generated code for the nuclear coupled cluster problem, we have also done a hand implementation of such a code for comparison purposes.

Technical Approach

The development of high-performance parallel programs for scientific applications is usually very complicated and is exacerbated by the increasing complexity of modern computer systems. High-level or domain specific languages (HLLs) are one approach to addressing some of these issues. They allow scientists to express their computational problems in a form that is close to how they are derived and might be expressed in scientific publications. They can also be used as inputs for tools that process the high-level problem expression to produce actual computer code. This clearly addresses the "human scalability" problem in modern scientific computing; it also provides an excellent opportunity to address performance issues.

In traditional software development, the implementer makes many decisions, often quite empirically, affecting performance and scalability as they manually translate from the high-level expression of the problem into code. The use of tools to generate the actual code also provides the opportunity to thoroughly and automatically analyze the various possible implementations, taking into account details of the target hardware platform, and even the specific problem. The result is highly optimized and tuned algorithms to solve the problem originally specified in the high-level language.

In this project, we have developed tools and techniques to apply this approach to a class of computations that are both CPU and memory intensive, spanning several scientific domains. Our primary focus has been a class of methods used in quantum chemistry and nuclear physics known as coupled cluster methods. These methods are among the most accurate in routine use in the quantum chemistry,[1,2] while in the area of nuclear physics they are not widely used. They are formulated as contractions of large tensors (the multidimensional analog of matrix multiplication), with a given method involving tens to hundreds of different tensor contraction terms, giving rise to the extreme complexity of both implementation and performance optimization of such codes.

Our approach is to treat the problem from a computer science viewpoint, creating a specialized optimizing compiler, which we call the *Tensor Contraction Engine* (TCE),[3] which takes as input a high-level domain-specific language for describing the tensor contraction equations of coupled cluster theory, and generates code in a traditional programming language (in this case Fortran) which can then be integrated with the rest of the chemistry computational infrastructure for execution. The TCE must make the same design decisions a human would make for a hand implementation, but it has the advantage of being able to more rigorously and exhaustively explore the

options and take into consideration details of the target computer architecture and performance model, as well as details of the particular target problem, if appropriate. Modules of the TCE, depicted schematically in Fig. 1, include the following:

- Algebraic transformations, which transform the input tensor contractions into a mathematically equivalent form that minimizes the number of operations required to evaluate them.
- Memory minimization, which reduces the amount of memory required to store intermediate results during evaluation of the input.
- Space-time transformation, which makes trade-offs between recomputing quantities and storing them on disk.[4]
- Data locality management, which optimizes utilization of storage hierarchies, such as cache/memory or memory/disk.[5,6]
- Data distribution and partitioning determines the best layout of the data for efficient parallel processing.[7,8]

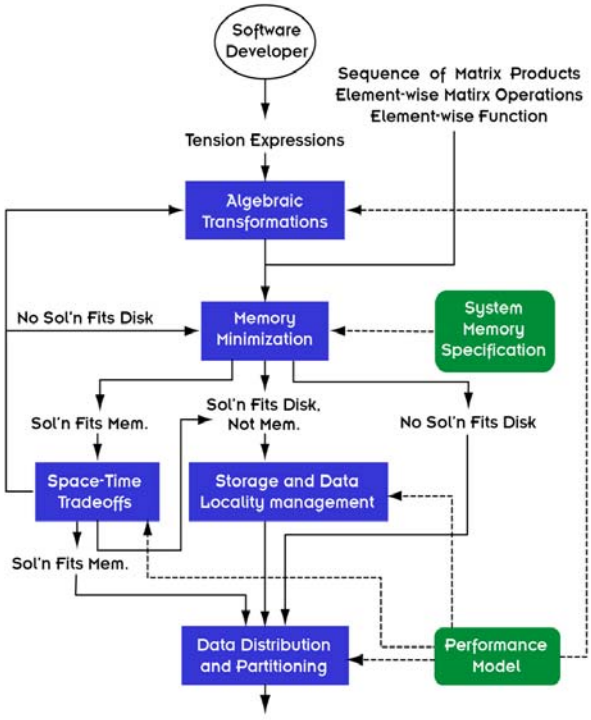


Fig. 1. The general architecture of the Tensor Contraction Engine (TCE). The user expresses their problem in a specially designed high-level language, and it is processed by the TCE to produce Fortran (in this case) source code implementing the input expression. TCE optimizations include algebraic transformations to minimize the overall operation count, memory minimization to reduce storage for intermediates, space-time trade-offs between storage and recomputation of intermediate quantities, data locality management, and parallel data partitioning.

It is worthwhile to note that most of these modules can make use of performance model information for the target computer system to determine the most effective optimizations.

Two versions of the TCE have been implemented. The "prototype" TCE is used primarily for fast extension of the automatic code generation capability and to examine issues around supporting various new types of coupled cluster calculations of interest to chemists. It implements few of the optimizations mentioned above, in the interest of simplicity and extensibility. The "optimizing" TCE is the "production" tool and includes all of the optimizations described above. The framework of the optimizing TCE is designed to facilitate generalization to other scientific domains, where similar needs exist, but the details of methodology and applicable optimizations will differ.

One of the special challenges regarding the use of automatically generated software for a complex problem is verifying the correctness of the code. In the quantum chemistry area, there are a number of existing handwritten implementations of various coupled cluster methods, so validation of the code generation process is relatively straightforward. In nuclear physics, however, coupled cluster methods are not in routine use, so that in order to be able to verify the correctness of the generated code, it was necessary to make a careful hand implementation to compare against. This capability has opened up some additional opportunities for research in the nuclear physics area, which are described below.

Another challenge for automatically generated software is its integration into existing software environments. In its current form, the TCE generates only the code to actually perform the tensor contraction operations, which must then be integrated into an existing chemistry package to provide the convergence checking and iterative control for the method itself, as well as providing various inputs required by the coupled cluster method. While it is possible to extend the high-level domain-specific language to include the necessary flow control and other capabilities required of a complete programming language, this approach leaves the developers of the tool with a task of maintaining an entire programming language rather than a much smaller, specialized language. An alternative solution involves embedding the HLL inside a traditional programming or scripting language, taking advantage of the existing infrastructure for the host language while providing the extra domain-specific capabilities of interest to the user. In anticipation of the need to provide a more complete language environment for the TCE, we undertook a small case study of the embedding of an HLL for expressing multigrid problems[9] in the functional programming language Haskell.[10] The results of the study are described below.

Results

The TCE, though still under very active development by our outside collaborators, has already proven very capable and effective at raising the productivity of software developers. The prototype TCE has been used to produce implementations of more than 20 different coupled cluster-based methods, including energies, lambda equations, dipole moments, MBPT methods through fourth order, CC and CI methods through CCSDTQ, EOM-CCSD, local/AO-based CCSD, and relativistic methods. These capabilities have been included in the latest release of the NWChem computational chemistry package.[11] Figure 2 demonstrates that reasonably efficient parallel code for coupled cluster methods can be generated automatically by the prototype TCE, in this case achieving a speedup of 9 (compared with ideal speedup of 16) between 16 and 256 processors.

The development of the optimizing TCE has been extremely challenging, both from the computer science standpoint and due to the complexity of the domain-specific (chemistry) details, but we have produced a very capable system in a flexible and extensible framework, which we plan to use as the basis for work on HLLs in other domains. From the computer science viewpoint, the optimization algorithms have virtually all been developed for this project, because they go beyond traditional compiler optimizations, resulting in a significant number of papers. We are now working on validating correctness and performance of the optimizing TCE on the simpler CCD and CCSD methods before broadening out. However Fig. 3 shows promise of the approach. This graph shows the effect of just the space-time trade-off optimization for one particular term from a Laplace-factorized CCSD(T) method:

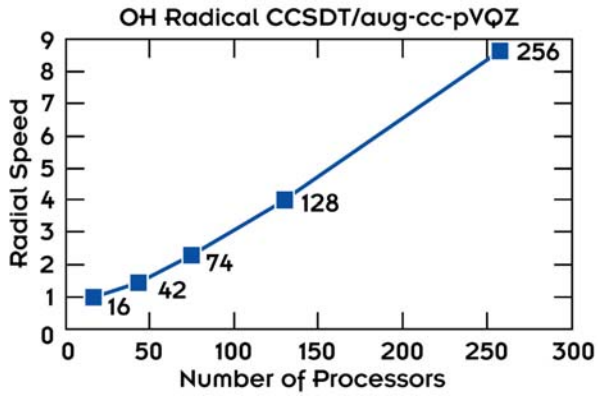


Fig. 2. Parallel scalability of code automatically generated by the prototype Tensor Contraction Engine for a CCSDT calculation on the OH radical using the aug-cc-pVQZ basis. Ideal scaling in this case would be 16 at 256 processors. These results were obtained on the HP Supercluster at PNNL's Molecular Science Computing Facility by So Hirata.

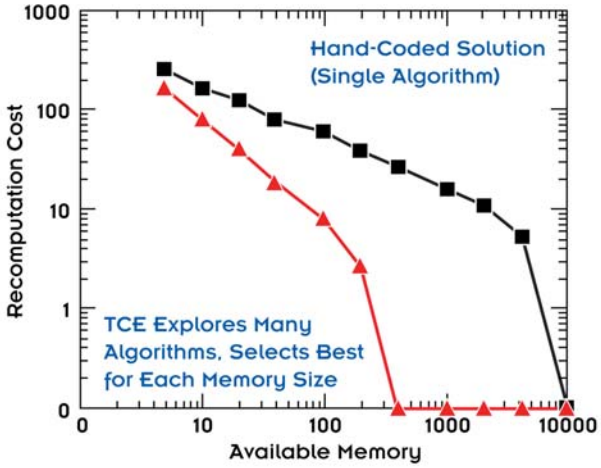


Fig. 3. Performance of the space-time trade-off optimization module compared to a typical hand-coded implementation of this term from the Laplace factorized triples formulation of the CCSD(T) method. The upper curve shows the amount of recomputation required as a function of available memory for a typical hand-coded implementation of the term. The lower curve shows results of the space-time optimization.

$$A3A = \tfrac{1}{2}(X_{ce,af}Y_{ae,cf} + X_{c\bar{e},a\bar{f}}Y_{a\bar{e},c\bar{f}} + X_{c\bar{e},\bar{a}f}Y_{\bar{a}e,cf} + X_{\bar{c}e,a\bar{f}}Y_{ae,\bar{c}f} + X_{\bar{c}e,\bar{a}f}Y_{\bar{a}e,\bar{c}f} + X_{\bar{c}e,a\bar{f}}Y_{\bar{a}e,c\bar{f}} + X_{\bar{c}\bar{e},a\bar{f}}Y_{\bar{a}e,\bar{c}\bar{f}} + X_{\bar{c}\bar{e},\bar{a}f}Y_{\bar{a}\bar{e},\bar{c}f} + X_{\bar{c}\bar{e},\bar{a}\bar{f}}Y_{\bar{a}\bar{e},\bar{c}\bar{f}})$$

$$X_{ce,af} = t_{ij}^{ce}t_{ij}^{af} \qquad Y_{ae,cf} = \langle ab\|ek\rangle\langle cb\|fk\rangle$$

The upper curve depicts the amount of recomputation required as a function of the system's memory for a typical implementation of this term. The lower curve is the result from the TCE optimization. Because it considers many different "implementations" of the term, the space-time optimizer is able to select variants for each memory size that minimize the recomputation cost, whereas in the hand coded case, only a single algorithm has been selected and implemented based on a more intuitive consideration of the problem.

We have given numerous talks on the TCE in both the computer science and chemistry community, which have been received with great interest in both communities. Many quantum chemists, in particular, have expressed interest in using and contributing to the development of the TCE. We will be holding a workshop in conjunction with the 2004 Sanibel Symposium in February, at which we will make the first formal release of the TCE software to the chemistry community, which we hope will begin a long community-wide collaboration to improve and extend this technology.

Work on nuclear coupled cluster theory has been very productive as well. We carried out calculations on the nuclei ^{4}He and ^{16}O, which are magic in both the neutron and proton shells. We used an oscillator basis with up to

360 single-particle orbits, and the Idaho-A nucleon-nucleon interaction, which is based on chiral perturbation theory. This interaction preserves the underlying symmetries of the QCD Lagrangian and precisely reproduces the nucleon-nucleon experimental phase-shift data. We renormalized this interaction for our model space and then performed the CCSD calculations within the same space.[12] We also performed during this same time calculations of the non-iterative triples corrections for these two nuclei, and for ^{4}He we investigated the EOM-CCSD method (an approach developed in quantum chemistry), which allows us to calculate excited states in nuclei.[13] By comparing with exact diagonalization results in a small model space (consisting of 80 single-particle orbits) with the triples-corrected CCSD results. We found that CCSD and the triples corrected versions compare extremely well with the exact answers obtained from exact diagonalization. The nuclear coupled cluster code used for these calculations has also been developed from scratch as part of this project, in order to serve as a reference to validate generated code in this domain. The code is now running in parallel at over 200 Mflops/processor on the largest jobs (on 256 processors), with the computational loops themselves running at 800 Mflops/processor and the rest of the time due to parallel communications overheads. In the near future, we plan to begin investigating the role of the three-body nuclear potential, which will require the capabilities of the optimizing TCE to produce implementations of this exceedingly complex method in a reasonable amount of time.

Finally, the HLL embedding work demonstrated the effectiveness of the approach by generating an implementation of the NAS parallel multigrid benchmark[14] that is performance competitive with the hand-coded reference implementation. The benchmark algorithm was expressed in just 60 lines of special-purpose high-level language, and another ten lines of code were required to express the stencil operations used. The generated code was 6000 lines of Fortran, which is comparable to the hand-coded version. One of the most interesting conclusions of this work is that there is a relatively limited set of features that are needed in a host language, and that these features can be satisfied by a language such as Python (a widely-used object-oriented scripting language),[15] as opposed to the extensive and complex features of a functional language such as Haskell. A paper describing this work is in preparation.

Summary and Conclusions

We have demonstrated the feasibility and utility of using high-level domain-specific languages together with specialized optimizing compilers to address both productivity and performance issues for complex scientific problems. This work provides a framework that we plan to extend to other problems and other domains. We have also made significant advances in the use of coupled cluster methods in nuclear physics, including introducing new methods developed by the chemistry community into nuclear physics.

Both aspects of this work provide significant benefit to the DOE. The problem of software productivity and the complexity of modern scientific software are issues which are rapidly reaching a crisis level; DOE is particularly effected by the problem because it has the largest concentration of computational science research in the United States. Our work shows the efficacy of one approach to dealing with productivity and complexity issues. Our work in nuclear physics will have the benefit of making it easier and more computationally efficient to compute the properties of nuclei to higher degrees of accuracy.

We are currently working on several proposals to secure follow-on funding. Dean is leading a proposal together with Piotr Piecuch and Karol Kowalski of Michigan State University that would explore the use of some of the coupled cluster-based methods developed by Piecuch's group in nuclear physics simulations. Harrison is leading a proposal, together with Bernholdt and other ORNL researchers to the Basic Energy Sciences Chemistry program that would include TCE-related work focused on quantum chemistry. In addition, Bernholdt, Sadayappan, Baumgartner, and Ramanujam plan to develop a proposal to DOE and/or NSF that would focus on the computer science issues of extending TCE technology into other domains. Finally, Bernholdt, Harrison, White, Sadayappan, and others have begun working with DOE MICS (computer science) program manager Fred Johnson to develop a new research program to address the larger issue of software productivity and complexity.

We are investigating funding opportunities through the NIH Bioengineering Consortium (BECON), in particular, the Exploratory/Developmental (R21) Bioengineering Research Grants program. Our collaborator, Eunok Jung, is visiting Lenhart in February 2004, and the plan of the proposal will be discussed at that time.

The implementation of the control strategy, obtained in our simulations for the optimal chest pressure profile, would have great benefit for saving lives in the civilian and military sectors. This biomedical innovation would be important to DOE and DOD.

References

[1]T. J. Lee and G. E. Scuseria. Achieving chemical accuracy with coupled cluster theory. In S. R. Langhoff (Ed.), *Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy,* pp. 47–109, Kluwer Academic, 1997.

[2]J. M. L. Martin. In P. v. R. Schleyer, P. R. Schreiner, N. L. Allinger, T. Clark, J. Gasteiger, P. Kollman, and H. F. Schaefer III (Eds.), *Encyclopedia of Computational Chemistry*, Vol. 1, pp. 115–128, Wiley & Sons, Berne (Switzerland), 1998.

[3]G. Baumgartner, D. E. Bernholdt, D. Cociorva, R. Harrison, S. Hirata, C.-C. Lam, M. Nooijen, R. Pitzer, J. Ramanujam, and P. Sadayappan, "A High-Level Approach to Synthesis of High-Performance Codes for Quantum Chemistry.," in *SC'02* (electronic publication), IEEE, 2002.

[4]D. Cociorva, J. Wilkins, G. Baumgartner, P. Sadayappan, J. Ramanujam, M. Nooijen, D. Bernholdt, and R. Harrison, "Space-Time Trade-Off Optimization for a Class of Electronic Structure Calculations," *Proc. Of the 2002 Conference on Programming Language Design and Implementation,* Berlin, Germany, June 2002.

[5]D. Cociorva, J. Wilkins, C. Lam, G. Baumgartner, P. Sadayappan, and J. Ramanujam, "Loop Optimizations for a Class of Memory-Constrained Computations," in *Proc. 15th ACM Intl. Conf. on Supercomputing,* 2001.

[6]D. Cociorva, J. Wilkins, G. Baumgartner, P. Sadayappan, J. Ramanujam, M. Nooijen, D. Bernholdt, and R. Harrison, "Towards Automatic Synthesis of High-Performance Codes for Electronic Structure Calculations: Data Locality Optimization," pp. 237–248 in *Proc. Of the Intl. Conf. on High Performance Computing,* Lecture Notes in Computer Science, Vol. 2228, Springer-Verlag, 2001.

[7]D. Cociorva, G. Baumgartner, C.-C. Lam, P. Sadayappan, and J. Ramanujam, "Memory-Constrained Communication Minimization for a Class of Array Computations," *Languages and Compilers for Parallel Computing,* 2002.

[8]D. Cociorva, X. Gao, S. Krishna, G. Baumgartner, C.-C. Lam, P. Sadayappan, and J. Ramanujam, *Global Communication Optimization for Tensor Contraction Expressions under Memory Constraints*, Tech. Report, Dept. of Computer & Information Science, Ohio State University, 2002.

[9]P. Wesseling, An Introduction to Multigrid Methods, http://www.mgnet.org/mgnet-books-wesseling.html.

[10]The Haskell Home Page, http://haskell.org.

[11]High Performance Computational Chemistry Group. *NWChem, A computational chemistry package for parallel computers, Version 4.5 (2003)*, Pacific Northwest National Laboratory, Richland, Washington, http://www.emsl.pnl.gov/docs/nwchem/.

[12]D. J. Dean and M. Hjorth-Jensen, "Coupled-cluster approach to nuclear physics," submitted to *Phys. Rev. C* (2003).

[13]K. Kowalsk, D. J. Dean, M. Hjorth-Jensen, T. Papenbrock, and P. Piecuch, "Coupled-cluster calculations of ground and excited states of nuclei," submitted to *Phys. Rev. Lett.* (2003).

[14]D. Bailey, J. Barton, T. Lasinski, and H. Simon, *The NAS Parallel Benchmarks*, Technical Report RNR-91-02, NASA Ames Research Center, Moffett Field, California, January 1991, http://www.nas.nasa.gov/Software/NPB/.

[15]The Python Home Page, http://www.python.org.

Cellular Algorithms for Next-Generation High-Performance Cellular Architectures

G. A. Geist, P. Agrawal, W. A. Shelton, and C. Engelmann
Computer Science and Mathematics Division

The world of high performance computing is rapidly changing, and a new type of highly scalable computer called cellular architecture is emerging as a potential next generation of computer for scientific simulation. In order to position ORNL for such petascale systems, we developed a new class of scientific algorithms that is able to scale to tens of thousands of processors and at the same time be tolerant of failures of some of these processors. We also developed a simulator for 100,000 processor computers that can support C, Fortran, and Java applications. The simulator can mimic single and multiple processor failures. Our research included collaboration with application scientists from materials and biology, as well as the Blue Gene group in IBM.

Introduction

ORNL has experienced tremendous growth in computing power in the past year and a half, going from 150 Gflops to over 5000 Gflops. If we are going to maintain this growth, we must begin preparing for the next generation of computers that will take us from 10 Tflops to over 100 Tflops.

This proposal has three goals:

1. Infrastructure—position ORNL for a 100 Tflops computer.
2. Genomes to Life—give ORNL a unique computational edge in the GTL initiative.
3. Materials Nanoscience—offer the opportunity to push the scale of first principles simulations.

Technical Approach and Results

Algorithms with "natural fault tolerance" have the property that they get the right answer even if one or more of the tasks involved in the calculation fail without warning or notification to the other tasks. In a sense, a naturally fault-tolerant algorithm just ignores failure and continues computing until the solution is found. Natural fault tolerance is an attractive property for cellular algorithms. For machines with 50,000 to 100,000 processors, the failure of a processor is predicted to occur every few seconds. Therefore, many failures will happen during a long scientific simulation.

In this project we have accomplished several important milestones. First, we have shown the existence of algorithms with both natural fault tolerance and the ability to scale to 100,000 processors. Second, we have created a cellular architecture simulator that can simulate a 100,000 processor system with arbitrary communication interconnect and the ability to mimic random failures. Third, we have presented two international and three domestic talks on this research at conferences.

We developed a meshless finite difference algorithm that is scale invariant and naturally fault tolerant. It can be used as a general template to create any number of scientific applications that can be formulated as a "function over some local region." To illustrate this, we took this algorithm and formulated it to solve Poisson's equation. The application gets the correct answer even if a hundred random parallel tasks are killed during the solution. The second challenge was to show the existence of natural fault-tolerant algorithms that are a function over the global region. Global information is often needed to determine convergence or maximum energy or things of this nature in large scientific applications. We were able to create a naturally fault-tolerant global maximum algorithm. At the beginning, all tasks had some random value. At the end of the algorithm, all tasks (that were still alive) knew the largest value and who had it. The algorithm was robust no matter which tasks died during the solution. We then developed a naturally fault-tolerant multigrid algorithm. We explored four different methods to reduce global synchronization within and between different grids. Only two were found to have the fast convergence property of serial multigrid.

During the first year we obtained and installed a Blue Gene simulator from IBM. We found that it simulated one Blue Gene node very accurately but was not designed to simulate hundreds (much less thousands of processors). We began development of a simulator that could simulate at least 100,000 processors and could kill any one or set of simulated processors while an application was running on it. In the first year we completed this simulator, which is itself a parallel application that runs on Linux clusters.

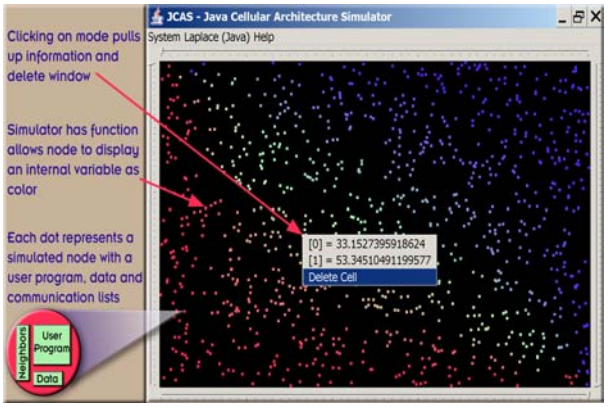


Fig. 1. Basic features of our simulator that can study the effects of faults and application recovery on computers with over 100,000 processors.

Using this simulator we were able to test the Poisson, global maximum, and multi-grid applications on 500,000 processors (see Fig. 1).

In materials nanotechnology, the major accomplishment in cellular algorithms was the development of a new K-space, O(N) first-principles screened Korringa, Kohn, and Rostoker (KKR) electronic structure method that has the potential to take advantage of next-generation architectures based on a cellular topology. This new method is more accurate than the locally self-consistent multiple scattering (LSMS) method. Equally important, we have been able to treat 2048 atoms on a single processor; thus, this new computational technique is capable of treating tens to hundreds of thousands of atoms. Similar to the LSMS method, the screened KKR constructs the so-called screened structure constants in real space by communicating with a local set of nearest neighbor atoms. The screened structure constants are then transformed to K-space via a Fourier transform, providing spectral or pseudo-spectral accuracy. Thus, the communication pattern is similar to the LSMS, which is an ideal algorithm for next-generation architectures based on a cellular topology. On the other hand, the LSMS method is purely a real-space method that lacks spectral accuracy and, thus, the energetics may not be of high enough accuracy to properly describe the underlying physics.

In computational biology, the challenge of folding a protein, through computer simulations, lies in finding the global minimum of the protein conformational energy landscape; sufficient evidence exists that the folded structure corresponds to a region close to the energy minimum. The size of protein conformational energy space increases very rapidly with the size of protein, and the conformational space for a biologically relevant protein is extremely large. It is impossible to exhaustively explore the complete conformational space even for a small protein in a reasonable time frame, given its multidimensional nature. The computational resources required for realistic protein folding simulations and resources available today fall short by several orders of magnitude. IBM's Blue Gene will drastically improve the available computational resources for large simulations. Therefore, we describe here the use of IBM's Blue Gene architecture to address the problem of protein folding by efficiently reducing the multi-dimensionality of the protein conformational energy landscape and matching it to the Blue Gene's architecture topology.

We are developing a new method to rapidly simulate the process of protein folding by reducing the multi-dimensionality of the protein conformational energy space to be searched. This new method is based on the understanding of protein dynamics in solution and using the knowledge of the physical principles which govern the process of protein folding. This method will take advantage of the Blue Gene architecture with thousands of closely connected processors. The method is being designed to be highly scalable with respect to the size of protein molecule and the number of processors.

In the proposed method, the multidimensionality is reduced by describing the modeled protein system by a set of essential dynamics variables. Such a description allows propagation of independent molecular dynamics trajectories to explore distant areas of the conformational space concurrently. Therefore, this approach is suitable for parallel computers with a large number of processors. The essential dynamics information obtained from simultaneous molecular dynamics trajectories are processed to direct the search into new regions of the conformational space. It is important to note that proposed method is different from standard molecular dynamics and Monte Carlo method as it is allows distant areas of the protein conformational space to be searched rapidly. In the current form, this method is implemented by using an iterative scheme. Each iterative step is composed of energy minimizations and molecular dynamics runs, which explore several local regions of protein conformational space. Based on the analysis of regions sampled during each iterative step, a set of new regions is explored in the next iteration, until global minimum is reached.

To take advantage of the Blue Gene's architecture with a large number of processors, the protein conformational space is mapped onto the topology of the machine. This is performed by dividing the protein conformational energy landscape into separate regions using the essential dynamics variables and then independently exploring these regions for the local minimum using a CPU block (see Fig. 2). The size of CPU block will be case dependent; the size of protein and scalability of the modeling program being the main factors. A small calculation at the beginning of each case will be

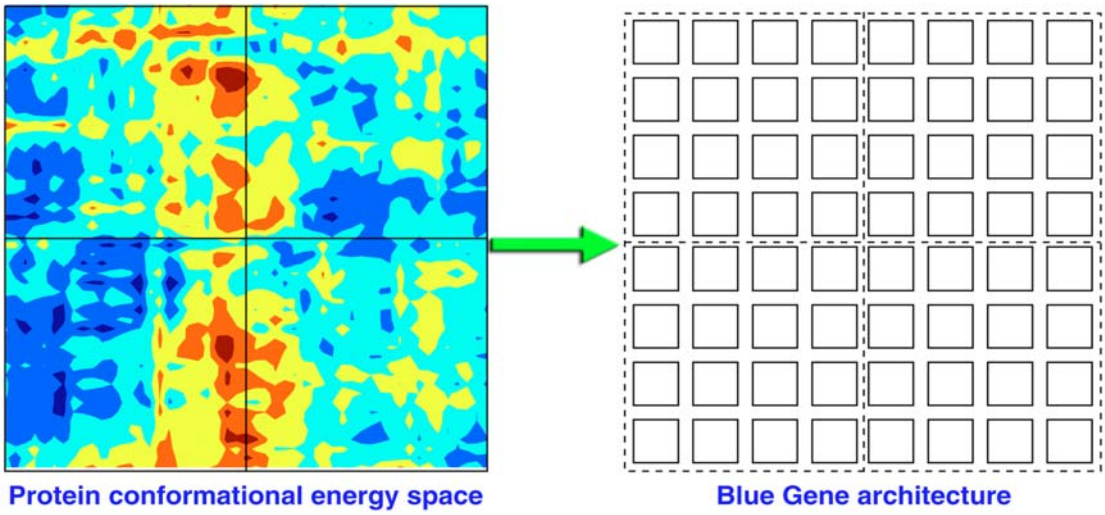


Fig. 2: Mapping the protein conformational energy surface onto the Blue Gene architecture.

able to determine the number of CPUs in each block for optimized performance.

Summary and Conclusions

As noted above, this project has been successful at strategically placing ORNL in position for petascale systems. We have been able to establish the existence of three broad classes of scientific solution methods that are naturally fault tolerant: finite difference, multi-grid, and global operations (such as global maximum). We have presented this work at invited talks at two international conferences and three conferences in the United States. We have also published the findings in two papers.[1,2]

The Blue Gene simulator obtained from IBM was found to be too low level for scalable application development. It could simulate Blue Gene nodes at the bit level but could only simulate up to 32 processors. In contrast, the simulator we developed can simulate up to 500,000 processors using a 64-node Linux cluster. It supports Fortran, C, and Java applications and is able to study the effects of processor failure on these applications.

Feasibility studies for using systems with 100,000 processors in ORNL key areas of biology and nanotechnology were completed and new algorithms developed that can take advantage of large-scale systems likely to exist at ORNL in the next three to five years.

Finally, this project has met its goals of being a catalyst in bringing in new funding to ORNL. One of the objectives of this project was to strategically position the Laboratory for a 100-Tflops computer. A proposal was prepared and submitted to DOE in July 2002 that if funded would place a Cray X1 computer into ORNL by the end of this project and would set the stage for placing a $100M computer into ORNL by 2005. The proposal, which is in response to the Japanese Earth Simulator, leverages the expertise we are gaining in super-scalar algorithms from this LDRD and ORNL's leadership in scalable systems software.

The project also led to a proposal to the DOE Genomes to Life initiative, which was subsequently funded for $19 million, of which $5.7 million will come to ORNL.

The third development activity was getting an IBM CRADA in place to help augment the research into cellular algorithms in general and Blue Gene in particular. This CRADA involves both funds-in from IBM as well as matching funds from DOE and is expected to bring over $2.4 million into ORNL over the past three years.

References

[1]A. Geist and C. Engelmann, "Development of Naturally Fault Tolerant Algorithms for Computing on 100,000 Processors," *Journal Parallel Computing*, December 2002.

[2]A. Geist, "Progress Towards a Petascale Virtual Machines," pp. 10–15 in *Recent Advances in Parallel Virtual Machine and Message Passing Interface*, ed. Laforenza and Dongarra, Springer LNCS 2840, September 2003.

Scalable Tools for Petascale Distributed-Data Analysis

G. Ostrouchov, N. F. Samatova, D. J. Erickson, and R. J. Toedte
Computer Science and Mathematics Division

This project concerned the development of scalable algorithms to perform complex analyses of massive distributed-data sets. The focus was on scalable dimension reduction computations and visualization for astrophysics and climate simulation data, and for biological data. Our results include several new algorithms for dimension reduction that operate on both distributed and streaming data, as well as breakthrough theoretical results that pave the way for faster and more robust dimension reduction computations on massive data sets.

Introduction

The current and future landscape of data analysis is shaped by the relative innovation speed in four basic components of modern computing: processor, memory, disk, and network. Trends recently reported in *Scientific American* show exponential growth in all components with disk capacity outpacing the rest by a factor of 2 per year. The result is that data sets are becoming massive, distributed, dynamic, and essentially immoveable. Nearly all data analysis software fails on such data sets because of excessive memory requirements and the need to bring the data to a central location. Further, the massive distributed-data problem exists not only in the usual sense over a wide area network but also on a single storage device, where data sets are fragmented into smaller files in response to memory constraints. Terabyte data sets from climate and astrophysics simulations are examples of such data, often residing on storage systems like the High-Performance Storage System (HPSS) as a collection of thousands of files. Biology data sets can be massive not only because of the number of genomes and genome components but also because of increasing availability of annotation tools that can produce very-high-dimensional feature sets for the genomes and genome components.

Dimension reduction techniques are fundamental to discovery and visualization of structure in high-dimensional data. These techniques are still underutilized in science even on smaller data sets, and their use will grow as we learn how to pose useful dimension reduction analyses in more applications. We also point out that dimension reduction methods are not only of interest for visualization of structure, but they are also enabling tools for other data analysis methods by reducing the amount of data for further processing. Pushing the frontier of dimension reduction tools to massive distributed data sets enables progress across many data science applications.

Climate model data archives generated at ORNL are approaching 40 terabytes and are increasing exponentially. These model-generated outputs are typically centuries in length, and the time resolution of the archived data sets is 6 hours to 1 month. The spatial resolution is usually 18–70 vertical levels and ~250-km horizontal grids for the global simulations. Several hundred "variables" are computed for each grid point in three dimensions for every 20 min, for centuries. Since general circulation models are highly nonlinear, the need to have scalable dimension reduction is a critical aspect of climate model interpretation.

Under the SciDAC activity lead by ORNL, supernova modeling continues with increasingly sophisticated multidimensional simulations with ever improving physics, particularly multidimensional neutrino (radiation) transport. The ultimate goal is to ascertain the key ingredients that constitute the core collapse supernova mechanism. Dimension reduction methods will be used to uncover spatial structure and relationships in the multidimensional simulation.

Technical Approach

Dimension reduction (DR) techniques are computationally complex but provide the most effective general means of discovering structure in high-dimensional data. For example, the result of an application of principal component analysis (PCA) to climate simulation data is shown in Fig. 1. First, it matters what projections are selected as some have strong structure and some look random. Second, each structured projection has an interpretation. For example, (2,4) in the middle illustrates that global warming affects winter temperatures the most and summer temperatures the least. Winter months are on the left, and summer months are on the right. And third, each component is also visualized as a map, giving insights

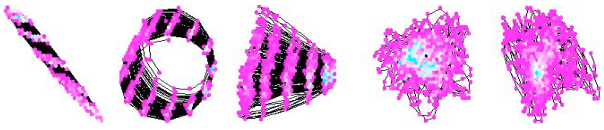


Fig. 1. Monthly surface temperature principal component dynamics of a 120-year climate model simulation under transient CO_2 increase (a 2500-dimensional time series representing 2500 locations in the northern hemisphere). Pair wise dynamics of the top six components are shown as they seasonally rotate through 12 months; all are two-dimensional projections of the same, roughly 2500-dimension object. Components shown are (1,2), (2,3), (2,4), (5,6), and (4,5), where (2,4) in the middle illustrates that global warming affects winter temperatures (left) the most and summer temperatures (right) the least.

into which regions are associated with certain component dynamics. Figure 2 displays the second principal component, where the coefficients are a temperature contrast between land and sea. PCA is by far the most popular and simplest of the DR techniques [known as empirical orthogonal functions (EOF) in climate, Karhunen-Loeve transform in signal processing, and is also related to normal modes in engineering and molecular structure simulation].

Algorithms for parallel matrix computation, and also algorithms for out-of-core matrix computations, provide some insights for organizing distributed DR computations. However, there are important differences. First, we cannot assume that we distribute the data to optimize performance of our algorithm; rather, the data are already distributed according to a given application and it is our task to minimize data transfers required by our computation. This can be considered a disadvantage relative to the parallel and out-of-core computations. Second, DR is a form of approximation and that presents an opportunity and an advantage over exact parallel matrix computation through local DR and reduced data transfer rates. Finally, in DR the matrix factorization itself is of interest rather than the solution of a matrix problem.[1]

We considered data distributed over several locations as well as data streaming from a running simulation experiment. Our approach to analysis of distributed data was driven by the concept of performing local analyses

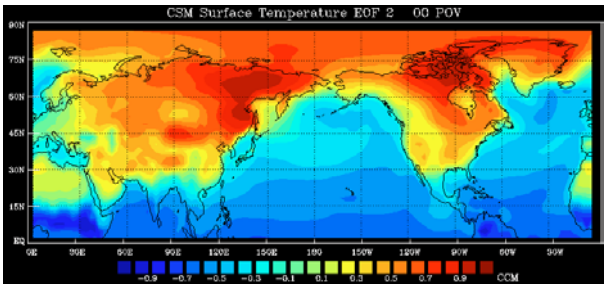


Fig. 2. Second principal component is shown as a map with positive coefficients in shades of red and negative as shades of blue.

on local data that are combined into a global analysis with minimal data transfers. We assume that the data are already distributed according to the needs of a given application and any data transfers must only involve the minimum information needed to compute a required global DR result. Such methods also, in part, address the issues of streaming data. As more data are generated, updating can be accomplished by considering the new data as separate distributed data sets that are sequentially combined into the global analysis. Within our focus on DR techniques, we view the distributed data as components of a massive virtual matrix, where columns are the features defining the dimensionality of the problem, and rows are the items. How such a matrix is constructed is highly varied among applications, although common concepts exist. One such concept is the construction of empirical orthogonal functions in climate simulation.

Results and Accomplishments

Two dimension reduction (DR) methods were selected for the distributed data problem. One is principal component analysis (PCA), which is currently the most popular DR method, and the other is FastMap, a distance-based method that can be viewed as a fast approximation to PCA. We considered the application of these algorithms to data distributed across several locations as well as streaming data. Because fast algorithms are particularly relevant in the massive data set problem, we considered theoretical properties of FastMap. This provided the basis for a new class of adaptive algorithms that are both fast and robust. This class of adaptive algorithms will likely provide the breakthrough performance that is needed for massive petascale data sets. Finally, we also considered very-high-dimensional data sets generated with annotation tools from biological data and developed the statistical basis for extracting relevant dimensions. We describe all these results.

Distributed PCA. We developed an algorithm for the application of PCA to data distributed over several locations.[2] Our algorithm does not require that the local data is transferred to the central location. Only an approximation of each local covariance matrix along with a vector of means is centralized. The basis of our algorithm is that the global data covariance can be partitioned into a sum of "within" locations covariances and a "between" locations covariance. We approximate the "within" local covariances with principal components and compute the "between" covariance centrally. Each local approximation carries the linear optimality properties of PCA. Compared to an algorithm that centralizes the data, our algorithm reduces data transfers from O(np), where n is items and p is features, to O(sp), where s is the number of data locations. The number of items is typically one or more orders of magnitude larger than the number of data

locations. For example, our algorithm computes the top 50 empirical orthogonal functions across *s* climate simulations stored at *s* locations, each with 2400 time steps (200 years of monthly averages) and a spatial resolution of 7000 grid points using data transfers of only (s)×2×50×7000 + 50 data items. Centralizing the data requires transferring (s)×2400×7000 data items, an amount roughly 24 times larger. For large simulations, combining so much data at one location may not be possible.

FastMap Properties and a New Class of DR Algorithms. FastMap is a DR technique that operates on distances between objects. Implicitly, the technique assumes that the objects are points in a high-dimensional Euclidean space. It selects a sequence of orthogonal axes defined by distant pairs of points (called pivots) and projects data onto the orthogonal axes. We showed that FastMap picks all of its pivots from the vertices of the convex hull of the data points in the original implicit high-dimensional Euclidean space.[3] This provided a connection to results in robust statistics, where the convex hull is used as a tool for multivariate outlier detection and for robust estimation methods. The connection sheds a new light on some properties of FastMap, in particular its sensitivity to outliers, and provides an opportunity for a new class of DR algorithms that retain the speed of FastMap and exploit ideas in robust statistics. The main lesson from robust statistics is that the most distant points are often not the best choice for defining a projection axis. The key to new fast and robust methods is a replacement of FastMap's pivot points heuristic by something that considers more than just the maximum distance from a point. Many other properties are available within *O(np)* complexity, thus retaining FastMap speed. In fact, methods using these properties will be more robust than standard PCA. There are many directions that this methodology can take, and undoubtedly many such algorithms will be proposed, particularly for DR in massive data sets. We have started this with several algorithms for distributed and streamline data.

DFastMap: Distributed FastMap Algorithms. Although FastMap is an order of magnitude faster than PCA, it too requires that the data are in a central location. We have developed two variants of FastMap that perform most of their computation locally near each distributed data set and centralize only some minimal information to compute a global result that is comparable to centralizing the data. The intuition behind our approach is as follows. FastMap tends to select each pair of pivots so that they are widely separated and among the extreme points of a data set. We present two approaches,[4] one is iterative and the other direct. Our connection of the convex hull to FastMap pivots[3] gives an explanation of why an application of DFastMap to distributed data performs as well as the serial FastMap on a centralized data set. The union of local convex hull vertices includes all convex-hull vertices of the centralized data set. DFastMap centralizes the pivots, arguably a very good subset of the local convex hull vertices.

Streamline FastMap Algorithm. This data model—often called data streams—includes scientific simulation output, satellite images, network traffic data, etc. Processing data streams presents both practical and theoretical challenges; it often requires immediate results as streams flow in. We developed a novel dimension reduction algorithm Xmap,[5] based on the FastMap heuristic. Xmap exploits the connection between FastMap and the convex hull of data in Euclidean space. Xmap effectively identifies a small representative subset of past data and maintains well-approximated lower-dimensional map of data received up to any given time.

Streamline Reservoir Sampling Algorithm. Random sampling is an accepted basis for estimation from large data sets that outstrip available computer memory. When the data comes as a stream, its total size is potentially infinite and usually only one pass through the data is possible. Reservoir sampling is a method of maintaining a fixed-size random sample from streaming data. Recently, several methods for reservoir sampling without replacement were introduced. We introduced a new method for reservoir sampling with replacement.[6] We first proved that the proposed method maintains a random sample with replacement at any given time. Then we introduced a refined version that significantly speeds up the overall sampling procedure. Sampling with replacement is relevant for methods such as bagging and bootstrapping, both necessary for uncertainty assessment in complex estimation settings.

High-Dimensional Annotation Data Reduction for Detecting Interactions. Annotation tools in biology produce massive data sets of labels or indicators. Such data are high-dimensional but are not points in Euclidean space as are the data necessary for the PCA and FastMap methodologies discussed earlier. However, such label data can be reduced to a potentially very large multidimensional table of counts. The ways in which this is done can vary depending on a particular application. We investigated such constructions for protein interaction data and developed methodology for reducing such tables to a few significant cells that are key for predicting interaction.[7]

Scalar Field Data Reduction and Reduced Views. We have also developed novel dimension-reduced views of instability propagation in the entropy field of a supernova simulation, shown in Fig. 3. These views provide a compact summary of various characteristics of an entire simulation and provide a new effective means of comparison between two- and three-dimensional simulations.

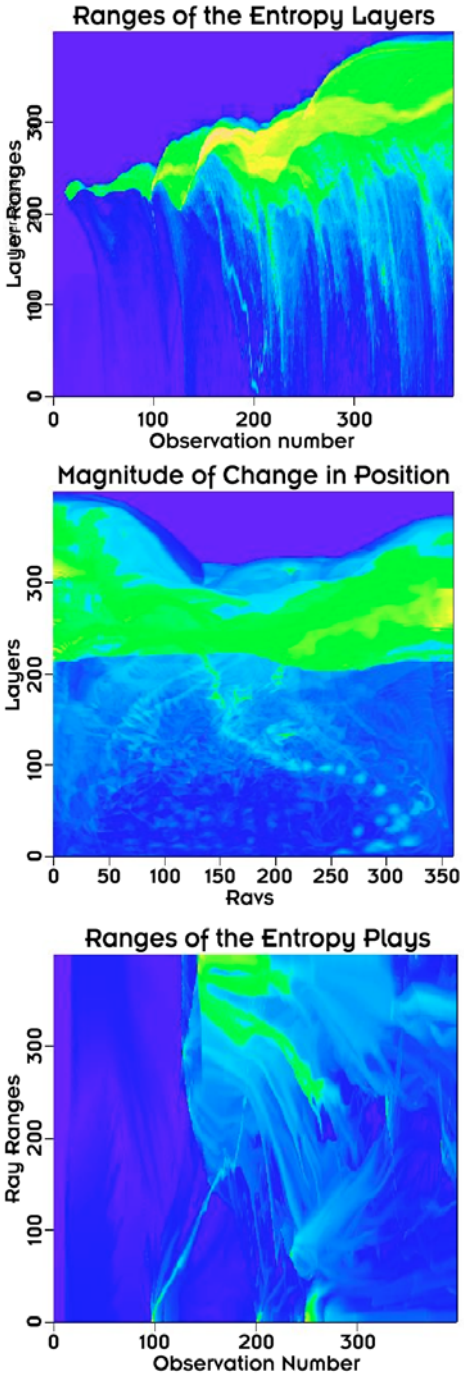


Fig. 3. Three orthogonal views of entropy instability in a 400 time-step 2-d supernova simulation are shown with polar coordinates presented as Cartesian. Spatial and temporal propagation of the instability can be seen from three points of view. These views remain preserved after 30x data reduction based on adaptive principal components.

Summary and Conclusions

The most promising result of this project is the theoretical connection of a distance-based DR to the convex hull of a high-dimensional data set. This provides the basis for a new class of adaptive algorithms that are both fast and robust. This class of adaptive algorithms will likely provide the breakthrough performance that is needed for massive petascale data sets. Strong results were also obtained on several algorithms for complex analyses of massive distributed data sets. Among these is a new distributed version of principal component analysis, which is the most popular DR technique. Other distance-based DR algorithms were developed for distributed as well as streaming data. Dimension reduction methods are not only of interest for visualization of structure but they are also enabling tools for other data analysis methods by reducing the amount of data for further processing. Pushing the frontier of DR tools to massive distributed data sets enables progress across many data science applications, including data from climate and astrophysics simulations.

This research puts the ORNL at the forefront of scalable algorithm development for DR computations. This capability is unique among the DOE laboratories and positions ORNL for future work in complex analysis and visualization methods for massive data sets. This also opens the door to developing useful analysis sequences that involve dimension reduction in applications that produce massive quantities of data.

Our work has resulted in six refereed publications of which three[2,4,5] have already appeared and the other three[3,6,7] are currently under review. In addition, we have co-sponsored the Statistical Data Mining Conference, June 22–25, 2002, in Knoxville jointly with the University of Tennessee Department of Statistics, where we also organized a special session on Distributed Data Mining. Several presentations of most of our results were made at other international conferences, including the Spring Research Conference on Statistics (2002 and 2003) and the Workshop on High Performance Data Mining at the Second (and Third) SIAM International Conference on Data Mining (2002 and 2003).

References

[1]L. Hubert, J. Meulman, and W. Heiser, "Two purposes for matrix factorization: a historical appraisal," *SIAM Review* **42**, 68–82 (2000).

[2]Y. M. Qu, G. Ostrouchov, N. F. Samatova, and G. A. Geist, "Principal Component Analysis for Dimension Reduction in Massive Distributed Data Sets," pp. 4–9 in *Workshop on High Performance Data Mining at the Second SIAM International Conference on Data Mining*, 2002.

[3]George Ostrouchov and Nagiza F. Samatova, "On FastMap and the convex hull of multivariate data: Toward fast and robust dimension reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, submitted for publication, 2003.

[4]F. N. Abu-Khzam, N. F. Samatova, G. Ostrouchov, M. A. Langston, and A. G. Geist, "Distributed Dimension Reduction Algorithms for Widely Dispersed Data," pp. 174–178 in *Parallel*

and Distributed Computing and Systems, Editors: S.G. Akl and T. Gonzalez, ACTA Press, 2002.

[5]B.-H. Park, N. Samatova, G. Ostrouchov, and G. A. Geist, "Fast dimension reduction algorithms for multivariate streamline data," pp. 1–6 in *Proceedings of the 6th International Workshop on High Performance Data Mining: Pervasive and Data Stream Mining*, 2003.

[6]B.-H. Park, G. Ostrouchov, and N. F. Samatova, "Reservoir-Based Random Sampling with Replacement from a Data Stream," in *Third SIAM International Conference on Data Mining*, submitted. April 22–24, 2004, Kissimmee, Florida, submitted for publication, 2004.

[7]B.-H. Park, G. Ostrouchov, G.-X. Yu, A. Geist, A. Gorin, and N. F. Samatova, "PICUPP: Protein Interaction Classification by Unlikely Profile Pair," *Journal of Computational Biology*, submitted for publication, 2003.

Scalable Visualization Tools and Technology

S. P. Dickson
Computer Science and Mathematics Division

There is a critical need for visualization tools and technology that support continued scientific progress in light of the immense growth of scientific dataset size and complexity, as generated by large-scale software simulations. This looming shortfall affects a wide variety of scientific domains, including, but not limited to, astrophysics, biology, climate modeling, material science, and fusion research. The growing emphasis on ultrascale software simulation exacerbates this problem, and visualization paradigms are a key technology for enabling the next generation of scientific discovery. Yet, the ad hoc manner in which visualization has typically been applied is highly inadequate for upcoming DOE Office of Science computational programs. The latest visualization tools and infrastructure must be carefully integrated and applied toward scalably viewing and intuitively interpreting these complex and massive data sets. This project has focused on the evaluation and development of state-of-the-art visualization technologies, including the integration of visualization hardware and software into a functional production-scale user facility. A broad survey of the visualization field has been made, and two distinct scientific visualization facilities have been designed. Several collaborations have also been formed with scientific research groups at ORNL to help educate them in their use of visualization for scientific discovery.

Introduction

Visualization is often applied to scientific datasets to provide an intuitive visual method of presentation and exploration, especially for more massive and complex data collections. Software simulation experiments at ORNL are already poised to generate data in the terabyte (fusion, climate, and astrophysics) to petabyte (biology) range in the near term. An elaborate data analysis and navigation infrastructure is required to enable *any* practical level of scientific discovery for these increasingly difficult scenarios. Yet, such an infrastructure consists of many challenging subproblems, including large-scale data storage, networking, data reduction, analysis and mining, and of course, visualization tools and algorithms and scalable high-resolution graphical displays. Unfortunately, adequate visualization support is often "assumed" when considering the needs of large-scale simulation experiments. Nonetheless, the technology and systems required to efficiently visualize such increasingly immense amounts of data are far from "turn key" solutions.

ORNL has a respectable and growing competence in many computer-related research areas, including high-performance scientific simulation, parallel and distributed computing, and large-scale data mining and analysis, but there are serious technology gaps with respect to scalable visualization tools and the associated technology. While many useful visual images of science have been generated over the past several decades, the visualization technology and facilities at ORNL were antiquated and in serious need of improvement. Further, most local scientific research teams have not been using visualization, even where there could be great benefit from doing so. Other research groups that have applied visualization technology have typically done so with simple, low-resolution techniques, using an informally organized infrastructure.

A vital and unified research program must therefore be systematically initiated at ORNL in several areas, including application-specific scientific visualizations, parallel rendering, scalable visualization and analysis pipelines, interactive data exploration, and large-scale display management. The purpose of this LDRD activity has been to promote such growth in state-of-the-art visualization at ORNL. The specific goals of this work were to (1) explore visualization ideas and platforms that support scientific application areas and (2) apply emerging technologies toward the development of new visualization hardware/software facilities at ORNL.

Technical Approach and Results

The two main thrusts of this work relate to opposite ends of the scientific visualization spectrum – exploring application-specific visualization solutions for key application areas and developing a general-purpose scalable visualization facility hardware design. The former requires interactions with domain scientists to determine

their general needs and how visualization might assist their efforts. The latter involves concrete design of a production visualization venue where scientists of all domains might come to explore their large datasets. Both of these efforts required an evaluation of the existing visualization technology, in both hardware and software.

Application-Specific Visualization

A variety of scientific application domains would benefit from visualization that is specially catered to the features and processes unique to associated physical models. Obvious examples include rendering molecules for chemical or materials simulations, but this idea could be extended to other fields such as biology, genomics (protein folding), climate, high-energy physics, and fusion science. Highly accurate and literal visual depictions can be employed to effectively communicate specific phenomena being modeled. Many new science research programs are also including a visualization component in their funding – a new concept gaining credibility in many SciDAC (DOE's Scientific Discovery through Advanced Computing) application areas, such as the Terascale Supernova Initiative (TSI) astrophysics center.

To determine the best tools and technologies to apply for application-specific visualization, and to avoid duplication of existing work in this area, a brief survey of several prominent visualization systems was performed. The following systems were explored: the VMD[1] system (University of Illinois) for visualizing the results of Molecular Dynamics, DIVERSE[2] (Virginia Tech) for visualizing data on multi-tile and immersive display systems, the Visualization Toolkit[3] (VTK, KitWare) for general-purpose scientific visualization, and ASPECT[4] (ORNL) for applying a variety of data-analysis functions in visualizing large-scale datasets. To explore the applicability of these tools to the variety of common research platforms used by scientists in different application domains, these visualization tools were installed and tested on both MacOS and Linux workstations. Testing was also performed using a prototype parallel visualization cluster dedicated to driving various display facilities at ORNL. Several experiments were performed to combine and integrate sets of these tools as would likely be required for production visualization scenarios; for example, VMD was combined with Chromium[5] (University of Virginia), a pixel compositing software system for parallel rendering and display.

Several meetings were arranged with various domain scientists at ORNL to determine their current uses of and potential needs for visualization. Based on interactions with the Computational Biology Group, several approaches were explored for visualizing the molecular models in the Protein Data Bank file format. Similarly, volume datasets from various mathematical models, and experimental data sources such as from X-ray computer tomography (CT), were visualized using the VTK system. An improved visualization was prepared of the Virtual Human work done by researchers in ORNL's Computational Sciences and Engineering (CSE) Division.

In collaboration with researchers in the Computational Biology Group, a view was developed of the *E. Coli* 1ML5 molecular model using the VMD software package on both Linux and MacOS-X systems (Fig. 1). Similar methods were explored to visualize the 1ML5 molecular model using VTK on a Linux system. As a result of this view development, several software bugs were identified in VTK's Protein DataBank (PDB) and VRML file readers; these problems were subsequently resolved in cooperation with VTK developers at Kitware. Protein Data Bank models were also visualized in VMD using Chromium on the prototype visualization cluster and PowerWall. Several other datasets were visualized, including a CT dataset of a porcine thorax using VTK, and data from the Virtual Human effort at ORNL. Much of this work was collected into a 20-min high-resolution movie to showcase ORNL science programs at the IEEE Supercomputing 2003 Conference.

Several external conferences, tutorials and workshops were attended to exchange information with peer researchers on the currently available visualization tools and determine the fastest growing areas of visualization research: the ACM SIGGRAPH and IEEE Visualization conferences, tutorials and workshops; the UTK Conference on Conformal Geometry; and the DOE Computer Graphics Forum.

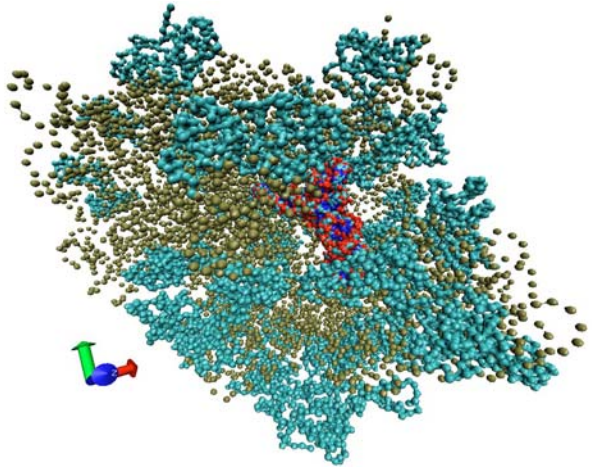


Fig. 1. View of the E. Coli 1ML5 *protein model.*

High-Resolution Visual Display Facility

The many powerful visualization tools are useless without a proper visualization facility infrastructure within which to execute them. A proper venue is needed for bringing together datasets from various application domains, to render and visually display the data for interactive exploration. When manipulating and visualizing such massive data sets, it is crucial to maximize the capabilities of the display hardware to provide the best opportunity for scientific discovery, by revealing fine-grained features and relationships in the scientific data. While ORNL has supported several useful visualization laboratories over the years, the current equipment was antiquated and unable to adequately handle the high-resolution needs of the large-scale data being produced. To this end, a new high-resolution visual display wall, known as the "Science Exploratorium," was designed.

Initially, a literature search[6,7] was performed to identify the proper hardware, software and facilities infrastructure necessary to construct a large "PowerWall" visualization theater. In addition to research papers, several conference sessions were attended regarding the principles of designing, constructing, and maintaining such a large-scale display. In conjunction with an ORNL Visualization Task Force, various display technologies, including screens, projectors, and mechanical mounting systems, along with rendering cluster and network infrastructure, were researched to design a production visualization facility for ORNL's new Computational Sciences Building.

The resulting design consists of three large adjacent glass screens, each 8 ft high and 10 ft wide, driven by an array of 27 powerful (high-lumen) rear-projection projectors (9 projectors per screen). A large 64-node computational cluster is dedicated to parallel rendering to drive the projectors for this "tiled" display wall, wherein each projector fills in the pixel illumination for one of many small rectangular areas, or "tiles," on a given screen of the display. The collection of projectors must be carefully synchronized, aligned, and color-matched to produce a smooth and uniform image across the entirety of the display wall (Fig. 2). The nature of these subtle adjustments is surprisingly complex and is still an active area of significant research.[8]

During the development and acquisition of this multi-million-dollar visualization facility, a simple prototype system was constructed using three large screens with only one projector per screen. This scaled-down version, at one-ninth the resolution, enabled much of the software infrastructure to be explored in advance of the final system installation. Many of the visualization tools explored in the first phase of this project were applied to drive the prototype PowerWall facility using a small cluster test bed. Several datasets from different application domains were selected for these initial experiments, as discussed in the following section, already enabling useful exploration of several scientific results (Fig. 3). In recognition of this effort, the Visualization Task Force received a Significant Event Achievement award for their contributions to the development of the new facilities.

Fig. 2. Projector array for tiled display walls.

In conjunction with the primary "Exploratorium" theater design, a smaller facility was designed for use in a conference room in the new Research Office Building at ORNL. This system consists of 6 projectors that drive a display in tiles on a 12-ft-wide by 8-ft-high glass screen. Based on the experiences from these two visualization facility designs, recommendations were also offered on the design of audio-visual facilities for the new Joint Institute for the Computational Sciences (JICS) Building at ORNL.

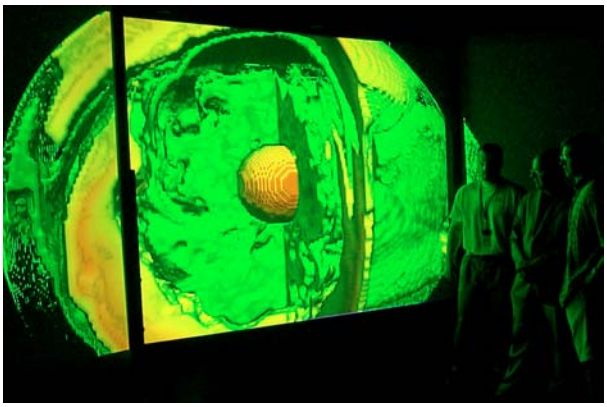


Fig. 3. Prototype powerwall visualization display.

Summary and Conclusions

Visualization provides an intuitive representation of the results of scientific computation, leading to useful insight into the science being explored. Visualization is often the point of reference through which scientists on an investigative team discuss their individual contributions toward the final results. The knowledge gained by this project, through research into PowerWall design and construction, has guided the planning and implementation of new visualization facilities in the Computational Sciences and Research Office Buildings. These facilities, even prior to full completion, have become a focal point for a variety of scientific meetings and collaborations. This result alone is significant.

The efforts of this project have also highlighted several scientific achievements at ORNL via visual presentation of their results. Observations of the activities in the Computational Sciences Building point to a broader use of visualization in exploring science and sharing project results. The expertise gained in scientific visualization will be of benefit to many existing and future scientific projects at ORNL and our collaborators.

References

[1]Visual Molecular Dynamics (VMD), Theoretical and Computational Biophysics Group, University of Illinois, http://www.ks.uiuc.edu/Research/vmd/

[2]DIVERSE Cross-Platform Open-Source Interface for Developing Virtual Reality Applications, Virginia Tech., http://diverse.sourceforge.net/

[3]The Visualization Toolkit (VTK), Open Source Software System for 3-D Graphics, Kitware, Inc., http://public.kitware.com/VTK/

[4]Scalable Data Management (SDM) Center, Office of Science, Scientific Discovery through Advanced Computing (SciDAC) Initiative, ASPECT Tool for Exploratory Analysis and Data Mining, http://sdm.lbl.gov/sdmcenter/

[5]G. Humphreys, M. Houston, Y.-R. Ng, R. Frank, S. Ahern, P. Kirchner, and J. T. Klosowski, "Chromium: A Stream Processing Framework for Interactive Graphics on Clusters," ACM Transactions on Graphics, Vol. 21, no. 3, (SIGGRAPH 2002), http://www.cs.virginia.edu/~gfx/pubs/cr/

[6]NCSA "Display Wall-in-a-Box," July 19, 2001, National Computational Science Alliance, http://www.ncsa.uiuc.edu/News/Access/Stories/IAB/

[7]Princeton University, "Scalable Display Wall." http://www.cs.princeton.edu/omnimedia/

[8]Raskar Ramesh, Paul Beardsley, "A Self Correcting Projector," Mitsubishi Electric Research Laboratories TR-2000-46, Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR) 2001, January 2002.

Scaling Climate Models for Future Computer Architectures

A. S. Bland, J. B. Drake, P. H. Worley , and J. B. White III
Computer Science and Mathematics Division

The 40-TFLOP/s Japanese Earth Simulator (ES40) has challenged the U.S. climate-modeling community and computer industry to accelerate the development of both the models and the computers required to run them, to keep our country competitive in both climate science and policy decisions. This project will characterize the performance of current U.S. models and suggest research areas for improvements. Researchers will use representative kernel codes to understand the factors that limit performance and scalability and then work to mitigate these factors. The second phase will be to instrument and evaluate the latest full climate model, CCSM2, to improve its performance. The objective is to prepare ORNL to assist DOE in developing and deploying an ultra-scale computer system for the U.S. science community.

The objective of this project is to prepare ORNL to assist DOE in developing and deploying an ultra-scale computer system for the U.S. science community. The approach to this is to begin scaling the components of the Community Climate Systems Model (CCSM) up to the resolution and physical fidelity that is expected to be used on this new class of computer systems and to port these codes to a new generation of high-end computer architectures.

This project has three major technical thrusts:

1. Benchmarking performance of a dynamical kernel on a high-resolution model to identify hardware and software issues with current and planned architectures.
2. Evaluation and implementation of performance enhancements for global spectral atmospheric models.
3. In-depth scientific performance study of the CCSM2 at high resolution on IBM p690 and Cray X1 systems.

The performance benchmark targets a spectral model of atmospheric dynamics. We will match the resolution of T1279L96 used in the ES40 benchmark that achieved 26 TFLOP/s. As currently configured, both the IBM p690 and Cray X1 at ORNL are currently insufficient to run this problem resolution. However, we have run experiments using problem sizes with large horizontal resolutions (up to T680 on the Cray and T1279 on the IBM) with small numbers of vertical levels and problem sizes with smaller horizontal resolutions and up to 92 vertical levels, determining the processor and memory requirements necessary to run the target problem size.

The experiments on the Cray were preceded by extensive performance tuning of the kernel code PSTSWM to improve its vectorization. We have improved the performance by a factor of 10 over the original version of the code on a single processor, achieving approximately 50% of peak for the largest horizontal resolution. Experiments indicate that maximum serial performance is achieved by computing on a few vertical levels at a time. This is not the approach used in the current production spectral model (nor in the Earth Simulator code) and has motivated investigating alternative parallelization strategies in the SciDAC project concerned with development of the production atmospheric model. The next step will be to run the full T1279L96 problem resolution on the upgraded Cray X1. Preliminary optimizations of the parallel algorithms for this large run have already been completed using the current system.

A high-resolution ocean configuration has been targeted using the POP code. The ES40 is showing performance of hundreds of Gflops in a 60-node configuration for POP at 1/10-degree resolution. The Cray X1 exceeds the ES40 performance for the low-resolution case, and we expect to find competitive performance with the high resolution case as well.

Results to date show that the Cray X1 architecture is superior to the IBM p690 in both performance and scalability for the CCSM. Future work will focus on completing scaling to T1279L96 resolution for benchmarking and maximizing performance of the full model for high-resolution configurations.

Advanced Computational Methods

R. J. Harrison, D. E. Bernholdt, and E. D'Azvedo
Computational Science and Mathematics Division

We are developing, evaluating, and deploying novel numerical methods for the solution of non-linear integral-differential equations on the next generation of massively parallel computers. Initial application is to computational chemistry, but the resulting methods and tools are expected to be widely applicable. In computational chemistry, we are developing methods that are both free of the basis-set error that plagues conventional methods and scale correctly with the system size. In a complementary effort, we are also developing a new and general framework for composing programs for the next generation of massively parallel computers with $O(10^5)$ processors.

In principle, all of chemistry can be understood and predicted from solution of the Schrödinger and/or Dirac equations. In practice, we cannot solve these 3-N dimensional (N, the number of electrons) differential equations and must resort to a variety of approximations. These approximations limit both the accuracy that may be obtained and the size of system that may be studied. We aim to remove several of these limitations with the objective of enabling more reliable and accurate computations on larger systems.

Conventional computational chemistry adopts the linear combination of atomic orbitals (LCAO) approximation and constructs effective one-electron molecular wave functions from a superposition of atomic wavefunctions. This gives very compact wave functions with desirable physical properties, since molecules are indeed composed of atoms. However, the LCAO approximation is not suitable for high-accuracy calculations and, because the atomic wave functions have extended support, the cost of calculations scales non-linearly with the size of the system. Many-electron wave functions are conventionally constructed from linear combinations of products of one-electron molecular wave functions (orbitals); however, this expansion converges very slowly. By using multiresolution methods in multiwavelet bases, we have already constructed fast algorithms for the solution of the density functional equations to arbitrary finite precision. Novel separated representations for integral operators make this approach feasible in three and higher dimensions. The use of multiwavelet bases with disjoint support enables high-order convergence even in the presence of singularities. This near-spectral convergence may be extended to many-body wave functions in several ways, and we plan to explore several of these options.

In collaboration with Prof. R. Butler (Middle Tennessee State University), we are also designing and implementing parallel tools for the support of irregular and hierarchical distributed data structures using one-sided access mechanisms on massively parallel computers. These will be built upon the Global Array (Harrison and Nieplocha) and ARMCI (Nieplocha) libraries that have already been optimized on essentially all current high-performance computers and workstation clusters. They both already exploit shared-memory within SMP nodes. We will add a primitive but efficient distributed hash table that can support a wide variety of data structures from tuple spaces to nested trees.

Creating New Climate Drivers and Interactions in Global Climate Models

D. J. Erickson,[1] M. Branstetter,[1] M. Sale,[2] W. M. Post,[2] A. King,[2] L. Gu,[2] F. Pan,[2] and S. Hadley[3]
[1]*Computer Science and Mathematics Division*
[2]*Environmental Sciences Division*
[3]*Engineering Science and Technology Division*

This project involves the identification of important climate drivers not already present in current climate models. The drivers to be perturbed are (1) carbon-cycle feedbacks, (2) hydrological-cycle feedbacks and (3) economic feedback responses to climate change. The scientific analysis of the significant feedbacks in the Earth system is the key to reliability and uncertainty estimation of future climate prediction and is critical in making meaningful policy decisions. The programmatic objective is to make ORNL the central DOE institution for computational climate modeling and to leverage the significant present and future ORNL computational resources to obtain future funding in the burgeoning field of numerical climate, carbon and water cycle modeling, and economic-climate feedbacks.

The technical R&D objectives of this LDRD are the evaluation and creation of new climate drivers/feedbacks in the climate modeling system. The installation of state-of-the-art global climate models and an assessment of feedbacks and critical biogeochemical physical parameterizations in the global climate models operating on ORNL computational platforms are central to this LDRD. The deliverables will be several refereed journal articles, meeting presentations, and a clear delineation of future (2–5 years) funding opportunities. This LDRD is intended to position ORNL climate science at the forefront of the U.S. climate modeling efforts.

We have a wide range of accomplishments to report at this juncture.

We have ported the U.S. state-of-the-art global Climate Model (CCSM2) to the CCS/CSM high-performance computing center. We have evaluated the control run of the CCSM2 (Branstetter and Erickson, 2003) and have performed a variety of sensitivity tests of the CCSM2.

An additional milestone is that we have, with colleagues at LLNL, ported the only fully coupled climate-carbon cycle model in the United States to ORNL computers. The effect of acclimation has been coded in to IBIS terrestrial biosphere submodel, and an experimental run has been staged. We have three additional test runs of the CCSM2 staged to explore new scientific characteristics of the CSM2.

In the area of hydrologic cycle feedbacks, we are developing new modeling procedures to improve how soil moisture dynamics and lateral transport of water is represented in the CCSM2 code. Our initial computational experiments completed this year indicate that the code for soil wetting and drying is not realistic and that there may be errors in the land surface modules of CCSM2. We are now in the process of recoding the important parts of the land surface model, including a new simplified algorithm and supporting global data for estimating TOPMODEL baseflow parameters and soil profile characteristics. This new approach is based on work completed by Pan this year that evaluated a number of different methods for calculating topographic index and flow direction.

We have several refereed publications published, in press, and in preparation.

Our program development strategy is initially to build an obvious expertise/group that is active in global climate modeling science. At the national CCSM2 workshop in Breckenridge, Colorado, in June 2003, ORNL had 14 scientific/computational scientists representing our efforts. We had several lectures, posters, and workshop participation on a variety if climate modeling issues.

Extensive team/partnership building with DOE program sponsors, NASA sponsors, sister DOE National laboratories, and university colleagues has been a focus of this LDRD. Erickson has initiated and taken responsibility in contacting and visiting scientists at LANL (January 2003), LLNL (December 2002), and DOE Headquarters [1.25-hour presentation to Patrinos, Elwood, Farrell, Amthor (May 2003)]. As the funding increases on the computer hardware side continue, we are positioned to secure funding from DOE HQ as the climate science budget grows in concert with the hardware applications. King is pursuing collaborations with Jon Foley, University of Wisconsin, author of the IBIS terrestrial biosphere model.

We have, at present, five proposals submitted to NASA, six proposals submitted to DOE HQ, and one submitted to USAID.

Biomolecular "Locks and Keys"—High-Performance Computing for Investigation of Recognition Principles in the Complexes of Biological Macromolecules

A. Gorin,[1] R. Day,[1] V. A. Protopopescu,[1] Y. Li,[1] D. Xu,[2] and M. Mascagni[3]
[1]*Computer Science and Mathematics Division, ORNL*
[2]*Life Sciences Division and University of Missouri*
[3]*Florida State University*

We propose to develop and implement a biomolecular modeling platform with new computational capabilities in three tightly linked areas: (1) modeling of diverse and heterogeneous biological macromolecular complexes; (2) algorithm design that will utilize a new generation of supercomputer architectures; and (3) incorporation of novel types of experimental and database information. None of the planned capabilities currently exist, and successful realization of at least some of them is likely to dramatically benefit the whole field of computational simulations in biology. The developed platform will be prototyped and applied to the study of two problems, which are known to be notoriously complex but of critical importance in modern biological research: structure of membrane protein assemblies and recognition principles in DNA-protein complexes. The proposed effort is complementary to other projects in the Computer Science and Mathematics, Life Sciences, and Environmental Sciences divisions and fits well into the framework for developments in nanotechnology and high-performance computing at ORNL.

The objectives of this project are to design new capabilities in the computational modeling of biological complexes in three areas: (1) biomolecular modeling platform, (2) exploration of Monte-Carlo methods, and (3) incorporation of NMR residual dipolar couplings (RDC) measurements into biomolecular modeling.

Here we report current progress and the FY 2004 research plans, including two additional research tasks. During the first year, we implemented a software library for collective variable algorithms and explored several Markov Chain Monte-Carlo schemas, proposing a new one—the Accelerated Simulated Tempering. In the application area we developed the combinatorial assignment procedure (CAP), which is a new method to utilize RDC for robust automatic assignment and structural characterization of biological complexes.

Task 1. Implementation of the Collective Variable Algorithms

We have developed and implemented (both in C and in Perl) algorithms for constructing molecular models using collective variables. Specifically, we have completed implementation of several types of helical parameters (local, global, symmetric, and linked to the single rotation axis), dihedral angle variables, including pseudo rotation angles, and routines for the manipulation of the rigid sub-fragments (such as secondary-structure elements). Algorithms and subroutines for the force field parameter files (reading, writing, calculating energy components) are progressing well and are expected to be finished by the end of the fiscal year. Plans for the next fiscal year include work on the closuring algorithms and command language for the modeling platform. With the completion of these components (end of FY 2004), we will be in a position to implement our platform on Cray X1 supercomputer (in collaboration with Cray team) and apply it for the studies of protein-protein recognition.

Task 2. Monte Carlo Sampling Methods and HPC implementations

During FY 2003, we have implemented and tested four major types of Markov Chain Monte-Carlo algorithms, namely, (a) Metropolis Monte-Carlo, (b) Simulated Annealing Monte-Carlo, (c) Simulated Tempering and (d) Weighted Monte-Carlo. The most promising results in terms of convergence speed were obtained with Simulated Tempering. Based on our studies, we have proposed a new method—"Accelerated Simulated Tempering Scheme." We construct an alternative ladder that combines temperature and size of step, (t_i, s_i), where $1 = t_1 < t_2 < \ldots < t_m$, and $s_1 < s_2 < \ldots < s_m$. Then, we adjust the original Greyer and Thompson's algorithm by "leaning: the temperature ladder. To this end by set at each ladder asymmetrical transition function, forcing the system to have a lower probability of visiting the higher-level temperature ladders. Plans for the next fiscal year include

an implementation of the conformational stack algorithm (to accumulate together results of the several parallel Monte-Carlo processes) and testing of the implemented algorithms on two real-life biological applications: our molecular modeling platform (models of protein-protein complexes) and Rosetta program (models of single-domain proteins).

Task 3. Integration of the Residual Dipolar Couplings data

Residual Dipolar Couplings (RDC) measurement is one of the most promising new developments in NMR solution techniques. Recently, we have recognized that RDC-based technologies assisted by high-performance Computing can be applied for solving hard problems in structural ribonomics and proteomics. Briefly, the CAP assignment method works in the following way. RDCs are first measured on a target protein molecule. This leads to the assignment of chemical shift values with a corresponding RDC value. Unlike chemical shifts, the magnitude of a given RDC value reflects the average orientation of the bond vector and hence can readily be related to a given structure. For a given assignment guess, the RDC values will either agree or disagree with the given structure. For the ideal case, the best agreement, measured as the root mean square deviation (rmsd) between measured and "best-fit" calculated RDCs, will occur for the correct set of assignments. To enhance this unique mapping, one can also work with a group of RDCs (backbone NH and $C\alpha H\alpha$) that have been correlated using 3D experiments [HNCA, HN(CA)HA]. All possible RDC assignment permutations can therefore be explored for consistency with a given target structure. During next fiscal year, we will apply our algorithm on a wide variety of biological systems including large RNAs, single- and multi-domain proteins.

Toward Common Components for Computational Nanoscience

T. C. Schulthess,[1] M. E. Summers,[2] D. E. Bernholdt,[1] and W. Elwasif[1]
[1]*Computer Science and Mathematics Division*
[2]*Computational Science and Engineering Division*

An open-source repository of computational materials science codes that are portable, extensible, and well maintained on modern high-performance computing architectures is needed to meet the challenges and take advantage of the opportunities posed by nanoscience. The costs of building and maintaining such a repository will be prohibitive, unless new approaches are found to integrate legacy code systems in such a repository. Here we propose to start building a suite of software analysis and modification tools to semiautomatically integrate legacy codes into modern software environments for computational material science. Recent developments of an object-oriented and generic toolset for computational materials science have proven an efficient avenue for building future applications codes for nanoscience. We propose to leverage this fact to develop versions of toolset codes that are compliant with the Common Component Architecture, an emerging approach to component-based software development, which explicitly recognizes the requirements of high-performance computational science codes, including those in nanoscience. It will also demonstrate how the efforts to develop software for computational nanoscience can be integrated with other software projects of the SciDAC initiative.

In order to meet future expectations, computational materials science (CMS) must make efficient use of the most advanced high-performance computing (HPC) infrastructure that is available. The strategy for advancing CMS within ORNL's Center for Computational Science (CCS) is through the participation of domain scientists in the development of computational methods, algorithms, and simulation software and early evaluation of new computer hardware. To enable this, a Materials Research Institute within CCS (CCS-MRI) has been created with the mission to build and maintain a community of leading computational materials scientists and to integrate them into the development of high-performance computing infrastructure of the CCS. One of the tasks of the CCS-MRI is to develop a repository of computational materials science codes that are optimized and maintained on CCS's high-performance computing infrastructure. Besides legacy electronic structure codes, the repository project includes the *ψ-Mag* toolset, a highly extensible generic library of data structures and algorithms that are commonly used in materials simulations. Furthermore, an important part of the SciDAC initiative is the development of a Common Component Architecture (CCA), which aims at providing specification for the development and use of components as basic building blocks of high-performance simulation codes. CCA is designed to make the incorporation of existing code and libraries into components straightforward.

The vision pursued by this project is to combine automated source-code analysis and manipulation tools (SCAMT) with the paradigms of the CCA and the generic software design techniques used in the *ψ-Mag* toolset, in order to build and maintain the software repository of the CCS-MRI. The SCAMT that are developed here will be used for the maintenances of existing codes as well as the transformation of legacy codes into CCA-compliant components. Similarly components of generic libraries, such as *ψ-Mag*, BOOST (www.boost.org), or ALPS (alps.comp-phys.org), will be made CCA compliant. Future software development in CMS that makes use of these tools will be more efficient and well tailored to high-performance computer architectures. Furthermore, the task of integration and maintenance of external codes and libraries into component-based frameworks is representative of widespread issues faced by CCA in other domains as well. At present this project is the only one in the CMS domain using CCA, presenting a unique opportunity to establish ORNL leadership in this area while establishing the groundwork for the development of CMS community codes.

In the first year of this project, we focused on developing the basis for SCAMT and with making the generic *ψ-Mag* tools CCA compliant.

Developing the basis for SCAMT, A F77/F90 Parser and user interface for CMS source code analysis. We followed our original plan and developed an open source

parser for Fortran 77 and Fortran 90 with which abstract-syntax tree (AST) representations of the source codes are built and subsequently used in the analysis tools. The specifications for the AST—there is no commonly accepted standard for ASTs—are being developed along with the analysis tools to optimize the parser systems for the particular requirements of CMS code development and maintenance. Our parser is presently complete enough to generate ASTs for the codes that are presently used in the CCS-MRI (the particular codes we tested are LSMS, SIC-LSD, and LKKR). For scientists who are working on the CMS codes but are not interested in the bare AST, we have developed a Web-based interface which displays source code along with the information provided by the AST in terms of pop-up menus and hyper links between source code segments.With this tool it will be straightforward to isolate parts of legacy codes that are to be used in CCA-compliant components. Making these tools Web-based also simplifies collaborative development of CMS software for the CCS-MRI repository over the internet.

Making the ψ-Mag Tools CCA Compliant. The *ψ-Mag* toolset is representative for numerous generic libraries based on C++ templates that are gaining importance in CMS. CCA uses the Scientific Interface Definition Language (SIDL) to interface with such libraries. The major issue in the present task was to delineate the parts of the *ψ-Mag* toolset that can be mapped onto the CCA-compliant components using SIDL in its present form from those parts that will require modification of SIDL. We have found that it is possible to turn generic classes and functions in the *ψ-Mag* toolset into SIDL objects by (1) creating an instance of the *ψ-Mag* object inside the SIDL implementation and (2) providing a C++ wrapper for each SIDL interface that exactly matches the corresponding *ψ-Mag* concept. Automation of this process appears straightforward, either through a user-generated specification file that extends SIDL to describe C++ typdefs and templates, or through a parser which is able to extract the appropriate information from the C++ source code. However, a wholesale export of *ψ-Mag* into SIDL is not desirable because generic libraries use templates to provide information about objects they implement at compiler time—CCA presently follows the object-oriented model that uses primarily run time polymorphism. We are examining several solutions on how to handle such problems.

During the second year of this project, we will

(1) develop example SCAMT such as an automatic precision conversion tool,
(2) combine components with competing parallelization strategies in one code, and
(3) based on task 1.1, develop CCA-compliant CMS components from legacy codes.

Intelligent Consequence Management for Energy Assurance

J. H. Sorensen,[1] O. Franzese[2], F. Southworth[2], B. Bhaduri[3], and C. Liu[3]
[1]*Environmental Sciences Division*
[2]*Engineering Science and Technology Division*
[3]*Computational Sciences and Engineering Division*

This project proposes to develop a prototype system to demonstrate the feasibility of intelligent consequence management for energy security. An intelligent consequence management system will utilize state-of-the-art sensor technology and wireless communications coupled to real-time simulation and decision support models that will greatly enhance management of crises in rapidly unfolding events. This system will have major impacts on how we can reduce the consequences of threats to our infrastructure. The concept applies to many energy security issues including power disruption from natural threats such as hurricanes and ice storms, energy emergencies such as blackouts, power outages or natural gas disruptions, and terrorist-induced damage against energy facilities or buildings. Specifically we will develop a real-time evacuation model for simulating the evacuation of an area threatened by the release of a hazardous material from an energy system and couple the model with a protective action expert system. This will integrate key ORNL strengths including emergency management, transportation, building technology, sensor technology, simulation, and high-performance computing.

We are developing components of an intelligent consequence management (ICM) system that utilizes state-of-the-art sensor technology and wireless communications coupled to real-time simulation and decision-support models that will greatly enhance management of crises in rapidly unfolding events. Among the properties of intelligent consequence management are the following:

- New sensor networks or links to existing sensor networks designed to detect and monitor the threats of concern.
- High-speed communications and data exchange.
- Real-time simulation models running on high-speed machines.
- Advanced decision support tools that can process data and simulation outputs into a format useful to decision-makers.

The chosen application to demonstrate ICM is a real-time evacuation management system. The problem of determining optimal evacuation paths out of an area at risk can be modeled as a network flow problem. The objective of the network flow evacuation problem is to route a given amount of people from a set of source nodes to a set of exit nodes in the least time, without violating the capacity constraints. Current evacuation models use static traffic assignment and assume the conditions of the network at the beginning of the simulation prevail throughout the evacuation. A better representation of the problem of determining the optimal a priori evacuation paths would be a dynamic network flow model.

A prototype dynamic assignment model was developed at ORNL prior to the development of real-time traffic monitoring for vehicular networks. An experimental application to use traffic data was also developed. This code was integrated with an existing evacuation simulation code ESIM to produce a dynamic evacuation code called DSIM. In each time period of the simulation, new traffic assignments are calculated that are based on the distribution of traffic in the previous time period. Predictions of future traffic behavior are based on current traffic patterns, which are continuously updated at discrete time periods.

The DSIM code more closely simulates driver behavior in congested traffic than a static model. If people encounter stalled traffic, they will take an alternative route based on their knowledge of the area, maps, or traffic guides. Seeking alternative routes that represent underutilized capacity decreases the time needed to evacuate. Initial tests of DSIM show that evacuation times are predicted to be shorter than when modeled with a static assignment. The dynamic assignment makes it feasible to use real-time traffic sensor data to drive the model. The real-time data would be supplied by sensors that assess travel times and flows along the different links on the network as well as queues at chocking points.

Distributed Intrusion Detection and Attack Containment for Organizational Cyber Security

S. G. Batsell,[1] N. Rao,[2] and M. Shankar[1]
[1]*Computational Sciences and Engineering Division*
[2]*Computer Science and Mathematics Division*

The sheer number and sophistication of cyber attacks are making our nation's critical information infrastructure increasingly vulnerable. Important techniques for cyber security include the use of intrusion detection and distributed containment. Most available intrusion-detection schemes are either single host based or local network based and consequently are unable to protect systems and networks effectively when the attacks propagate autonomously in a distributed fashion. The goal of this project is to address the growing demands on intrusion detection and containment by developing an integrated cyber-security framework to identify and contain attacks within large-scale distributed organizational networks. The framework will be distributed, autonomous, and capable of detecting new and rapidly spreading attacks.

Summary

Federal agencies such as the DOE and other large nongovernmental organizations rely heavily on information infrastructures, especially through connection with the Internet to conduct their operations. For example, the entire energy production and transportation industry uses networked computer systems to monitor, control, manage, and barter information and goods. Cyber attacks can significantly impact their operations and potentially create denial-of-service scenarios affecting millions of Americans. A cyber attack jeopardizes the ability to provide services, enforce privacy, and protect information; it undermines the level of trust one has in both the infrastructure and in an agency's ability to perform its mission.

While recent years have seen significant advances in cyber security research, the task of combining diverse detectors of intrusion remains challenging and open from an analytical and foundational perspective. In this project, we employ known results from data fusion to correctly weight the signals of the intrusion sensors based on their detection and miss probabilities. An advantage of the technique of data fusion is that it enables us to build upon and reuse preexisting sensors by incorporating them into new detection schemes without sacrificing accuracy. Following the detection of an intrusion at a host, the system must take preventive steps and distribute this information to other hosts so that they can proactively take action against a spreading intrusion. Containment measures such as closing router ports and isolating compromised hosts have high overheads, thus requiring that fusion and information distribution steps are timely.

In our first year, we completed modules that utilize network-level sensors to trace the physical paths of packets within a domain by using simulated sensors. By building propagation graphs that capture attack paths, and by utilizing graph-search algorithms to perform packet-level content search, we isolate the attack sources. When feasible, such tracing leads us to the attacker. In particular, this method detects if the attacks are from within or from outside the organizational network by suitably collecting the sensor data. We are planning to integrate these methods with actual network sensors in the future.

We have also developed a new approach of fusing intrusion data from diverse sensors to improve detection performance.[1] We determined, through quantitative analysis, the response-time parameters required of the distributed system implementation we are developing. Based on peering and fusing information from distributed sensors of intrusion, the detection and containment architecture can deter rapid self-propagating attacks such as worms. We calculate the real-time constrained responses needed to stifle the spread of an intrusion when the protection is enforced locally within an intranet. Our analysis also suggests the response-time characteristics required in a wider-area distributed network. We are currently developing our research results for further publication and continuing development on the prototype implementation.

Reference

[1]M. Shankar, N. Rao, and S. Batsell, "Fusing Intrusion Data for Detection and Containment," in IEEE *MILCOM 2003*, Boston, October 13–16, 2003.

Image to Intelligence Archive: Intelligent Agent–Based, Large-Scale Spatial-Data Management and Analyses

T. E. Potok,[1] B. L. Bhaduri,[1] K. W. Tobin,[2] P. J. Palathingal,[1] E. A. Bright,[1]
S. S. Gleason,[2] and T. P. Karnowski[2]
[1]*Computational Sciences and Engineering Division*
[2]*Engineering Science and Technology Division*

Enormous volumes of remote sensing image data are being produced on a daily basis throughout the world. This geographical data is analyzed to create scientific, military, and intelligence information. There are two significant challenges that exist in producing this information: (1) managing the vast amount of available data which is steadily increasing and (2) automating the manual processes that are currently needed to produce and search this type of information. We have developed a prototype that addresses these issues by combining novel approaches from three distinct research areas, namely, software agents, georeferenced data modeling, and image analysis.

Image and remote sensing information is of critical importance to the scientific community and for national security purposes. Staggering increases in the volume of remote sensing image data are being produced and made widely available. For example, the United States Geological Survey (USGS) has been archiving enormous volumes of data and faces exponential near- and long-term growth in digital data.[1]

We have focused on two key issues in processing this type of data: (1) managing the vast amount of increasing data and (2) automating the manual processes that are currently used to produce and search this image information. The current work in the field has a strong focus on managing and performing analysis on a closed set of images.[2–4] Many of the proposed techniques do not scale well and do not consider how new images will be found and incorporated into the system. Likewise, the ability for a scientist or an image analyst to effectively use these systems to find geographic information is also limited. What is needed is the ability to select a geographic location for search, then the ability to find regions within that location, and then to find features within that region. For example, an image analyst may be looking over the country of Iraq for regions that contain small cities and then truck convoys within those cities. Developing tools that support this type of analysis is the objective of this project.

We have developed a prototype system that merges capabilities from three separate areas: software agents, georeferenced data modeling, and image analysis. Our system uses software agents to gather new images from previously unknown sources, update image archives using "better" images, and parallelize aspects of the image region and feature extraction process. The georeference data modeling is able to normalize the images placed into the archive, while the image analysis technology provides the ability to segment an image into geographic regions, then within those regions, to find various features. With this system, a scientist or an imagery analyst is able to determine changes in regions of images based on items of interest, for applications such as damage assessment, automatic target recognition, and a wide variety of intelligence, security, and monitoring purposes.

The architecture of our system is an integration of the three components, namely (1) software agent technology, (2) geoconformance modeling, and (3) image analysis. A high-level outline of the architecture is shown in Fig. 1. The agent subsystem is responsible for searching for and retrieving spatial data. Intelligent software agents search for images on the internet in places where they are likely to be found. They store these images in a repository that also maintains metadata about the image in XML format. The images in the repository are modeled and normalized by the geoconformance modeling subsystem. Finally the image analysis module takes the images and generates regions of interest. A graphical user interface (GUI) interacts with the image analysis module to enable users to effectively search and retrieve regions of interest from these vast amounts of spatial data.

With our prototype system, we have demonstrated the ability to manage vast amounts of image data and automate the manual processing of this data. This is a significant step forward in image analysis capabilities. We have presented this novel approach to a number of potential sponsors who are very interested in the concept, including the Department of Defense, International Atomic Energy

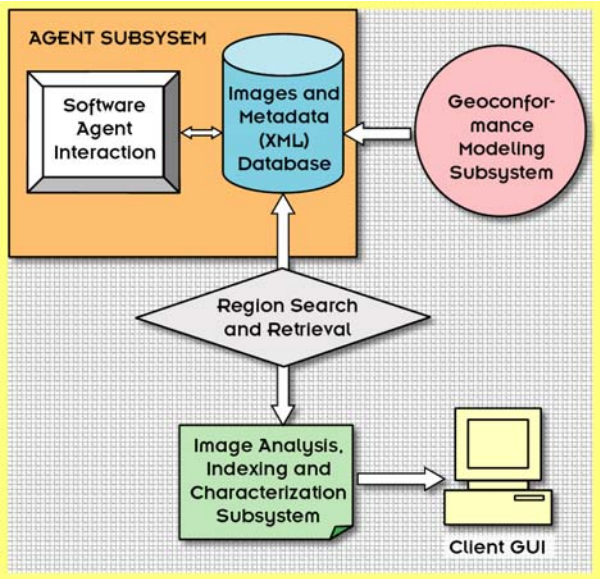


Fig. 1. An outline of the I2IA Architecture.

Agency, and the Intelligence Community. We are planning on furthering the capabilities of the prototype to include a demonstration using ORNL reservation data and to publish the results of our work.

References

[1]United States Geological Survey, "http://vangogh.cse.buffalo.edu:8080/digital/digital/node4.html," *Multimedia Research Group, Department of CSE, SUNY at Buffalo* (2000).

[2]B. Zhu, M. Ramsey, T. D. Ng, H. Chen, and B. Schatz, "Creating a Large-Scale Digital Library for Georeferenced Information," *D-Lib Magazine* **5**, 7/8 (1999).

[3]T. M. Lehmann, B. Wein, J. Dahmen, J. Bredno, F. Vogelsang, and M. Kohnen, "Content-Based Image Retrieval in Medical Applications: A Novel Multi-Step Approach," *in Proceedings SPIE* **3972**, 312–320 (2000).

[4]O. Derrick, S. Raha, J. Kerievsky, and K. Kaila, http://htmlparser.sourceforge.net/.

COMPUTER AND COMPUTATIONAL SCIENCES

Seed Money Fund

Cardiopulmonary Resuscitation Using Optimal Control

E. Jung,[1,2*] S. M. Lenhart,[1] and V. A. Protopopescu[1]
[1]*Computer Science and Mathematics Division*
[2]*Konkuk University, Seoul, South Korea*

Optimal control (OC) techniques are applied for the first time to a model of standard cardiopulmonary resuscitation (CPR). Pattern of control of the pressure on the chest is implemented on a validated blood circulation model consisting of a system of seven difference equations. The externally applied chest pressure acting as the "control" is the nonhomogeneous forcing term in this system. The OC approach seeks to maximize the blood flow, as measured by the pressure differences between the thoracic aorta and the right heart and superior vena cava.

Introduction

Each year, more than 250,000 people die in the USA from cardiac arrest. Despite widespread use of CPR, the survival of patients revived from cardiac arrest remains poor. The rate of survival for CPR performed outside the hospital is 3%. For patients who have CPR in the hospital, the rate of survival is 10–15%. The practical technique of CPR has changed little since the 1960s. The technique of OC proposes a strategy for improving resuscitation rates using a validated circulation model developed by Babbs.[1]

Technical Approach

We apply OC techniques for the first time to a CPR model. The idea is to develop a new CPR strategy to improve resuscitation rates. The chosen CPR model consists of seven difference equations for the hemodynamics of adult human circulation. For the OC application, the circulation model is extended to include the control functions as discrete inputs. As a control, we choose the pattern of the pressure applied on the chest. The controls at the current time step and the previous time step enter the model in the calculation of pressures at the next time step. This type of input control is unusual and requires an innovative adaptation of the discrete version of Pontryagin's Maximum Principle.[2,3] To make the chest pressure profiles medically reasonable, we assumed that the admissible controls are zero at the beginning and end of the time interval.

In Babbs' model,[1] heart and blood vessels are represented as resistance-capacitive networks, pressures in the chest and vascular compartments as voltages, blood flow as electric current, and cardiac and venous valves as diodes (electrical devices that permit flow in only one direction). The following pressure variables are used: pressures in abdominal aorta, inferior vena aorta, carotid, jugular, thoracic aorta, right heart and superior vena cava, and thoracic pump. There is one difference equation for the time evolution of each pressure variable. Valves are incorporated into the equations for the pressures in jugular, thoracic aorta, right heart and superior vena cava, and the thoracic pump.

The OC seeks to maximize the blood flow as measured by the pressure differences between the thoracic aorta and the right heart and superior vena cava. The goal is to maximize this pressure difference summed over the time steps subject to a stabilizing control size minimization. [1] The flow is measured in systemic perfusion pressure (SPP), a measure of the blood flow between the thoracic aorta and the right heart and superior vena cava. The "control to pressure" solution map and the "control to the goal" map are differentiated to characterize the optimal control. Existence, uniqueness, and explicit characterization of the OC were completed for this control problem. The OC is completely and explicitly characterized in terms of the solutions of the circulation model and the solution of an adjoint model. Uniqueness of the OC can also be shown under the assumptions that the time step size is small enough. For this discrete CPR model, the "natural" time step in this discrete CPR model depends on the other coefficients in the system. Whenever warranted, uniqueness is obtained by proving the uniqueness of the solution of the optimality system, obtained by coupling the original circulation system, the adjoint system, and the control characterization. Thus when

*This work was conducted in collaboration with Eunok Jung of Konkuk University, South Korea, who was an ORNL employee at the time of the initial work on this project. C. F. Babbs from Purdue University (consultant M.D. on this project) was involved in the setup and analysis of the OC problem.

the optimal control is numerically calculated through solving the optimality system, the uniqueness guarantees that we find the unique OC.

The numerical implementation of this control solution involves an iterative method with a forward sweep of the circulation model followed by a backward sweep of the adjoint model with a control characterization update afterwards. Since the circulation model has initial data and the adjoint system has data at the final time, the coupled system cannot be solved forward together. The iterative method starts with a guess for the control values, and then the control is updated after each forward sweep and backward sweep. The forward and backward sweeps are repeated until convergence of the iterates is achieved. Since we are using difference equations, the sweeps are "time stepping" through the interval. The state and adjoint equations give the next values from the previous values and control values from the last two steps. The resulting numerical code was run for a range of parameters which are reasonable for this CPR model. The "optimal" chest pressure profiles and the corresponding SPP were calculated on each run.

Results and Accomplishments

The OC procedure applied to Babbs' discrete CPR system resulted in an explicit characterization of the OC. The OC gives the chest pressure profile to be followed in the CPR process. Our OC chest pressure profile results in improvement over the standard CPR procedure used on this model.[1] The improvement is measured in SPP. In Fig. 1, we show a typical chest pressure example profile, "optimal control," for a certain set of parameters, which is appropriate for the standard heart model. The time scale

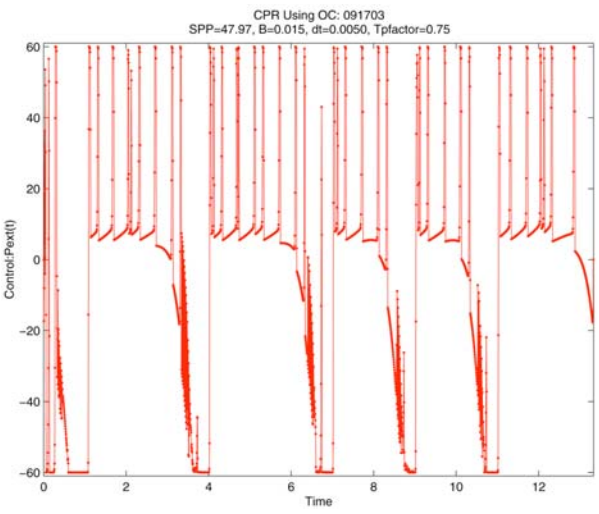


Fig. 1. Graph of the "optimal" chest pressure profile is given over a time scale in seconds. The term dt *is the size of the time step. The coefficient B is the stabilizing factor and "Tp factor" is the strength of the cardiac pump. The SPP is the systemic perfusion pressure (mmHg).*

is in seconds. The terms *dt* gives the size of the time step. The coefficient *B* is the stabilizing factor, and *Tp* factor is the strength of the cardiac pump. The SPP for this example is much higher than the SPP from the standard CPR technique as seen in Babbs' work;[1] we see about 20% improvement in SPP in our various examples. The fluctuation seen in this profile is typical of many of the examples we ran and indicates that rapid changes in pressure levels can make a significant improvement in SPP. This profile can be considered as a type of CPR with active compression and decompression (ACD) of the chest. The SPP for our example compares favorably with the SPP calculated from the standard ACD procedure.[1]

While OC of discrete difference equations has been used for many physical models,[4] applying these control techniques to a CPR or heart model has not been done before. From a technical standpoint, the control entering into the system at two moments of time (current and immediate past time steps) and giving input to the pressure system at the next time introduces a novel feature and a non-trivial difficulty that has been addressed and solved here.

This improved CPR strategy can be implemented either by an emergency helper or by a portable pressure device, which can achieve the optimal chest pressure profile, as indicated by the OC. The device could be further optimized by using more detailed circulation models based on partial differential equations and even customized to account for various age, sex, and weight groups within the general population. Extensions to nonstandard CPR techniques can also be treated with this OC technique. This type of portable device should interest medical equipment firms as well as the Army, and we expect to submit external proposals to continue this type of work.

We have submitted a Report of Invention[5] documenting the novel idea of applying OC techniques for the first time to a model of CPR. In that report, we note the novelty and relevance of the transition from differential equations to difference equations for this application and focus on the implementation of the improved strategy on a portable CPR device. Finally, as a part this project, E. Jung and C. F. Babbs submitted a paper on "valveless pumping."[6] Medical evidence shows that valveless pumping may play an important role in CPR.

Summary and Conclusions

Based on the results of this research, an increase of 20% in blood flow (quantified by SPP) is to be expected as a result of applying optimized CPR strategies. This may represent the difference between life and death for many people who undergo cardiac arrest. Moreover the technique can be easily implemented using portable devices.

References

[1]C. F. Babbs, "CPR Techniques that Combine Chest and Abdominal Compression and Decompression: Hemodynamic Insights from a Spreadsheet Model," *Circulation* **100**, 2146–2152 (1999).

[2]L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mischchenko, *The Mathematical Theory of Optimal Processes,* Wiley , New York, 1962.

[3]F-T Fan and C-S Wang, *The Discrete Maximum Principle: A Study of Multistage Systems Optimization,* Wiley, New York, 1964.

[4]S. P. Sethi and G. L. Thompson, *Optimal Control Theory, Applications to Management Science and Economics,* Kluwer, Boston, 2nd edition, 2000.

[5]S. Lenhart, V. Protopopescu, and E. Jung, *Optimal Control of CPR Procedure*, Report of Invention, Oak Ridge National Laboratory, October 2003.

[6]C. F. Babbs and E. Jung, "Valveless Pumping: the Reflected Wave Hypothesis," submitted to *American Journal of Physics*.

Automated Image Analysis for Functional Genomics

J. R. Price,[1] Y. Liu,[2] D. Aykac,[1] K. Chourey,[2] and S. S. Gleason[1]
[1]*Engineering, Science, and Technology Division*
[2]*Life Sciences Division*

The primary goals of this project were (1) to construct a database of mitotic spindle imagery from wild-type and knockout (mutant) mouse cells and (2) to develop proof-of-concept, automated image analysis tools to characterize and compare the observed structures. The mitotic spindle is a crucial structure in cell-cycle progression, serving as the machinery for chromosome segregation. Furthermore, its tubulin-based structure is the target of many anti-cancer drugs. In this project, we collected fluorescent confocal microscopy images of mitotic spindles from wild-type, mutant, and drug-treated fixed cells. Quantitative image-based features and software to compute these features were developed. The software was then applied to the imagery data set, and the acquired sets of quantitative features were compared statistically. The results demonstrate that the spindle structure can indeed be characterized by the features computed via automated image analysis and, furthermore, that these features can vary statistically between the different cell types studied. These results imply that the proposed methodology may prove significantly beneficial in the search for new anti-cancer drug targets and in determining the roles of various proteins in cell-cycle progression, spindle checkpoint, and/or chromosome segregation.

Introduction

Data related to functional genomics research is often captured in image form by a variety of instruments such as the confocal microscope, the atomic force microscope, and/or the deconvolution microscope. Currently, much of the analysis of this data is performed manually—a well-trained person examines the image and attempts to make qualitative observations concerning the contents. In order to conduct statistically meaningful experiments, it is necessary that many, many images be produced and analyzed quantitatively. It is, however, prohibitively time-consuming for a human to manually analyze a large number of images and virtually impossible to do so quantitatively. To advance functional genomics understanding, automated image analysis will be required in a variety of experimental endeavors.

Previously, Yie Liu (co-principal investigator) found that VPARP (a protein complex) interacts with alpha, beta, and gamma tubulins in vivo (unpublished data). The interaction with the spindle and the specific roles of VPARP, however, are poorly understood. It has also been reported[1] that a tumor suppressor gene known as p53 is a passenger of the centrosome and that the absence of p53 could lead to abnormal centrosome amplification. As alpha and beta tubulins are core components of the microtubules and the mitotic spindle, and since the centrosome plays such a crucial role in the spindle formation and function, the aim of this project was to determine if quantifiable differences between wild-type, VPARP-knockout, and p53-knockout spindles could be detected and quantified. Furthermore, the image processing tools developed to analyze the images from these experiments can be used for similar analyses of other spindle-associated proteins which may prove to be potential targets for future cancer treatment strategies.

Technical Approach

In this project, five different cell types were established for study. These cell types, along with the corresponding shorthand names we will use for them in the remainder of this report, were as follows:

1. Primary wild-type (WT),
2. Primary VPARP-knockout (VPARP-KO),
3. p53-knockout (p53-KO),
4. p53-KO, treated with Taxol for 24 hours at 0.01 uM concentration (p53-KO-TaxLo),
5. p53-KO, treated with Taxol for 24 hours at 0.1 uM concentration (p53-KO-TaxHi).

These cells were fixed and then probed with fluorescent-labeled, anti-tubulin antibodies that attach to the spindles of metaphase cells. The spindles were then imaged via confocal microscopy. A three-dimensional (3D) stack of images was acquired for each spindle, where each image represents one "slice" of the spindle volume. Each element in such a 3D array will be referred to as a voxel, and the value of the voxel represents the fluorescence intensity and, hence, the amount of tubulin, at the

corresponding location. Forty different spindles were imaged, which included eight different spindles for each of the five cell types in the list above. One example of a spindle image is shown in Fig. 1. Note that, for visualization purposes, all layers in Fig. 1 have been summed in the z-direction to make a single two-dimensional (2D) image from the three-dimensional (3D) stack. Brighter areas indicate regions of higher tubulin concentration.

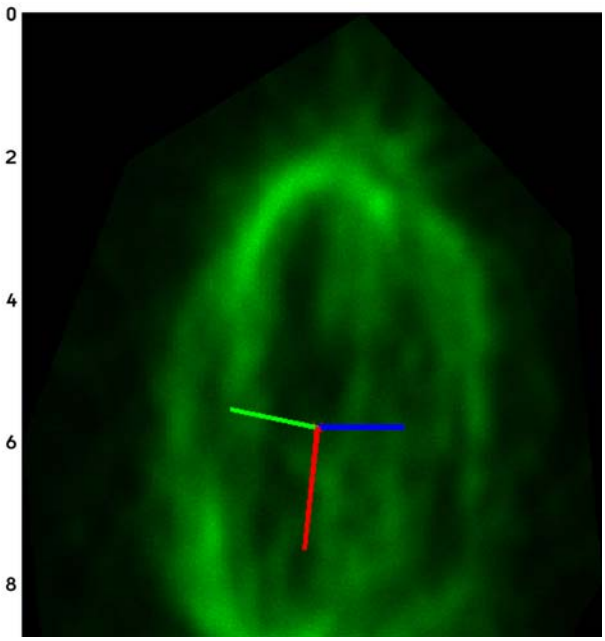


Fig. 1. Two-dimensional rendering of 3D spindle stack. The blue, green, and red lines indicate the first, second, and third principal axes, respectively. Units on both axes are in microns.

Each of the image stacks was then processed to produce a set of 49 descriptive features. The first eight features are independent of the spindle shape or geometry and depend only upon the voxel values. The remaining 40 features are morphology based and are computed with respect to an intrinsic coordinate system computed for each spindle. This intrinsic coordinate system has its origin at the spindle center of mass and axes aligned with the principal axes of inertia.[2] The center of mass and principal axes of inertia (and other mass-based features) are found by treating each voxel as a point mass, with mass equal to the voxel value, at the voxel location in the volume. Principal axes of inertia (projected onto the 2D viewing plane) are shown in the example spindle in Fig. 1. Below are brief descriptions of the 49 computed features.

1. Spindle mean. Average voxel value for the entire 3D stack.
2. Spindle standard deviation. Standard deviation of all voxel values in the stack.
3. Histogram skewness. Third central moment divided by the cube of the standard deviation. Gives a measure of the symmetry of the curve.
4. Histogram kurtosis. Fourth central moment divided by the fourth power of the standard deviation Gives a measure of how peaked the curve is.

5–8. Principal component coefficients of the histogram. We performed principal component analysis[3] on the 256-point histograms from the 40 spindle stacks. The coefficients of the histograms projected onto the first four principal components are kept for comparison.

9–12. Percentiles of mass projected onto first principal axis. These numbers are simply the intervals, in microns, along first principal axis which enclose 25%, 50%, 75%, and 95% of the total spindle mass.

13–16. Principal component coefficients of mass projected onto first principal axis. We project the spindle mass onto the first principal axis. Each point of the resulting curve, an example of which can be seen in see Fig. 2(a), represents the fraction of total spindle mass in the interval about that point. We use 200 intervals over the range of –10 to +10 μm about the center of mass.

17. Skewness of mass projected onto first principal axis.
18. Kurtosis of mass projected onto first principal axis.

19–22. Percentiles of mass projected onto second principal axis.

23–26. Principal component coefficients of mass projected onto second principal axis. An example curve of the mass projected on the second principal axis is shown in Fig. 2(b).

27. Skewness of mass projected onto second principal axis.
28. Kurtosis of mass projected onto second principal axis.

29–32. Percentiles of mass projected onto third principal axis.

33–36. Principal component coefficients of mass projected onto third principal axis. An example curve of the mass projected onto the third principal axis is shown in Fig. 2(c). Note the bimodal nature of this curve. This is due to the fact that the third principal axis—which is the axis about which the spindle has lowest moment of inertia—lies along the pole-to-pole axis of the spindle. The two peaks are related to the increased tubulin concentration at the spindle poles.

37. Kurtosis of mass projected onto third principal axis.

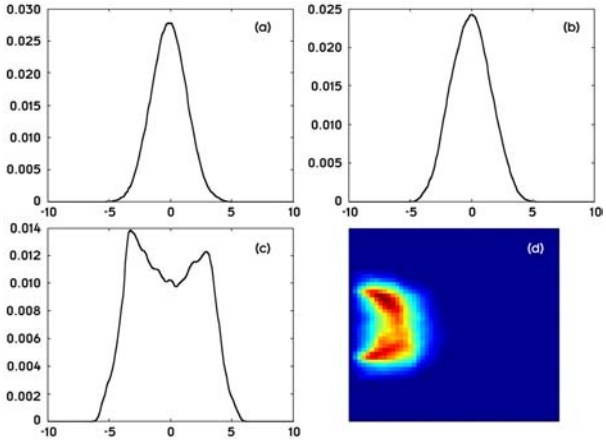


Fig. 2. Examples of projected mass measurements for spindle of Fig. 1. Horizontal axes for (a)–(c) are in microns; vertical axes represent the fraction of total mass. (a) Mass projected onto first principal axis. (b) Mass projected onto second principal axis. (c) Mass projected onto third principal axis. (d) Mass projected onto planes parallel to first and second principal axes at intervals along third principal axis. The vertical axis represents location along the third principal axis in microns, where zero is the center of mass and the horizontal axis represents radial distance, in microns, from the third principal axis. The color scale represents fraction of total mass.

38. Skewness of left half of the mass projected onto the third principal axis.
39. Kurtosis of left half of the mass projected onto the third principal axis.
40. Skewness of right half of the mass projected onto the third principal axis.
41. Kurtosis of right half of the mass projected onto the third principal axis.
42–5. Radial mass percentiles of mass projected onto plane of first and second principal axes. These are radial intervals (in microns) in the plane of the first and second principal axes which contain 25%, 50%, 75%, and 95% of the total spindle mass, respectively.
46–49. Principal component coefficients of mass projected onto planes parallel to the plane of the first and second principal axes. We select 50 equally spaced intervals along the third principal axis ranging from –10 to +10 μm from the center of mass. In each of these intervals, we compute the fraction of total mass in 50 equally spaced radial intervals from 0 to +10 μm from the third principal axis. The result is a 50 × 50 2D array that is a function of radius (in the plane) and distance from center of mass along the third principal axis. An example of this projected mass is shown as an image in Fig. 2(d). Note the two peaks which again correspond to the spindle poles. We perform principal components analysis on this on the 2500-point vectors obtain by raster scanning the 50 × 50 array and keep only the first four coefficients.

Software was developed in the MATLAB environment to automatically extract all of the above features from the image stacks.

Results and Accomplishments

The 49 features described above were computed for each of the 40 spindles that were imaged. The values of these features were then compared for different pairs of data sets using one-way analysis of variance (ANOVA). The p-values from these ANOVA tests are given in Table 1.

Analyzing the first data column of Table 1, we note that for the WT-vs-p53-KO comparison there are 27 features that are different at $p < 0.03$. As we would expect to see only about one or two features different at $p < 0.03$ (since $0.03 \times 49 = 1.47$), this is a clear indication that the WT and p53-KO spindles are significantly different with respect to the computed features. In the second data column, we compare p53-KO to p53-KO-TaxLo. Here we see that there are 13 features that vary at $p < 0.03$. Again, the features indicate a quantifiable difference in the spindles. In the third data column of Table 1, we compare the p53-KO-TaxLo to p53-KO-TaxHi. Only one feature is different at $p < 0.03$, and since we might expect one or two features at $p < 0.03$ purely by random chance, this indicates that there are essentially no quantifiable differences between these two sets with respect to the features computed. This indicates that both the high and low concentrations of Taxol induce quite similar effects in the spindles. In the final data column, we compare WT to the VPARP-KO cells. In this case, we find five features that differ at $p < 0.03$. This evidence indicates that the wild-type and VPARP-knockout spindles do indeed exhibit differences, but there are fewer such differences in this case than WT-vs-p53-KO case.

Summary and Conclusions

We recall, as stated in the abstract, that the two primary objectives of this project were (1) to construct a database of mitotic spindle imagery from wild-type and knockout (mutant) mouse cells and (2) to develop proof-of-concept, automated image analysis tools to characterize and compare the observed structures. As evident from the previous sections, we have successfully achieved both of these primary objectives. Furthermore, the results of the analysis have demonstrated that both p53-KO and VPARP-KO spindles exhibit quantifiable differences from WT spindles. These results were obtained via software and algorithms developed to perform automated analysis of 3D confocal microscopy images. Since the mitotic spindle is a crucial structure in cell-cycle progression, and since

Table 1. Collection of p-values from one-way ANOVA
Features different at p <0.03 are shaded

Feature	WT vs. p53-KO	p53-KO vs. p53-KO-TaxLo	p53-KO-TaxLo vs. p53-KO-TaxHi	WT vs. VPARP-KO
1	0.001100	0.216930	0.742420	0.000967
2	0.016648	0.125270	0.497540	0.001426
3	0.000008	0.167500	0.171880	0.172870
4	0.000102	0.122130	0.183460	0.672430
5	0.020666	0.000016	0.023851	0.599550
6	0.014652	0.499240	0.968500	0.001207
7	0.830570	0.554860	0.092413	0.231380
8	0.830030	0.540580	0.902870	0.295980
9	0.015389	1.000000	0.794190	0.790010
10	0.002535	0.900400	1.000000	0.798600
11	0.006921	0.769620	0.953030	0.435070
12	0.020876	0.613750	1.000000	0.099915
13	0.005776	0.468370	0.896140	0.456690
14	0.011732	0.261670	0.908000	0.118680
15	0.227180	0.825320	0.378690	0.513280
16	0.263140	0.439310	0.102700	0.149360
17	0.057692	0.060344	0.423860	0.778920
18	0.508910	0.077402	0.912250	0.011427
19	0.485960	0.644610	0.770330	0.316980
20	0.730720	0.415440	0.174470	0.350070
21	0.249250	0.748570	0.289560	0.629520
22	0.009242	0.147420	0.836600	0.950480
23	0.801310	0.230810	0.197920	0.475720
24	0.000016	0.094417	0.748320	0.316810
25	0.635920	0.813610	0.964130	0.605840
26	0.865520	0.705520	0.407840	0.939310
27	0.339630	0.577150	0.430680	0.945780
28	0.000715	0.000318	0.098707	0.020092
29	0.000040	0.007702	0.391660	0.781440
30	0.000008	0.003143	0.243640	0.825680
31	0.000014	0.003072	0.519390	0.866930
32	0.000159	0.002471	0.859190	0.798970
33	0.000001	0.002628	0.421900	0.885090
34	0.870750	0.021608	0.724500	0.945770
35	0.084633	0.028553	0.571950	0.988380
36	0.271430	0.052767	0.521250	0.149250
37	0.000370	0.663360	0.218600	0.089845
38	0.002539	0.045054	0.171260	0.134810
39	0.001925	0.061579	0.144530	0.128000
40	0.000015	0.000283	0.971010	0.871750
41	0.000012	0.000322	0.981600	0.864620
42	0.071363	0.549460	0.659140	0.948180
43	0.287360	0.852210	0.607640	0.790310
44	0.938410	0.906100	0.621420	0.677790
45	0.163670	0.131070	0.824360	0.348410
46	0.000017	0.009114	0.310020	0.962760
47	0.008446	0.418190	0.682500	0.679810
48	0.146150	0.016600	0.530830	0.969500
49	0.641360	0.206300	0.044541	0.707530

there are multiple proteins that interact with the spindle in unknown ways, the developed software and algorithms will be broadly applicable in investigating the effects of other associated proteins on spindle morphology. Understanding the interactions of such proteins with the spindle may uncover future anti-cancer drug targets.

This project has demonstrated proof-of-concept for the application of automated image analysis to enable high-throughput analysis of mitotic spindle imagery. The next phase of this work is to apply the developed techniques to the analysis of living cells. The mitotic spindle is a dynamic structure that we have currently only examined in a fixed state. It is likely that more—and perhaps quantitatively more pronounced—differences between wild-type, knockout, and/or drug-treated cells will be observed as we examine the spindle dynamics through its formation and function in living cells. The analysis spindles in living cells will be the focus of follow-on efforts we are currently pursuing with the National Institutes of Health. Specifically, we are currently preparing a multi-year proposal to be submitted to the "Innovations in Biomedical Computational Science and Technology" program (PAR-03-106) in February 2004. We believe that the successful preliminary results obtained and the proof-of-concept established with this project will enable us to make a strong proposal.

References

[1]K. Fukasawa, T. Choi, R. Huriyama, S. Rulong, and G. F. Vande Woude, "Abnormal Centrosome Amplification in the Absence of p53," *Science* **271**, 1744 (1996).

[2]J. Foote and A. Raman, "A relation between the principal axes of inertia and ligand binding," in *Proc. of the National Academy of Sciences* **97**, 978 (2000).

[3]K. Fukunaga, *Statistical Pattern Recognition*, 2nd Ed., Morgan Kaufmann, 1990.

Ontology-Based Three-Dimensional Modeling for Human Anatomy

L. C. Pouchard and S. P. Dickson
Computer Science and Mathematics

The project focused on the problem of seamlessly presenting a human anatomical three-dimensional (3D) model with other types of systemic information ranging from physiological to molecular information while navigating the anatomical model. We proposed solutions that include a visual 3D interface with navigation features that integrate structured information display and the adaptation of an ontology of medical information where several types of indicators are formally and semantically linked. The anatomy serves as a visual entry point to medical indicators and other information about a patient. The ontology of medical information provides labeling to the highlighted anatomy parts in the 3D display. Because of the formal structure and links between anatomical concepts found in the ontology, the navigation is greatly enhanced: navigation based on both 3D visual images and text and formal links representing anatomical concepts within the model becomes possible.

Introduction

As the use of computer modeling spreads in medical research, smarter models of the human anatomy are needed in a broad range of applications, including clinical diagnosis, surgical planning,[1] cancer treatment using radiation therapy,[2] and other applications. In silico models constructed in computerized simulation are used to determine the effects of parameter variations and compare between actual clinical data and simulation results. Static models of human anatomy in two-dimensional (2D) slices from 3D geometric coordinates are now available to clinicians and researchers for visualization, diagnosis, measurements and other purposes. From scans such as X-ray Computed Tomography (CT) or magnetic resonance imagery (MRI) volumetric reconstructions of the human body are created. However, the volumetric reconstructions have not been used for 3D animations for scientific and medical purposes, in part to due to computational complexity and resource requirements.

In addition, researchers, clinicians, and practitioners who analyze computerized segments of the anatomy do not have the ability to integrate different types of medical data. There is also a lack of easy-to-use capabilities to facilitate analysis and increase accuracy. At the time of data analysis and visualization, scientists must rely on their own mental models to integrate into a coherent whole information provided by different systems and represented in several media: multiresolution imaging, property measurements (physiological, mechanical, etc.), standardized annotations of anatomy, characterized links between parts of the anatomy, sensor data, and others.

Faced with an explosion of parameters to consider when making decisions, researchers and practitioners need new computerized environments where various media types and tools are seamlessly integrated and easily manipulated at the point of service.[3] Tools for analysis and diagnosis guidance may be critical for the quick assessment of an emergency situation. This is particularly true when split-second assessment of an injury is required, for instance by paramedics in the field of battle.[4]

Based on this research, a proposal to the DARPA Virtual Soldier Program (VSP) Phase I[5] was submitted and selected to participate under sponsorship of the United States Army Medical Research and Materiel Command (USAMRMC) (DOE Proposal Number 1868-HH48-X1). Whereas results described below were innovative and proved highly applicable to the work in the VSP, the scope of the analysis was restricted by effort and resources.

Technical Approach

Animated human models have been developed in the cinematic industry with an emphasis on good appearance rather than accuracy of representation. The level of detail and accuracy of 3D models appropriate for the needs of scientific representation has been one of two research directions investigated in this project. It included accurately and precisely identifying anatomical parts with 3D volumetric coordinates at a level of precision not achieved previously, visualizing parts of the anatomy with this new level, and easy-to-use adding navigation capabilities. The foot was chosen for visualization as an example of an anatomical part for its concentration of interacting small muscles and bones, and variety of tissue (Fig. 1). The chosen development environment for 3D animation is Maya™,[6] a highly flexible environment with solid community support and numerous extensions.

Fig. 1. Foot volumetric data.

A multilevel information infrastructure was also investigated. This infrastructure permits associating each modeled part of the anatomy and its corresponding geometric coordinates to a web of textual information and properties contained in an ontology. An ontology is a declarative, symbolic model that formalizes (anatomical) concepts, terminology, and abstract relationships between these concepts and axioms or logical rules on how relationships inform concepts. All items in an ontology are called frames. An ontology is best imagined as a web of information structured with links and semantic definitions. The information may include text, parts of speech, concepts, images, or mathematical coordinates that are linked with each other by logical relationships described in rules. It is by design human and machine-readable, nonexhaustive, and able to accommodate new types of information and concepts thanks to its formal specifications when discoveries are being made. Some purposes of ontologies include creating standards on a particular topic, declare machine-readable specifications, and re-use of components. The Digital Anatomist Foundational Model (FMA) is the domain ontology we used to accommodate the needs of the future sponsor.

The FMA is developed at the University of Washington and contains approximately 100,000 terms representing 70,000 concepts, 120 logical relationships iterated over a million times between the terms.[7] As an example, in the FMA, tissues are related to each other through abstract relationships such as "has boundary," "is a tributary of," "is contiguous to," and many others. The FMA is based on the *Terminologia Anatomica*. In addition, it also includes physiological and some cell-related information that is characterized with relationships to anatomical entities.

The approach also included search and retrieval mechanisms that were both appropriate to the visualization environment and would return accurate and meaningful answers for the scientist. As a simple mouse-over within the visualization by the user translates into a computerized query to the FMA, the depth of ontological navigation was taken into account and set. Query and return formats to the visual interface were also investigated because the format of a query and its display affect the number of nodes traversed in the ontology.

Results and Accomplishments

The results found here are based on heuristics and the researchers' experience, and do not include metrics. They are still valuable because they guide the choice of computer architecture for integrating visual, textual, and logical information. This research proved necessary for developing a middleware architecture for the Virtual Soldier Program and of is expected to continue under the USARMC sponsorship.

Testing Search Mechanisms

The FMA is available through a web interface that uses the Foundational Model Explorer (FME).[8] A text string such as "metatarsal bone" in the query box and a check on the sub-class for the type of query returns all the information available on this anatomical part contained in the FME. Figure 2 shows the results of a query for the metatarsal bone as a subclass of bone in the FME. The relationships show that "metatarsal bone" is part of the foot, has itself three parts, has a boundary, an inherent 3D shape, and no synonym. (By contrast, blood as an anatomical part has a definition, synonyms, a boundary, but no 3D shape). Figure 3 shows the web of concepts that constitute the results of a query for the talus as part of the foot. But the heel of the foot is a subclass of foot with a synonym of heel region and calcaneal region. This

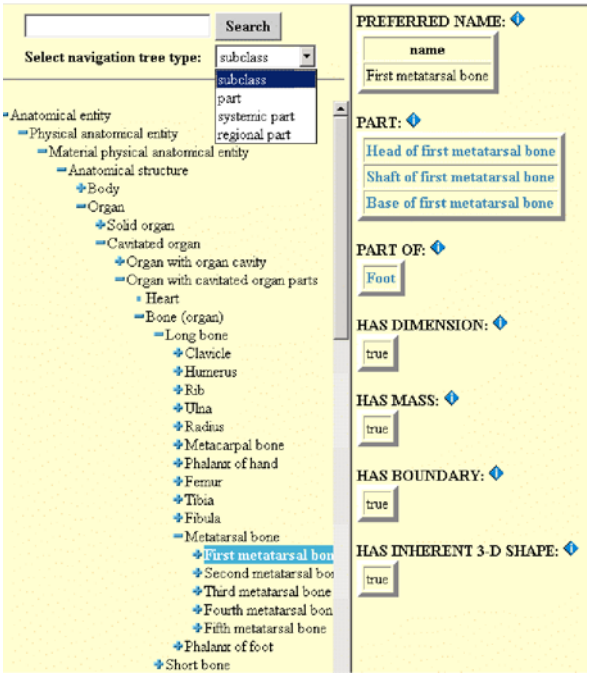


Fig. 2. First metatarsal bone in the FMA.

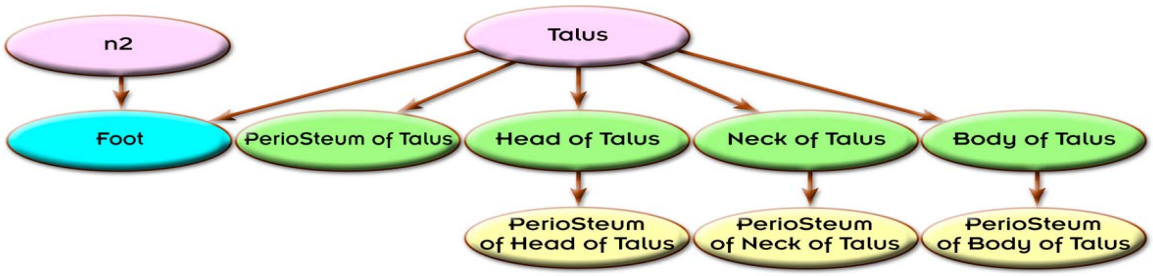


Fig. 3. Some nodes and edges in the talus as part of the foot.

example illustrates the complexity of relationships in an anatomical description of what is commonly thought of as a part of the body. The FMA is mostly used for educational purposes at present, but our part of research illustrates other usages such as use in diagnosis when it is integrated to a visual environment.

The choice of a computer architecture that would enable searches to the ontology by mouse-over 3D modeling (i.e., the integration between the ontology and the visualization environment) was guided by search mechanisms. Searches in an ontology are more complex than searches in a database because ontologies allow more general query formulations and more precise definitions at the same time (and thus are more flexible). The strength of an ontology comes from the possibility to do searches using rules rather than string matches as in a relational or object-oriented database. As exemplified above for the FMA, a search on a string can be specified as "string as part of" or "string as subclass" or "string as relationship" where relationship is any of the relationships contained in the FMA. Two caveats to point out are the complexity of queries to construct and the necessity to know the relationships included in the ontology of choice. However, this last caveat is also valid for any database search and well documented in the literature of information retrieval.

A search in an ontology can represent a large number of queries to a database. Query language and formats for the returned answers were considered as they dictated the choice of architecture for integrating with visualization and accuracy of representation. We tested the Ontology Querying Agent for the Foundational Model of Anatomy (OQAFMA) query language developed with the FMA,[9] and the Protégé query mechanisms[10] for suitability to a visualization environment.

Analysis and Discussion of Search Mechanisms

OQAFMA is based on STRUQL, a modified version of SQL to include rule-based searches. For a given query, OQAFMA traverses the entire tree of the FMA, and returned results are guaranteed to bring closure to a search. This means that nodes satisfying the criteria are returned, not only at the first level of the FMA tree where strings are found, but also at all branches of the tree down to the last leaf satisfying the criteria. This is called bringing closure in a search. By contrast, a search using Protégé search mechanisms does not bring closure as it returns only the first level of nodes in the tree. To obtain closure, parsing each return to a search and sending new searches based on relationships found between returned frames until a search returns no result is necessary with Protégé. While OQAFMA appears a more economical and elegant solution, its returns are formatted as an XML document. Because XML cannot easily represent a tree hierarchy and a web of relationships and rules for combining frames (due to its ancestry), OQAFMA results are not suitable for use in a visualization environment without further adaptation. In addition an OQAFMA query must be constructed in the Knowledge Interchange Language.[11]

Example: Search for talus using OQAFMA.
Query:
Where
X->":NAME"->"talus",
X->"part"+->Y,
Y->":NAME"->Parts
CREATE
TheTalus(Parts)

A good knowledge of KIF is required to formulate the query corresponding to a search in OQAFMA. Due to the million + iterations (i.e., meaningful combinations) possible with the frames in the FMA, searches must be reduced by the developers of the tool. A good rule of thumb is to reduce the number of queries according the level of detail in the visualization.

Results:

```
<results>
  <Talus>
    <Parts>Head of talus</Parts>
  </Talus>
  <Talus>
    <Parts>Neck of talus</Parts>
  </Talus>
  <Talus>
    <Parts>Body of talus</Parts>
  </Talus>
  .....
```

```
<Talus>
  <Parts>Periosteum of head of talus</Parts>
 </Talus>
 <Talus>
  <Parts>Periosteum of neck of talus</Parts>
 </Talus>
 <Talus>
  <Parts>Periosteum of body of talus</Parts>
 </Talus>
…..
<results>
```

As is clearly visible here when contrasted with Fig. 3, Periosteum of head, neck, and body of talus are treated as equal elements in an XML representation, and the tree hierarchy exemplified in Fig. 3 has disappeared.

Visualization

Architecture: The system is prototyped as a collection of Maya plugins.

DH_selectEntity is declared in Maya as a global procedure which is called when the user selects an entity from the Digital Human customized Marking Menu. DH_selectEntity maps from the Foundational Model of Anatomy (FMA/Protege) entity name to the name of the corresponding geometry node in the Maya scene graph (DAG) then selects the appropriate geometry node for focus.

Global procedure DH_selectionChange is a custom Maya Embedded Language (MEL) command executed upon a change of the selection list. DH_selectionChange maps from the selected geometrical entity in the Maya Scene Graph (DAG) to the corresponding ontological name for the node. DH_selectionChange builds and sends the query to the Foundational Model of Anatomy (Protégé server) and collects the results. DH_selectionChange then edits the marking menu to reflect the relational information relevant to the current focus context.

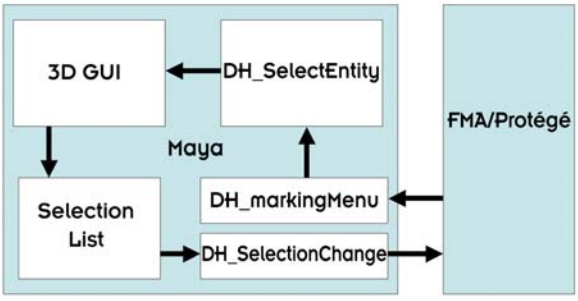


Fig. 4. Maya plugin architecture.

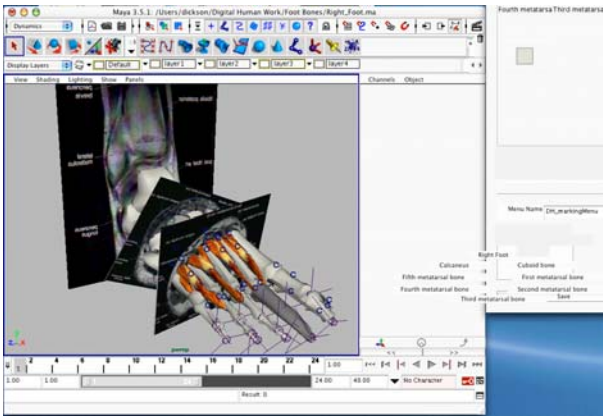


Fig. 5. Knowledge-guided 3-D user interface.

Summary and Conclusions

The FMA and its various access interfaces was obtained from the University of Washington (UW) and resides on an ORNL server. Collaboration with UW has been developed for maintenance at ORNL, and for purposes as needed in the VSP. At this time, ORNL is not expected to become a mirror site. Access to the FMA data and ORNL visualization architecture from team members of the VSP is already being planned within the framework of DOE proposal 1868-HH48-X1. Search mechanisms have been tested for seamless access by computer applications and re-usable Maya and Protégé API modules have been developed. Searches have been integrated to a 3D visualization environment for navigation.

Further research directions have been open for access to ontology architectures, ontology searching languages, and visualization that will be explored in the VSP. In particular, the need for metrics for ontology searches and level of detail in visualization have been highlighted.

References

[1]J. D. Westwood et al., editors, *Medicine Meets Virtual Reality 2000*, IOS Press, Amsterdam (2000).

[2]F. M. Kahn, *The Physics of Radiation Therapy*, 2nd Ed., Williams & Wilkins, Baltimore (1994).

[3]R. M. Satava, "Cybersurgeon: Advanced simulation technologies for surgical education," *Medical Simulation and Training* **1**, 6–9 (1996).

[4]R. M. Satava, "Medical Virtual Reality,"pp. 100–105 in *Medicine Meets Virtual Reality: 4 Health Care in the Information Age – Future Tools for Transforming Medicine,* IOS Press, Amsterdam.

[5]The Virtual Soldier Program. http://www.virtualsoldier.net.

[6]The Maya Family. ©2000–2003 Alias Systems, a division of Silicon Graphics Limited. All rights reserved.

[7]C. Rosse, L. G. Shapiro, and J. F. Brinkley, "The Digital Anatomist Foundation Model: Principles for Defining and Structuring its Concept Domain," *J Am Med Inform Assoc*. American Medical Informatics Association (AMIA) 1998 Symp. Suppl. 1998:820–824. http://sig.biostr.washington.edu/projects/fm/index.html.

[8]FME, http://sig.biostr.washington.edu/projects/fm/FME/index.html.

[9]OQAFMA. Forthcoming paper by Peter Mork. http://quad.biostr.washington.edu/noqafma/index.html.

[10]Protégé 2000, http://protege.stanford.edu/.

[11]M. Genesereth and R. E. Fikes, *Knowledge Interchange Format*, Version 3.0. Reference Manual, Technical Report Logic-92-1. Computer Science Department, Stanford University, Stanford, California, January 1992.

The Global Optimization Problem for Remote Sensing: A Guaranteed, Efficient Solution

J. C. Wells, V. A. Protopopescu, and C. D'Helon
Computer Science and Mathematics Division

We establish proof of principle that a broad class of continuous global-optimization problems (GOPs) can be solved efficiently, with a guarantee, by exploiting available additional information in the problem. We demonstrate that our proposed algorithm is considerably more efficient than current algorithms and apply our methodology to a remote sensing problem in the atmosphere, namely, monocular passive ranging (MPR). The expected deliverable will be a robust, efficient, and validated software tool, which will be useful in solving GOPs efficiently, where the need for the optimal—as opposed to a suboptimal—solution is critical.

Objective

Independent of specific approach and algorithmic efficiency, current solutions for continuous GOPs,[1] share a common drawback in that none of them can guarantee that the global minimum will be found in a reasonable time.

We aim to develop a new algorithm that can efficiently solve the continuous GOP arising from monocular passive ranging[2] and guarantee the result. MPR determines the instantaneous range to a target relying on remote sensing in the atmosphere, using a single sensor. Our choice of application is motivated by yet unsolved challenges facing the Department of Defense Missile Defense Agency (MDA), instances of which require the optimal, as opposed to a suboptimal, solution of a GOP.

Our new algorithm guarantees to find the global minimum by systematically using additional information about the error function between the predicted and observed radiances in MPR. This information may include the size of the basin of attraction for the global minimum, the separation between the global minimum and the next lowest local minimum, the general shape of the error function, and the uncertainties inherent in the sensor measurements. Indeed, some degree of information is usually available in many applications, but present algorithms either do not or cannot use it. This new idea enables us to map the continuous GOP onto a discrete search, by eliminating all basins of attraction except that of the global minimum.

The performance of our approach can be assessed in terms of a set of benchmark objective functions to show that the new algorithm is more efficient than present algorithms, as well as in terms of how much additional information is sufficient to guarantee that the global minimum is found quickly and accurately, for the error function in MPR.

By the end of October, we successfully implemented our algorithm to solve continuous GOPs. An authoritative benchmark for calculating atmospheric radiation transmission, MODTRAN,[3] was used to generate the error function between the predicted and observed radiances in MPR. Multiple minima in the error function were observed for fluctuating ozone profiles, since the ozone density is non-monotonic as the altitude increases. The algorithm was also parallelized on the IBM Power 4 supercomputer within ORNL's Center for Computational Sciences, in order to address independent fluctuations in multiple ozone layers, and thus expand the dimensionality of the error function.

Varying degrees of additional information about the error function were assumed in order to investigate how much information is required to eliminate all basins of attraction except that of the global minimum. Our results to date (for up to four dimensions) indicate that the error function available in MPR is likely to contain sufficient additional information to guarantee that the global minimum is found in a reasonable time. We have investigated the time scaling of the new algorithm, as well as the probability of finding the global minimum, and found that information related to the size of the basin of attraction for the global minimum is of prime importance for an efficient guarantee.

We have compared the performance of our algorithm with that of the TRUST algorithm[4] for a set of seven standard benchmark functions. The results show that our approach has the ability to outperform TRUST, in terms of the number of function evaluations, at least for two-dimensional objective functions. As the number of dimensions grows, our algorithm requires an exponential increase in the number of function evaluations. We have yet to investigate the possibility that our algorithm can

outperform TRUST for high-dimensional problems, in the case that the global minimum has a relatively narrow basin of attraction. In any case, our approach guarantees that the global minimum will be found, something that TRUST cannot achieve in principle. Thus, in general, a completely conclusive comparison between the two algorithms is impossible.

Our algorithm is applicable to a broad class of continuous GOPs, arising in a wide range of applications that are relevant to strategic ORNL thrusts in national security, materials science, and computational biology. Many of the problems in these areas require precise, guaranteed identification of the global minimum and will benefit from this research.

There is also an immediate need and opportunity to apply our algorithm to the challenges facing national defense. Remote sensing in the atmosphere is currently a critical field of research within strategic and theater missile defense. Monocular passive ranging is a prime example of a problem that will benefit from our optimal GOP solution for accurate ranging, with a guarantee.

References

[1]R. Horst and H. Tuy, *Global Optimization*, 2ed., Springer-Verlag, Berlin, 1993.

[2]W. Jeffrey, J. S. Draper, and R. Gobel, "Monocular Passive Ranging," *IRIS Targets, Backgrounds and Discrimination Conference*, 1994.

[3]MODTRAN website: http://www-vsbm.plh.af.mil/soft/modtran.html.

[4]J. Barhen, V. Protopopescu, and D. Reister, "TRUST: A Deterministic Algorithm for Global Optimization," *Science* **276**, 1094 (1997).

BIOLOGICAL SCIENCES AND TECHNOLOGY

Director's R&D Fund

Elucidating Eukaryotic Gene Regulatory Networks

B. H. Jones,[1] J. R. Snoddy,[1] S. Kirov,[1] E. J. Michaud,[1] D. Schmoyer,[2] M. Leuze,[2]
L. C. Pouchard,[2] and D. K. Johnson[1]
[1]*Life Sciences Division*
[2]*Computer Science and Mathematics Division*

The availability of fully sequenced genomes from an increasing number of species creates an unprecedented opportunity to integrate experimental and computational approaches for elucidating gene regulatory networks (GRNs). For this project, we exploit ORNL's ability to analyze differential gene expression on a semigenome-wide scale using cDNA microarrays and to use large-scale computing to perform comparative genome sequence analysis among multiple genomes and implement algorithms to recognize functional elements in DNA. Our goal is to establish an ongoing ORNL program in GRN discovery.

Introduction

Coordinate regulation of gene expression provides the basis for activation and repression of specific cellular pathways and the adaptive response to changing cellular conditions. Although the coding regions of each gene involved in any given pathway are now/will soon be known as mouse and human genome sequence are annotated, the molecular mechanisms that mediate shared regulation and determine which subsets of genes interact in specific pathways are poorly understood. These molecular mechanisms, consisting of transcription factor proteins and the regions of DNA to which they bind, form the architecture of gene regulatory networks, which in turn underlie basic physiology and adaptive responses within an organism.

While GRNs are important in the cellular physiology of both unicellular and multicellular life, the GRNs in multicellular animals are likely to differ from and be more complex than those in unicellular microbes. In the Metazoa, the changes created by cascades of developmental GRNs during the life cycle of an organism help build up the complex body structure and behaviors seen within that multicellular organism.[1] The diverse and complex processes that are found in a human, a fruit fly, a nematode, or a mouse could be seen as largely the result of cascading changes of GRNs. Ultimately these networks work to transform one-dimensional genomic information into complex, multi-dimensional biological phenotypes. Because of their core importance to several aspects of biology, GRN discovery and characterization are a major goal of the DOE Genomes to Life Program. Other agencies, as well, are interested in investing more in specific areas of research that will arise from this project.

The conventional approach to identifying genes involved in specific cellular pathways and the regulatory elements that control expression of those genes has been largely experimental. Although eventually effective, the methods employed are laborious, tedious, and limited by a need for a priori knowledge of the genes and proteins involved in the pathway of interest. However, genomic sequencing has been paralleled by the development of high-throughput techniques such as microarrays that yield expression information for thousands of genes simultaneously. These techniques provide the identities of many, often hundreds, of genes that behave similarly under a specific condition(s), behavior that is due in part to the presence of common transcription factor binding sites in the noncoding regions of each gene. This "guilt-by-association" approach to identify sets of co-regulated genes continues to gain popularity as an entry point for computational studies into molecular mechanisms of co-regulation.[2] This wealth of both expression and DNA sequence data, coupled with the complexity arising from the fact that ~95% of the mammalian genome is non-coding, highlights the need for computational approaches to interpret emerging data and to make predictions that can be extended to the entire genome.

Technical Approach

Optimally exploiting the wealth of sequence information for this application requires development of new tools and integration of the diverse areas of computing and experimental biology. Accordingly, this proposal integrates expertise in functional genomics, computer sciences and computational biology resident at the laboratory. Our general scheme is as follows:

1. Identify genes with shared expression patterns from microarray experiments.
2. Develop and implement computational tools to analyze the regions of the genes likely to contain regulatory elements, using a local database.
3. Identify potential regulatory elements for subsets of the genes.
4. Use bioinformatics to collect functional and pathway information for each subset of genes.

Details of the experimental and computational approaches are outlined separately below:

Experimental Approach

Microarray data can be used to identify genes differentially expressed in a specific condition or mutant mouse model, to group genes according to shared regulatory patterns across multiple experimental conditions, and to classify unknown samples based on their expression profiles. We used microarrays as a high-throughput way to identify sets of genes with shared expression patterns under specific conditions. We chose the following two experimental models: primary mouse keratinocytes during the time course of differentiation and a unique ORNL mouse model of hair loss, near naked hairless (Hr^n). These mice carry a regulatory mutation in the hairless (*hr*) gene that manifests as a sparse (Hr^n/+) or nonexistent (Hr^n/Hr^n) coat of hair shortly after birth.[3] The *hr* gene encodes a putative transcription factor, making the phenotype of *hr* mutants an excellent starting point for the studies included herein. Collectively these two models allowed us to examine gene expression related to aspects of skin biology, an area of long-standing interest in the ORNL mouse program.

We printed our microarrays in-house using a collection of cDNAs enriched for genes involved in hair and skin development and physiology, cell cycle control, DNA repair, apoptosis, and oncogenesis. We used cDNA clones that were originally obtained from Research Genetics and were amplified and sequence verified at ORNL in an earlier project. Microarrays were fabricated from these clones according to standard protocols. Each array represented ~1000 unique mouse genes, and each gene was spotted in triplicate.

Bioinformatics Approach

The work described herein requires vast amounts of DNA sequence accession and manipulation. To facilitate this effort as well as to make it possible to easily access multiple types of information about genes, we decided to create a local version of the National Center for Biotechnology Information (NCBI) database and to expand our database to include other information about genes, such as gene ontology (GO) assignment. Such a database would allow querying and processing sets of genes in the high-throughput manner necessary for this project.

Computational Approach

A common core task in the elucidation of gene regulatory networks is to identify the transcription factor (TF) binding sites that are shared by co-regulated genes. Identification of TF binding sites is complicated by the fact that these motifs are often small (8–15 bp), and that a single TF protein can bind to variations in a consensus sequence for that protein.[4] In addition, many true binding sites exist in a single gene, making it difficult to visually compare all of the sites from collections of co-regulated genes. Furthermore, the task is compounded in eukaryotes, in which collections of binding sites referred to as modules or composite elements, rather than individual sites, act synergistically to affect enhancement or suppression of transcription. Therefore the task of identifying the sets of DNA binding sites that mediate gene response to even a single biochemical factor is daunting.

Two main strategies were employed for computational identification of TF binding sites. The first approach involves searching for known TF binding sites in DNA sequence of interest. This tactic is based on consensus binding-site sequences archived in a library of binding sites, such as TRANSFAC (http://transfac.gbf.de/TRANSFAC/). While known binding sites can sometimes be identified with a high degree of confidence, this approach is limited to sites present in the database, and those that have been very well annotated experimentally. An alternative approach, and the one we employed herein, is to develop algorithms that search for de novo binding sites. This method allows the identification of novel sites not yet described, as well as known sites.

Results and Accomplishments

Our experimental effort has led to the identification of sets of genes that we believe to be important in the etiology of hair loss in the ORNL mouse model Hr^n. We first used microarrays to identify a set of genes that were differentially expressed in skin of adult Hr^n mutant mice compared to their wild-type littermates. Further characterization of this list of genes identified many known to be important in the development and maintenance of the hair follicle. We followed these results by repeating the analyses in young animals (7 days of age) in which the lack of hair phenotype was just beginning to manifest, with the goal identifying the initial genes that respond to the hairless protein and are primary rather than secondary responders to the mutation. Many of the genes identified in adult skin were also differentially expressed in skin of young mutant animals, underscoring their potential role in development of the phenotype. Of particular interest is

that many of the genes we identified as differentially expressed reside in gene clusters. These clusters have been proposed to be regulated as a group with specific timing of expression for each gene related to its position within the cluster, much like has been demonstrated for the Hox gene clusters. This raises the possibility that the hairless protein acts as a regulatory switch for the clusters of genes that are required for normal hair follicle development. This possibility is being pursued and will likely become the focus of an upcoming proposal to NIH. Collectively, this set of genes with altered regulation in response to a mutation in *hr* should contain many genes that share mechanisms of co-regulation, making them an excellent input set for the computation efforts. These genes are currently being parsed through the analysis algorithms, and we expect that the results will be included in an upcoming manuscript.

To facilitate the extraction of DNA sequence and the processing and querying of sets of genes, we created a gene key database, GeneKeyDB, to relate our genes of interest to standard databases such as NCBI's RefSeq and LocusLink (see Fig. 1). The current GeneKeyDB is largely a relational representation of the LocusLink data with some additional tables to identify the genomic locations of the genes and interactions with other databases at ORNL. Unlike the NCBI organization of this data, the current database allows for processing sets of genes by computerized analysis, and not just inferencing by

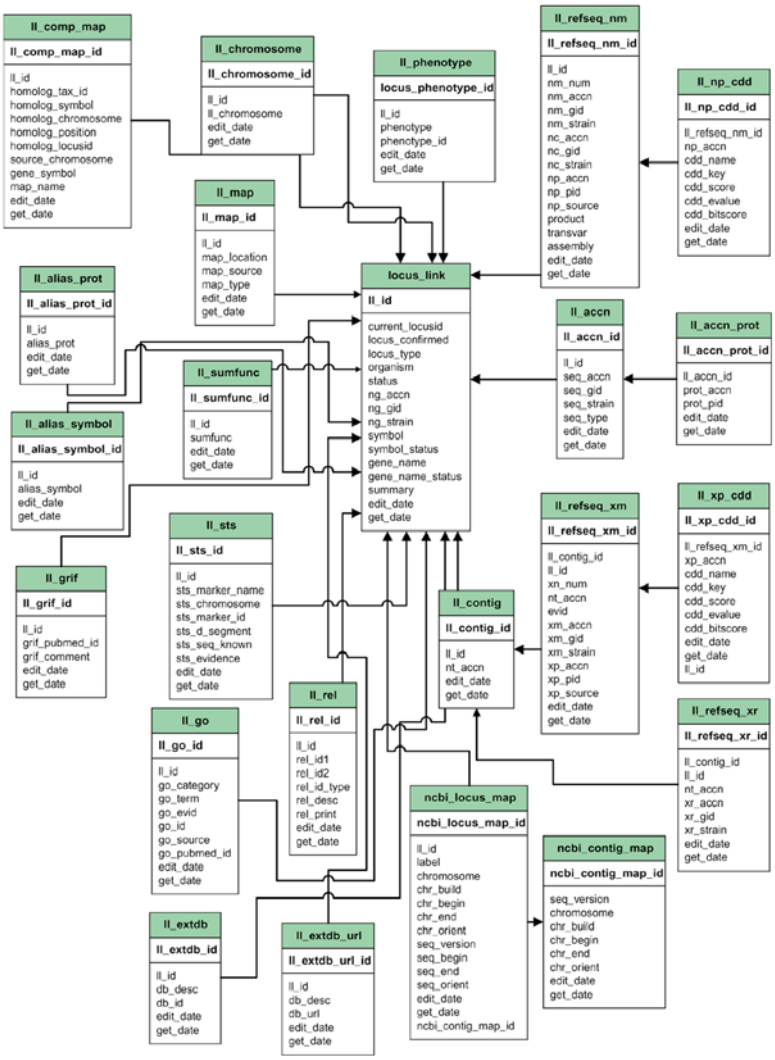


Fig. 1. Schema of the GeneKeyDB database created as a repository for data from NCBI and other sources and as a source for processing sets of genes with high-throughput.

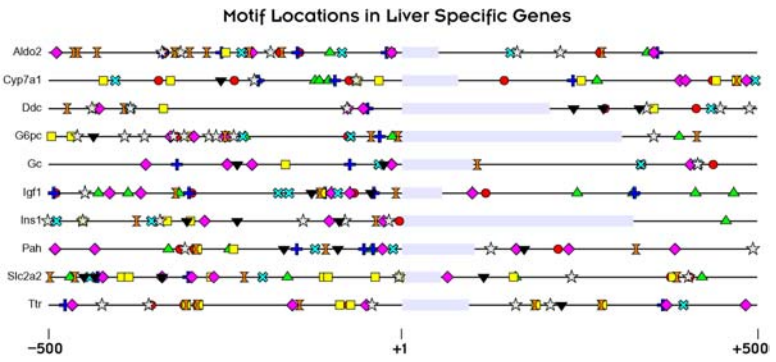


Fig. 2. Example output of the motif discovery toolkit applied to a set of genes co-regulated in liver, as described by Krivian and Wasserman.[5]

analyzing one gene at a time by humans using web pages and web links. The current GeneKeyDB links to our expression array database (GIMS) and the ORNL gene annotation database (GGPIW).

The suite of novel computational tools that we have developed for the motif discovery toolkit constitute a new approach for identifying functional DNA in non-coding regions by sequence comparison. Our approach differs from other motif-finding methods in a number of important ways, which collectively make it unique:

- it is applicable to any collection of related genomic sequences,
- it does not depend on the global alignment of these sequences,
- it does not require a motif to be present in each sequence of the collection, and it is able to discover multiple copies of a motif in a single sequence,
- it can find multiple motifs in a single run, and
- it is a general computational method; it does not require specific biological information about the sequences being examined.

Our motif discovery toolkit contains tools for a large number of tasks, including filtering sequences, finding short inexact matches, combining short matches, constructing graphs to represent relationships between motifs, identifying interesting graph structures, producing position weight matrix and profile hidden Markov models of motifs, scoring sequences relative to a motif model, and searching for inexact inverted and direct palindromes. An example of the toolkit as applied to a set of genes co-regulated in liver[5] is shown in Fig. 2.

A significant feature of our motif discovery toolkit is its extensive use of graph algorithms. The integration into the motif discovery toolkit of a number of innovative graph algorithms, some of which execute three orders of magnitude faster than competitors, will allow consideration of problems much larger than any previously addressed. Additionally, although the motif discovery toolkit uses global search methods, it examines only those patterns actually found in the input sequences, rather than all 4^n possible patterns for an n-base sequence, thus greatly improving efficiency. As a result, the graphs we examine are generally smaller than would be the case if an edge between two patterns were weighted only by the number of matches. Our metric tends to connect only those patterns for which there is a biologically significant relationship. Collectively, these new ideas make the toolkit well suited to the task of motif discovery.

Summary and Conclusions

In summary, we have developed a set of novel computational tools that can be customized for a variety of applications within the general area of identifying DNA regulatory elements. We have also created a local database that allows for rapid retrieval and processing of sets of genes from genomic databases. Experimentally, we have identified an interesting set of genes that we believe to be disregulated as a direct consequence of a mutation in the transcription factor *hr*, and these genes serve as an input dataset for further analysis using our motif discovery toolkit. The results of our work have been submitted in manuscript form to The Eighth Annual Conference on Research in Computational Molecular Biology (RECOMB 2004). A manuscript focusing on the biology and molecular mechanisms of the Hr^n mutants is in preparation. The infrastructure, collaborations, and tools that we have developed during this project provided the basis for two proposals submitted in FY 2003 to DOE Genomes to Life and to the Joint DMS/NIGMS Initiative to Support Research Grants in the Area of Mathematical Biology, NSF 02-125. In addition, the resources developed in this project will serve as the basis for continuing collaborative efforts between the Life Sciences and Computer Science and Mathematics Divisions, as well as with the Department of Computer Science at the University of Tennessee, Knoxville.

References

[1]V. F. Hinman, A. T. Nguyen, R. A. Cameron, and E. H. Davidson, "Development of Gene Regulatory Network Architecture Across 500 Million Years of Echinoderm Evolution," Proceedings of the National Academy of Science, *PNAS* **100**, 13356 (2003).

[2]J. Q. Quackenbush, *Science* **302**, 240 (2003).

[3]K. F. Stelzner, *Journal of Heredity* **74**, 193–6 (1983).

[4]W. W. Wasserman and W. Krivian, *Naturwissenchaften* **90**, 156–66 (2003).

[5]W. Krivian and W. W. Wasserman, *Genome Research* **11**, 1559–66 (2001).

High-Throughput Analysis and Modeling of Protein Complexes

F. W. Larimer,[1] Y. Xu,[1] D. Xu,[1] P. F. Britt,[2] R. L. Hettich,[2] and G. Hurst[2]
[1]*Life Sciences Division*
[2]*Chemical Sciences Division*

We developed methods to accelerate the study protein complexes by covalently crosslinking the protein subunits of a complex, digesting the crosslinked complex enzymatically to produce crosslinked pairs of peptides, and subsequently identifying these crosslinked pairs of peptides using mass spectrometry. These crosslinked peptides thus define a potential interaction face of the subunits making up the complex. We have examined a biotin-labeled, heterobifunctional (lysine-reactive N-hydroxysuccinimide ester and non-selective photoactivatable groups) crosslinking reagent, sulfosuccinimidyl [2-6-(biotinamido)-2-(p-azidobenzamido)-hexanoamido]ethyl-1,3'-dithiopropionate (sulfo-SBED), to understand the chemistry, side reactions, affinity isolation, and mass spectrometric behavior of sulfo-SBED crosslinked peptides.

Introduction

A major goal this project was to develop techniques for studying protein complexes by covalently crosslinking the protein subunits of the complex, digesting the crosslinked complex enzymatically to produce crosslinked pairs of peptides, and to identify these crosslinked pairs of peptides using mass spectrometry. Because of generally low yields from crosslinking reactions, we proposed to use affinity-labeled crosslinking reagents that would allow selective extraction of crosslinked peptide pairs from the many non-crosslinked peptides that would result from the proposed scheme. A commercially available reagent, sulfo-SBED (Fig. 1), was chosen because it contains a biotin affinity group, which has a strong affinity for the protein avidin. Separation media derivatized with avidin are also commercially available.

Technical Approach

To understand the chemistry, side reactions, affinity isolation, and mass spectrometric behavior of sulfo-SBED crosslinked peptides, we reacted sulfo-SBED with a small peptide, neurotensin. Bovine hemoglobin was chosen as a model protein complex for developing crosslinking methodology because it is commercially available in reasonable purity and at low cost. Because of the complications that we encountered in identifying crosslinked peptide pairs from the sulfo-SBED crosslinking of bovine hemoglobin, it was not possible to study crosslinking of mixtures of protein complexes.

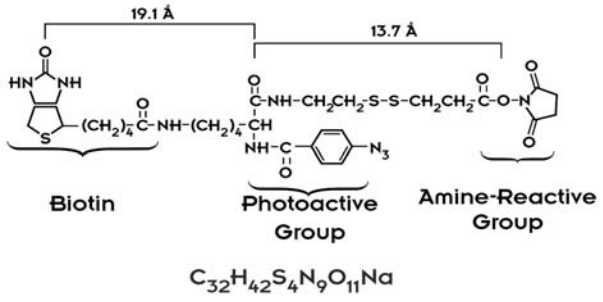


Fig. 1. Sulfo-SBED (Pierce).

Results and Accomplishments

Reaction of sulfo-SBED with a peptide model system. As described in a manuscript submitted to the *Journal of the American Society for Mass Spectrometry*, we studied the reaction of the crosslinking reagent sulfo-SBED with neurotensin, a small peptide.[1] This manuscript describes side reactions such as "one-sided" products between the crosslinking reagent and the peptide, and oxidation of the biotin sulfur in sulfo-SBED. A "double-blocking" protocol is also described. This protocol was found to reduce non-specific binding of non-crosslinked peptides to the avidin affinity chromatography medium used to isolate crosslinked peptides. We demonstrated the detection of neurotensin/sulfo-SBED reaction products in the presence of a background of non-biotinylated peptides.

Reaction of Sulfo-SBED with bovine hemoglobin. Using techniques similar to those described in the sulfo-SBED/neurotensin manuscript, we crosslinked bovine hemoglobin with sulfo-SBED. Hemoglobin is protein complex containing four protein subunits—A, B, C, and D. Subunits A and C are of identical amino acid sequence (the α chain), as are subunits B and D (the β chain). Each subunit contains a heme group, noncovalently bound

between two histidine residues. The spatial relationships among the four subunits in hemoglobin are known, and several three-dimensional structures for this protein are available in the Protein Data Bank.

Detection of Intact Sulfo-SBED Crosslinked Hemoglobin

To characterize initially the reaction between sulfo-SBED and bovine hemoglobin, we examined the behavior of the intact crosslinked protein prior to digestion by trypsin. Figure 2 shows the affinity isolation of the crosslinked complex, as well as a non-crosslinked control. In this experiment, sulfo-SBED crosslinked hemoglobin was fractionated using monomeric avidin column chromatography.[2] Unreacted hemoglobin is not retained by the column, and elutes in fractions 1–3. Sulfo-SBED crosslinked hemoglobin is eluted from the column in fractions 8–10 via addition of a low-pH buffer.

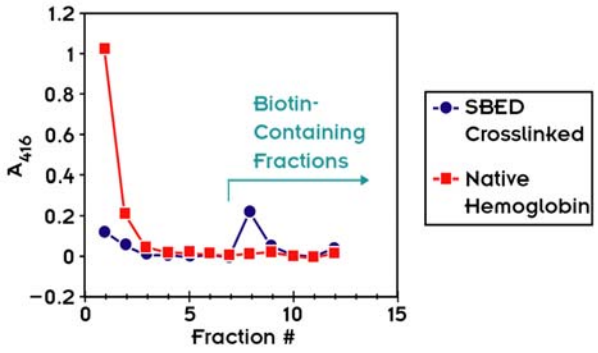


Fig. 2. Avidin fractionation of sulfo-SBED crosslinked hemoglobin.

Various fractions from the avidin separation of sulfo-SBED crosslinked and control hemoglobin were analyzed using matrix-assisted laser desorption time-of-flight mass spectrometry (MALDI-TOF-MS), as shown in Fig. 3. The sulfo-SBED-treated sample shows addition of the crosslinking reagent to both the α and β chains [Fig. 3(d)]. After affinity isolation, however, only the α chain and its sulfo-SBED reaction product appear [Fig. 3(f)]; little or no β chain is observed. These measurements indicate that the separation shown in Fig. 2 indeed yields crosslinked hemoglobin in fraction 8.

Significantly, it appears that the crosslinking reaction allows hemoglobin to retain sufficient tertiary structure to maintain the non-covalently bound heme groups. Evidence for this is that detection of hemoglobin in the fractions eluted from the avidin column was performed using optical absorbance at 416 nm, a characteristic absorption band associated with the heme group. A significant distortion of the hemoglobin structure by the crosslinking reaction could cause loss of the heme group.

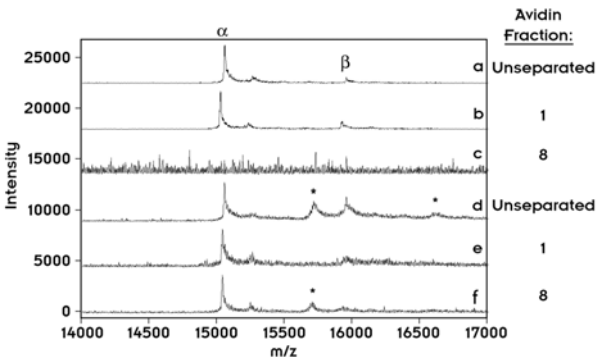


*Fig. 3. MALDI-TOF spectra of hemoglobin. (a,b,c) not crosslinked. (d,e,f) sulfo-SBED crosslinked. * = addition of sulfo-SBED.*

Detection of Crosslinked Peptides in Trypsin-Digested, Crosslinked Hemoglobin

Using conditions under which hemoglobin could be crosslinked using sulfo-SBED, and the procedures described in the sulfo-SBED/neurotensin manuscript, the next step was trypsin digestion of sulfo-SBED crosslinked hemoglobin, followed by avidin isolation of crosslinked peptide pairs, and mass spectrometric identification of these crosslinked peptide pairs. Despite the improvements described in the manuscript, it still proved difficult to detect and identify crosslinked peptide pairs. The probable reason for this is the complexity of the peptide mixture resulting from the digestion, and the low amounts of crosslinked peptides in this mixture. Initial yields of crosslinking reactions are typically 1–10% (or less). Furthermore, digestion results in an *increase* in the number of species present in the mixture, by a factor of roughly 10–50. While the avidin separation should mitigate this problem, there remains a considerable background of non-crosslinked peptides in mass spectra of avidin-purified digests.

Figure 4 shows a narrow range of the MALDI-TOF mass spectra of avidin-purified trypsin digests of a non-crosslinked bovine hemoglobin sample and a sulfo-SBED crosslinked hemoglobin sample. Although non-specific peptides appear in both spectra, one peak at mass-to-charge

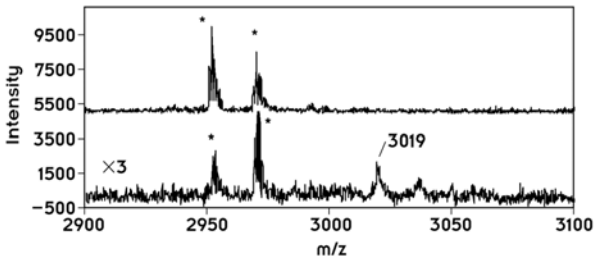


*Fig. 4. MALDI-TOF mass spectra of tryptic digests of (a) non-crosslinked hemoglobin, (b) sulfo-SBED crosslinked hemoglobin. * represents peptides binding non-specifically to avidin affinity medium.*

Table 1. Comparison of putative crosslinks from mass spectrometry with PDB model of hemoglobin 1HDA

Observed (m/z)	Calculated (m/z)	Peptide 1[a] (subunit) RESIDUES positions	Peptide 2[a] (subunit) RESIDUES positions	Distance range[c] (Å)
1924.7	1924.9	(B) MLTAEEK β1–7[b]	(A) GHGAK α57–61	32–37
	"	(D) MLTAEEK β1–7[b]	(A) GHGAK α57–61	41–45
	"	(B) MLTAEEK β1–7[b]	(C) GHGAK α57–61	42–46
	"	(D) MLTAEEK β1–7[b]	(C) GHGAK α57–61	31–36
3019.8	3020.5	(B) AHGKK β61–65	(D) VKAHGKKVLDSFSNGMK β59–75	45–57
"	"	(B) AHGKK β61–65	(D) VKAHGKKVLDSFSNGMK β59–75	54–59
"	"	(B) AHGKK β61–65	(D) VKAHGKKVLDSFSNGMK β59–75	52–57
"	"	(B) AHGKK β61–65	(D) VKAHGKKVLDSFSNGMK β59–75	51–56
"	"	(D) AHGKK β61–65	(B) VKAHGKKVLDSFSNGMK β59–75	45–57
"	"	(D) AHGKK β61–65	(B) VKAHGKKVLDSFSNGMK β59–75	54–59
"	"	(D) AHGKK β61–65	(B) VKAHGKKVLDSFSNGMK β59–75	52–57
"	"	(D) AHGKK β61–65	(B) VKAHGKKVLDSFSNGMK β59–75	51–56
"	3020.5	(B) VKAHGKK β59–65	(D) AHGKKVLDSFSNGMK β61–75	22–32
"	"	(B) VKAHGKK β59–65	(D) AHGKKVLDSFSNGMK β61–75	18–28
"	"	(B) VKAHGKK β59–65	(D) AHGKKVLDSFSNGMK β61–75	25–30
"	"	(B) VKAHGKK β59–65	(D) AHGKKVLDSFSNGMK β61–75	28–33
"	"	(D) VKAHGKK β59–65	(B) AHGKKVLDSFSNGMK β61–75	12–30
"	"	(D) VKAHGKK β59–65	(B) AHGKKVLDSFSNGMK β61–75	12–29
"	"	(D) VKAHGKK β59–65	(B) AHGKKVLDSFSNGMK β61–75	24–28
"	"	(D) VKAHGKK β59–65	(B) AHGKKVLDSFSNGMK β61–75	25–29
3524.0	3523.8	(A) VLSAADKGNVKAAWGK α1–16[b]	(B) LLVVYPWTQR β30–39	14–21
"	"	(A) VLSAADKGNVKAAWGK α1–16	(B) LLVVYPWTQR β30–39	15–26
"	"	(A) VLSAADKGNVKAAWGK α1–16	(B) LLVVYPWTQR β30–39	18–30
"	"	(A) VLSAADKGNVKAAWGK α1–16[b]	(D) LLVVYPWTQR β30–39	26–31
"	"	(A) VLSAADKGNVKAAWGK α1–16	(D) LLVVYPWTQR β30–39	32–38
"	"	(A) VLSAADKGNVKAAWGK α1–16	(D) LLVVYPWTQR β30–39	33–40
"	"	(C) VLSAADKGNVKAAWGK α1–16[b]	(B) LLVVYPWTQR β30–39	26–31
"	"	(C) VLSAADKGNVKAAWGK α1–16	(B) LLVVYPWTQR β30–39	32–38
"	"	(C) VLSAADKGNVKAAWGK α1–16	(B) LLVVYPWTQR β30–39	33–40
"	"	(C) VLSAADKGNVKAAWGK α1–16[b]	(D) LLVVYPWTQR β30–39	15–21
"	"	(C) VLSAADKGNVKAAWGK α1–16	(D) LLVVYPWTQR β30–39	15–26
"	"	(C) VLSAADKGNVKAAWGK α1–16	(D) LLVVYPWTQR β30–39	18–30
"	3523.9	(B) AAVTAFWGKVK β8–18	(B) LLGNVLVVVLARNFGK β104–119	9.5–23
"	"	(B) AAVTAFWGKVK β8–18	(D) LLGNVLVVVLARNFGK β104–119	41–58

Table 1 (continued)

Observed (m/z)	Calculated (m/z)	Peptide 1[a] (subunit) RESIDUES positions	Peptide 2[a] (subunit) RESIDUES positions	Distance range[c] (Å)
"	"	(D) AAVTAFWGKVK β8–18	(B) LLGNVLVVVLARNFGK β104–119	41–58
"	"	(D) AAVTAFWGKVK β8–18	(D) LLGNVLVVVLARNFGK β104–119	9.5–23
"	3523.7	(B) VLDSFSNGMKHLDDLK β66–81	(A) MFLSFPTTK α32–40	22–32
"	"	(D) VLDSFSNGMKHLDDLK β66–81	(A) MFLSFPTTK α32–40	28–37
"	"	(B) VLDSFSNGMKHLDDLK β66–81	(C) MFLSFPTTK α32–40	28–37
"	"	(D) VLDSFSNGMKHLDDLK β66–81	(C) MFLSFPTTK α32–40	22–32
3539.1	3539.8	(B) VKAHGKKVLDSFSNGMKHLDDLK β59–81	(A) YR α140–141	43–45
"	"	(B) VKAHGKKVLDSFSNGMKHLDDLK β59–81	(A) YR α140–141	43–45
"	"	(B) VKAHGKKVLDSFSNGMKHLDDLK β59–81	(A) YR α140–141	43–45
"	"	(B) VKAHGKKVLDSFSNGMKHLDDLK β59–81	(A) YR α140–141	45–47
"	"	(D) VKAHGKKVLDSFSNGMKHLDDLK β59–81	(A) YR α140–141	22–24
"	"	(D) VKAHGKKVLDSFSNGMKHLDDLK β59–81	(A) YR α140–141	26–27
"	"	(D) VKAHGKKVLDSFSNGMKHLDDLK β59–81	(A) YR α140–141	26–27
"	"	(D) VKAHGKKVLDSFSNGMKHLDDLK β59–81	(A) YR α140–141	36–37
"	"	(B) VKAHGKKVLDSFSNGMKHLDDLK β59–81	(C) YR α140–141	22–24
"	"	(B) VKAHGKKVLDSFSNGMKHLDDLK β59–81	(C) YR α140–141	26–27
"	"	(B) VKAHGKKVLDSFSNGMKHLDDLK β59–81	(C) YR α140–141	26–27
"	"	(B) VKAHGKKVLDSFSNGMKHLDDLK β59–81	(C) YR α140–141	36–37
"	"	(D) VKAHGKKVLDSFSNGMKHLDDLK β59–81	(C) YR α140–141	43–44
"	"	(D) VKAHGKKVLDSFSNGMKHLDDLK β59–81	(C) YR α140–141	43–45
"	"	(D) VKAHGKKVLDSFSNGMKHLDDLK β59–81	(C) YR α140–141	44–45
"	"	(D) VKAHGKKVLDSFSNGMKHLDDLK β59–81	(C) YR α140–141	45–47
"	3538.6[d]	(B) GTFAALSELHCDKLHVDPENFK β82–103	(A) GNVK α8–11	38–43
"	"	(D) GTFAALSELHCDKLHVDPENFK β82–103	(A) GNVK α8–11	67–71
"	"	(B) GTFAALSELHCDKLHVDPENFK β82–103	(C) GNVK α8–11	44–50
"	"	(D) GTFAALSELHCDKLHVDPENFK β82–103	(C) GNVK α8–11	67–73

[a]Residue containing NHS–reactive group is underlined.
[b]NHS reactive group is N–terminal amine of protein.
[c]distance between Cα in NHS–reactive residue and Cα in residues of the other peptide.
[d]Mass match based on oxidized peptide pair.

ratio (m/z) 3019 appears only in the crosslinked sample. Table 1 lists other m/z values for peaks observed similarly in mass spectra of digested crosslinked samples but not observed in the mass spectra of the corresponding non-crosslinked control experiment.

To determine possible identities of the peaks representing potential crosslinked pairs identified by mass spectrometry, a comparison was carried out between observed m/z values and calculated m/z values for all possible complete and incomplete tryptic fragments, inter- and intra-subunit sulfo-SBED crosslinked peptide pairs, and known side products of the sulfo-SBED reaction with bovine hemoglobin.[3] (There are approximately 65,000 possible products.) Table 1 also lists the resulting matches between observed and predicted m/z values; several matches were found within a small m/z error (<1 Da) of each observed peak. These possible matches were then evaluated to determine whether they were consistent with the known tertiary and quaternary structure of bovine hemoglobin, using the Protein Data Bank (PDB) structure 1HDA. The estimated maximum spanning length of sulfo-SBED between alpha carbons of two amino acid residues is approximately 29Å. Therefore, two peptides that have been crosslinked by sulfo-SBED must be separated by 29Å or less in the quaternary structure of hemoglobin. Table 1 lists distances between alpha carbons of relevant amino acid residues in possible crosslinked peptide pairs, as calculated from the PDB 1HDA structure. Because the m/z selection procedure described above cannot distinguish between the two α subunits (A and C) and the two β subunits (B and D) of hemoglobin, all possible combinations of crosslinked peptides are included in Table 1. The two reactive groups of the crosslinking reagent sulfo-SBED are an N-hydroxysuccinimide (NHS) ester, which attaches specifically to the primary amine groups on the side chains of lysine residues and at N-termini of proteins, and a photoactivated arylazide that inserts non-specifically into C-H and N-H bonds. Thus one end of the sulfo-SBED crosslinker is attached to a lysine residue, or to a residue that occupied the N-terminus of a hemoglobin subunit; the other end of the sulfo-SBED crosslinker may be attached to any residue. The distances shown in Table 1 represent a range of distances between a lysine (or terminal) residue on one peptide and the alpha carbons of *all* residues in the other peptide. If the maximum span of sulfo-SBED falls within this range, the pair of peptides could plausibly be crosslinked. If the range of distances separating the peptides is greater than the length of sulfo-SBED, it is unlikely that the pair could be crosslinked without distortion of the hemoglobin structure.

While further proof in addition to the mass spectrometric identification and comparison with the PDB model is required, we have identified the following peptides in bovine hemoglobin that were crosslinked by sulfo-SBED:

β 59–65 of subunit B to β 61–75 of subunit D
β 59–65 of subunit D to β 61–75 of subunit B
α 1–16 of subunit A to β 30–39 of subunit B
α 1–16 of subunit A to β 30–39 of subunit D
α 1–16 of subunit C to β 30–39 of subunit B
α 1–16 of subunit C to β 30–39 of subunit D
β 8–18 of subunit B to β 104–119 of subunit B
β 8–18 of subunit D to β 104–119 of subunit D
β 66–81 of subunit B to α 32–40 of subunit A
β 66–81 of subunit D to α 32–40 of subunit A
β 66–81 of subunit B to α 32–40 of subunit C
β 66–81 of subunit D to α 32–40 of subunit C
β 59–81 of subunit D to α 140–141 of subunit A
β 59–81 of subunit B to α 140–141 of subunit C

Summary and Conclusions

The original proposal outlined a program for identifying and characterizing protein complexes by combining crosslinking, mass spectrometry, and computation. Much of the originally proposed work was not feasible because of the reduced scope of the project necessitated by a budget substantially below that requested. Many of these originally proposed ideas would be of considerable benefit in developing an approach for characterizing protein complexes via an approach involving crosslinking and mass spectrometry. Synthesis of more suitable crosslinking reagents would avoid the problem of the lack of specificity of the photoreactive group in sulfo-SBED. Exploration and development of tandem mass spectrometric approaches, which could potentially give partial amino acid sequence information on each of the crosslinked peptides, would be of great benefit. Finally, more suitable computational methods for analyzing experimental data would be a necessity for carrying out crosslinking/mass spectrometry experiments with a reasonable throughput.

References

[1]G. B. Hurst, T. K. Lankford, and S. J. Kennel, "Mass Spectrometric Detection of Affinity Purified Crosslinked Peptides," *Journal of the American Society for Mass Spectrometry*, accepted for publication (2004).

[2]G. B. Hurst, M. Sega, J. L. Stephenson, Jr., P. K. Lankford, and J. A. Lewis, "Tandem mass spectrometry analysis of crosslinked proteins and protein complexes," Proceedings of the 50th ASMS Conference on Mass Spectrometry and Allied Topics, Orlando, FL, June 2–6, 2002.

[3]M. B. Strader, G. B. Hurst, T. K. Lankford, R. L. Hettich, and S. J. Kennel, "Affinity Purification of Crosslinked Peptides," Proceedings of the 51st ASMS Conference on Mass Spectrometry and Allied Topics, Montreal, Canada, June 9–12, 2003.

Elucidating the Functions of Genes and Pathways That Contribute to Genomic Instability, Cell Death, and Malignancies in Mouse Models with Telomere Dysfunction

Y. Liu
Life Sciences Division

The objective of this project was to build a laboratory capability and explore research tools at ORNL for dissecting complex biological systems, using telomere or telomerase as an initial model. We now have a fully operating lab and have obtained preliminary results. We have initiated genomic screening of chemical induced mutant mouse bank previously generated by the Life Sciences Division for mutations in genes involved in telomerase or telomere regulation. To understand the protein network associated with telomerase or telomere, we isolated putative proteins that may associate with these complexes. Mammalian expression vectors have been used to express tagged telomerase core protein or telomere binding protein in an immortalized cell line. Protein or protein complexes that are copurified with tagged proteins have been isolated and further identified by mass spectrometry. We have analyzed two mutant mouse models that are deficient in telomerase-associated proteins. Our framework allows us to understand the possible function of these proteins in telomere maintenance, which is critical in cell viability and genomic stability. Dysfunction of telomerase or telomere complexes is frequently observed in cancer and premature aging syndromes. Elucidating the molecular mechanism that regulates these complexes would benefit our understanding in these human disorders and their treatment.

Introduction

Telomere is a DNA-protein complex localized on the end of each chromosome, the function of which is to cap and protect chromosomes against degradation or fusion.[1] Telomere integrity in cells thus plays an essential role in the control of genomic stability. Loss of genetic material at chromosome ends ("telomere shortening") is frequently observed in elder populations and premature aging syndromes.[2] Furthermore, telomere dysfunction contributes to genomic instability that leads to cell death or cancer development.[3] Accumulating evidence suggests that telomere integrity depends on the ability to maintain telomere length and/or the ability to mask the telomeres from being recognized as damaged DNA.[4] However, the mechanisms by which short telomeres lose their capacity to suppress DNA damage responses are not well understood. In this proposal, we have planned to search for new mouse models, use the existing mouse models, and identify novel proteins to study the molecules that are important in telomere length maintenance and genomic stability. These molecules may serve as excellent therapeutic targets that may reduce cell viability in cancer cells.

Several key proteins or protein complexes are critical in telomere length maintenance. Telomerase is critical in telomere length maintenance by replenishing telomere loss due to "end replication problem" as the result of DNA replication in almost all eukaryotes. Loss or decrease in telomerase activity would lead to loss in telomeric DNA, which, in turn, triggers genomic instability or cell death.[1] However inadequate expression of this complex is also observed in 90% of human cancers and may be associated with malignant transformation or cancer progression in human.[2] These findings implicate telomerase and telomere homeostasis as key elements in the proliferation of normal and cancerous cells. Elucidating molecular machinery in regulating telomerase activity is critical in our understanding telomere length maintenance and malignant transformation.

Mammalian telomerase is a large ribonucleoprotein complex and contains two core components, telomerase reverse transcriptase and telomerase RNA, the latter serving as an integral template for de novo synthesis of telomeric DNA by telomerase. [4] These molecules are minimal components that are required in reconstituting telomerase activity in vitro.[5–7] Loss function in either component would result in loss of telomerase activity in vivo.[8–10] In addition to the core components, several mammalian telomerase-associated proteins have been identified . The function of these proteins in telomerase regulation or telomere length maintenance is poorly understood and need to be characterized. Elucidating these molecular events will certainly benefit our understanding in telomerase or telomere length regulation and thus cancer development and treatment.

Besides telomerase, telomere binding proteins and telomere-associated proteins are also key factors in

telomere protection, telomere length maintenance, and chromosome integrity.[11] Several factors are involved in the DNA damage response and also appear to be critical for telomere integrity. For example, cells expressing mutant telomere binding proteins or cells deficient in DNA repair proteins show dysregulated telomere length and chromosome instability in yeast or mammals.[11] Further studies are required to understand at the molecular level how these known proteins or yet unidentified telomere binding or associated-protein proteins protect telomere.

To elucidate telomerase-associated proteins or telomere binding/associated proteins in telomere length regulation, we have proposed to (1) search the established chemical-induced mutant mouse bank for new mouse models that carry point mutations at genes involved in telomerase and/or telomeres regulation, (2) search for novel telomerase or telomere-associated proteins, and (3) elucidate function of telomerase associate protein or DNA repair proteins in telomere length regulation in vivo.

Technical Approach

We have applied the new high-throughput mutation-scanning technique, temperature-gradient capillary electrophoresis (TGCE), to identify point mutations induced by N-ethyl-N-nitrosurea (ENU) in the mouse genome.[12] TGCE detects the presence of heteroduplex molecules formed between a wild-type gene segment and the corresponding homologous segment containing an induced mutation. Partially denatured heteroduplex molecules are resolved from homoduplexes by virtue of their differential mobilities during capillary electrophoresis conducted in a finely controlled temperature gradient. Heteroduplex analysis of three DNA fragments of telomere binding protein 1 and one DNA fragment of telomerase RNA ranging from 500–600 bp in size was applied to identify ENU-induced mutations of a bank of mutant mice.

To identify protein-protein interaction, we used mass spectrometry (MS) coupled with rapid methods for enrichment of protein complexes from total protein lysate. The approach involves expressing tagged proteins (as the bait) in an immortalized cell line, isolating tagged proteins and their binding proteins or protein complexes from cell lysate using anti-Flag immobilized beads, followed by stringent washes. The entire affinity-isolated complex is eluted off the beads through Flag peptide competition and is then digested by trypsin. Mass spectrometry can then separate the resulting array of peptides according to their mass. The combinations of mass and existing trysin cleavage sites provide identifications for the proteins in the complex through bioinformatics database search (Fig. 1).

To study the telomerase activity, we applied an in vitro– based PCR technique called telomerase amplification protocol. This protocol is a highly sensitive

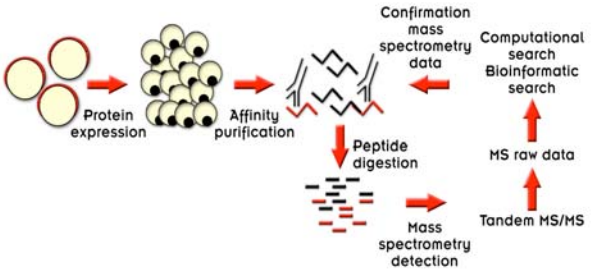


Fig. 1. Schematic presentation of protein tagging and mass spectrometry (MS) detection of protein complex.

assay for detecting telomerase activity in cells. The assay includes two steps. In the first step of the reaction, telomerase adds a number of telomeric repeats onto the oligonucleotide substrate. In the second step, the extended products are amplified by PCR which generates a ladder of products. Finally, the products are separated according to their size through electrophoresis in PAGE gel and the intensity of the products (reflecting the amount of telomerase activity in cells) are visualized through gel photography or imaging.[13]

To measure the average length of telomere repeats at chromosome ends in individual cells, we applied a flow cytometry method combined with fluorescence in situ hybridization (flow FISH) with labeled peptide nucleic acid (PNA) telomeric DNA probes.[14] The signal intensity of telomere fluorescence in each cell reflects the average length of telomeres in that cell. The average of telomere length in the total cell population was calculated from the mean fluorescence obtained from 10^4 cells/sample.

Results and Accomplishments

We initiated, in collaboration with Drs. Edward Michaud and Cymberline Culiat, the genomic screening of the ORNL Cryopreserved Mutant Mouse Bank (CMMB) for mutations in genes involved in regulating telomere, telomerase, or genomic stability. This bank contains DNA from 4000 C57BL/6JRn mice, each carrying a load of ENU-induced mutations inherited from their father. In this bank, every gene in the mouse genome is mutated multiple times, permitting the isolation of an allelic series of mutations for any gene of choice. We use TGCE for identifying single base-pair mutations and then verified new mutations by sequencing. We have intensively searched several databases (NCBI, Ensemble, and UCSC) for genomic sequences of four genes (two telomere-binding proteins and two telomerase core components). Such information is required for the genomic screening of ENU-induced mutations in CMMB. We initially screened for mutations at two genes (one telomerase core component and one telomere-binding protein). By the end of the project, we screened 1421 mutant mice, which total

over 2,131.5 kilobases mouse genomic sequences for mutations at these two genes. Thirty-six putative mutations were identified from the initial screen, which will be verified by sequencing analysis in the future.

To identify telomerase or telomere-associated proteins, we used an immortalized cell line (293T) that expresses a moderate level of Flag-tagged full-length human telomerase catalytic protein or 293T that transiently express a telomere binding protein. These tagged proteins served as the bait protein for isolating their interacting proteins in cells. The tagged proteins and their binding proteins or protein complexes were purified from cell lysate through anti-Flag immobilized beads (Fig. 2) and

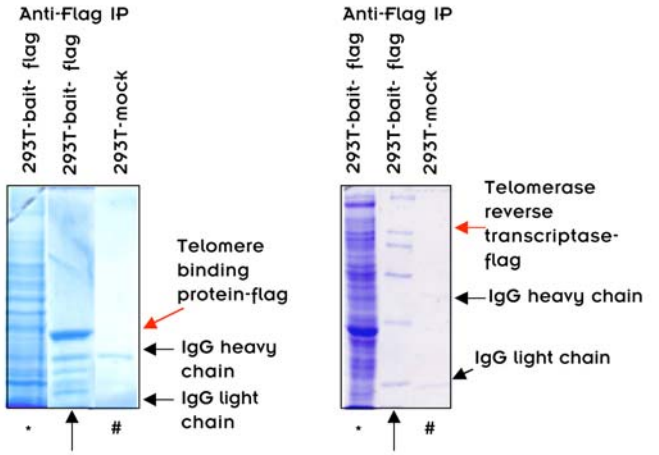


Fig. 2. Affinity purification of telomerase or telomere binding protein complexes using anti-Flag M2 beads. Quality of affinity purification was evident when comparing the same samples before or after affinity purification. After the purification, purified samples were digested by Trypsin and then injected into a mass spectrometer for mass identification.

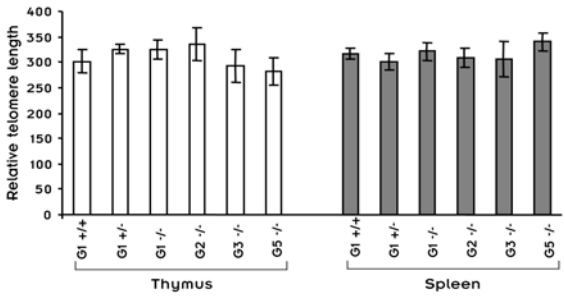
further identified by MS in collaboration with Dr. Greg Hurst of the Chemical Sciences Division, ORNL. These procedures allowed us to identify the tagged proteins and a number of potential binding proteins. Several previously identified telomerase-associated proteins, for example, Hsp70 and Hsp90, small nuclear ribonucleoprotein,[15, 16] were identified through this procedure, demonstrating that our approach was successful. In addition to these known proteins, we were able to identify 20 proteins that potentially associated with the bait proteins. Further studies are required to confirm the accuracy of MS data and the biological significance of these putative telomerase or telomere associated proteins in telomerase function or telomere length regulation.

Previously, Wen Zhou identified a protein component of large cytoplasmic structure, vault, also associated with telomerase activity in vivo (Zhou et al., unpublished data), suggesting that it is a telomerase-associated protein. To elucidate the function of this protein, namely, *V PARP*, in telomerase or telomere length regulation in vivo, we studied the telomerase activity and telomere length in mouse deficient for *V PARP*. This work was carried out in collaboration with Dr. Lea Harrington at the Ontario Cancer Institute, Canada. The knockout mice were viable and fertile for up to several generations. These mice resulted in no apparent change in telomerase activity (data not shown) or telomere length [Fig. 3(a)]. One possibility could be other telomerase associated proteins, such as telomerase-associated protein 1 (*TEP 1*) might share a redundant role with *V PARP* in vivo, as both proteins are commonly shared by two ribonucleoprotein complexes, telomerase and vault. To test this hypothesis, we further generated double knockout mice that are deficient for both *V parp* and *Tep 1*. These mice are also viable and fertile and did not show any detectable change in telomere length at the first generation [Fig. 3(b)]. These data provide direct genetic evidence that neither murine *Tep 1* nor *V parp* is essential for normal mouse development, telomerase catalysis, and/or telomere length maintenance in vivo. A manuscript based on above data is in preparation.

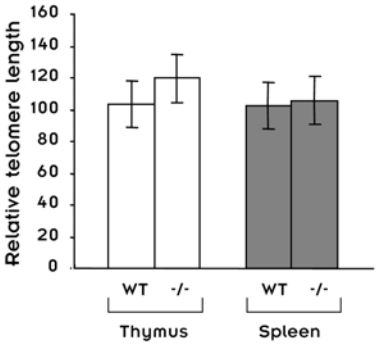
Summary and Conclusions

Our high-throughput screening of CMMB for mutations at the genes regulating telomere length enabled us to identify 36 potential mutations in this mutant mouse bank. Using protein tagging combined with protein affinity purification and MS, we were able to identify several known telomerase-associated proteins and 20 putative telomerase or telomere binding proteins. Using a genetic approach, we demonstrated the possible function of two telomerase-associated proteins in telomerase function or telomere length regulation in vivo. Further studies including verifying mutations at CMMB and characterize putative telomerase or telomere-associated proteins in telomere length regulation would enable us to establish new research tools and understand the mechanism in regulating telomerase or telomere length at the molecular level. A manuscript based on some of our finding is in preparation.

Our projects exploited a leading resource in mouse mutagenesis for generating animal models and in protein



Telomere length was measured by an fluorescent based in situ hybridization method using fluorescent labeled telomere DNA as probe. The intensity of the telomere probe was detected by flow cytometer, which indicate the length of telomere DNA in cells. The average telomere fluorescence in thymocytes and splenocytes derived from wild type (+/+), heterozygous vparp (+/-), or different generations (G_1 and G_3) of V parp deficient mice. In each set, data were pooled from at least five individual mice (error bars represent the standard deviation).



Telomere length was measured by an fluorescent based in situ hybridization method using fluorescent labeled telomere DNA as probe. The intensity of the telomere probe was detected by flow cytometer, which indicate the length of telomere DNA in cells. The average telomere fluorescence in thymocytes and splenocytes derived from wild type (+/+), or knockout mice deficient for both V parp and Tep 1. In each set, data were pooled from at least five individual mice (error bars represent the standard deviation).

Fig. 3. No detectable change in telomere length in mV parp *null or* mV parp /mTep 1 *double null mouse. (a) Telomere length analysis in mouse thymocytes and splenocytes from different generations of* mVparp-*deficient mice by Flow-FISH. The average telomere fluorescence in thymocytes and splenocytes derived from different generations (G_1 and G_3) of* mV parp *knockout mice. (b) Telomere length analysis in thymocytes and splenocytes derived from wild-type and* mV parp / mTep 1 *double-deficient mice by Flow-FISH. In each set, data were pooled from at least five individual mice (error bars represent the standard deviation).*

tagging technique for discovering the nature of the protein-complex "machines" important for telomere maintenance and genomic stability. Such results will form the basis for research directions compatible with DOE's Genomes to Life and Health Effects/Genome programs, particularly in the realms of analysis of protein machines (in this case, telomerase and telomere complexes). Our effort has been expanded to several relevant interesting projects through the internal and international collaboration and formed the basis for establishing new tagging vector system for the GTL program. This project has applications in several programs at Department of Defense and the National Institutes of Health including the National Institute on Aging.

References

[1]C. W. Greider and E. H. Blackburn, "Telomeres, telomerase and cancer," *Sci. Am*. **274**, 92 (1996).

[2]M. W. Djojosubroto et al., "Telomeres and telomerase in aging, regeneration and cancer," *Mol Cells* 15(2), 164–75 (2003).

[3]R. Oulton and L. Harrington, "Telomeres, telomerase, and cancer: life on the edge of genomic stability," *Curr Opin Oncol* 12(1), 74–81 (2000).

[4]E. H. Blackburn, "Switching and signaling at the telomere," *Cell* **106**(6), 661–73 (2001).

[5]T. L. Beattie et al., "Reconstitution of Human Telomerase Activity In Vitro," *Curr Biol* **8**(3), 177–80 (1998).

[6]T. L. Beattie et al., "Functional Multimerization of the Human Telomerase Reverse Transcriptase," *Mol Cell Biol* **21**(18), 6151–60 (2001).

[7]S. L. Weinrich et al., "Reconstitution of Human Telomerase with the Template RNA Component hTR and the Catalytic Protein Subunit hTRT," *Nat. Genet.* **17**(4), 498–502 (1997).

[8]M. A. Blasco et al., "Telomere Shortening and Tumor Formation by Mouse Cells Lacking Telomerase RNA," *Cell* **91**(1), 25–34 (1997).

[9]Y. Liu et al., "The Telomerase Reverse Transcriptase is Limiting and Necessary for Telomerase Function In Vivo," *Curr Biol*. **10**(22), 1459–62 (2000).

[10]X. Yuan et al., "Presence of Telomeric G-Strand Tails in the Telomerase Catalytic Subunit TERT Knockout Mice," *Genes Cells* **4**(10), 563–72 (1999).

[11]T. de Lange, "Protection of Mammalian Telomeres*," Oncogene* **21**(4), 532–40 (2002).

[12]Q. Li et al., "Integrated Platform for Detection of DNA Sequence Variants Using Capillary Array Electrophoresis," *Electrophoresis* **10**, 1499–1511 (2002).

[13]N. W. Kim and F. Wu, "Advances in Quantification and Characterization of Telomerase Activity by the Telomeric Repeat Amplification Protocol (TRAP)*," Nucleic Acids Res*. **25**(13), 2595–7 (1997).

[14]N. Rufer et al., "Telomere Length Dynamics in Human Lymphocyte Subpopulations Measured by Flow Cytometry," *Nat. Biotechnol*. **16**(8), 743–7 (1998).

[15]S. E. Holt et al., "Functional Requirement of p23 and Hsp90 in Telomerase Complexes," *Genes Dev* **13**(7), 817–26 (1999).

[16]A. G. Seto et al., "Saccharomyces Cerevisiae Telomerase is an Sm Small Nuclear Ribonucleoprotein Particle," *Nature* **401**(6749), 177–80 (1999).

Identification and Characterization of Genes and Protein Components in Cell-Cycle Control and Cancer Development

Y. Wang
Life Sciences Division

A "substrate-trapping" strategy to isolate target proteins of cell-division-cycle 14 phosphatase (cdc 14) has been developed, and the substrate-trapping mutants of cdc 14 have been produced. Different epitope-tags of cdc 14 proteins have been engineered, and purified proteins have been made in *E. coli* and mammalian cells. The combination of immunoprecipitation and mass spectrometric analyses has allowed us to identify several potential cdc 14 substrates or interacting proteins. The biological significance of some interacting proteins has been characterized and others are currently under investigation. In collaboration with Dr. Mike Tyers in Samuel Lunenfeld Research Institute, Canada, a systematic functional annotation of a mammalian gene B23 in *S. cerevisiae* gene-deletion set has been performed. We also initiated a collaborative project with Dr. Klas Wiman, Karolinska Institute, Sweden, to study the gene expression profile regulated by the p53 tumor suppressor gene using a high-throughput microarray assay.

Introduction

Tumor cells often acquire damage to genes that directly regulate their cell cycle. A great majority of DNA damage in cancer patients is the result of environmental insults, such as exposure to irradiation and chemical mutagens. Understanding the molecular mechanisms that regulate these events will help us to assess and prevent the risks of cancer associated with the exposure of environmental hazardous materials, which is consistent with DOE's long-term mission. The fundamental task of the cell cycle is to ensure that DNA is faithfully replicated once during the synthesis (S) phase and that identical chromosomal copies are distributed equally to two daughter cells during mitosis. Disruption of this fundamental process could lead to disasters to our life, such as the development of cancer, developmental failure or degenerative diseases.[1]

Numerous genetic interactions and biochemical studies positioned cell division cycle 14 (cdc 14) gene at the bottom of the mitotic exit cascade in budding yeast, where it triggers mitotic exit by inactivating the active cyclin-dependent kinase. Activation of the Cdc 14 is mediated by the release of cdc 14 from nucleolus and tightly controlled through FEAR (Cdc 14 early anaphase release) and MEN (mitotic exit network) pathways. Cdc 14 is a dual-specificity phosphatase conserved from yeast to *C. elegans* to mammals. In mammals, it has been found to play an important role in centrosome splitting at the onset of mitosis, and mutations of the gene have been detected in human breast carcinoma cell lines. The molecular mechanisms underlying the role of cdc 14 in cell cycle regulation and tumor development is largely unknown.[2] Our goal is to isolate and characterize the Cdc 14 phosphatase substrates and interacting proteins in collaboration with the Mass Spectrometry Group at ORNL. This study will enable us to dissect the biological significance of the Cdc 14 protein networks in cell cycle control and cancer development. Our study will not only help us to understand the mechanism of cancer development but also to provide possible protein targets for anti-cancer drug development.

Nucleophosmin/B23 mapped to human chromosome 5q35 is a nucleolar phosphoprotein that is more abundant in tumor cells than in normal resting cells. It has been identified as a substrate of cyclin-dependent kinase 2 (CDK2)/cyclin E in centrosome duplication. B23 is associated with unduplicated centrosomes, and dissociated from centrosomes by CDK2/cyclin E-mediated phosphorylation. B23 has also been found to be involved in at least three distinct forms of hematologic malignancy. The N-terminal region of B23, which contains oligomerization domain, is fused to anaplastic lymphoma kinase in anaplastic lymphoma with t(2;5), retinoic acid receptor, in acute promyelocytic leukemia with t(5;7), and myeloid leukemia factor 1 in myelodyplastic syndrome with t(3;5) translocations.[3] An in-depth annotating the molecular mechanisms of B23 will be of benefit not only to our understanding of the cell cycle regulation but also to the development of cancer.

The p53 tumor suppressor protein is expressed at low levels in normal cells but accumulates in response to DNA damage, leading to cell-cycle arrest and/or apoptosis. The wild-type p53 protein has sequence-specific DNA binding, transcriptional transactivation, and repression activity. Point mutations of p53 found in human tumors abolish its sequence-specific DNA binding activity, as a rule. Recent work demonstrated p53 exerts its effect through a diverse network of transcriptional changes, which makes exploration of the molecular mechanisms that mediate p53-triggered cell-cycle arrest and apoptosis more challenging and difficult.[4] As the first step to solve these puzzles, we have proposed to isolate novel genes that are directly regulated by p53 through microarray approach.

Technical Approach

Transient protein-protein interaction and low-affinity "bait" proteins are the two major blockades in isolation of substoichiometric protein complexes by an immunoprecipitation/mass spectrometric approach. If a target protein is a substrate of an enzyme, it is, in general, extremely difficult to "fish out" the substrates using the enzymes as bait, because of the fugitive enzyme-substrate interactions. Based on the crystal structures of protein tyrosine phosphatase 1B (PTP1B), we have identified a conserved aspartate residue in human Cdc 14 (cell division cycle 14) proteins. Site-directed mutation of this conserved aspartate residue to alanine increases binding affinity to its substrates several-fold while it dramatically decreases their catalytic capacity. These mutants have been used for substrate-trapping followed by mass spectrometric identification. Proteins that are substantially interesting are further evaluated by their subcellular localization through fluorescent imaging, cell cycle distribution, and genetic complementation analyses.

Seventy percent of yeast genes are functionally classified and more than 80% of the genetic pathways in yeast can be found in human. Each of the 5100 nonessential genes in the 6200-gene *S. cerevisiae* genome has been systematically deleted and bar-coded, with the composite deleted genomes assembled into an ordered array of live yeast strains.[5] Overexpression of human genes in wild-type yeast cells either can be lethal or can have no effect. Subsequent crosses with deletion strains, in synthetic complementation or dosage-lethality experiments, respectively, can identify the analogous genetic pathway to which the human gene belongs, thus providing a gene-based method to filter the function of mammalian genes. To systematically annotate the function of the B23 gene in yeast gene-deletion arrays, the B23 gene was first cloned into a yeast expression vector and transformed into wild-type yeast cells. Haploid MATα yeast cells carrying the human B23 were mated with each of MATα deletion mutant by an automated robot. The presence or absence of yeast colonies was chosen as a standard cutoff to map the genetic interacting pathways using an automatic colony analyzing software.

The advanced development of cDNA microarray technology enabled us to study thousands of gene expression profiles instead of one or two genes at a time. To study the expression profiles of p53-regulated genes, mRNAs were isolated and reverse-transcribed into cDNA from experimental and control groups. The cDNAs of the two groups were differentially labeled and then hybridized onto the mouse 22k cDNA chips from Vanderbilt University. Differentially expressed genes were analysed through an array reader using Gene-Pix software, and the presence of consensus p53 binding sequences (two consensus decamers RRRCWWGYYY separated by a gap of 0–13 nucleotides) in the genes of our interest was searched through the NCBI public database.

Results and Accomplishments

We have created point mutations converting the wild-type Cdc 14 into substrate-trapping mutants. The mutants were first tested in complementation assay in yeast and shown to be incapable of rescuing their yeast orthologue defects (Fig. 1), suggesting they could be used as substrate-trapping mutants. These mutants were in-frame cloned into a myc-epitope tagged expression vector and introduced into mammalian cells by stable transfection. Immunoprecipitaion of Cdc $14A^{D/A}$ mutant protein with anti-myc antibody (9E10) from Cdc 14 $A^{D/A}$ expressing cells allowed us to pull down several potential Cdc 14A

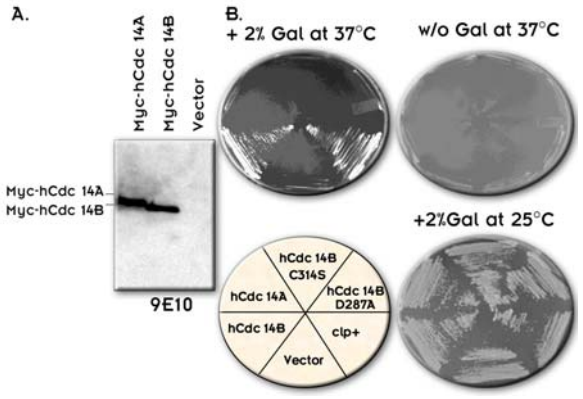


Fig. 1(a) Western blot analysis of Myc-tagged human Cdc 14A and Cdc 14B expression in cdc 14-3 *cells transformed with pRSHisGal-hCdc 14A, pRSHIsGal-hCdc 14B and pRSHisGal vector alone respectively. (b) Expression of human Cdc 4B but not Cdc 14A rescues the lethality of* cdc 14-3 *cells at non-permissive temperature. The temperature-sensitive mutant* cdc 14-3 *cells containing different plasmids as indicated were tested for colony formation on XY plates with and without galactose at 37°C and 25°C. Human Cdc* $14B^{C314S}$ *and Cdc* $14B^{D287A}$ *are catalytic dead mutants.* Clp^{+} *is the* S. pombe *ortholog of* S. cerevisiae *Cdc 14p.*

substrates or interacting proteins. Following SDS-PAGE separation, in-gel peptide digestion with trypsin and mass spectrometric analysis, we were able to narrow down to several potential Cdc 14 substrates (Fig. 2). As a proof of principle, we focused on B23 because of recent findings on its involvement in the control of centrosomal duplication. We pulled out B23 full-length cDNA from a cDNA library and cloned it into a Flag-epitope vector. Co-transfection of both B23-Flag and Myc-Cdc 14A showed that B23 and Cdc 14A could indeed interact with each other in vivo by reciprocal immunoprecipitation (Fig. 3). Other groups have shown unphosphorylated threonine 199 of B23 binds to unduplicated centrosome and inhibits centrosomal duplication. We believe our Cdc14 A is the genuine phosphatase that may

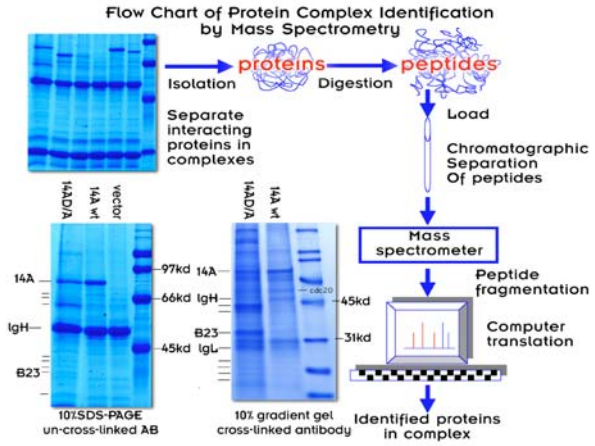


Fig. 2. A flow-chart of immunoprecipitation/mass spectrometric assay to isolate and identify potential associated protein complexes. Substrate-trapping mutant proteins were expressed in cells. After affinity purification, associated protein complexes were separated on SDS-PAGE gel. Individual associated proteins were isolated and digested to generate individual peptides. Peptides were loaded to mass spectrometer for peptide identification. Upon data bank search, individual peptides were assigned to individual proteins. Potential substrates of Cdc 14s were indicated.

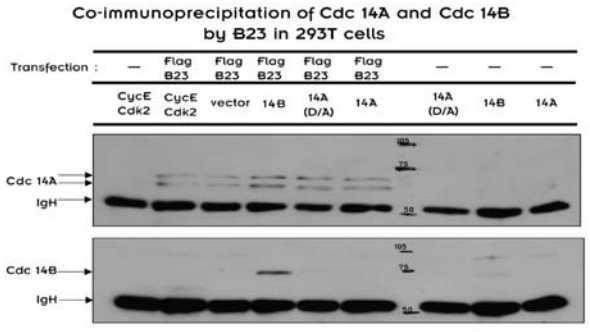


Fig. 3. In vivo interaction of Cdc 14 with B23. Flag-tagged B23 protein complexes were affinity precipitated with anti-Flag antibody. Cdc 14 proteins were detected by anti Cdc 14 antibodies, confirming the mass spectrometric finding.

dephosphorylate Thr199 of B23, which in turn controls centrosomal duplication in S phase and prevents centrosomal reduplication during mitosis. Our preliminary results indicate they might co-localize on centrosomes in G1 and anaphase. We also constructed GFP-B23 and Red-Cdc 14 vectors in order to study the dynamic or temporal interaction of the two proteins in a living cell. Due to the lack of a live cell imaging system at ORNL, this approach will not be possible for the time being. As an expansion of the project, we also purified centrosomal fractions from Jurkat cells (a T cell line) by sucrose gradient centrifugation. Centrosomal fractions were verified by the expression of γ-tubulin on Western blot. We showed that B23 and Cdc 14 co-migrate in the same centrosomal fractions (Fig. 4). These centrosomal fractions can be used for the isolation of the mysterious unknown centrosomal components using $B23^{T199A}$ mutant, that may be directly regulated by the unphosphorylated B23. A manuscript based on the above studies is in preparation.

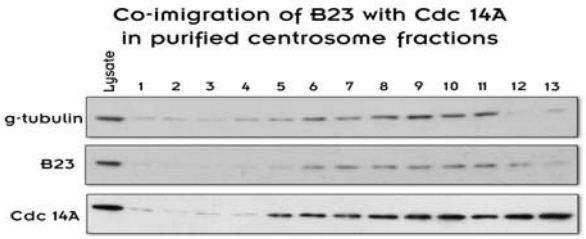


Fig. 4. Co-immigration of B23 with Cdc 14A in purified centrosomal fractions. Centrosomal fractions were purified from T-cells through ultracentrifugation in sucrose cushion. γ-tubulin was used as the centrosome marker. Centrosome fractions that containing B23 also expressed Cdc 14A.

In order to annotate the functions of human B23 gene in yeast, we have performed, in collaboration with Dr. Mike Tyers, Toronto, Canada, two rounds of synthetic dosage lethality screening using both wild-type B23 and $B23^{T119A}$ mutant (a critical phosphorylation site mutant) in ~5100 viable yeast gene-deletion mutant arrays. When a stringent bioinformatic filter was applied, we found that a total of 66 (33 for wild-type B23, 5 for $B23^{T199A}$ mutant and 28 for both) out of 5100 genes were mapped into B23 genetic pathways (Fig. 5). In addition to the verification of known functions of B23 involving ribosome biosynthesis and spindle pole body (centrosome) splitting, we identified several novel roles of B23 such as histone acetylation and actin organization. We are currently investigating the potential relationships between B23 and the mammalian orthologues of the newly discovered B23-related yeast genes in mammalian systems. We will use the data obtained from this study as a proof of concept to conduct a high- throughput annotation of genes on human chromosomes 5, 16 and 19, an ultimate goal of functional genomic program sponsored by DOE.

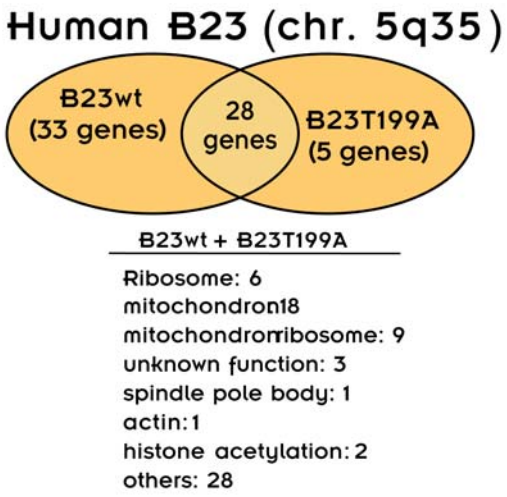


Fig. 5. Summary of B23 genetic pathways based on the yeast gene-deletion array screening. A total of 66 genes were identified. Twenty-eight genes were involved in both wild-type B23 and mutant B23^{T199A}, 33 exclusively interacted with wild-type B23 and 5 only with mutant B23. Actin network assembly and histone acetylation are the new genes that have never been found to associate with B23 before.

Using a temperature-sensitive mouse mutant p53 construct, we have previously shown that wild-type p53 can trigger both G1 arrest and apoptosis in a v-myc retrovirus-induced p53 negative mouse T lymphoma line J3D. We have also found that bcl-2 can prevent p53-induce apoptosis but not G1 arrest, whereas HPV16E7 blocked p53-induced G1 arrest but not apoptosis.[6] These findings let us believe simultaneous expression of both bcl2 and HPV16 E7 would prevent both apoptosis and G1 arrest even in the presence of activated wild-type tsp53. Moreover, this triple-gene expressing cell line could be used to isolate genes that are potentially directly regulated by p53 through cDNA microarray approach, because the downstream signals (the cellular effects of p53) are supposed to be blocked. The experiments were performed on mouse 22k cDNA chips from Vanderbilt University. As a proof of concept, we made a stable line expressing tsp53, bcl2, and HPV16 E7. Western blotting and RT-PCR showed that these cells express all three exogenous proteins. By comparison of the gene expression profiles using cluster analysis at different time points at non-permissive temperature, we have identified several novel genes directly regulated by p53. As a control for the temperature shift, we have also generated stable cell lines expressing E7 and bcl2 from the p53 negative parental J3D cells. Candidate genes that are differentially expressed are currently being confirmed by Northern blot and quantitative Real-PCR analyses. Bioinformatic analysis has enabled us to confirm the presence of p53 consensus sequences (two consensus decamers RRRCWWGYYY separated by a gap of 0–13 nucleotides) in the promoter regions of our newly identified genes. We are currently investigating the ability of those genes in inhibiting or promoting p53-induced G1 arrest and apoptosis in p53$^{-/-}$ and p53$^{+/+}$ cells. These studies will help us to understand how the p53 tumor suppressor protein modulates the cellular response to various genomic insults and to discover novel pathways that mediate the p53-triggered cell cycle arrest and apoptosis.

Summary and Conclusions

The initial LDRD funding has helped me to set up a new molecular biochemical laboratory and obtain preliminary as well publishable data. My lab is now fully functioning and able to perform most of the molecular and biochemical experiments. Using a protein-tagging strategy in combination with mass spectrometric peptide identification, we have identified several Cdc 14 substrates/interacting proteins and characterized their potential involvement in centrosome duplication and cytokinesis. We have identified at least two novel functions of the mammalian B23 gene based on yeast deletion set screening. Using the high-throughput microarray approach, we identified several novel genes that are potentially regulated by p53 tumor suppressor and function as checkpoint controls in response to DNA damage in cells.

This study of the protein complexes involved in the cell cycle control and cancer development is in line with DOE's Genome to Life protein complex efforts and long-standing interest in issues related to human susceptibility to the products of energy production. In fact, as a member directly working on Genome to Life project, I have experienced tremendous mutual benefits between our mammalian protein complex project and the Genome to Life project, though the latter is mainly focused on microbial biology. I believe our interaction and collaboration will continue in the very near future. Since the Cdc 14 regulatory networks, B23, and p53 tumor suppressor pathways are all instrumental in helping us understand cellular response to the exposure of environmental hazardous materials and cancer development, we believe that data generated in these studies will qualify us to apply for follow-up funding from the DOE Genomes to Life and Low Dosage Radiation programs, NASA Low Dosage Radiation program, and the NCI Cancer Genetics and Molecular Biology programs.

References

[1]J. Rouse and S. P. Jackson, "Interfaces between the detection, signaling, and repair of DNA damage," *Science* **297**(5581), 547–51 (2002).

[2]S.Trautmann and D. McCollum, "Cell Cycle: New Functions for Cdc14 Family Phosphatases," *Curr. Biol*. **12**(21), R733–5 (2002).

[3]M. Okuda, "The Role of Nucleophosmin in Centrosome Duplication," *Oncogene* **21**(40), 6170–4 (2002).

[4]A. Gudkov, "Microarray Analysis of p53-Mediated Transcription: Multi-Thousand Piece Puzzle or Invitation to Collective Thinking," *Cancer Biol. Ther.* **2**(4), 444–5 (2003).

[5]A. H. Tong et al., "Systematic Genetic Analysis with Ordered Arrays of Yeast Deletion Mutants," *Science* **294**(5550), 2364–8 (2001).

[6]Y. Wang et al., "Abrogation of p53-Induced G1 Arrest by the HPV 16 E7 Protein Does Not Inhibit p53-Induced Apoptosis," *Oncogene* **12**(12), 2731–5 (1996).

Comprehensive Molecular Probing of Live Biological Cells

T. G. Thundat,[1] R. J. Warmack,[2] T. L. Ferrell,[1] M. J. Doktycz,[1] M. D. Barnes,[3] and T. E. McKnight[2]
[1]*Life Sciences Division*
[2]*Engineering Science and Technology Division*
[3]*Chemical Sciences Division*

Imaging, characterizing, and identifying intra- and extra-cellular localization of proteins and protein complexes at nanometer resolution are important goals in nanobiology. The objective of this proposal is to design and develop nanoscale probes and a photonic force microscope (PFM) to image intra- and extra-cellular protein complexes. Such probes reside at the confluence of the smallest of man-made materials and those of natural systems, making them ideal probes for biological cells. Nanoscale probes offer a superlative approach to characterizing the molecular interactions taking place within living systems. The PFM using laser tweezers and nanoprobe optical monitoring will allow probing and imaging outside and inside the cell. This unique microscope will provide information about biomolecular complexes and cellular nanomechanics to complement conventional scanning probe and optical techniques, providing an unprecedented perspective of cellular processes.

Nanoscale imaging of molecules and cells is currently achieved using atomic force microscopy (AFM). In an AFM, a cantilever with a very small spring constant is scanned over the samples to create a force image of the object. Since cantilevers can detect extremely small forces such as van der Waals' forces between the cantilever tip and the sample, a resolution image can be obtained. However, the cantilevers have limited freedom as they are attached to a massive chip and cannot be used for imaging sharp features, for example, the underside of a cell. The PFM is an innovation that will allow extra-cellular probing and, for the first time, nanoscale intracellular probing of cells. The PFM retains the sensitivity of the AFM without the limitations of the cantilever spring and its inherent restriction to surface scanning (two-dimensional). The ability of the PFM to tether a dielectric bead with an optically adjustable force provides full three-dimensional imaging of structures, revealing the nature and magnitude of molecular forces. The extremely small force constant of the optical trap allows detection of very small forces. The ability to detect extremely small forces translates into very high resolution in imaging. Advancing these probes beyond surface analyses will enable extensive characterization of cellular processes in real time. This basic work is expected to promote fundamental discoveries in biomedical imaging and bioengineering programs hoped to impact medical applications.

Progress in the first year included the design, acquisition, assembly, and operation of the PFM, providing state-of-the-art optical imaging of cells plus guided probe analysis of internal and external nano-mechanics. Theoretical studies of light scattering and fluorescence were performed to substantiate work to be performed in FY 2004. We were successful in trapping a bead that executes Brownian motion in the optical trap of the photonic force microscope. The Brownian motion amplitude decreases significantly once the bead is trapped in the optical trap. We were also successful in navigating the trapped bead in a fluid medium by translating the optical trap. Efforts are under way to image a living cell using optically trapped bead. Our next studies will concentrate upon demonstrating the efficacy of these unique nanoprobe methods in conjunction with other advanced imaging studies that will solve issues of interest to DOE and NIH.

The role of high-resolution imaging has applications directly relevant to the Department of Energy's (DOE) Genomes to Life program. It will provide a more complete imaging capability for the exterior and interior of a living cell, bound proteins, receptor sites, etc. Such information will greatly enhance our understanding of living cells and the way cells work. The tools afforded by this work are also quite relevant to Nanoscale Science, Engineering, and Technology Research.

This project also has applications in several NIH programs, especially the National Institute for Biomedical Imaging and Bioengineering. The ability to real-time image cells to external stimuli has significant applications in clinical applications, biomedical engineering, and drug discovery. The combination of unique nanoprobe techniques with other imaging and probing instrumentation at ORNL opens the possibility of becoming an NIH center that would attract and foster very fruitful collaborations with visiting cellular biologists.

BIOLOGICAL SCIENCES AND TECHNOLOGY

Seed Money Fund

Simulation of Biomolecular Complexes: Advanced Computational Sciences at the Molecular Level

B. E. Hingerty,[1] T. Vo-Dinh,[1] S. Broyde,[2] and M. Cosman[3]
[1]*Life Sciences Division*
[2]*Biology Department, New York University*
[3]*Biology & Biotechnology Research Program, Lawrence Livermore National Laboratory*

We have developed advanced theoretical modeling to study the interactions between bioreceptors with target species at the molecular level. (1)The pyridyloxobutylating (POB) agents derived from metabolically activated tobacco-specific nitrosamines can covalently modify guanine bases in DNA at the O^6 position. The adduct formed, O^6-[4-oxo-4-(3-pyridyl)butyl]guanine ([POB]dG), results in mutations that can lead to tumor formation, posing a significant cancer risk to humans exposed to tobacco smoke. A combined NMR-molecular mechanics computational approach was used to determine the solution structure of the [POB]dG adduct within an 11mer duplex sequence d(CCATAT-[POB]G-GCCC)·d(GGGCCATATGG). (2) Six DNA hybridization structures were calculated using the program DUPLEX to compare the effects of triplet base repeats and the effects of a single base substitution on DNA structure and energy. (3) Protamine is a small (~50 amino acids), highly basic (pI ~ 12), arginine-rich (~50%) protein that binds to and condenses DNA in the sperm of most vertebrates. The NMR studies have been completed for one complex, and work on two more peptide/DNA hairpin complexes, as well as the full-length protamine/DNA hairpin complex, is under way. We expect this work to continue while we obtain outside funding.

Introduction

Biological molecules have evolved exquisite signaling and detection mechanisms to communicate with one another. Recently, much effort has been put forth to develop improved bioprobes and to mimic these natural systems in developing robust artificial receptors, or biosensors, as sensitive and highly specific detectors. A first step toward this goal would be the ability to understand the fundamental principles by which these molecules recognize and bind to each other. Thus the aim of this proposal is to develop new computer algorithms and programs to better enable modeling of DNA/protein or protein/protein complexes, which in turn will allow rational design of smart biosensors, with applications toward detectors for diseases, toxins, and other agents.

Technical Approach

Although commercially available software packages are capable of determining biomolecular structures and interactions, they rely heavily on acquiring accurate experimental data. Thus biomolecular interactions that are transient, weak, or otherwise difficult to characterize experimentally but are of critical importance to function are not easily studied. There is, therefore, a need to develop software that can independently provide information about these difficult systems and allow the development of biosensors capable of detecting biologically relevant agents under native conditions. We began by studying a peptide/DNA hairpin complex. Our recently improved sigmoidal distance-dependent dielectric function,[2] which enables our force field to model structures and energetics of carcinogen-DNA adducts to within 1–2 kcal/mol of the experimental NMR benchmarks, was employed at this stage to implicitly model a solvent. Counterions were modeled via reduced partial charges on the non-linked phosphate oxygens according to counterion condensation theory. We had previously developed methods for determining structures of carcinogen-damaged DNA.[1] Our newly derived algorithms were tested with an experimental peptide/DNA complex and against commercially available packages. The experimentally derived NMR distance restraints for the peptide/DNA sample were provided by Dr. Monique Cosman (Biology & Biotechnology Research Program, Lawrence Livermore National Laboratory). Our model building program, DUPLEX,[1] using improved potentials that accurately reproduce B-DNA double helical structure,[2] in concert with commercial software, was used to perform docking experiments whereby the bioreceptor is complexed with the specific analyte under question. A typical example would be a DNA/protein complex, an antibody/antigen complex, or a DNA hybridization study. The torsion space minimization procedure has a much greater radius of convergence than Cartesian minimization

routines, so stable minima are more likely to be found far from the starting position.

Results and Accomplishments

(1) The **pyridyloxobutylating (POB) agents** derived from metabolically activated tobacco-specific nitrosamines can covalently modify guanine bases in DNA at the O^6 position. The adduct formed, O^6-[4-oxo-4-(3-pyridyl)butyl]guanine ([POB]dG), results in mutations that can lead to tumor formation, posing a significant cancer risk to humans exposed to tobacco smoke. A combined NMR–molecular mechanics computational approach was used to determine the solution structure of the [POB]dG adduct within an 11mer duplex sequence d(CCATAT-[POB]G-GCCC)·d(GGGCCATATGG) (Fig. 1). Gaussian 98 was used to determine quantum mechanically optimized geometry and charges for the G-POB adduct, and DUPLEX was used in the combined energy/NMR restrained optimization. In agreement with the NMR results, the POB ligand is located in the major groove, centered between the flanking 5'-side dT·dA and 3'-side dG·dC base pairs, and thus in the plane of the modified [POB]dG·dC base pair, which is displaced slightly into the minor groove. The modified base pair in the structure adopts wobble base pairing {hydrogen bonds between [POB]dG(N_1) and dC(NH_4) amino proton and between [POB]dG(NH_2) amino proton and dC(N_3)}. A hydrogen bond appears to occur between POB carbonyl oxygen and the partner dC's second amino proton. The modified guanine purine base, partner cytosine pyrimidine base, and POB pyridyl ring form a "triplex" via this unusual hydrogen-bonding pattern. The phosphodiester backbone twists at the lesion site, accounting for the unusual phosphorus chemical shift differences relative to those for the control DNA duplex. The helical distortions and wobble base pairing induced by the covalent binding of POB to the O^6-position of dG help explain the significant decrease of 17.6°C in melting temperature of the modified duplex relative to the unmodified control. This work is to be published in the December 2, 2003, issue of *Biochemistry*.[3]

(2) The illustrated six **DNA hybridization structures** were calculated using the program DUPLEX to compare the effects of triplet base repeats (Fig. 2) and the effects of a single base substitution (Fig. 3) on DNA structure and energy. The results indicate that once such slipped structures form, very little additional energy is required to produce either triplet repeat extensions or mutations and deletions.

(3) **Protamine** is a small (~50 amino acids), highly basic (pI ~ 12), arginine-rich (~50%) protein that binds to and condenses DNA in the sperm of most vertebrates. It inactivates the sperm genome by condensing it into toroids until it can be uncondensed and reactivated following fertilization in the egg. Previous studies have shown that the basic arginine residues in protamine bind in the major groove of the DNA helix and interact predominately with the acidic phosphodiester backbone of both strands of DNA in a sequence-independent manner to form a neutral, insoluble complex. In order to gain insight into how protamine interacts with DNA at the molecular level, we

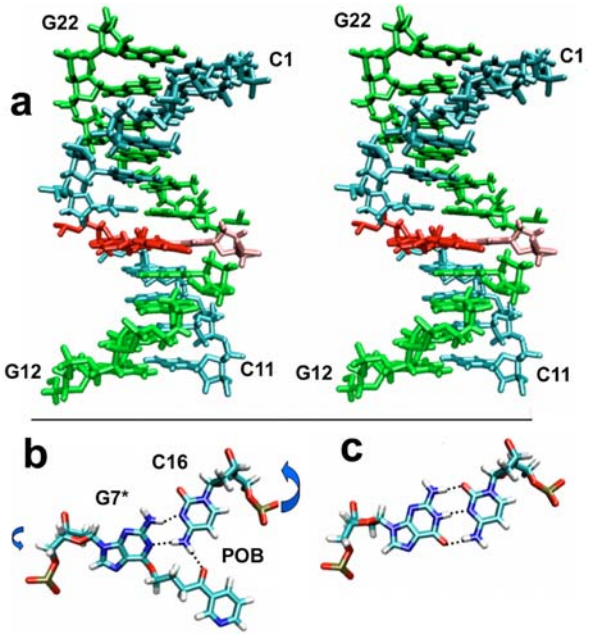


Fig.1. (a) Stereo view looking in the major groove of the lowest energy structure with the modified strand (C1-C11) in blue, complementary strand (G12-G22) in green, [POB]dG7 in red, and dC16 in pink. (b) The bases of the modified guanine, partner cytosine, and POB pyridyl ring form a "triplex" through an unusual hydrogen-bonding pattern (dashed lines). In order to accommodate the larger aromatic surface area while still retaining the normal overall dimensions of the DNA helix, the phosphate backbone at the dC16-dA17 step and, to a lesser degree, the [POB]dG7-G8, twist (arrows), in agreement with the phosphorus chemical shift differences (Fig. 5d). (c) The dG7·dC16 base pair in the control DNA duplex is shown for comparison.

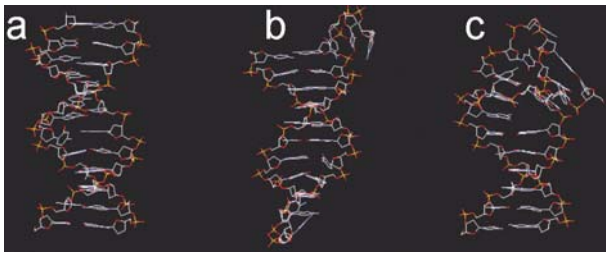


Fig. 2. Triplet repeats are seen in such diseases as fragile X syndrone, mytonic distropy, and Huntington's disease: (a) control duplex, d(CTGCTGCTGCTG)·d(CAGCAGCAGCAG), energy = –495.01; (b) 3′ three-base single strand overhang on each strand, energy = –465.42; and (c) a duplex containing a 3′ three-base single strand overhang on one strand and a central three-base loop out region in the other strand, energy = –456.84.

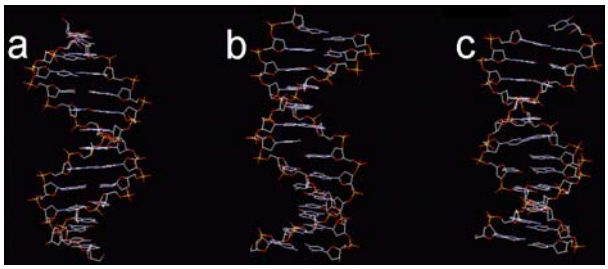


Fig. 3. Single-base substitutions effect DNA structure as shown above: (a) control duplex, d(GGGGGGGGGGGG)·d(CCCCCCCCCCCC), energy = –542.99; (b) a central substitution of a guanine with adenine in one strand, energy = –450.61; and (c) deletion of a single cytosine opposite one guanine, with a 3′-C overhang, energy = –448.09.

have prepared soluble complexes composed of short (14–16 amino acids) protamine segments and DNA hairpins to study by a combined NMR/computation method. The complex remains soluble because the protamine binds and neutralizes the duplex stem of the DNA hairpin, while the hairpin loop retains its negative charges, thereby keeping the complex charged as well.

The NMR studies have been completed for one complex (Fig. 4), and work on two more peptide/DNA hairpin complexes, as well as the full-length protamine/DNA hairpin complex, is under way. The addition of peptide parameters to DUPLEX and refinement of the DNA hairpin structure calculated using DUPLEX is also work in progress. We will also use the package CYANA, which has been optimized for proteins with DUPLEX in these calculations.

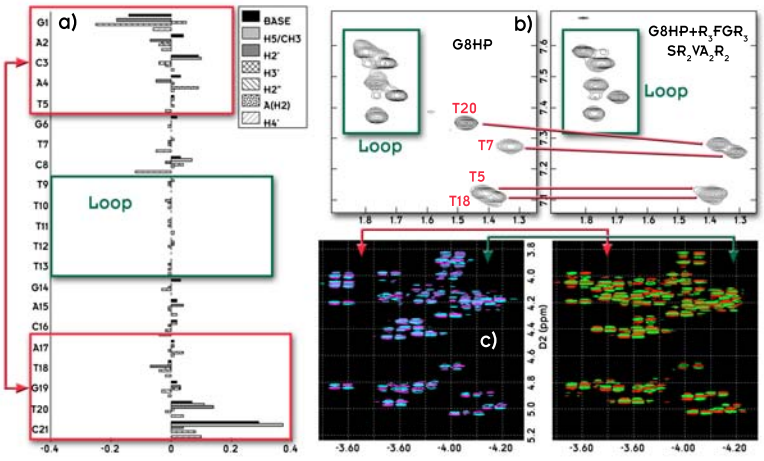


Fig. 4. (a) Summary of DNA chemical shift perturbations (control-complex) for the d(GACATGTCTTTTTGACATGTC) (dG8HP) +$R_3FGR_3SR_2VA_2R_2$ complex. (b) Expanded region of the 300-ms NOESY spectrum of the G8HP+$R_3FGR_3SR_2VA_2R_2$ complex showing the thymine methyl region. Note that the thymines in the hairpin loop do not shift, while T20 and T5 in the duplex stem in the complex shift relative to their positions in the control. (c) ^1H-^{31}P-heteronuclear correlation spectra of the control (left) and complex (right). The arrows indicate changes in the spectra that occur upon binding.

Summary and Conclusions

The [POB]dG adduct posed several challenges in structural determination. There are numerous degrees of freedom of rotation due to a longer butyl chain and pyridyl ring, which can interact with the DNA nonspecifically, resulting in a heterogeneous mixture consisting of a number of conformers. Yet, despite these expected challenges, it was possible to determine a single conformation for the [POB]dG adduct within an 11mer duplex sequence. A combined NMR-molecular mechanics computational approach was necessary to provide a detailed structural view of the types of distortions that can be induced into DNA by a bulky and flexible alkylating agent, such as POB, at the O^6 position of guanine (Fig. 1) because of the broad overlapped lines in the NMR spectra due to the motions. We had to work iteratively between the NMR analysis and the computational results in order to obtain the best structures with lowest energies and best fit to the data. This work resulted in one publication.[3]

The DUPLEX program has been given the ability to analyze hybridized double strands of DNA. DUPLEX is capable of analyzing the three-dimensional structures and energetics for such DNA hybridization studies. DUPLEX also has the ability to study hairpin loops such as those found in the stems of tRNA and in the DNA-protamine complex.

From the chemical shift perturbation studies of the G8HP+$R_3FGR_3SR_2VA_2R_2$ and the changes induced upon complex formation in the intensities of the proton, carbon, and phosphorus resonances we observe in the DNA, as well as proton and carbon in the peptide, we can say that the peptide binds to the 5-base pair DNA duplex terminal end, and the peptide is oriented so that its N-terminal end is located at the DNA duplex terminal end. In addition, we can say that the conformation of the peptide in the complex adopts an extended conformation and that a helical conformation, as has been postulated to be the case by others, is not observed. From NMR temperature studies, we have been able to form a hypothesis about why the binding is specific to the terminal end of the DNA duplex based on dynamic factors. We also show evidence that the PHE residue is intercalating between the two terminal DNA base pairs. However, the peptide and DNA are too short, resulting in a very weak binding constant (~200 µM) for the complex, and thus not allowing us to determine enough contacts between the peptide and the DNA to

allow us to obtain a high-resolution structure. Thus computational calculations will be critical in helping to determine the structures of these peptide/DNA hairpin complexes. This work is still in progress as of this report.

Once we are successful with a solution of the DNA/protamine complex by combined NMR-computational methods, we feel confident that we will be successful in obtaining NIH or DOE funding, especially with the expansion of the Genomes to Life (GTL) initiative at DOE. Much of this is in relation to the future direction of biological research since the completion of the Human Genome sequence.

References

[1]S. Broyde and B. E. Hingerty, "Effective Computational Strategies for Determining Structures of Carcinogen-Damaged DNA," *J Comput Phys* **151**, 313–332 (1999).

[2]L. Wang, B. E. Hingerty, A. R. Srinivasan, W. K. Olson, and S. Broyde, "Accurate Representation of B-DNA Double Helical Structure with Implicit Solvent and Counterions," *Biophysical J.* **83**, 382–406 (2002).

[3]L. A. Peterson, C. Vu, B. E. Hingerty, S. Broyde, and M. Cosman, "Solution Structure of an O^6-[4-oxo-4-(3-pyridyl)butyl]guanine Adduct in an 11mer DNA Duplex: Evidence for Formation of a Base 'Triplex'," *Biochemistry* (in press).

Development of a Generic Computational Method for Biological Data Clustering

D. Xu, V. Olman, and Y. Xu
Life Sciences Division

Data clustering is fundamental to mining large-scale biological data. We further developed a general clustering method based on the representation of multidimensional data as a minimum spanning tree. The presentation opens a new door for clustering, which converts a clustering problem to a problem of partitioning a one-dimensional profile (a sequential representation) derived from the Prim algorithm for constructing a minimum spanning tree. Such a method facilitates partitioning and identification of "dense" clusters from a noisy background. We have successfully applied our methods to a number of data mining problems for high-throughput biological data, including cluster identification in gene expression profiles, regulatory binding site identification, two-hybrid protein-protein interaction data, and signal transduction pathway construction.

Introduction

The mass production of biological data is unprecedented, especially with high-throughput protocols at the genome scale. Types of data include (1) sequence data generated by the Human Genome Project and other sequencing efforts, (2) structural data produced by the NIH Structural Genomics Initiative, (3) gene expression data from microarray experiments, (4) mass-spectroscoopy data for proteins, and (5) two-hybrid data.

In addition to the experimental data, huge amounts of computational data are also being produced through annotations and predictions. The rates of data generation far exceed our current capability of analysis and interpretation. Analyzing these data is highly challenging, since they are often noisy and incomplete. New ideas and approaches are urgently needed to establish greatly improved capabilities for biological data analysis. Data clustering is fundamental to mining a large quantity of biological data. Many problems in mining and analysis of high-throughput data are clustering or they can be formulated as clustering problems. The main objective of this project was to develop generic clustering methods to help biological discovery using the high-throughput biological data. The research tasks included two parts: (1) developing algorithms for a general-purpose clustering method that can be applied to various biological data and (2) testing the algorithms on actual biological clustering problems.

Technical Approach

Our methods are based on the concept of minimum spanning tree. A *spanning tree* is a connected graph that contains every vertex but does not contain any cycle. A *minimum spanning tree (MST)* is a spanning tree with the minimum total distance between vertices. An MST can be efficiently constructed by classical Kruskals algorithm. In essence, our MST-based approach tries to capture our intuition about a cluster so that distances between neighbors within a cluster should be smaller than any inter-cluster distances. We have shown that each cluster of the multidimensional data corresponds to one subtree of the MST. This property can generally convert a multi-dimensional clustering problem to a tree-partitioning problem. In addition, MST-based clustering has higher tolerance to noise in the data than other clustering methods. Due to these properties, the MST representation is very suitable for general clustering problems.

Based on our early development in minimum spanning tree–based clustering, we discovered a new approach to identify clusters from noisy background based on a rigorous relationship between data clusters and a sequential representation derived from the Prim algorithm for constructing a MST of a data set.[1] This approach converts a clustering problem to a problem of partitioning a one-dimensional profile (a sequential representation), which facilitates easy identification of "dense" clusters from a noisy background. A statistical model was developed to assess the confidence level of a cluster. We have applied the method to a number of data mining problems for high-throughput biological data, including cluster identification in gene expression profiles and regulatory binding site identification. The results have shown that our generic clustering methods can identify underlying patterns from large-scale biological data effectively and efficiently.

Results and Accomplishments

MST-Based Clustering Methods and Implementation in the EXCAVATOR Package

We implemented the sequential representation–based cluster identification method, as described above. We also explored different features, such as different distance measurement between gene expression profiles, different objective functions, etc. We incorporated all the new developments into the computer package, EXCAVATOR[2] (Fig. 1, http://compbio.ornl.gov/structure/excavator/), for clustering gene-expression profiles. This program has a number of unique features, including capabilities for (i) data-constrained clustering, (ii) identification of genes with similar expression profiles to prespecified seed genes, (iii) cluster identification from a noisy background, and (iv) computational comparison between different clustering results of the same data set. EXCAVATOR can be run from a Unix/Linux/DOS shell, from a Java interface, or from a Web server. The clustering results can be visualized in colored figures and two-dimensional plots. Moreover, EXCAVATOR provides a wide range of options for data formats, distance measurements, objective functions, clustering algorithms, methods to choose number of clusters, etc. The effectiveness of EXCAVATOR has been demonstrated on several experimental data sets. Its performance compares favorably against the popular K-means clustering method in terms of clustering quality and computing time.

Application of the MST-Based Clustering Methods in Regulatory Binding Site Identification

Transcriptional binding sites are short fragments in the upstream regions of genes, to which transcription factors bind to regulate the transcription of genes into mRNA. Computational identification of transcriptional binding sites remains a challenging and unsolved problem though a great amount of effort has been put into the study of this problem. We have recently applied the sequential representation-based cluster identification method for identification of binding sites from a set of upstream regions of genes[3,4] that could possibly be transcriptionally co-regulated and hence might share similar transcriptional binding sites. By utilizing two key features of such binding sites (i.e., sequence similarities among them and their relatively high frequencies compared to other sequence fragments), we have implemented the cluster identification algorithm and a new statistical significance analysis method as a computer package CUBIC for regulatory binding site identifications. Extensive testing on CUBIC has been carried out.

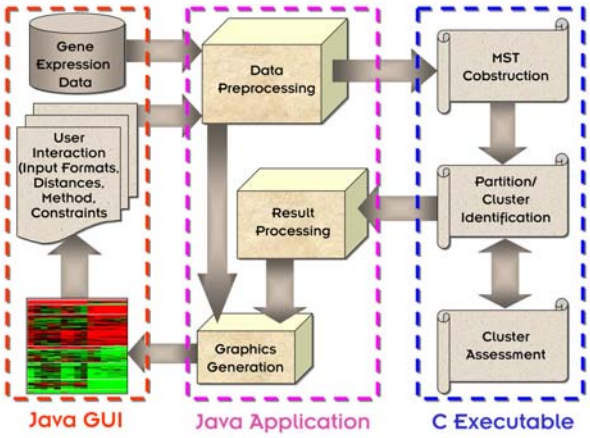


Fig. 1. The flowchart for the design of EXCAVATOR.

Application of the MST-Based Clustering Methods for Pathway Prediction

To further validate our generic clustering methods, we applied our methods in studying a particular biological system, that is, the signal transduction pathway in the amino acid and peptide transport in yeast *S. cerevisiae*. In collaborating with experimentalist Dr. J. Becker at the University of Tennessee, together with Yu Chen, a graduate student from the University of Tennessee doing his thesis under the supervision of Dong Xu, we constructed a global pathway model[5] using various information including data from DIP (http://dip.doe-mbi.ucla.edu), BIND (http://binddb.org), and gene expression data. Regulatory region analysis (clustering of binding site motifs) and gene expression clustering analysis validate the pathway model by showing how the selected transcription factors control the amino acid transporters and how the cross-talks between the amino acid transport pathway and the other related pathways are achieved (Fig. 2). We also extend this work to general signal transduction pathway construction using high-throughput biological data.[6,7] A Web server (http://compbio.ornl.gov/structure/expression/pathway) has been constructed for public use. This work demonstrates the power of our analysis methods for biological discovery from high-throughput biological data.

Summary and Conclusions

In this project, we developed MST-based clustering methods and successfully applied our methods to a number of data mining problems for high-throughput biological data, including cluster identification in gene expression profiles, regulatory binding site identification, and signal transduction pathway construction. The results of this work were incorporated into our software packages, that is, EXCAVATOR for clustering gene expression profiles, CUBIC for regulatory binding site identification, and a pathway prediction Web server. We also applied our methods in constructing the amino acid transport pathway model in yeast.

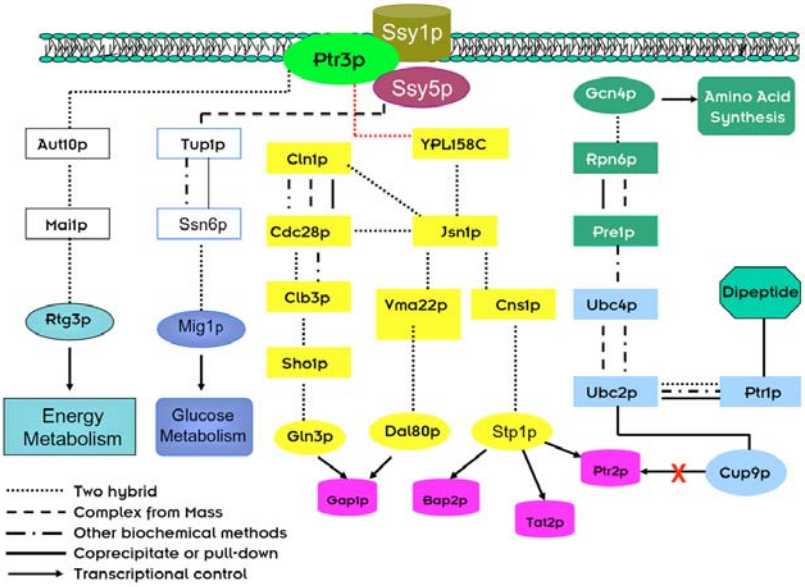


Fig. 2. Our constructed model for the signal transduction pathway of amino acid and peptide transporters in yeast, where the different shades of gray scale indicate different pathways. The different line patterns represent protein-protein interactions as identified by various experimental techniques. The ovals represent transcription factors, and the boxes represent intermediate proteins between the SPS sensor and transcription factors.

The above work resulted in seven publications listed as references. Based on our results, we have submitted an NSF proposal (Ying Xu, Victor Olman, and Dong Xu, entitled "Biological Data Clustering Using Minimum Spanning Trees"), which has been funded. This work also provided some foundation for other DOE-funded research projects, in particular, the U.S. Department of Energy's Genomes to Life program (www.doegenomestolife.org) under project "Carbon Sequestration in *Synechococcus Sp.*: From Molecular Machines to Hierarchical Modeling" (www.genomes-to-life.org).

References

[1]V. Olman, D. Xu, and Y. Xu, "Solving Data Clustering Problem as a String Search Problem," in *Proceedings of Conference on Statistical Data Mining and Knowledge Discovery*, Hamparsum Bozdogan, ed., Chapman & Hall/CRC, in press.

[2]D. Xu, V. Olman, L. Wang, and Y. Xu, "EXCAVATOR: a computer program for gene expression data analysis," *Nucleic Acid Research* **31,** 5582–5589 (2003).

[3]V. Olman, D. Xu, and Y. Xu, "Identifications of Regulatory Binding Sites Using Minimum Spanning Trees," pp. 327–338 in *Proceedings of the 2003 Pacific Symposium on Biocomputing (PSB)*, 2003.

[4]V. Olman, D. Xu, and Y. Xu, "CUBIC: Identifications of Regulatory Binding Sites through Data Clustering," *Journal of Bioinformatics and Computational Biology* **1**, 21–40 (2003).

[5]Y. Chen, Y. Liu, K. M. Goldstein, J. M. Becker, Y. Xu, and D. Xu, "A Computational Study on the Signal Transduction Pathway for Amino Acid and Peptide Transport in Yeast: Bridging the Gap between High-Throughput Data and Traditional Biology," *Applied Genomics and Proteomics* **2**, 43–50 (2003).

[6]Y. Chen and D. Xu, "Computational Analyses of High-Throughput Protein-Protein Interaction Data," *Current Protein and Peptide Science* **4**, 159–181 (2003).

[7]Y. Chen, T. Joshi, Y. Xu, and D. Xu, "Towards Automated Derivation of Biological Pathways Using High-Throughput Biological Data," pp. 18–25 in Proceeding of the 3rd IEEE Symposium on Bioinformatics and Bioengineering (2003), IEEE/CS Press.

Nanostructures for Spatially Resolved Macromolecular Delivery to Viable Whole Cells

T. E. McKnight,[1,2] A. V. Melechko,[3] G. D. Griffin,[4] V. I. Merkulov,[1] M. A. Guillorn,[1,4] M. L. Simpson[1,4]
[1]*Engineering Science and Technology Division*
[2]*Condensed Matter Physics Division*
[3]*University of Tennessee, Center for Environmental Technology*
[4]*Life Sciences Division*

The goal of the project was to evaluate the potential for parallel delivery of DNA to groups of cells using arrays of vertically aligned carbon nanofibers (CNF). CNF arrays are comprised of individual, solid nanofiber elements, with tip diameters of tens of nanometers and lengths of several microns. The needlelike aspect of these fibers proves to be well suited for penetration and delivery of large molecules to the intracellular domain of targeted cells. Specifically, plasmid DNA is spotted and adsorbed onto localized subregions of nanofiber arrays. These arrays are then integrated with cellular matrices using a variety of methods including centrifugation of cells onto fibered substrates and even direct pressing of fibered substrates into monolayers of adherent cells. The subsequent desorption of DNA from the fibers into the intracellular, and potentially nuclear, domains of cells provides a highly efficient material delivery method that is easy to implement with a variety of cell lines and types.

Introduction

Among the many anticipated applications of CNF devices, their potential as cellular, subcellular, and even molecular-scale probes and actuators is perhaps the most intriguing. The size scale of nanofibers (20+ nm radius and micron lengths), their carbon-based nature, and the ability to synthesize them in vertically aligned arrays (Fig. 1) inspire their integration with cellular and tissue matrices to provide a massively parallel electrical, electrochemical, or biochemical interface for interrogation and control of cellular processes. While parallel interfacing has been reported with extracellular devices, nanofiber arrays provide an interface with the potential for massively parallel intracellular integration, where the fibers are directly inserted into live cells and used by the researcher to interrogate or manipulate the intracellular domains of these integrated cells.

In this project, we explored one approach to such a parallel cellular interface—the insertion, retention, and assimilation of nanofiber arrays in live cell matrices. We demonstrated this integration by using the fibers to introduce a new gene to the cellular targets, where gene expression (green fluorescent protein—GFP production) not only denotes successful fiber insertion but also continued viability of the interfaced cell. To further demonstrate this integration, we covalently bound the gene vector to the fibers and observed long-term expression via transcription off tethered plasmid within interfaced cells, thus verifying the continued presence of the nanofiber element within the intracellular domain and the continued biochemical manipulation of the interfaced cell.

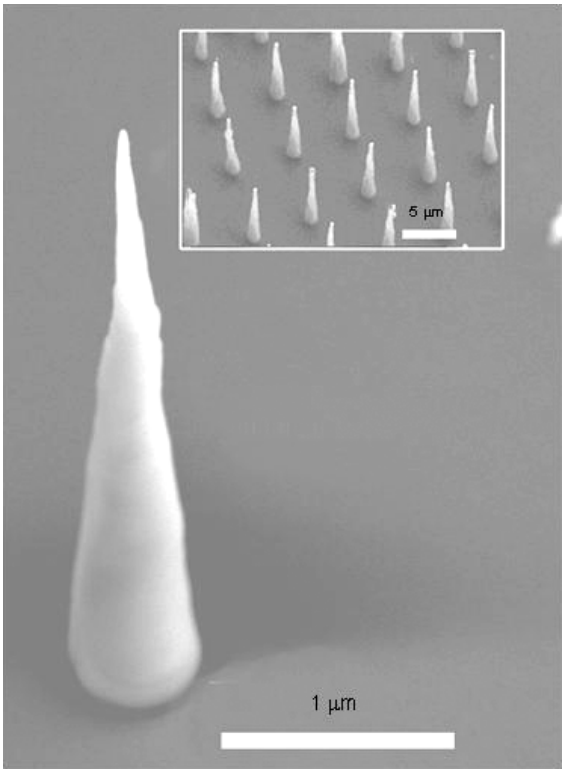


Fig. 1. Carbon nanofiber array photolithographically defined and synthesized in a plasma-enhanced chemical vapor deposition (PECVD) process. Fiber height, radius, location, and surface composition (carbon/nitrogen ratio) may all be defined based upon the initial catalyst thin-film layer and PECVD parameters, such as gas composition, temperature, and growth time. Here, a grid array of 5-µm-spaced, 10-µm-tall fibers with tip diameters of 30 nm was synthesized over a 50- × 50-mm region of a 100-mm silicon wafer.

Technical Approach

An enabling factor for the potential application of CNF as a cellular interface is the ability to synthesize nanofibers in a highly deterministic manner, at specific physical locations.[1] This ability provides the integration of nanoscale, and ultimately *molecular-scale*, elements (nanofiber tips) with mesoscale devices (i.e., silicon substrates), and, in turn, with the macroscale physical world of the researcher. Using a plasma-enhanced chemical vapor deposition (PECVD) process, carbon nanofibers may be synthesized at specific locations upon a substrate, as defined by the lithographic patterning of catalyst metals, such as Ni, Co, and Fe. In catalyst-"tip-based"[2] growth, the size and shape of a nanofiber may be well defined[3] and large arrays of vertically aligned nanofibers may be synthesized using large-area (100-mm wafer) processing techniques. In this effort, we synthesized nanofiber arrays from photolithographically defined nickel catalyst dots on 100-mm, n-type silicon wafers. Following synthesis, fibered substrates were cleaved into smaller pieces to enable convenient handling.

An overview of our methodology for integrating cells with fibers is shown in Fig. 2. In brief, cells were centrifugally pelleted out of suspension onto 3-mm × 3-mm fibered substrates oriented normal to the pelleting force. As such, cells were spun down and impaled upon the vertically aligned fibers grown on the substrate—similar to throwing a ball against a sparsely populated bed of nails [Fig. 2(a)]. Optionally, following the spin, the fibered substrate and the mass of cells upon it were gently pressed against a flat, wetted surface, such that the cells were further integrated with the fiber matrix [Fig. 2(b)]. The fibered substrate was then placed face up in growth media in a culture dish and incubated to allow cell recovery and proliferation [Fig. 2(c)]. The success of this method to provide for macromolecular delivery to intact cells was dependent upon adsorption and retention of plasmid vectors on the fiber surfaces and cell survival to the manipulations.

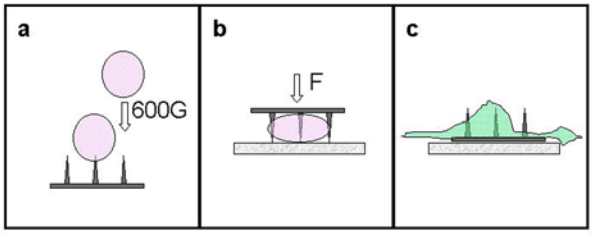


Fig. 2. Schematic of process for integration of adherent cells to a nanofiber array. (a) Cells are spun out of a suspension in phosphate buffered saline down onto a fibered substrate. (b) Following the spin down, the substrate may optionally be pressed against a wetted, flat surface. This further integrates the cells with the fibers, potentially pressing fibers deeper into the cell. (c) Following these manipulations, the fibered substrate and the cells upon it are placed face up in a culture dish, covered with growth media, and incubated for at least 24 h prior to fluorescent imaging.

Adsorption and Retention of Plasmid Vectors

Successful delivery of plasmid vectors to targeted cells using this fiber-mediated method requires that plasmid vectors be retained by fibers during interactions with the cell but remain available for transcriptional activity following the integration. We investigated a variety of coating methods to provide adsorption of DNA onto fibers, retention during cell interaction, and desorption following the integration procedure. We initially experimented with suspensions of plasmid in dilute solutions of agarose (0.1%) but also evaluated a variety of surface coatings on the fiber (sputtered-gold, plasma-enhanced chemical-vapor-deposited SiO_2), thermally reactive coatings (poly-*n*-isopropyl acrylamide, or pNIPAM), and DNA complexing agents (spermidine and $CaCl_2$ precipitation). Plasmids and plasmid cocktails were spotted directly onto fibered substrates with a micropipettor. Plasmid retention on fibers was evaluated qualitatively by staining with propidium iodide and observing the degradation of fluorescence of fibered regions over time when immersed in aqueous buffers. In general, the agarose method provides a very diverse coating methodology, relatively independent of the surface charge effects that often limit DNA/surface interactions. Dried spots of DNA in 0.1% agarose provide fairly uniform coating of fibered surfaces, only slightly diminishing the high-aspect feature of the nanofibers. The agarose also helps to provide a slow release of the DNA from the fiber surface. Unfortunately, the presence of agarose on fibers during interaction and penetration of fibers into cells appears to considerably stress cells—resulting in poor subsequent growth and markedly different morphology. It is unclear if this is a cytotoxic effect of the agarose, or merely an artifact of the ability of the cell to reseal following interaction with an agarose-coated fiber. Surface treatments of fibers can also appreciably affect DNA retention. In general, the native carbon nanofiber is a negatively charged surface. The negatively charged CNF surface will tend to electrostatically repel the anionic phosphate backbone of non-complexed DNA in dilute solutions. Dilute solutions of DNA will coat fibers upon evaporation but will quickly be shed upon rehydration. This is unacceptable for material delivery, as plasmid vectors will be lost to solution upon wetting with the cellular suspension. However, this repulsive effect can be decreased by complexing the DNA with cations (i.e., Ca^{++}), or by modifying the surface charge of the fiber with coatings—such as a sputtered gold or tungsten.

DNA may also be complexed to fiber surfaces using a variety of techniques that have been developed in the microprojectile bombardment community, where plasmid-

coated microprojectiles are physically accelerated onto plates of recipient cells. In these methods, DNA is typically complexed to the particle using a cationic lipofective reagent (i.e., spermidine) and a highly concentrated precipitant ($CaCl_2$). Of all the methods employed, the agarose and spermidine/$CaCl_2$ treatments resulted in the best retention of plasmid DNA on native CNF. However, each method merits further investigation, particularly the controlled release methodologies.

Finally, in addition to post-synthesis modification of CNF, the fibers themselves may be synthesized with higher nitrogen content by simply altering the ratio of carbonaceous gas and the nitrogenated etch gas during PECVD synthesis. Highly nitrogenated fibers may be synthesized that can bear an overall positive charge in plasmid solutions, and thus electrostatically retain plasmid DNA without requiring any additional processing steps.

Another feature of nanofiber tips is a short segment below the catalyst of uncontaminated carbon, typically approximately 100 nm long. Elemental (electron dispersion X-ray) analysis reveals that even with a highly nitrogenated source gas mixture which will produce heavily nitrogen-doped fibers, the extreme tip of the fiber will be virtually nitrogen free. Modification of this region can thus provide another effective handle for DNA retention. One example is the modification of these carbon-rich regions with an oxygen plasma, thus producing oxide functionalities within the carbon tip, including carboxylic acids. With these handles, we have demonstrated that DNA may be covalently bound to the fibers using an EDC-mediated [1-ethyl-3-(3-dimethylaminopropyl) carbodiimide] condensation reaction between the fiber carboxylic acids and the DNA base amines.

Plasmid Delivery and Cell Survival

Delivery and expression of GFP coding plasmids was highly dependent on a number of factors including fiber geometry (length and diameter), spacing, and chemistry. In general, untreated carbon fibers were not successful with plasmid delivery. However, nitrogen-containing fibers, exposed catalyst fibers, and plasmid-derivitized (covalent) fibers all provided some amount of GFP+ expression in targeted cells (Fig. 3). The failure of untreated carbon fibers appears to be due to poor retention of DNA on the fiber surface.

Centrifugation of cells at 3000 rpm (150 G) for 1 min onto plasmid-coated and plasmid-derivitized nitrogen-bearing fibers resulted in transient expression of green fluorescent protein at efficiencies of approximately 1% (where efficiency is defined as GFP+ cells vs total number of cells on the fibered substrate). These efficiencies could be improved by approximately a factor of 5 if, following the spin, the cell/fiber matrix is pressed lightly against a flat, wetted surface such as a microscope slide or the bottom of a tissue culture dish. Efficiencies with fibers that had been spotted with DNA after emancipating the nickel catalyst at the tip resulted in even higher transfection efficiencies using this spin and press method, with typically 20–50% of cells on the substrate expressing GFP during subsequent evaluations 24+ hours following the transfection experiment. In each of these scenarios, transfection efficiency is based on GFP+ cells vs total cells on the fibered substrate, and not GFP+ cells vs total number of cells spun down during the experiment.

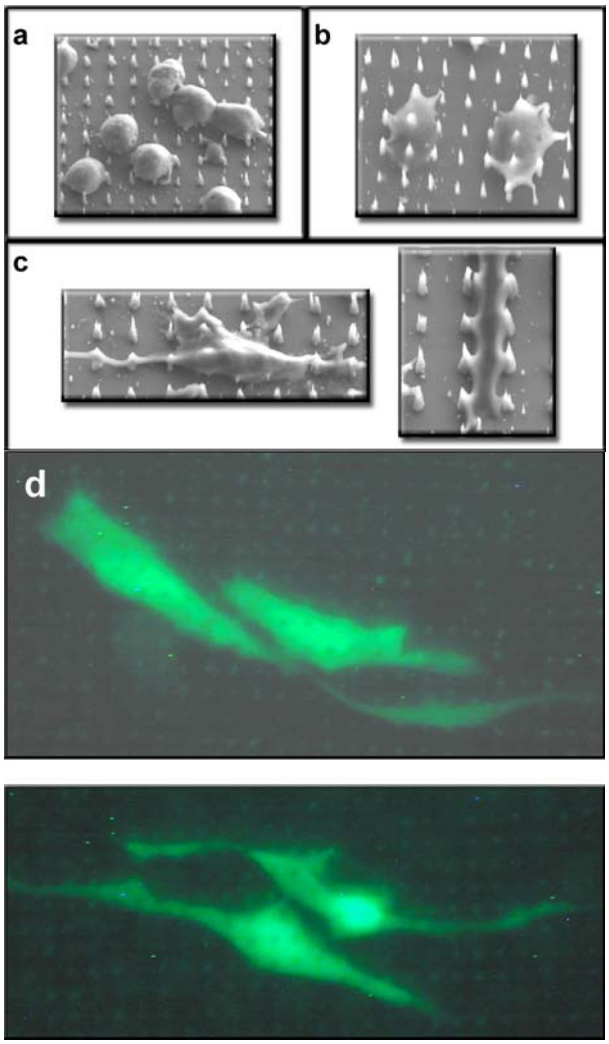


Fig. 3. Scanning electron micrographs of cells following (a) pelleting, (b) press, and (c) culture for 48 h. (d) Production of GFP from the delivered plasmid can be observed as early as 4 h following the integration. Here, GFP+ cells are observed with fluorescent microscopy at 48 h. Large colonies may be observed after several weeks of culture, apparently due to stable genomic insertion of the delivered plasmid.

Covalent attachment of plasmid DNA to fibers and integration with cell matrices also results in plasmid expression within integrated cells. In reaction volumes of 1 µg/mL of plasmid, transfection efficiencies were

approximately 1% using the method described in the paragraph above. These efficiencies could be dramatically improved (10–30% efficiency) by reducing the reaction volume (increasing plasmid concentration) during EDC condensation of DNA to fiber scaffolding.

Expression of GFP following cellular integration with nanofiber arrays has been followed for at least 30 days, indicating long-term survival of interfaced cells.

Results and Accomplishments

1. Demonstration of intracellular integration of arrays of CNF with cellular matrices and long-term biochemical modification of the integrated cells (expression of GFP from delivered plasmid vector).
2. Demonstration of a relatively high efficiency (up to 50%) macromolecular delivery system that is easy to implement with a variety of mammalian cell lines.
3. Development of nanofiber synthesis methods that result in nitrogen-doped fibers that are well suited for retention of anionic macromolecules (DNA) during cellular interactions and subsequent desorption of these molecules within the intracellular domain.
4. Development of a new CNF growth technique capable of producing arrays of single CNFs using only photolithographic definition of catalyst patterns. This eliminates the previous need for electron-beam lithography—thus providing a scalable nanofiber growth method, no longer limited by serial e-beam lithographic processes.
5. Demonstration of covalent linkage strategies for tethering macromolecules (plasmid DNA) to discrete regions of carbon nanofibers.
6. Demonstration of transcription from covalently bound plasmid tethered to nanofiber arrays.

Summary and Conclusions

This project has successfully demonstrated that CNF arrays may be integrated with cellular matrices to provide a massively parallel biochemical interface to live cells. This interface may be utilized as a macromolecular delivery vehicle, as demonstrated via the delivery and expression of plasmid DNA or, by leveraging previous work by our research group,[4] as an electrical/electrochemical interface. In addition to providing an adsorption/desorption–based delivery method, nanofibers may be synthesized or modified to feature rich surface chemistries that may be used for covalent attachment of macromolecules. We have demonstrated that covalently bound plasmid on fibers may be introduced to the intracellular domain—and transcribed while remaining tethered to the fiber scaffold. Transcription from tethered plasmids provides the ability to exploit cells as transcriptional factories, while minimizing the potential for escape of plasmid or genetically modified organisms to the environment. Finally, as physical (i.e., microinjection-based) approaches, these techniques should be applicable over a diverse array of cell types, potentially extending even into the plant and microbial kingdoms.

The following proposals have been generated as a direct result of this seed money effort:

NIH EB-03-011, *Synthetic Nanostructure Gene Delivery Vectors*, April 2003. $250K requested.

NIH PAR-03-045, *Nanoscale Architectures for Intracellular Biochemical Manipulation*, August 18, 2003. $1.8M requested.

NIH PA-03-058, *Nanostructured Gene Delivery Arrays,* October 1, 2003. $250K requested.

Joint USDA/DOE Biomass Program, *Enhanced Biomass Production: Novel Transformation Approaches to Engineering the Cellulose Content of Loblolly Pine,* collaborative proposal between ORNL, the Institute of Paper Science and Technology, North Carolina State University, and Clark Atlanta University, April 2003. $1.5M requested ($350K ORNL).

References

[1]V. I. Merkulov, D. H. Lowndes, Y. Y. Wei, G. Eres, and E. Voelkl, *Appl. Phys.Lett.* **76**, 3555 (2000).

[2]V. I. Merkulov, A. V. Melechko, M. A. Guillorn, M. L. Simpson, and D. H. Lowndes, "Alignment Mechanism of Carbon Nanofibers by Plasma-Enhanced Chemical Vapor Deposition," *Applied Physics Letters* **79**(18), 2970–2972 (October 2001).

[3]V. I. Merkulov, M. A. Guillorn, M. L. Simpson, D. H. Lowndes, and E. Voekl, "Shaping Carbon Nanostructures by Controlling the Synthesis Process," *Applied Physics Letters* **29**(8), 1178–1180 (Aug. 20, 2001).

[4]M. A. Guillorn, T. E. McKnight, A. V. Melechko, D. W. Austin, V. I. Merkulov, M. L. Simpson, and D. H. Lowndes, "Individually Addressable Vertically Aligned Carbon Nanofiber-Based Electrochemical Probes," *J. Appl. Phys.* **91**, 3824 (2002).

Prediction of Sepsis Onset in Trauma Victims Using Advanced Nonlinear Analysis of Time-Serial Physiological Data Including Heart Rate Variability

L. M. Hively,[1,2] V. A. Protopopescu,[3] and N. B. Munro[4]
Collaborators: M. D. Karlstad,[2] W. M. Vassy, M.D.,[2] B. Enderson, M.D.,[2]
B. J. Daley, M.D.,[2] and W. Holland, R. EEG. T[5]
[1]*Computational Sciences and Engineering Division*
[2]*Department of Surgery, UTMCK*
[3]*Computational Science and Mathematics Division*
[4]*Life Sciences Division*
[5]*Department of Neurology, UTMCK*

The primary cause of death in trauma patients is sepsis and septic shock. This situation has not improved significantly over the past 30 years for several reasons. First, reliable predictors for sepsis do not exist to allow early interventions. Second, the patient's condition often deteriorates rapidly after sepsis onset, even with aggressive medical interventions, frequently causing the failure of multiple organ systems and eventually leading to death. Also, recent medical improvements have been offset by the dramatic increase in antibiotic-resistant infections, which are hard to diagnose and harder still to fight. This work is the first step towards sepsis forewarning by the acquisition and analysis of time-serial physiological data. Physiological (ECG) data from endotoxin-exposed rats were acquired by collaborators at the University of Tennessee Medical Center in Knoxville (UTMCK). Oak Ridge National Laboratory (ORNL) analyzed the data via nonlinear dynamical methods, which previously have provided up to an hour's forewarning of epileptic seizures from scalp brain waves. The results show a clear superiority of the phase-space dissimilarity measures in comparison to traditional nonlinear measures, in terms of robustness, timeliness, and discriminating power.

Introduction

Despite considerable advances in treatment, sepsis remains a common source of morbidity and mortality in surgical patients. Early treatment of sepsis is paramount for minimizing complications and preventing death. Unfortunately, timely *identification* of sepsis development prior to clinically detectable onset remains an outstanding challenge. Complex systems analysis of physiological data is a new and powerful computational method to discover interactions and emergent properties of physiological systems. These methods provide a means to study systemic properties of disease and injury in critical care patients for evaluation and differentiation of their complex nonlinear systems. Such complex nonlinear systems are ubiquitous; they include weather patterns, the stock market, biological systems including the central nervous system, and the complex host response to trauma, sepsis, or shock. For example, a change in heart rate (HR) variability, measured as the change in interbeat intervals, is thought to be due to changes in the sympathetic and parasympathetic control of the heart. Standard complex systems analysis methods have shown that reduction in HR variability is associated with increased mortality rate after myocardial infarction,[1] sepsis, and critical illness in pediatric patients.[2,3] Blood pressure variability is also reduced in septic shock.[3] Griffin and Moorman[4] showed reduced baseline HR variability and short-lived decelerations in HR up to 24 h before clinically observable deterioration in a prospective study of infants at risk of sepsis. Heart rate variability has been shown to increase during recovery from sepsis in pediatric patients.[1,2] Advanced nonlinear analysis of electrocardiogram (ECG) data has been shown to forewarn of cardiac fibrillations[5,6] but has not been applied to sepsis detection or prediction. Improved interpretations of physiological data have recently been reenergized by advances in nonlinear and chaotic dynamics[7–10] (and references therein). Here we show that advanced nonlinear analysis of ECG data can be used for pre-symptomatic detection of sepsis.

Technical Approach

UTMCK initially acquired four-channel ECG patient data at a sampling rate of 500 points per second. ORNL's analysis revealed 1- to 2-h gaps during off-line procedures, after which the nonlinear signatures were dramatically different. Consequently, the team agreed that human data were inappropriate for this study, due to the confounding effects of patient treatments.

Subsequently, UTMCK acquired the same data from 23 male Long-Evans rats, which were anesthetized via 2% inhaled isoflurane. The experimental protocol began with a 30- to 60-min acclimation period in a nose-only inhalation chamber, followed by exposure of 17 rats to aerosolized *Salmonella enteriditis* endotoxin for 30 min. Subsequent monitoring of the physiological response continued for 1 to 2 h. The bacterial endotoxin induces an inflammatory response, which should be detectable as changes in the ECG data under our hypothesis. Six other rats were exposed only to deionized water for 30 min, as controls. Table 1 characterizes the 23 datasets. The immune response was verified subsequently by lung lavage and counting of neutrophils in the lavage fluid. Further verification was accomplished by histopathological examination of lung sections after removal at euthanization.

Data were analyzed by the ORNL-developed phase space dissimilarity measures (PSDM) approach.[5,11] The method first removes confounding artifacts (breathing artifacts in ECG data) by a zero-phase, quadratic filter. The artifact-filtered data then are converted to a discrete distribution function (DF) that captures the essential features of the underlying dynamics. Condition change assessment compares test case DFs to a base case (normal state) DF via novel measures of dissimilarity. Forewarning is indicated by several sequential occurrences of the dissimilarity measures above a threshold. While traditional nonlinear measures compare averaged (global) quantities, the enhanced discrimination power of our measures is achieved by focusing on the absolute difference between the two DFs. This last feature enables biomedical event forewarning from (near) real-time, noisy biomedical data of numerous types.

Table 1. Characterization summary

Date	Exposure	ID	T0	T1	T2	T3	SZ(MB)
08/15	Endotoxin	9V	15:23:20	15:53:14	16:23:54	17:24:37	131325
08/20	Endotoxin	9W	13:35:35	14:35:55	15:05:55	17:05:37	227619
08/22	Endotoxin	9X	12:46:34	13:46:31	14:16:29	16:16:47	227818
08/23	Endotoxin	9Y	12:39:38	13:39:33	14:09:38	16:09:41	227637
08/29	Endotoxin	9Z	13:59:28	14:59:35	15:29:32	17:29:27	227565
09/10	Endotoxin	A1	12:09:22	13:09:17	13:39:17	15:39:21	227565
09/12	Endotoxin	23	12:55:38	13:55:31	14:25:31	16:25:34	221531
09/13	Endotoxin	A4	13:26:49	14:26:46	14:56:46	16:56:50	227601
10/08	Endotoxin	A5	12:57:13	13:57:18	14:27:08	16:27:11	227547
10/09	Endotoxin	~1	11:17:26	12:17:20	12:47:39	14:31:13	227565
10/10	Di H_2O	~2	12:31:15	13:31:30	14:01:07	16:01:25	227764
10/15	Di H_2O	A9	15:04:07	15:44:01	16:14:08	18:14:43	206554
10/17	Di H_2O	AA	11:39:41	12:39:36	13:09:36	15:09:40	227565
10/23	Di H_2O	BD	11:34:31	12:34:34	13:04:28	15:04:29	191108
10/24	Di H_2O	AE	12:29:39	13:29:34	13:59:34	15:59:37	227547
10/28	Endotoxin	AF	11:21:31	12:21:28	12:51:25	14:51:29	227547
11/05	Di H_2O	AG	14:29:12	15:29:12	15:59:12	17:59:15	227529
11/11	Endotoxin	AH	13:10:21	14:11:09	14:40:28	16:40:20	227565
11/13	Endotoxin	AI	11:29:35	12:29:30	12:59:30	14:59:37	228071
11/14	Endotoxin	AJ	12:46:12	13:46:06	14:16:07	16:16:10	227547
11/19	Endotoxin	AK	12:33:33	13:33:28	14:03:30	16:03:43	227764
11/20	Endotoxin	AL	12:19:27	13:20:37	13:52:41	15:49:25	227547
11/21	Endotoxin	AM	11:37:52	12:37:47	13:07:47	15:07:50	227457

DATE = 2002 date of this experiment in MM/DD format.
EXPOSURE = either endotoxin or de-ionized water.
ID = dataset identifier.
T0 = starting time of the experiment in HH:MM:SS format.
T1 = beginning of the exposure period.
T2 = end of the exposure period.
T3 = finish time of the experiment.
SZ = size of the dataset in megabytes (MB).

Results and Accomplishments

The 17 endotoxin-exposed animals showed significant post-exposure condition changes in ECG dynamics, as compared with 6 control animals. Figure 1 shows results for one endotoxin exposure case. The raw ECG data in subplot (a) have 14 distinct heartbeats with additional quasi-periodic (nonlinear) features. No indication of condition change is displayed by either the correlation dimension in subplot (b), or by the Kolmogorov entropy in subplot (c). Condition change is shown clearly by both $U(\chi^2)$ in subplot (d) and $U(L)$ in subplot (e), which remain low, then rise abruptly after the exposure onset, remain high for the next 20 minutes, then decrease slowly, presumably as the immune response counteracts the bioagent effects. Table 2 summarizes the analysis results for all 23 animals.

The phase-space dissimilarity measures provide indication in at least one of the four ECG channels for all 17 of the endotoxin-exposure cases. The recovery response is consistent with other physiological observations during the test. Pulmonary inflammation in endotoxin-exposed animals was verified by increased neutrophils in bronchoalveolar lavage ($13.2 \pm 4.2\%$ versus $3.6 \pm 1.8\%$, $p < 0.02$). The phase-space dissimilarity measures give no indication in at least one of the ECG channels for all six of the non-exposure cases. The earliest detection of condition change occurs within 1 h of the exposure onset in 16 datasets. Dataset #19 (AI) shows condition change after 91 min. The criterion for condition change requires several sequential occurrences, N_{OCC}, above a threshold value, U_{TH}, for a specified number, N_{SIM}, of renormalized PSDM in any one of the four ECG channels. This criterion is a very weak one[11] because no one channel provides adequate indication. Moreover, Table 2 shows instances of correct indication in one channel that is accompanied by contrary indications in other channel(s), including later indication time(s), as well as false positive (FP) and/or false negative (FN) indications. While this initial success is encouraging, much additional work is necessary to provide channel consistency, as well as robustness and reliability that is needed for clinically useful diagnostics.

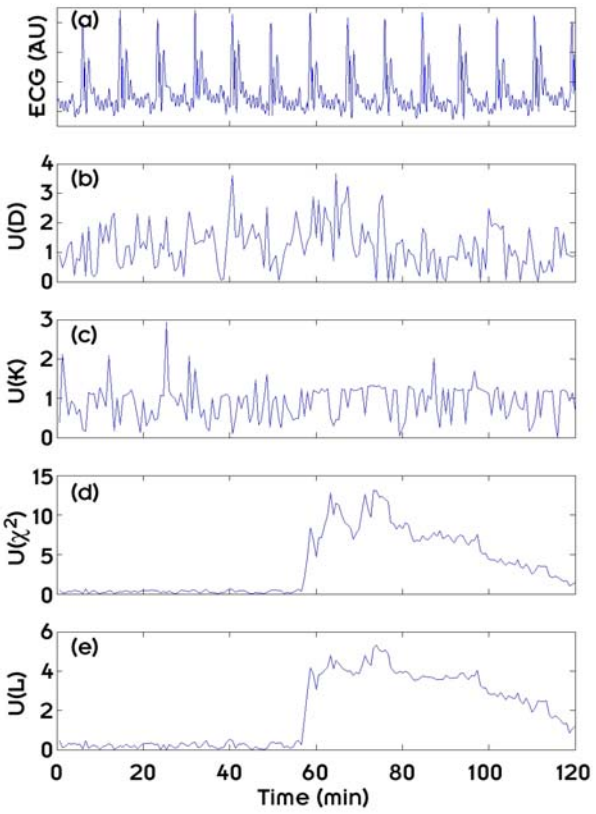


Fig. 1. Results for data set #33209V, showing time-serial plots for (a) 2.4 s of ECG data collected at 500 Hz; (b) correlation dimension, D; (c) Kolmogorov entropy, K*; (d)* U(χ^2)*; and (e)* U(L)*. The phase-space dissimilarity measures in subplots (d) and (e) were computed for d = 2, S = 2, λ = 30, after removal of breathing artifacts. Each cutset has* N *= 20,000 points, corresponding to 40 s.*

Table 2. Summary of results

CS	ID	T3-T2	T3-T0	Indication time			
				CH1	CH2	CH3	CH4
1	9V	1794	5332	270	FN	FN	FN
2	9W	3620	10707	1153	FP	5238	FN
3	9X	3597	10750	FP	FP	488	FP
4	9Y	3595	10750	705	2769	FP	4059
5	9Z	3607	10707	1768	FN	FP	435
6	A1	3595	10750	FP	FN	FP	1952
7	23	3593	10406	1997	FN	FP	FP
8	A4	3597	10750	FP	FP	832	FP
9	A5	3605	10707	FP	2458	FN	2071
10	~1	3594	10750	FP	FP	319	4275
11	~2	0	10707	FP	TN	FP	TN
12	A9	0	10750	TN	TN	FP	TN
13	AA	0	10750	TN	FP	FP	TN
14	BD	0	8729	TN	TN	TN	FP
15	AE	0	10707	FP	FP	TN	FP
16	AF	3597	10750	FP	FP	FP	3240
17	AG	0	10707	FP	FP	TN	TN
18	AH	3648	10707	4350	FN	FN	781
19	AI	3595	10707	FN	5478	FN	FN
20	AJ	3594	10750	FP	FN	FN	663
21	AK	3595	10750	FP	FN	361	FN
22	AL	3670	10750	7080	802	FP	FP
23	AM	3595	10664	490	FP	FP	FN

CS = case number (1 through 23).
ID = identifier from Table 1.
T3-T2 = exposure time (seconds) from Table 1.
T3-T0 = total duration (seconds) of the experiment from Table 1.
CHn = ECG channel number, with the value of n between 1 and 4.
FN = false negative (no indication when exposure occurred).
FP = false positive (indication of change without exposure, or indication of change before exposure).

Summary and Conclusions

Using ORNL's previously developed model-independent analysis to detect condition change (dissimilarity) in nonlinear time series, we examined the ECG data from endotoxin-exposed rats. The phase space indicators of condition change measure the difference between suitably discretized forms of the invariant distribution functions on the attractor for the base case and test case, as χ^2 statistics and L_1 distance. Use of these metrics magnifies the differences between the process dynamics, and avoids the inner cancellation effects due to averaging that occur when calculating the traditional nonlinear measures (TNM). As a result, significant changes in model dynamics are clearly detected by the PSDM, as the parameters vary. On the other hand, these changes are either barely detected or undetected by the TNM, such as the Kolmogorov entropy or the correlation dimension (see subplots b–c in Fig. 1). These preliminary results show that advanced nonlinear analysis of rat ECG data is a sensitive and specific identifier of inhalational endotoxin exposure. This novel tool represents a significant step toward identifying early sepsis in the hospitalized patient. Specifically, the novel nonlinear measures of rat ECG show small values of the PSDM in the normal state, followed by a notable monotonic rise and/or by successive occurrences of the PSDM above the "normality threshold" that indicates forewarning or onset of sepsis, as shown in Fig. 1. The renormalized PSDM are consistently more robust and timely than those provided by the TNM. An important next step is multichannel analysis to strengthen the criteria for correct forewarning.

ORNL stands to benefit substantially from this work. This project has laid the groundwork for further investigation of changes in nonlinear dynamics associated with sepsis onset for the purpose of early identification and treatment of sepsis. This work has fostered a serious collaboration between ORNL and investigators at UTMCK including Dr. Michael Karlstad and several trauma surgeons. A proposal to the National Institute of General Medical Sciences (NIGMS) of the NIH will be prepared. Dr. Scott Somers of NIGMS has indicated to Dr. Karlstad that such a proposal would be received with great interest.

References

[1]D. Toweill et al., "Linear and Nonlinear Analysis of Hemodynamic Signals During Sepsis and Septic Shock," *Crit. Care Med.* **28**(6), 2051–2057 (2000).

[2]M. S. Ellenby et al., "Uncoupling and Recoupling of Autonomic Regulation of the Heart Beat in Pediatric Septic Shock," *Shock* **16** (4), 274–277 (2001).

[3]D. T. Doweill and B. Goldstein, "Linear and Nonlinear Dynamics and the Pathophysiology of Shock," *New Horizons* **6** (2), 155–168 (1998).

[4]M. P. Griffin and J. R. Moorman, "Toward the Early Diagnosis of Neonatal Sepsis and Sepsis-Like Illness Using Novel Heart Rate Analysis," *Pediatrics* **107**(1), 97–104 (2001).

[5]L. M. Hively and V. A. Protopopescu, "Detection of Changing Dynamics in Physiological Time Series," *Proc. ANS Topical Conf. on Nucl. Math. Comput. Sci.* (Gatlinburg, Tennessee) April 6–10, 2003.

[6]J. E. Skinner et al., "Application of Chaos Theory to a Model Biological System: Evidence of Self-Organization in the Intrinsic Cardiac Nervous System," *Integr. Physiolog. Behav. Sci*. **31,** 122–146 (1996).

[7]J.-P. Eckmann and D. Ruelle, "Ergotic Theory of Chaos and Strange Attractors," *Rev. Mod. Phys.* **57**, 617–656 (1985).

[8]E. A. Jackson, *Perspectives of Nonlinear Dynamics*, Cambridge University Press, Cambridge, vol. 1989; vol. 2, 1990.

[9]S. H. Strogatz, *Nonlinear Dynamics and Chaos*, Addison-Wesley, Reading, 1994.

[10]H. D. I. Abarbanel, *Analysis of Observed Chaotic Data*, Springer Publ., New York, 1996.

[11]L. M. Hively and V. A. Protopopescu, "Channel-Consistent Forewarning of Epileptic Events from Scalp EEG," *IEEE Trans. Biomed. Engr.* **50**(5), 584–593 (May 2003).

Genetic Variation in Cellular Responses to Low-Dose Radiation

E. M. Rinchik,[1,2] K. T. Cain,[1] L. Branstetter,[1] and P. R. Hoyt[1,3]
[1]*Life Sciences Division, Oak Ridge National Laboratory*
[2]*Department of Biochemistry, Cellular, and Molecular Biology*
[3]*Graduate School of Genome Science and Technology, The University of Tennessee*

The molecular responses of cells exposed to biological stimuli or insults (such as radiation, or a chemical or biotoxin) vary with the type of exposure, dose and dose rate, the type and genetic constitution of exposed cell(s), and the tissue/organ site of the responding cell. Our research goal is to use genomic technologies to address how individual genetic constitution affects tissue responses to insults, starting with response to low-dose radiation (LDR). This Seed project began standardizing conditions for validating the whole-genome, molecular LDR response, using microarray technology and quantitative gene-expression techniques, in tissues from genetically diverse inbred mouse strains. These experiments will build a conceptual and technical framework for future applications of mouse genetics to studying variation in gene responses that may underlie differences in human susceptibility to toxic exposures.

Introduction

A primary mission of the DOE's Life Sciences Program historically has been the assessment of hazards from radiation. Cells exposed to radiation undergo a number of complex molecular responses that result in some kind of cellular consequence or action. Although some responses are shared between high- and low-dose exposures, other responses differ. The DOE Low-Dose Radiation Research Program (http://lowdose.org/) is interested in the cellular and molecular consequences of exposure to low-dose (or low-dose-rate), low–linear energy transfer radiation (LDR). It has become increasingly clear over the past 10 years that there are specific cellular and molecular responses to LDR (defined here as ≤10 cGy) and that the tools of molecular biology, as well as newer and evolving tools of genomics, can be exploited to study the highly complex LDR response. In this project, we focused on how an individual's molecular tissue response to LDR may vary because of differences in individual genetic constitution.

Much of the work previously done on the molecular LDR response has been done in vitro, on cell lines in tissue culture; some, but not nearly enough, has been done in whole animals or complex tissues. It is known that exposure of cells to low doses of ionizing radiation induces or represses certain gene products, some with very rapid kinetics (i.e., within hours).[1–3] A number of different responses, feed-in points, and effector mechanisms may characterize LDR responses, and responses in tissues in whole animals may differ from those in cultured cells. Moreover, variation in any of a number of gene products, even minor variation, due to differing genetic constitutions among unique individuals, may have significant effects on molecular responses as well as on overall outcomes.

This project was designed to provide preliminary data for using techniques and resources of mouse genetics and genomics to investigate issues of genetic susceptibility (in this specific case, to low-dose radiation). The project concentrated on (1) building a substantial bank of tissues from genetically different animals exposed to both high- (200-cGy) and low- (10-cGy) doses of acute X rays, a significant investment that could be drawn on for all future LDR-related genomics projects; (2) standardizing global genomic measurements of the LDR response reproducibly on site, at high throughput; and (3) gaining experience with data-analysis issues to identify bioinformatics needs relevant to DOE and other federal agencies interested in low-dose radiation exposure.

Technical Approach

We used DNA microarray technology[4] and other quantitative techniques for measuring gene expression among individuals (represented by different inbred or mutant strains of mice). If we could identify such LDR-response differences, we could subsequently define gene-expression profiles associated with LDR, and, importantly, learn how these profiles can vary from individual to individual. We began the project by using DNA microarrays with 1152 genes ("1K chip") enriched in skin-specific genes and oncogenes to analyze RNA samples (representative of the gene-expression compartment) from irradiated and control tissues. As the project progressed,

we introduced the use of oligonucleotide microarrays representing 22,000 mouse genes, as well as initiated development of quantitative polymerase chain reaction (PCR) (qPCR) technology to verify individual microarray results and to assay specific genes previously known to be involved in cellular stress-response pathways.

Results and Accomplishments

Tissue Bank

One project goal was to create a tissue bank from both high- and low-dose-exposed C57BL/6J and DBA/2J tissues. Almost 2000 individual samples were collected, and all exist in storage either as tissues preserved for subsequent RNA preparation, or as prepared RNA samples. Table 1 provides a general summary of what constitutes this tissue bank as of the end of the project.

Initial Microarray Experiments

At the beginning of the project, we grew, amplified, purified, and verified 1152 clones and controls to be used as the "1K" microarray for this project. This array included approximately 40 skin-specific genes, and at least 100 genes involved in the progression of cancer.

Preliminary results showed that exposing mice to a 10-cGy dose of acute X rays did produce a change in the expression of genes in skin. Results of typical microarray data utilizing the ORNL "1K" chip for 10-cGy exposures in C57BL/6JRn (B6) mice are shown in Fig. 1.

Two general conclusions can be drawn from these studies. First, a transcriptional response due to LDR was detected using our microarrays, as indicated by the subset of genes differentially expressed at >95% confidence levels. The second conclusion is that the LDR response involves unexpected cellular systems in vivo as judged by the lack of "predicted" genes, known to respond in vitro[1–3] (e.g., *Cdkn1a*). We then moved to compare (by microarray) transcriptional responses in dorsal skin among distinct strains B6, C57L/6-p^{un}/p^{un} (abbreviated PUN), and C3H-*a/a* (abbreviated C3H-a). B6 and PUN differ by only one gene, with PUN having diluted amounts of eumelanin pigment in the skin. We were very interested in assaying the effects of pigmentation on the LDR response in skin, as it is known that melanin pigment can protect against higher doses of X rays in cells in culture. (We also bred a large population of B6-Tyr^{c-2J} animals, which are essentially B6 mice with no pigment, but could not analyze these within the project period; however, they are included in the tissue bank in Table 1.) B6 and C3H-a are both black (having black eumelanin but no yellow phaeomelanin in the skin) but have entirely different genetic backgrounds and could differ by thousands of minor variants. Thus, we could compare LDR responses due to variation in genes other than those involved in pigmentation. Preliminary results from these studies are summarized in Table 2.

At 95% confidence levels, each mouse strain displayed a nonidentical, but overlapping, set of LDR-responsive genes. When the strains were compared, C3H-a showed the highest number of genes differentially expressed and also showed the highest number of unique LDR responses (genes responding only in C3H-a). An interesting aspect of these data was the high number of response genes with opposite transcriptional effects. For example, there were 13 common genes found to respond in both B6 and C3H-a mouse strains, using 10-cGy radiation. However, 14 of those genes were upregulated in one strain, while being downregulated in the other. While the mechanism of LDR response is beyond the scope of this study, these differences certainly warrant further study. It should be pointed out that the extensive unique response in C3H-a as compared to the B6 strain could be the result of many genetic differences, or could simply be due to one major difference that reverberates through one or more LDR-response pathways. Such preliminary data will be important in considering future projects.

These data were also examined by looking at those genes with the highest levels of differential expression in order to ascertain their usefulness as strain-specific, robust biomarkers for LDR response, and to provide preliminary data for future funding opportunities. These results are summarized in Table 3.

Table 1. General summary of current bank of tissues from mice of various genotypes exposed to both low (10-cGy) and high (200-cGy) doses of acute X rays

Number of samples in tissue bank	1939 total samples
Strains represented	A/J; C57BL/6J (B6); B6-Tyr^{c-2J}/Tyr^{c-2J}; B6-*Trp53–/Trp53–*; BALB/cJ; BALB-*Tyr+/Tyr+*; DBA/2J; C3H/HeJ, C3H/Rl; C3H/HeJ-*a/a*; B6-p^{un}/p^{un}; C3H-p^{J}/p^{J}; B6C3F1
Doses represented	Zero cGy (sham exposures); 10 cGy; 200 cGy
Tissues represented	Dorsal skin, spleen, testis—all strains Bone marrow, thymus—some strains
Time sampled after exposure	3.5 h—all strains and tissues 1 h—some strains, tissues, and exposures
Tissue preservation for RNA	All strains for all exposures and times

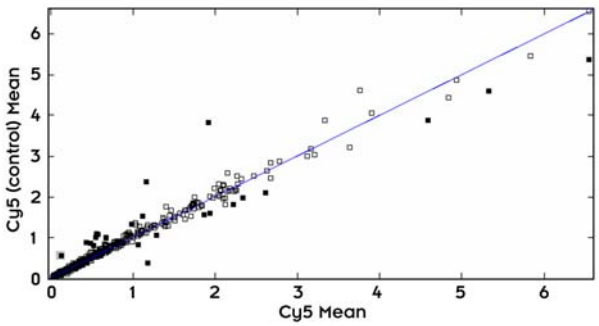


Fig 1. Scatterplot of differential gene expression in C57BL/6JRn mouse skin exposed to 10 cGy of acute X rays. Data were generated using the ORNL "1K" microarray. Control (sham exposed) B6 mouse transcriptional response data are on the y-axis, while the exposed B6-mouse skin data are plotted on the x-axis. The diagonal line represents a 1:1 ratio of gene expression (no response). Genes whose response falls above the diagonal are repressed by the exposure, while induced genes are indicated below the diagonal line. Filled squares are the genes whose response was determined to be significant at the 99% confidence level.

When specific gene systems are analyzed in detail, we noticed that certain genes that might have been thought to have been induced (such as *Cdkn1a*, a gene induced by p53 in a damage/stress response) were not induced in skin according to our microarray results. In retrospect, these data are consistent with other array data reported by other laboratories doing microarray research on the LDR response, reported at the DOE Low Dose Radiation Research Program Workshop IV, held recently in Washington, D.C. Detecting these types of differences in in vivo tissues and perhaps among genetically diverse individuals is what was driving our whole project and indicates potentially enticing data that will need to be confirmed by repetition of microarrays and by other quantitative gene-expression techniques. Because the in vivo response could be dramatically different from the already published in vitro response, we decided, for all mouse microarray applications (not just for LDR), to transition to larger 22,000-gene oligonucleotide arrays. The larger arrays would provide a more complete view of the LDR response, and the use of oligonucleotide probes may also provide additional information about alternate splicing events (if they occur). These arrays would validate the results seen using the ORNL "1K" chip and eliminate possible false data resulting from chimeric cDNA elements that sometimes occur in cDNA libraries.

Summary and Benefits

Using in-house microarrays, a transcriptional response to LDR of 10 cGy was detected and analyzed. In general, when looking at mouse skin, approximately 10% of the genes on the array may be statistically determined to respond to LDR at 3.5 h post exposure. Different strains showed different, but overlapping, responses, with C3H-a showing the greatest number of response genes. The gene-expression experiments, as well as the considerable tissue bank that was amassed during the project period, have provided important preliminary data as well as an important resource for future work. Although some genes predicted to be involved in low-dose response were detected, many genes were not, indicating that the in vivo process has unexpected complexity. The project was successful at determining whether strain-specific LDR response genes might be identified, and a number of presumed genotype-based response differences observed in one early experiment funded by this project were included as preliminary results in a proposal submitted to the DOE's Low-Dose Radiation Research Program.

We continue to believe that an important post-genome niche for ORNL will lie in studying the genetics of susceptibility among individuals to exposures of all types. This focus area would thus leverage ORNL's Mouse Genetics Program into solving problems in human susceptibility, beginning with investigating LDR-response variation in whole-animal tissues. This project clearly helped facilitate this goal. These capabilities for measuring

Table 2. Summary of strain comparisons for differentially expressed genes using 10-cGy radiation exposure

Mouse strain	LDR-responsive genes	Common genes showing opposite transcriptional effects
B6	61	
PUN	82	
C3H-a	179	
Common to all three strains	10	
Common to B6 and PUN	3	2
Common to PUN and C3H-a	33	14
Common to B6and C3H-a	13	8
In B6 ONLY	34	
In PUN ONLY	37	
In C3H-a ONLY	124	

Table 3. Summary of strain comparisons of easily detectable differentially expressed genes
(Genes showing 2X or greater differences)

Strain	Total	Induced	Repressed
B6	5	4	1
PUN	11	4	7
C3H-a	22	4	18

gene expression, as well as the building of a large bank of LDR-exposed tissues from a number of mouse strains, which can be tapped far into the future, have set the stage for determining exposure, sampling, and molecular-screening criteria for future larger projects.

Likewise, we feel that the approach to studying genetic differences among individuals exposed to an agent (either radiation, chemicals, or biotoxins), pioneered in this project, will position ORNL to study genotype-based differences in responses to other types of exposures that will be relevant to funding by the National Institutes of Health [for cancer susceptibility (NCI), susceptibility to chemical exposures (NCI, NIEHS)], and possibly by the Departments of Defense and Homeland Security (for bioterror agents).

References

[1]A. V. Prasad, N. Mohan, B. Chandrasekar and M. L. Meltz, "Induction of Transcription of 'Immediate Early Genes' by Low-Dose Ionizing Radiation," *Radiat. Res.* **143**, 263 (1995).

[2]S. A. Amundson, K. T. Do, and A. J. Fornace Jr., "Induction of Stress Genes by Low Doses of Gamma Rays," *Radiat. Res.* **152**, 225 (1999).

[3]S. A. Amundson, R. A. Lee, C. A. Koch-Paiz, M. L. Bittner, P. Meltzer, J. M. Trent, and A. J. Fornace, Jr., "Differential Responses of Stress Genes to Low-Dose-Rate Gamma Irradiation," *Mol. Cancer Res.* **1**, 445 (2003).

[4]P. Hegde, R. Qi, K. Abernathy, C. Gay, S. Dharap, R. Gaspard, J. E. Hughes, E. Snesrud, N. Lee, and J. Quackenbush, "A concise guide to DNA microarray analysis," *Biotechniques* **29**, 548 (2000).

Developing a New Core Competency for ORNL: Macromolecular Neutron Crystallography

G. J. Bunick
Life Sciences Division

The project compared the neutron and the ultra-high resolution synchrotron X-ray structures of D-xylose isomerase (XI) in order to ascertain the relative merits of each technique in locating the critical hydrogen atoms and water molecules in macromolecular structures. The working hypothesis was that neutrons could locate the positions of hydrogen atoms from moderate resolution data that might lead to an understanding of the enzymes mechanism, whereas X rays might not be able to provide the same information. Neutron diffraction data were measured by the Protein Crystallography Station (PCS) at the Los Alamos Neutron Sciences Center (LANSCE). Subsequent solution of the structure validated the hypothesis. XI is the largest protein studied to date by time-of-flight (TOF) neutron diffraction. The second objective of the proposed research was to validate the operational capabilities of the PCS. The project also demonstrated the use of spallation-generated neutrons for macromolecular crystallography and thereby provided scientific justification for the future macromolecular neutron diffractometer (MaNDi) at the Spallation Neutron Source (SNS).

Introduction

The hydrogen bond lies at the heart of biochemistry, with hydrogens comprising roughly half of the atoms in biological materials such as proteins and DNA. Hydrogen ions are involved in the molecular actions of virtually all enzymatic reactions. Within macromolecules, biochemical mechanisms can be elucidated and modeling studies facilitated from knowledge of hydrogen positions, waters of solvation, and ions.

The principal techniques currently used for experimental structure determination of macromolecules are X-ray diffraction and nuclear magnetic resonance (NMR). It is often stated that a combination of these two techniques can provide all the information needed to determine the atomic positions within a macromolecule, including the positions of hydrogen atoms. This is simply not the case.

Neutron diffraction can provide important and unique information for macromolecular structure-function studies. The ability of neutrons to reveal the positions of hydrogens even at moderate resolution (2.0–2.3 Å) is the foundation of the scientific justification for neutron diffraction of biological samples. The protonation and ionization states of atoms, which are necessary to determine the atomic charges and pK_as of proteins, can be determined from a neutron structure. Neutron structures can also provide the orientation of polar groups such as hydroxyls and amides, which can reveal hydrogen-bonding patterns. Water structure is essential to understanding protein structure and function. Neutron density maps can provide detailed information about the water structure in protein active sites and the solvation shells around proteins. From neutron diffraction studies, details of enzyme catalytic mechanisms and protein molecular recognition sites can be determined. Knowing such details are of great value for rational drug design, ligand design, and rational protein design.

Technical Approach

The ultra-high-resolution X-ray structure of XI required that the crystal be flash-cooled below the glass transition temperature of water to reduce radiation damage. Previous X-ray diffraction studies of XI did not utilize flash cooling and synchrotron radiation.[1] A suitable cryoprotectant was found and a flash cooling procedure was developed for XI crystals grown using ammonium sulfate as the precipitant.[2] The X-ray measurements were treated by standard crystallographic procedures to produce an electron density map of the protein. The program Xfit of XtalView[3] was used to fit and interpret the electron density map. The program Refmac5 of the CCP4[4] suite was used to refine the X-ray structure.

Large crystals were required for the neutron diffraction studies because of the low flux that is a characteristic of all neutron sources in comparison to X-ray sources. For example, our laboratory rotating-anode X-ray generator provides an order-of-magnitude-higher flux than any neutron source. Figure 1 shows a large single crystal of XI that was grown in a counter diffusion cell (CDC). This device was developed by the National Aeronautics and Space Administration (NASA) for the

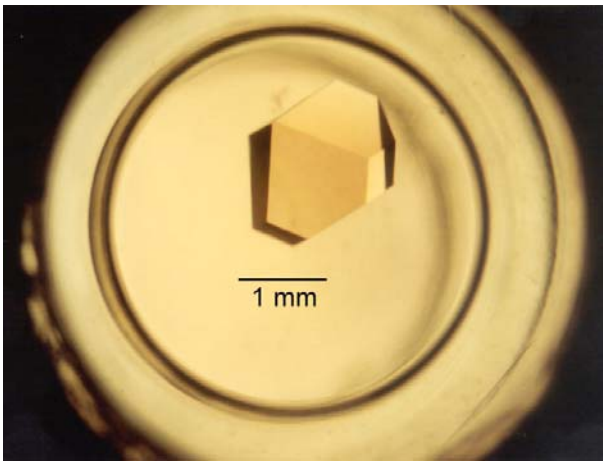


Fig. 1. A single large crystal of XI grown in a counter diffusions cell designed for macromolecular crystallization in microgravity. The crystal was approximately 3 mm × 2 mm × 1.5 mm when it was transferred to a capillary for the neutron diffraction measurements.

growth of macromolecular crystals in microgravity. The concentrations of solutions in the two chambers of the CDC had to be determined experimentally in order to grow large high-quality crystals.

The integration of the measured reflections and the subsequent neutron intensity data scaling, merging, and wavelength normalization required modifications to several different software programs in order to utilize the time-of-flight information encoded in the data. The XI data set was used to debug the programs. The statistics of the data set suggest that other programs will have to be tried to apply absorption corrections to the data. The analysis of the XI data points to the areas where better algorithms for processing TOF neutron data will be needed in the future.

The neutron structure was visualized and fitted using XtalView[3] and refined using the program SHELX.[5] SHELX is the only crystallographic program able to control the refinement process atom by atom and to handle the replacement of hydrogen atoms by deuterium atoms as the refinement proceeds.

Results and Accomplishments

Neutron diffraction data (Fig. 2) were measured on the PCS at LANSCE and the data were integrated, wavelength normalized, and scaled. The quality of the neutron diffraction data were then assessed. The values of R_{merge} for the XI data for most resolution bins are larger than values reported to date in neutron studies of other proteins. The two main differences between this study and previous ones are the size of the system under investigation and the wavelength range of neutrons used. With a unit cell volume of approximately 1×10^6 Å,[3] this study

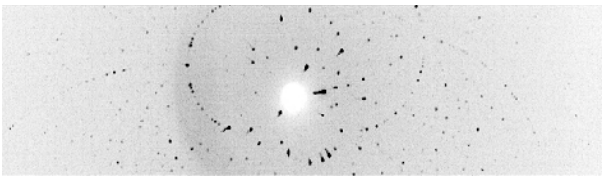


Fig. 2. Neutron diffraction data from one crystal setting for the native XI crystal. The image is an overlay of all time intervals for the setting. Therefore, weaker reflections from longer wavelength diffraction can be lost in the background from the short wavelength time slices. However, since the data from each time slice is analyzed separately, these reflections can be measured against the background within their time slice. This reduces the time needed to measure the signal for each reflection and improves the signal/noise ratio for the weak data.

represents by far the largest system studied to high resolution with neutron diffraction.

Previous neutron data sets on other proteins have been collected at steady-state neutron sources using either monochromatic neutron beams or quasi-Laue neutron beams with a wavelength band of approximately 1 Å, whereas the TOF neutron data set has been collected over a large wavelength range of 0.7 Å to 7 Å. In order to investigate the effects of using a large wavelength band on the data merging statistics, the data were normalized and then merged over small wavelength bands of 1 Å. The results show that over equivalent wavelength ranges the accuracy of the data is at least as good as previous data collected on other neutron instruments. It is postulated that the behavior of the data merging statistics as a function of the wavelength bandwidth may be explained by variations in incoherent scattering from hydrogen, which result in an effective absorption factor that is a function of both wavelength and crystal shape. The scaled intensities of equivalent reflections collected at the same wavelength will differ because of the different associated path lengths through the crystal and the size of this difference will change with wavelength. The unusually high redundancy in the data may help compensate for this effect and result in accurate averaged intensities. However, it is clear that methods should be developed for properly correcting for this effect so that the accuracy of the data can be further improved.

The ultimate test of the quality of the data set is its ability to locate single deuterium atoms. A 2Fo-Fc sigma-A map was calculated using phases from the room temperature structure determined from a 1.55 Å X-ray data set[1] with all of the water molecule atoms and labile hydrogen atoms omitted, and using neutron amplitudes in the resolution range 10 Å–2 Å. There are clear indications of the protonation states of residues around the catalytic site. The direct observation of these states is of crucial importance in determining the details of catalytic mechanism. Figure 3 shows density at the end of a lysine

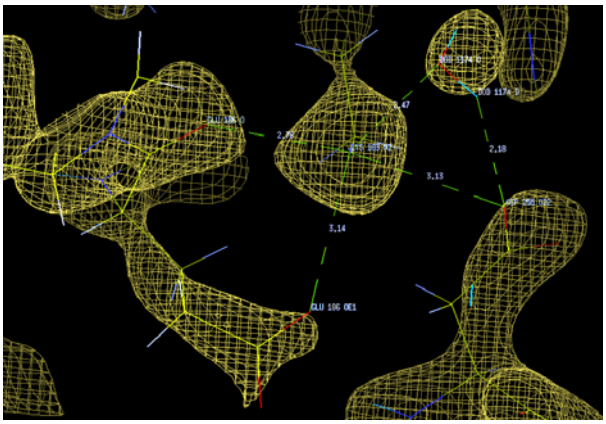


Fig. 3. The terminal nitrogen atom of lysine 183 is triply deuterated in the neutron density map contoured at 2 σ. Hydrogen bonds are made from the nitrogen atom to the main chain oxygen of Glu186, a carboxylate oxygen of Glu186, and a D_2O molecule. The D_2O molecule makes a hydrogen bond to Asp255, which may share one of the deuterons on the terminal nitrogen with Glu186 OE1. Lysine 183 may function as a hydrogen donor during catalysis.

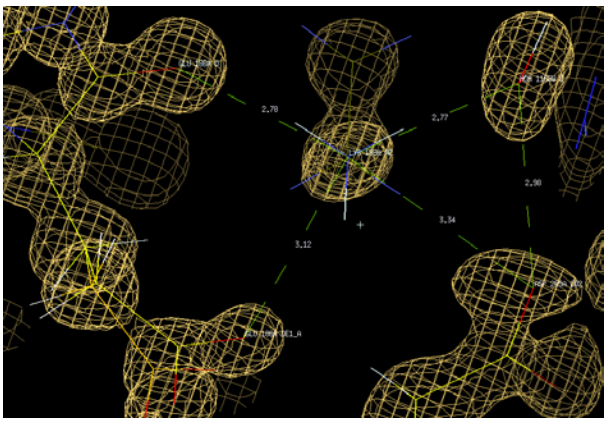


Fig. 4. Electron density for the terminal nitrogen atom of lysine 183 contoured at 2 σ from the 0.94 Å X-ray structure. The protonation state of the terminal nitrogen atom cannot be determined in the electron density map but is clearly determined in the neutron density map of Fig. 3.

residue (Lys183) that defines the orientation of the terminal ND_3 group. This particular orientation allows the individual deuterium atoms in the ND_3 group to hydrogen bond to a main chain oxygen atom from Glu186, a carboxylate oxygen atom from Glu186, a carboxylate oxygen atom from Asp255, and a water molecule. The proximity to the catalytic site suggests that this residue is important. Figure 4 shows the same lysine residue in the 0.94 Å resolution X-ray structure. The protonation state of the lysine cannot be determined from the electron density map.

Summary and Conclusions

The metalloenzyme D-xylose isomerase forms well-ordered crystals that diffract X rays to ultra-high resolution (<1 Å). However, structural analysis with X-ray diffraction data has been unable to differentiate between several postulated mechanisms that describe the catalytic activity of this enzyme. Neutrons with their greater scattering sensitivity to hydrogen atoms could help resolve this by determining the protonation states within the active site of the enzyme. As a comparison of the relative merits of X-ray and neutron structures for locating hydrogen atoms, neutron data to 1.8 Å were measured from the unliganded protein at the LANSCE PCS. Measurement of these neutron diffraction data represents several milestones: this is one of the largest biological molecules (a tetramer, MW ~160,000 daltons (Da), with unit-cell lengths around 100 Å) ever studied at high resolution using neutron diffraction. It is also one of the first proteins to be studied using time-of-flight techniques.[6]

The success of the initial diffraction experiments with D-xylose isomerase demonstrate the power of spallation neutrons for protein crystallography and should provide further impetus for neutron diffraction studies of biologically active and significant proteins.[7] This research provided results that were used in the scientific justification for the macromolecular neutron diffractometer (MaNDi) instrument to be constructed at the SNS. The need for a specialized suite of software to process and solve macromolecular neutron structures was apparent as the project proceeded. It was also determined from presentations of this research that the majority of X-ray crystallographers do not have the basic information to start a neutron study of a protein. A guidebook on the subject will be most useful to help the X-ray crystallographer prepare for a neutron diffraction study.

Additional data are being measured on crystals of XI with bound substrates and inhibitors in order to provide the specific information needed to clarify the catalytic mechanism of this enzyme. Follow-on funding will be sought in the form of grants submitted through the University of Tennessee to NIH and possibly NASA or NSF.

References

[1]H. L. Carrell, H. Hoier, and J. P. Glusker, "Modes of Binding Substrates and their Analogues to the Enzyme D-Xylose Isomerase," *Acta Cryst.* **D50**, 113–123 (1994).

[2]B. L. Hanson, J. M. Harp, K. Kirschbaum, C. A. Schall, K. DeWitt, A. Howard, A. A. Pinkerton, and G. J. Bunick, "Experiments testing the abatement of radiation damage in D-Xylose Isomerase Crystals with cryogenic helium." *J. Synchrotron Radiation* **9**, 375–381 (2002).

[3]D. E. McRee, "XtalView/Xfit—A Versatile Program for Manipulating Atomic Coordinates and Electron Density." *J. Structural Biology* **125**, 156–165 (1999).

[4]"CCP4, Collaborative Computational Project, Number 4." *Acta Cryst.* **D50**, 760–763 (1994).

[5]G. M. Sheldrick and T. R. Schneider, "SHELXL: high resolution refinement," pp. 319–343 in *Methods in Enzymology*, 227, edited by C. W. Carter, Jr. & R. M. Sweet, Academic Press, San Diego, 1997.

[6]LANSCE Highlight, "High-Resolution Neutron Diffraction Data Collected on the Protein Crystallography Station at LANSCE Exceed Researchers' Expectations." (2003) http://lansce.lanl.gov/news/features/Xylose.htm

[7]C. Day, "Search & Discovery: Neutron Diffraction Overcomes Flux Limits to Resolve a Large Protein Structure," *Physics Today*, pp. 17–19, November 2003.

Novel Treatment for Biological Warfare Pathogens

C. E. Easterly,[1] G. D. Griffin,[1] M. Z. Hu,[2] T. L. Nichols,[3] and A. A. Vass[1]
[1]*Life Sciences Division*
[2]*Nuclear Science and Technology Division*
[3]*University of Tennessee Medical Center*

For the first time, engineered antimicrobial nanoparticles have been demonstrated as the basis for an extremely broad-spectrum therapy for life-threatening microbial agents such as might be used in a bioterrorist attack. This work was intended to test whether or not elemental or small-molecule antimicrobials in nanoparticle form can be therapeutic against pathogenic infections. If elemental or very-small-molecule biocidal ions can be transported to bacteria at the site of infection, these ions may provide novel therapies. "Stealth" liposomes appear to provide innate targeting to inflammation sites by taking advantage of natural processes that accompany inflammation-associated infections. The work reported herein entails a series of experiments designed to produce biocidal nanoparticles, make carrier-targeting long-circulating liposomes, and conduct proof-of-principle testing of liposome-encapsulated nanoparticles with bacteria-infected macrophages and bacteria-infected amoebae. Both of these systems are phagocytic, engulfing organisms. Two replicates of these experiments support our hypothesis that liposomes containing antimicrobial nanoparticles do in fact rescue macrophage cells infected with *Legionella.*

This project focuses on the need for therapies against biological warfare pathogens that have been selected or engineered to be pathogenic to humans and immune to modern antibiotics. (Medical use on resistant organisms frequently encountered in modern society is also anticipated.) Our objective was to demonstrate, with a proof-of-principle test, an approach to guide the development of a broad-spectrum therapy against such pathogens. A number of elemental as well as very small molecular ions are known to be toxic to pathogens, but after the discovery of penicillin and the development of subsequent generations of antibiotics, these materials have lacked interest by the medical community. Given the correct form and delivery mechanism, a number of elemental ions have the potential to provide killing mechanisms different from current antibiotics and, in combinations, their diverse modes of attack make future evasive development highly unlikely. Our task is to combine a novel killing methodology together with a carrier and targeting system and demonstrate the recovery or increased survival of pathogen-infected cells.

Technical Approach and Findings

The approach entailed the use of elements or very small molecules in ionic form known to be toxic to pathogens, delivered in a package that was also a targeting agent. This seed project focused on a first demonstration using nanoparticles as the source of killing ions and "stealth" liposomes as the delivery and targeting mechanism. Certain elemental and very-small-molecule ions have long been known to be toxic to pathogens, but the means for providing a constant-concentration internal dosage that is therapeutic but not overtly toxic to the host has not been developed. Nanoparticles offer a means for providing a constant concentration of ions, because the solubility of nanoparticles can be controlled, thus providing a steady source of ions. Thus the first step is to develop a nanoparticle that is stable and biologically active against pathogens. Next, appropriate liposomes must be constructed and loaded with nanoparticles. Finally, a series of biological experiments must be carried out. We have chosen to use *Legionella* as the biological warfare surrogate. The bacterium, *Legionella pneumophila,* is the causative agent of Legionnaires' disease and related respiratory ailments. It is a facultative intracellular pathogen with a similar pathogenicity to many biological warfare (BW) agents. The bacteria are able to infect, multiply within, and kill human macrophages in a fashion similar to BW agents. We have used two phagocytic organisms to demonstrate the potential therapeutic efficacy of the proposed therapeutic approach: J774, a mouse-macrophage cell line, and an amoeba (*Acanthamoebae*), which are noted for their phagocytic activity.

Synthesis of Nanoparticles

Nanoparticles made from elements and/or very small molecules are intended to be encapsulated inside liposomal vesicles. Since the sizes of the "stealth" liposomes will be in the range of 80–100 nm, the nanoparticles must be less than this dimension in order to be encapsulated within the liposome vesicles. An aqueous nanoparticle sol is preferred for compatibility with the liposome formation conditions. Typically, the production of nanoparticles involves the use of steric capping molecules or stabilizers (such as trialkylphosphine/amine, alkanethiols, long-chain unsaturated carboxylates, CS_2, or quaternary ammonium disulfide) to control the size and stabilize the nanoparticles.

Nanoparticles were produced using a process currently under patent application. By virtue of their solubility characteristics in biological media, these nanoparticles were able to maintain a killing concentration of ions for bacteria, but not their host cells (see below). The overall concentration of the killing agent in the sol was approximately 2 ppm, including ions and nanoparticles, as determined using laser-induced breakdown spectroscopy (courtesy of Madhavi Martin, Environmental Sciences Division). However, we have not yet determined the size vs number density for the nanoparticles, nor the concentration of ions in the sol.

Nanoparticle sols were stored in low-potassium scintillation vials because storage in other glass containers resulted in degradation of the nanoparticle size. This phenomenon is thought to be associated with the decay of K-40 in other containers, leading to radiation-induced deterioration of the self-assemblies to particles larger than the liposomes. We have monitored samples, and the prepared sols have remained consistent in size and concentration for at least 2 months.

Liposome Preparation

For intravenously introduced liposomes, it is generally accepted that the prolonged circulation time of polyethylene glycol (PEG)–coated liposomes as compared to conventional non-coated liposomes is a result of the PEG coating. Methods for preparation of PEG-coated liposomes described by Schiffelers et al.[1] were followed. In addition to prolonged circulation times, preferential localization of these liposomes at sites of infection or inflammation has been demonstrated in a variety of experimental models.[1] Schiffelers et al.[1] incorporated two antibiotics into their liposomes.

Liposomes were prepared using procedures described in the literature with slight modifications. Briefly, the following lipids were dissolved in a mixture of chloroform/methanol at molar ratios of 1.85:1:0.15: partially hydrogenated egg phosphatidyl choline, cholesterol, and 1,2-distearoyl-sn-glycero-3-phosphoethanolamine-N-[polyethylene glycol 2000]. The solvents were evaporated under nitrogen and the lipids dried and redissolved in 2-methyl-2-propanol, followed by shell-freezing and subsequent lyophilization. Following hydration, the lipid dispersion was sonicated to produce the liposomes. (Sonicator Model 112SP1G, 80KC, 80 Watts, Laboratory Supplies Co., Hicksville, New York).

Liposomes were then separated from unincorporated nanoparticles with a Sephadex G-25 column using distilled, deionized water as the eluant. This column allows larger particles to progress faster through the column than smaller particles. After passage of the liposomes, the next 3 mL were collected. Analysis of this volume indicated that the nanoparticle concentration was 25 times less than the concentration in the nanoparticle sol that was encapsulated by the liposomes. Thus the nonencapsulated material in the liposome preparation is less than 0.1 ppm.

Biotesting of Nanoparticles and Nanoparticle-Containing Liposomes

The first step was to evaluate the concentration of nanoparticle solution required to kill cultures of *Legionella*. Cultures of *Legionella* were grown to a concentration of 10^6/mL and exposed to the nanoparticles. *Legionella* and nanoparticles were combined in a ratio of 1:3 with a final overall concentration of about 2 ppm. After 30 min, all *Legionella* were killed as determined by re-examination of plates incubated 48 h after exposure to determine if any breakthrough growth occurred. Repeating the measurement with 10× diluted nanoparticle sol (about 0.2 ppm) produced about 99% kill after 1 h, but even after 48 h the kill was not 100%. Replication using similar nanoparticle concentrations showed complete killing with about 1 ppm within 1 h and very incomplete killing at 0.1 ppm. Thus, nanocrystalline materials in the low-ppm range provided good killing of *Legionella,* thus allowing for the final stage of this effort.

The next step was the proof-of-principle experiment. This involved the infection of cells with *Legionella* and subsequent exposure of the cells to the liposomes containing the nanocrystals. Two phagocytic cell lines were used: J774 mouse macrophages and *Acanthamoebae*. Prior experiments established the concentration of nanoparticles that was effective against *Legionella* and what concentration of nanoparticles and liposomes did not exhibit toxic effects on the cell lines. The two proof-of-principle experiments used liposomes that were hydrated in pure water containing about 2-ppm nanoparticles. Experiment one used a preparation of liposomes, which were on the order of 800 nm in size. Experiments were set up in duplicate with two flasks of *Acanthamoebae royreba* and line J774 set up as controls (nothing added). Both cultures were plated initially at densities of 10^4/mL in a

total volume of 2 mL. To the eight test culture flasks (four of *Acanthamoebae* and four of J774), 10μL of a 1.2 x 10^8/mL culture of viable *Legionella* (final concentration 1.2 x 10^6/flask) was added 2 days prior to treatment with the liposomes in order to ensure adequate time for phagocytosis. After 2 days, the media was changed in all flasks and 100 μL of the liposome preparation was added to two flasks of each set resulting in two control, two *Legionella*-infected, and two *Legionella*-infected plus liposome-treated flasks for each cell line. These flasks were monitored daily over a period of 2 weeks, and the media was changed as needed. After media changes, liposomes were re-added to ensure adequate treatment. Ultimately there were no survivors in the untreated samples of either organism, but all treated samples recovered.

Experiment two was set up in a similar manner, but used a liposome preparation which had an average 85-nm particle size. Because of slight differences in liposomal preparation, we added 160 μL of the liposome preparation to obtain the same quantity of nanoparticles as the previous experiment. After 8 days, the liposome-treated amoebic cultures were fully recovered, while the untreated test flasks had no obvious survivors. By day 12, the liposome-treated J774 line was nearly fully recovered while the untreated cells had no obvious survivors. The results of the two experiments with different size liposomes tend to indicate that size is important in how the cell lines react to the liposomes, mostly in terms of the initial recovery phase. For both experiments, both J774 and amoebae cultures that were infected with *Legionella* and subsequently treated with the liposomes recovered. Those cultures that did not receive liposomes were completely destroyed (Figs.1 and 2).We estimate that the maximum concentration of nanoparticles outside of the liposomes initially present in the treated flasks was less than 0.005 ppm. This is well below the 0.1 ppm that failed to provide complete killing in pure cultures of *Legionella*. Thus we have confidence that the effects seen with the liposome treatment are due to nanoparticles phagocytized by the macrophages and amoebae in close proximity to the engulfed *Legionella* that provide the recovery of the treated organisms.

Summary and Conclusions

We have demonstrated for the first time that it is possible to use engineered nanoparticles encapsulated within liposomes to act as a "stealth" bacteriocidal agent that can destroy bacteria in growing mammalian cells, and in a non-mammalian phagocytsing cell line, the amoeba. This is a substantial achievement, considering the large number of variables that these experiments entailed.

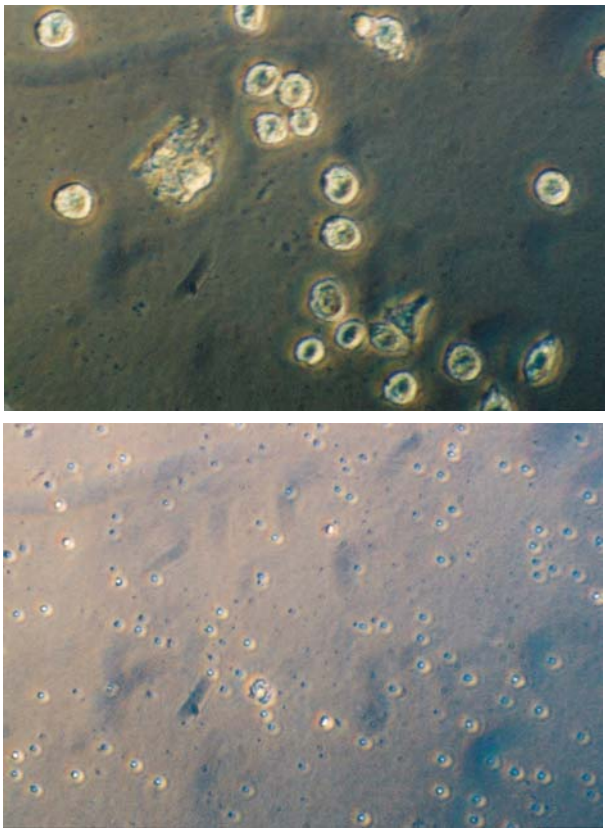


Fig 1. 400× view of J774 macrophage cells infected with Legionella (above) treated with nanoparticle-filled liposomes and (below) untreated.

- The nanoparticles must be stable.
- They must be of the correct size and physico-chemical composition to release ions in the correct active form resulting in an adequate concentration to kill engulfed bacteria but not the host cells.
- The liposomes must effectively incorporate the nanoparticles and simultaneously lend themselves to inclusion into phagocytic cells, either by membrane fusion or by engulfment by phagocytic mechanisms.
- The liposomes must release their bacteriacidal kill package intracellularly.
- The pathogen-toxic ions must traverse the cell interior to kill the bacteria.

These successful proof-of-principle experiments demonstrate the first of many potential antimicrobial materials that could be identified, prepared, and inserted into stealth liposomes.

Experimental results demonstrate an approach to solve one aspect of a critical problem—what could be done

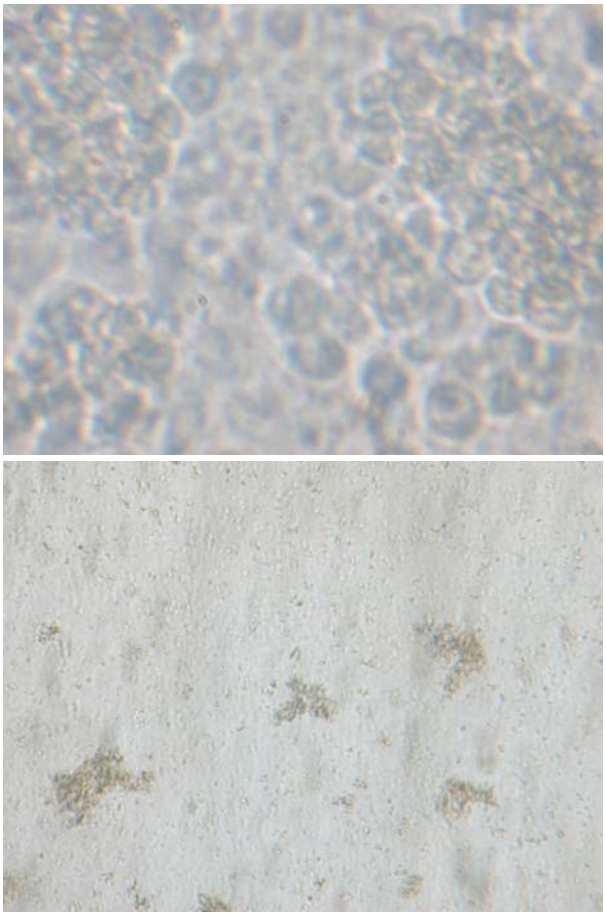


Fig 2. 400× view of amoebae infected with Legionella (above) treated with nanoparticle-filled liposomes, and (below) untreated.

in the event of a terrorist attack with bacteria that have been selected or engineered to be antibiotic resistant. Proof-of-principal therapy for such an organism using in vitro means has been demonstrated. Since the delivery mechanism (stealth, or long-circulating liposomes) has been demonstrated repeatedly in animal models using conventional antibiotics,[1] there is little risk in assuming that they will target inflammation sites carrying different kill materials.

Two of ORNL's key science objectives, nanotechnology and biotechnology, are integral to this project. Observations during the course of work suggest that additional nanoparticles may be engineered using methods similar to those developed during our work. In addition, a substantial number of elements or extremely small molecules may be engineered into nanoparticles for use as additional components in future "kill packages."

A system consisting of delivery, targeting, and a kill package with the kill package capable of eliminating microbial organisms immune to normal therapies should be useful for defense strategies against biological warfare agents as well as organisms found in hospitals that are insensitive to current antibiotic strategies. Given the current fluidity within the federal agencies, the exact target for further developing this concept is not known at this time, but DARPA is a likely candidate, as is the NIH. However, considerably greater quantification of the concept will be required to enable a successful NIH proposal.

References

[1] R. M. Schiffelers, G. Storm, M. T. Ten Kate, L. E. T. Stearne-Cullen, J.G. Den Hollander, H.A. Verbrugh, and I. A. J. M. Bakker-Woudenberg, "*In Vivo* Synergistic Interaction of Liposome-Coencapsulated Gentamicin and Ceftazidime," *J. Pharm. Exp. Therapeut*. **298**(1) 369–375 (2001).

Enhanced Biological Hydrogen Production Using Three-Phase Systems

K. T. Klasson
Life Sciences Division

This work focused on the proof of principle that a three-phase (gas-aqueous-oil) microbial system can yield higher hydrogen production rates than current biohydrogen techniques. The hypothesis to be tested was that improved conversion of CO and water to H_2 and CO_2 via the water-gas shift reaction should be noted in systems where a co-solvent is used. Some results obtained in this study were promising; however, the results were not conclusive, as different results were obtained in different systems.

Introduction

Biological production of hydrogen by microalgae has been known for at least 100 years and other systems have been known for decades. Overall, however, very little progress has been made to make this an applied technology. Current biohydrogen efforts continue to be focused on fundamental research, addressing the shortcomings of biological hydrogen production.

The main approaches to biological hydrogen production are

1. biological water-gas shift reaction,
2. fermentative hydrogen production from organic substrates,
3. photosynthetic hydrogen production from organic substrates, and
4. water-splitting by algae.

Each of these approaches has advantages and disadvantages, and it is clear from the literature that the potential for biological hydrogen has been demonstrated, although the technical feasibility and several key questions remain to be answered.

Of these approaches, the water-gas shift reaction holds the most immediate promise. The water-gas shift reaction ($CO + H_2O \rightarrow H_2 + CO_2$) has been proposed for production of hydrogen from carbon monoxide–containing synthesis gas. Synthesis gas can be produced from a variety of carbon-rich materials such as coal, petroleum products, and renewable biomass. Gasification is an established technology, and the water-gas shift reaction can be carried out by traditional catalysis or by biocatalysts. Limitations for the traditional catalysis are associated with catalyst cost and catalyst poisoning from sulfur present in many low-Btu coals and raw materials. The biological approach does not suffer from this limitation but suffers from rate limitations due to mass transfer, as the reactant is gaseous and reaction is carried out in an aqueous medium. Our proposed work was designed to demonstrate the enhancement of biological hydrogen production in mass transfer–limited systems by the introduction of third phase, silicone oil. The silicone oil would aid in creating new interfacial surface area. This would allow for rapid transfer of CO reactant in the gas phase to bacteria in the aqueous phase.

Technical Approach

Our work used a bacterial strain of *Rhodopseudomonas palustris* P4, kindly provided by Dr. Sunghoon Park, Pusan National University, South Korea. This organism can grow under a variety of conditions (e.g., photosynthetic or in the dark, aerobically or anaerobically, and on a variety of substrates) and can be induced to carry out the shift reaction by switching the growth mode to anaerobic and exposing the cells to CO.[1]

Growth conditions of the culture were established and the conditions for hydrogen production determined through the testing of different media compositions. Once these conditions were determined, cultures were grown and placed in the refrigerator for later use as seed inoculums. The size of these starter cultures varied from 50 to 700 mL.

The first set of experiments was conducted in sealed 160-mL serum bottles and used different quantities of inoculums, providing cell initial concentrations of 0 to 2.0 g dry cell weight/L. The gas phase was 0.14 atm CO in N_2. The conversion of CO was monitored, and the conversion rate and mass transfer properties were determined.

The second set of successful experiments included the addition of two co-solvents, silicone oil or hexadecane, at two different levels. The experiments were conducted in sealed 160-mL serum bottles, and the total liquid volume

and quantities of cells were the same in each experiment. The conditions were chosen so that mass transfer limitations were experienced.

The third set of successful experiments included the addition of three new co-solvents, perfluorocarbons FC-40, FC-72, or FC-77, obtained from 3M. The experiments were conducted in sealed 160-mL serum bottles, and the total liquid volume and quantities of cells were the same in each experiment. The conditions were chosen so that mass transfer limitations were experienced.

Experiments were also conducted in a larger-scale stirred batch fermenter. The total volume of liquid in this system was 1000 mL, of which 30% was either one of two solvents (silicone oil or FC-40) or just additional water (in the case of the control experiment). A continuous stream of gas (47% CO, 10% CO_2, 36% H_2, 2% N_2, and 5% CH_4) was purged through the system, and the conversion of CO was determined from the analysis of the inlet and exit gas streams. The agitation was kept constant until a constant conversion was noted and then increased to investigate if the system was mass transfer limited. (An increased conversion when agitation is increased suggests that mass transfer limiting conditions exist at the lower agitation rates.)

Results and Accomplishments

The conversion of CO in the first set of experiments is shown in Fig. 1, and the calculated mass transfer rate has been plotted in Fig. 2 as a function of the partial pressure in the gas phase of the serum bottles. From the results in Fig. 1, it is noted that CO conversion was not noted in the experiment without cells. From the results in Fig. 2 it can be deduced that the transport of CO from the gas (to the liquid and cells) became proportional to the partial pressure of CO at approximately 0.05 atm CO. It should also be noted in Fig. 2 that the rate/partial pressure relationship was the same for cell concentrations above

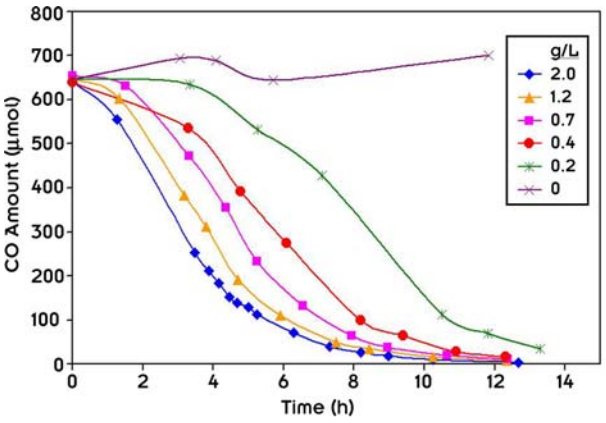


Fig. 1. CO concentration profiles in experiments with different cell concentrations.

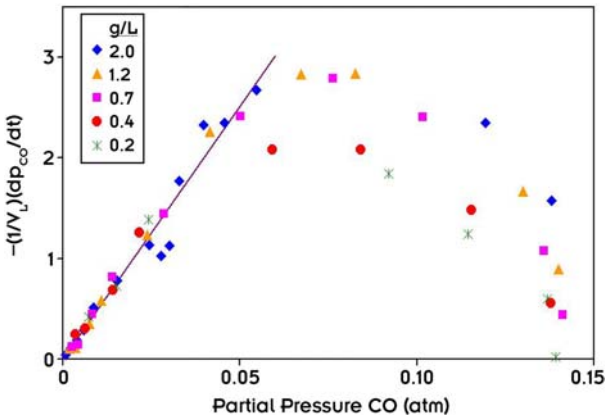


Fig. 2. CO transport rate as a function of CO partial pressure in experiment with different cell concentrations. The rate of consumption was calculated by numerical methods and curve fitting.[3]

0.4 g/L. The slope of the straight line at low partial pressures is proportional to the mass transfer coefficient.[2] Thus, it was concluded that mass transfer conditions in these types of serum bottle experiments should be occurring at cell concentrations of at least 0.7 g/L and at partial CO pressures below 0.05 atm. Experiments in serum bottles were conducted so that these conditions were met by using approximately 0.7 g dry cell weight/L and an initial CO partial pressure of 0.08 atm.

The conversion of CO in the absence and presence of different quantities of either silicone oil or hexadecane is shown in Fig. 3. From the results presented in Fig. 3 it is noted that the presence of either oil did not affect the rate of CO conversion; that is, the curves fall on top of each other for all cases. There was a shift in the CO conversion profile in the experiment with 30 volume % hexadecane, when after one sampling event this serum bottle was not returned to its agitated condition for approximately 30 min. Overall, this experiment proved our hypothesis wrong.

The conversion of CO in the absence and presence of 10 volume % of three perfluorocarbons is shown in Fig. 4.

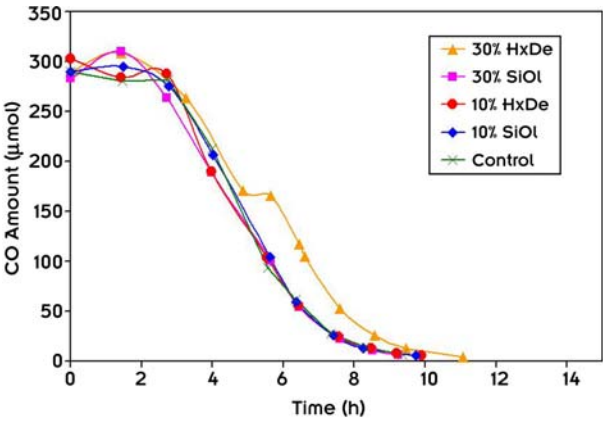


Fig. 3. CO conversion profiles in the absence and presence of either silcone oil (SiOl) or hexadecane (HxDe).

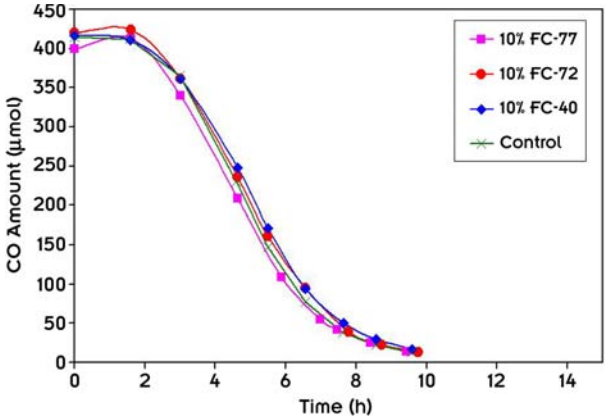


Fig. 4. CO conversion profiles in the absence and presence of different perfluorocarbons.

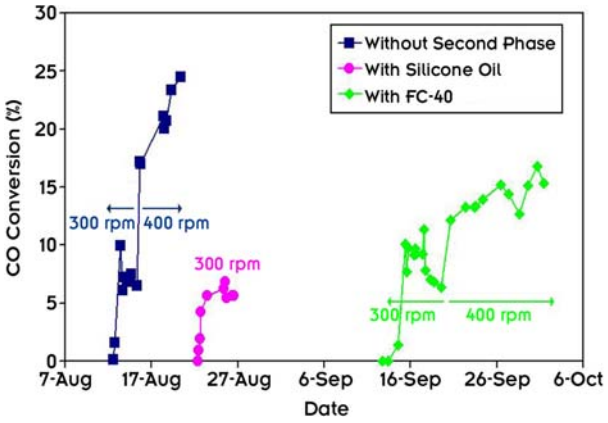


Fig. 5. CO conversion in a stirred batch fermenter without co-solvent (■), with silicone as a co-solvent (●), and with perfluorocarbon FC-40 as a co-solvent (◆).

From the results presented in Fig. 4 it is noted that the presence of either oil did not affect the rate of CO conversion; that is, the curves fall on top of each other for all cases. Overall, this experiment proved our hypothesis wrong. The rationale for investigating perfluorocarbons was that they have a density higher that water, while silicone oil and hexadecane have densities less than that water.

While all the small-scale experiments proved our hypothesis incorrect, it was also noted that the agitation and mixing of the organic and aqueous phases was less than complete. In order to achieve a more complete mixing, three experiments were conducted in a 1-L fermenter with mechanical agitation. The results are shown in Fig. 5. The data to the left correspond to an experiment without any co-solvent present. After several days of fermentation, the CO conversion rate had stabilized at approximately 7%. The agitation was then increased from 300 to 400 rpm and the conversion increased. This suggested that the conversion rate at 300 rpm was limited by mass transfer. The data in the center of Fig. 5 correspond to an experiment where 30% silicone oil was present and the agitation was 300 rpm. Again the CO conversion stabilized around 6 to 7%, indicating that silicone oil did not increase the CO conversion. The data to the right in Fig. 5 correspond to an experiment where 30% of perfluorocarbon FC-40 was present and the agitation was 300 rpm. The CO conversion this time was 9–10% for several days before dropping to around 7%. At this point the agitation was increased to 400 rpm to make sure the culture was not loosing its CO converting abilities. The conversion increased, indicating the culture was still performing its function. The large-scale fermenter data suggest that FC-40 addition had a positive impact on the CO conversion, indicating that the mass transfer was higher in this case and resulted in increased conversion of CO. It is, however, disturbing that the conversion dropped off in a manner which could not be explained.

Summary and Conclusions

Several experiments were performed as part of this investigation. Most of these experiments failed to prove the hypothesis that co-solvents will enhance the mass transfer in biological CO conversion systems. One experiment proved the hypothesis; however, not conclusively. As part of this project, several collaborative contacts have been made which may result in biohydrogen proposal writing activities.

References

[1]T. H. Lee, J.-Y. Park, and S. Park, "Growth of *Rhodopseudomonas palustris* Under Phototrophic and Non-phototrophic Conditions and Its CO-dependent H_2 Production," *Biotechnol. Lett.* **24**, 91–96 (2002).

[2]J. L. Vega, E. C. Clausen, and J. L. Gaddy, "Study of Gaseous Substrate Fermentations—Carbon-Monoxide Conversion to Acetate. 1. Batch Culture," *Biotechnol. Bioeng.* **34**, 774–784 (1989).

[3]K. T. Klasson, "Experimental Data Analysis: An Algorithm for Smoothing of Data and Determining Enzyme and Microbial Kinetic Rates," *Appl. Biochem. Biotechnol.* **63/65**, 339–348 (1997).

Robust Segmentation of Telomeres in Metaphase Fluorescence In Situ Hybridization Images

T. P. Karnowski and J. R. Price
Engineering Science and Technology Division

In this work we explored the use of graphical models to develop a robust segmentation and labeling algorithm for automated, high-throughput telomere quantification in fluorescence in-situ hybridization (FISH) imagery. We first segmented a set of test data and compiled statistics. We developed an independence diagram for the data. Finally, we compiled test results using an algorithm based on the independence diagram.

Introduction

Telomeres, which are the end sections of linear chromosomes, are vital to proper cell maintenance and play a role in cancer and aging. Fluorescence in situ hybridization (FISH) is a procedure used to image telomeres and chromosomes through the use of fluorescent probes that are bound (hybridized) to DNA. Generally, through the use of different fluorescent probes, the telomeres will be imaged by one color channel (e.g., red) and the chromosomes by another (e.g., blue). Scientists analyze these FISH images to examine telomere and/or chromosome characteristics to answer various biological questions. Figure 1 illustrates the basic idea of FISH. Figure 2 shows a composite image where the telomere probes are shown as red and the chromosome as blue.

While some semi-automated software exists to identify and label telomere and chromosome structures, this software suffers from user-interface issues that force scientists to discard significant amounts of experimental data. Furthermore, the image processing algorithms of the existing software are based upon heuristic techniques that do not properly exploit the known structural relationships between the telomeres and chromosomes. In 2002 we implemented a graphical user interface (GUI) for processing telomere/chromosome image pairs. For expediency, this software used only heuristic image processing algorithms. The software has been well received and noted as an improvement by prominent telomere researchers, but the currently employed, simplistic segmentation approach limits broad applicability, hence the need for this project.

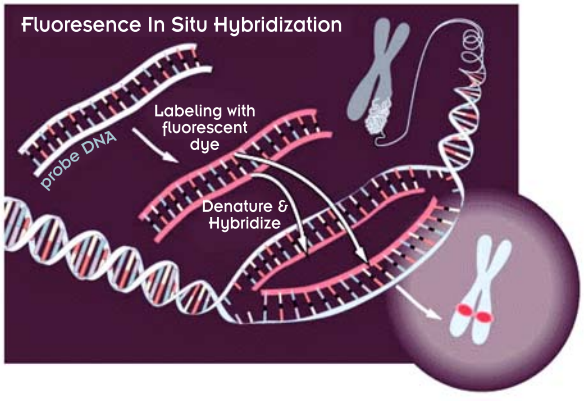


Fig. 1. Fluorescence in situ hybridization concept. (Image from http://www.accessexcellence.org/AB/GG/fish.html.)

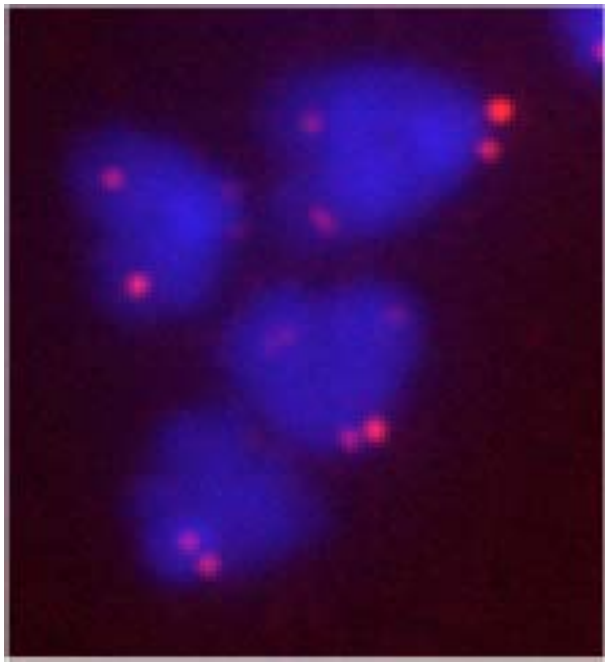


Fig. 2. Chromosomes and telomeres. The chromosomes are blue; the telomeres are red. There are four telomeres per chromosome.

Technical Approach

Our technical approach consisted of three main tasks.

Segment Test Data and Compile Statistics

We first segmented existing test data to determine important Bayesian statistics and relationships between pixel levels for both telomere and chromosome pixels of images.

Create Independence Diagram Models and Solution Methods

We next used the statistics compiled as well as our own knowledge of the chromosome and telomere relationships to create a basic independence diagram. Our survey of the literature in this field revealed that independence diagrams have been used in image processing for a variety of purposes, but their application is markedly different from that suggested by the artificial intelligence field.

Compile Test Results

We then applied our new algorithms and statistics to the test images and determined their effectiveness compared with the "ground truth" results of manual and assisted labeling.

Results and Accomplishments

Segment Test Data and Compile Statistics

The existing ORNL FISH image-processing tool was used to segment a set of seven sample telomere and chromosome image pairs. The tool processes the images through a sequence of heuristic algorithms that include image processing operations such as labeling, dilation, and linear filtering. MATLAB programs were written to take the test results and compile statistics, including histograms, to model conditional probabilities. The telomere and chromosome images are made using different staining compounds and illumination sources. As a result, one image causes only telomere material to illuminate and the other image only shows chromosomes. We can combine the two together effectively for display purposes by setting the red plane of an image to the telomere and the blue plane to the chromosome. We compiled histograms of the R,B values to model the probabilities that are required. The histograms are computed on each image and then they are all added together to get an approximation to the probability density function required. We also performed some interpolation between histogram bins, however, to handle cases where a particular R,B pair may exhibit zero probability due to the limited training data. This was done by convolving the histogram with an interpolating low-pass filter.

Create Independence Diagram Models and Solution Methods.

As the project progressed, we realized a very complex independence diagram could be constructed for the telomere—chromosome segmentation problem. The only really observable values are the individual pixel intensities of the telomere and chromosome, but the nature of the structures of the telomeres and chromosomes indicate clique statistics would also be extremely valuable. Unfortunately, the limited scope of this project prevented extensive study of this intriguing possibility, but a greatly simplified independence diagram as shown in Fig. 3 was developed. Even this simple independence diagram is difficult to use, however; it is a multiple connected graph due to the "loop" formed between the telomere and chromosome nodes with the (Red, Blue) value node. We simplified the diagram even further by using a method called "clustering", where we combine the loop to form the diagram in Fig. 4.

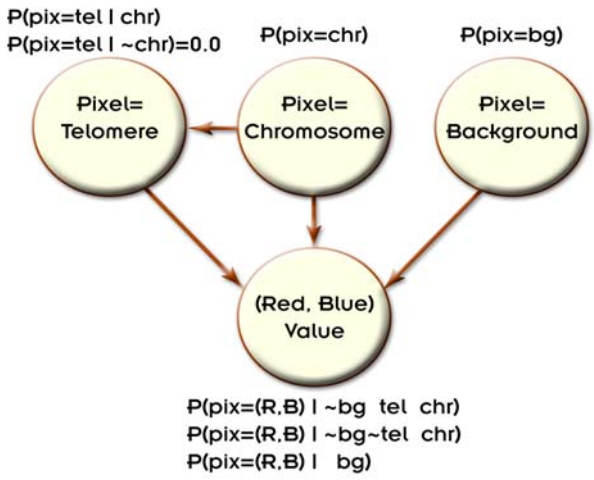


Fig. 3. Original simplified independence diagram with loop.

Design and Conceptualize Algorithms and Hardware for High Throughput

Our segmentation was simple. We merely take each pixel R,B value and determine which is most likely: background, telomere, or chromosome. However, we performed our testing with a hold-one-out procedure. In this process we take one of the seven telomere/chromosome training image pairs, "hold it out," then compile the sample probability functions on the remaining six images. We then label the "held out" image and compare its results to the manual labeled case. Some subsequent image processing is performed to remove noise, but the objective is to keep this to a minimum.

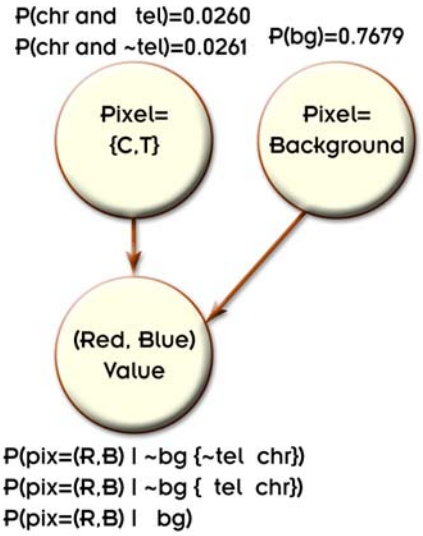


Fig. 4. Revised independence diagram with loop replaced by clustered node.

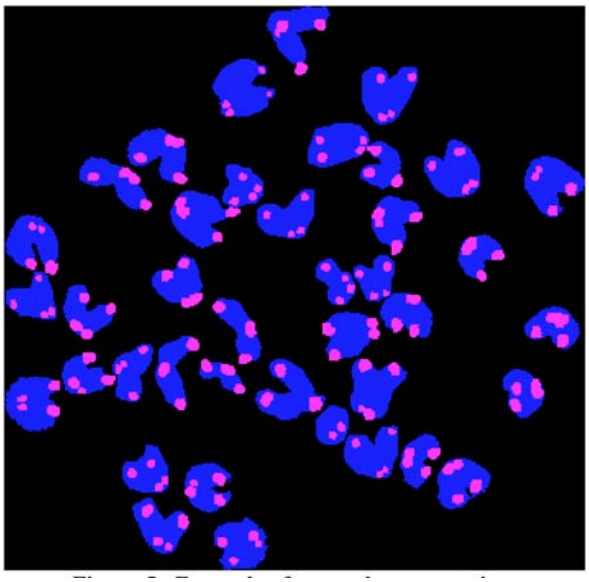


Fig. 5. Example of manual segmentation.

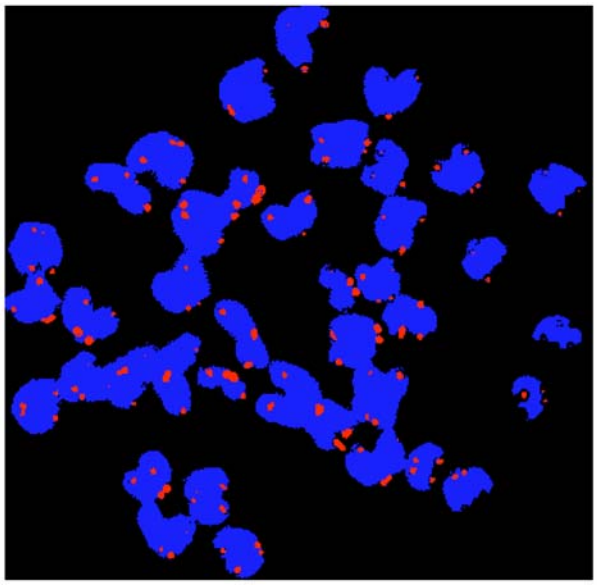


Fig. 6. Example of automatic segmentation. Many entities are joined; some telomeres are not found; and generally the shapes are not as large as the manual segmentation.

Summary and Conclusions

An image pair and example segmentation is shown in Figs. 5 and 6 after simple probabilistic processing. In lieu of more sophisticated processing, we simply removed all blobs from the chromosome and telomere-labeled images that had very small areas (less than six pixels). In Table 1 we show the results of processing. The "Num Chr" and "Num Tel" indicate the number of chromosomes and telomeres that were labeled in the original images. The "Found…" columns indicate the number of these labeled blobs that were intersected by a blob from the segmentation process. We then show the percent of manually labeled chromosomes and telomeres. Finally, we also compare the total area of the chromosomes and telomeres.

Our analysis revealed some advantages and some problems. First, locating chromosomes is rather successful, but finding telomeres is slightly less so, particularly for the first two and the eighth image. The failure to consistently find the telomeres may indicate a sensitivity to the parameters on the FISH process; in other words, images 1, 2, and 8 may have had different image acquisition or chemical processing conditions than the other images, making them somewhat unrepresentative of the true statistics of the image distributions. Secondly, we seem to consistently under-size the chromosomes and telomeres. Finally, our labeling does not split chromosomes or telomeres that are joined together.

These problems may be addressed through heuristic approaches or through more adequate models of the Bayesian relationships between telomeres and chromosomes. The expected area of the chromosomes and telomeres, for example, could help to separate the structures, as can the spatial distribution of the intensity levels within a "joined chromosome" or "joined telomere" region. Clique-based models, where we look at adjacent pixels in bounded areas, should improve these problems.

Our work will further two ultimate goals: (1) to disseminate ORNL image processing expertise, through software distribution and conference publication, to

Table 1. Results of segmentation algorithm

Image	Num Chr	Num Tel	Found Chr	Found Tel	% Chr	% Tel	Area Chr	Area Tel	Found area Chr	Found area Tel	% Chr area	% Tel area
1	39	155	39	100	100	64.51	45995	5199	15227	3520	33	68
2	39	140	39	103	100	73.57	63859	5230	21963	3291	34	63
3	41	161	41	161	100	100	58192	6972	11786	1783	20	26
4	41	164	41	164	100	100	58079	5925	12512	809	21	14
5	40	160	40	160	100	100	52293	6645	11650	1030	22	16
6	40	160	40	157	100	98.12	68486	6911	9434	1370	14	20
8	40	157	40	103	100	65.6051	26671	4991	5239	3736	20	75

scientists interested in analyzing intracellular structures and (2) to generate preliminary data and publications that can be used to support a planned NIH grant application (PA-03-106 Innovations in Biomedical Computational Science and Technology) in 2004. For the NIH grant application , we intend to address automated, high-throughput telomere analysis of interphase cell images. This project dealt only with metaphase cells. Although telomeres in metaphase cells are easier to quantify since they are usually well separated, the preparation and image acquisition is very difficult. Interphase cells, on the other hand, are much easier to prepare and acquire, and therefore provide a more tractable approach for the high-throughput processing that will be required for drug-screening and robust statistical studies. We believe that the planned follow-on effort will be of great value to the pharmaceutical industry for drug-screening applications.

CHEMICAL SCIENCES AND TECHNOLOGY

Director's R&D Fund

Nanoscale Photosynthesis, the Photophysics of Neural Cells, and Artificial Sight

E. Greenbaum,[1] T. Kuritz,[1] J. W. Lee,[1] F. W. Larimer,[2] I. Lee,[3] B. Bruce,[4] M. Humayun,[5] and S. Sadda[5]
[1]*Chemical Sciences Division*
[2]*Life Sciences Division Oak Ridge National Laboratory*
[3]*Department of Electrical Engineering*
[4]*Department of Biochemistry/Cell and Molecular Biology The University of Tennessee*
[5]*The Doheny Eye Institute/Keck School of Medicine University of Southern California*

The objective of this work was to impart photoreceptor activity to mammalian cells using the previously determined molecular photovoltaic properties of isolated Photosystem I (PSI) reaction centers. Incubation of WERI-Rb-1 retinoblastoma cells with functional PSI reaction centers that were isolated from spinach leaves and reconstituted into proteoliposomes resulted in a light-induced PSI-dependent increase in intracellular Ca^{2+}. The increase, due to Ca^{2+} uptake, was dependent on the presence of extracellular Ca^{2+} ions. Control experiments in the absence of PSI showed no photoactivity. Application of this hybrid photosynthetic-mammalian system to the problem of artificial sight is discussed.

Introduction

Two triumphs of 20th century photobiological research are elucidation of the molecular mechanisms of vision and photosynthesis. Although both processes are initiated by absorption of visible light, they are chemically and energetically different. In vision, light triggers a thermodynamically downhill reaction that is pre-loaded by dark metabolism. Photon absorption by rhodopsin activates a G-protein cascade leading to cyclic guanosine monophosphate (cGMP) hydrolysis that in turn closes cation-specific channels to generate a nerve signal. A single photon absorbed by a dark-adapted rod closes hundreds of cation-specific channels and leads to a hyperpolarization of about 1 mV that is sensed by the synapse and conveyed to other neurons of the retina.[1] In photosynthesis, on the other hand, absorption of photons by the special reaction center chlorophylls in Photosystems I and II trigger charge separations across the photosynthetic membrane. This charge separation generates a voltage that is the source of Gibbs energy for the thermodynamically uphill reactions of green plant photosynthesis: oxidation of water to molecular oxygen and reduction of atmospheric carbon dioxide to sugars.[2] We have shown that isolated Photosystem I (PSI) reaction centers retain their photovoltaic properties[3–5] and that they are single-molecule diodes.[6] Humayun et al. have demonstrated that the human brain can evoke the perception of light and resolve patterns in individuals who are blind due to macular degeneration or retinitis pigmentosa when retinal tissue is judiciously stimulated by passing current through platinum electrodes.[7,8] In these diseases, which are some of the worldwide leading causes of blindness, neural communication from the eye to the brain is still relatively intact: blindness is caused by loss of the first step in the visual process: photon absorption by rhodopsin followed by activation of the G protein cascade. We report here the insertion of photovoltaically active PSI reaction centers into retinoblastoma cell membranes and observation of light-activated influx of calcium ions into the cells.

The objective of the present studies was to determine if photoreceptor activity could be imparted to mammalian cells by direct insertion of PSI reaction centers into the cellular membranes. The goal of the work was aided by the fact that PSI is an integral membrane protein and the expectation that the mammalian cellular membrane would be a reasonably accommodating environment. The modes of action of PSI reaction centers that are inserted into mammalian cell membranes can be expected to fall into at least two categories. First, the reaction centers may trigger transmembrane photoredox chemistry that alters the intracellular chemical potential and initiates biochemical transformations that do not occur in the absence of light. This is similar to the operation of photosynthetic reaction centers in plants. Second, depending on areal density and proximity to voltage-gated ion channels, PSI might gate the ion channels and/or possibly increase ion permeability by a nonspecific electrical disturbance that increases ion permeability. The photoinduced electric field by PSI in a 5-nm membrane is 10^7–10^8 V·m. Theoretically, PSI reaction centers can depolarize or hyperpolarize a membrane depending on orientation of the electron transport vector of the

photoinduced movement of electrons from donor P700 to the acceptor FAB complex within the reaction center.

Technical Approach

PSI reaction center core complexes were isolated from spinach leaves according to the method of Shiozawa et al.[9] These multimolecular complexes contain 40–45 chlorophyll *a* molecules, β-carotene and cytochromes[9] bound to two PsaA-PsaB heterodimers and up to six other polypeptides.[10,11] Activity of the PSI reaction centers was confirmed by differential optical absorption spectroscopy (810–860 nm)[10] using a Walz Model 103 chlorophyll spectrometer. The sample was comprised of PSI core complexes that were reconstituted into proteoliposome[12] membranes and imaged with the combined techniques of tapping mode atomic force (AFM) and surface scanning probe microscopies (SSPM), illustrated in Fig. 1.[13] The one-to-one correspondence between the AFM and SSPM liposome images is evident. The images were obtained under illumination with a diode laser at 670 nm, near the absorption maximum of chlorophyll *a* (671 nm) in PSI-proteoliposomes. The AFM image of Fig. 1(a) illustrates the gross geometric structure of the PSI-proteoliposome, whereas the electrostatic SSPM image, Fig. 1(b), reveals a finer grained pebble-like structure in the surface potential map, suggesting a close-packing of the PSI reaction centers in the liposome membrane. Photoactivities of the native PSI preparation and the PSI-proteoliposome heterostructure compared favorably with previously reported results of other groups for PSI core particles isolated from spinach,[10] *Synechocystis* sp.,[11,14] and *Spirulina platensis*[15] using a variety of lipids and reconstitution techniques.

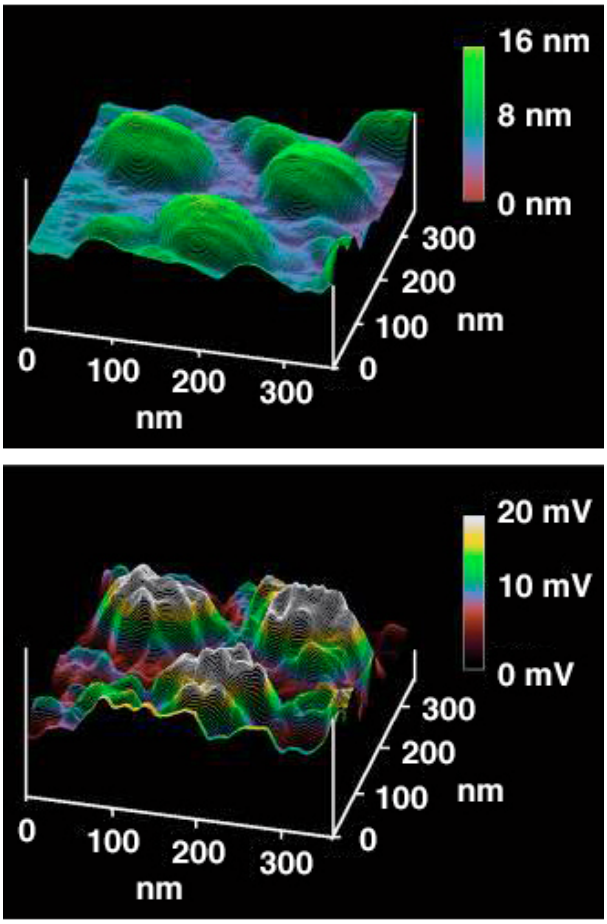


Fig. 1. PSI-proteoliposomes imaged by (a) tapping-mode atomic force (AFM) and (b) scanning surface probe (SSPM) microscopies.

Rigaud et al.[16] pointed out that Triton X-100-mediated reconstitutions resulted in 70–80% orientation of membrane proteins and complexes including reconstitution of lactose transport protein[17] and PSI.[11] Although protein orientation in lipid membranes cannot be predicted from first principles, empirical data demonstrate a dependence of protein orientation on lipid composition of the proteoliposomes. The theory of liposome fusion with biomembranes is far from settled. However, 30-years' experience with liposomes as vectors for drug delivery has resulted in extensive practical experience in the interaction of liposomes with biomembranes and development of numerous liposome-based agents—an increasing number of which are undergoing clinical trials.[18] The liposome preparations reported here used therapeutic model lipid formulations that are known to be incorporated by human and murine cells.[19]

Results and Accomplishments

A retinoblastoma WERI-Rb-1 culture was chosen as the test system for interaction of PSI reaction centers with mammalian cell membranes. WERI-Rb-1 is an undifferentiated, primitive multipotential neuroectoderm culture with the potential to differentiate into neuronal, photoreceptor and glia-like cells making it a reasonable model host with biochemical and immunochemical properties closest to the cells of the optic cup region.[20] Cellular response to light was monitored by visualization of intracellular Ca^{2+} transients.[21] Calcium ion movement is a signal for many physiological events, including photoreception in vision, and is an important factor in cell adaptation to environmental impacts. The culture of human retinoblastoma cells (WERI-Rb-1, ATCC) was anchored to the bottom of the poly-D-lysine precoated slide wells by overnight incubation in the growth medium (RPMI 1640, Invitrogen, Carlsbad, CA). The cells were counted and washed with Hanks saline with or without Ca^{2+} as appropriate, and the proteoliposomes (0.5–2000 ng phosphorus in 500 µl of Hanks saline) were added in different ratios to the cells. All subsequent operations were performed in strict darkness. The slide wells were

incubated for 1 h under growth conditions (37°C, 5% CO_2), the supernatant decanted, and after a wash with Hanks saline were replaced with 500-µl Hanks saline supplemented with 16-µM Fluo-3 AM (Molecular Probes, Eugene, OR) in dimethylsulfoxide (DMSO, Aldrich Chemical, Milwaukee, WI), and 0.02% Pluronic F-127 (Molecular Probes), followed by incubation at 37°C for 2 h. The cells were then washed in the dark with an excess of Hanks saline supplemented with 1 mM ascorbate to ensure that P700 was reduced prior to illumination.

Spectral properties of the system components were selected so that the excitation and emission peaks of the Ca^{2+}-binding dye Fluo-3 did not overlap with the absorption maxima of chlorophyll *a* in PSI particles in the proteoliposomes (Fig. 2). Light-induced Ca^{2+} ion transients were initiated by illumination at 670 nm (1.5 nm bandwidth) using a diode laser (OZ Optics, Ltd., Carp, Ontario). In the presence of an extracellular calcium concentration of 1.26 mM, we observed light-induced intracellular Ca^{2+} fluorescence transients in WERI-Rb-1 cells that were treated with PSI-proteoliposomes containing a concentration of 6.3 ± 2.6 pg [P] per cell. A 3-sec exposure to the 2.5–3.0 µE s^{-1} m^{-2} beam of 670 nm laser light stimulated an increase in intracellular fluorescence in the cells 500 sec after illumination and climbed to a maximum after 1000 sec. The cells then started rupturing and releasing the dye [Fig. 3(a)]. As expected, the dye release led to an abrupt decrease in fluorescence. The time course of calcium increase, bursting of the cells, and abrupt release of the dye illustrates the power and promise of photoreceptor activity by PSI reaction centers. For the WERI-Rb-1 cells chosen for these experiments, cell rupture occurs only at the high PSI-proteoliposome loading for the data of Fig. 3. Lower

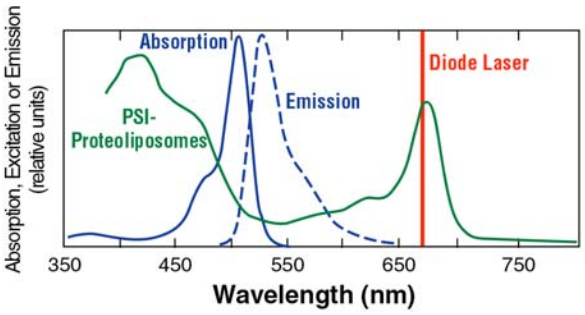


Fig. 2. Absorption, excitation, and emission spectra of the system components. For the Ca^{2+} specific dye Fluo-3, λ_{exc} = 490 nm; λ_{em} = 540 nm (Source: Molecular Probes Catalog, Eugene, Oregon). The green curve is the measured absorption spectrum of PSI proteoliposomes, λ_{abs}^{max} = 671 nm obtained with at Shimadzu Model UV-160 spectrophotometer. The vertical red line at 670 nm is the wavelength emission of the actinic diode laser used to trigger the Ca^{2+} influx (Source: Oz Optics, Ltd., Carp, Ontario). Unlike free chlorophyll in solution, chlorophyll bound to PSI reaction centers emits virtually no fluorescence at room temperature.

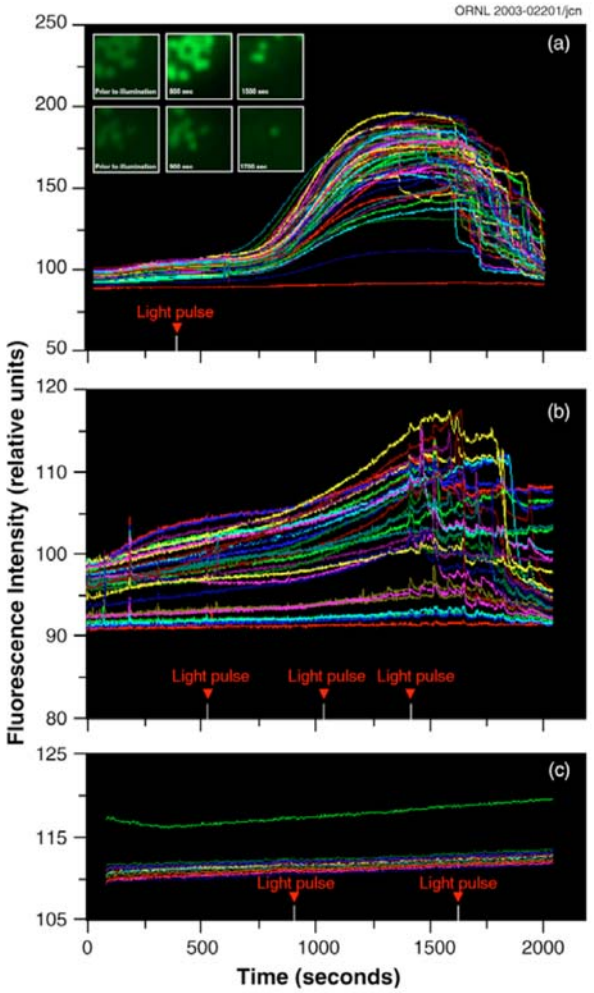


Fig. 3. Response to illumination at 670 nm by PSI-proteoliposome treated WERI-Rb-1 cells in the (a) presence and (b) absence of external Ca^{2+} ions; (c) control experiment with liposome-fused WERI-Rb-1 in the absence of PSI. Inset: Pseudocolor fluorescence image of WERI-Rb-1 cells fused with PSI proteoliposomes at selected time points in the presence (a) and absence (b) of external Ca^{2+} ions. Data were obtained with a Nikon Eclipse TS-100 microscope.

loading does not cause cell death. The reason these are promising results is that elementary electrostatics predicts that charging of the membrane capacitance and the corresponding photogenerated neural voltages can be controlled by appropriate titration of the PSI-proteoliposomes. This conclusion is also supported by the work of Packham et al.,[22] on the areal density-dependent photovoltage of bacterial photosynthetic reaction centers that were incorporated in planar bilayer membranes. Moreover, recent preliminary data indicate that rat retinal progenitor and ganglion cells do not die when treated with proteoliposomes at the concentrations that cause death in WERI-Rb-1.[23] Cell death apparently depends (1) on the nature of the cells; (2) on the dose of PSI-proteoliposomes (data not shown), and (3) on the environment.

The photoinduced increase in intracellular Ca^{2+} occurred simultaneously in all cells. Control experiments [Fig. 3(b)] in the absence of external calcium indicated that cells treated with PSI-proteoliposomes at a concentration of 7.2 ± 4.8 pg [P] per cell also responded to illumination. However, the intensity of fluorescence was about 10-fold less than in the presence of external Ca^{2+}. In the absence of external Ca^{2+}, the light-induced Ca^{2+} spikes in Fig. 3(b) were probably due to the release of intracellular stores. Additional control experiments with liposomes only (no PSI present) at a concentration of 75 pg [P] per a cell [Fig. 3(c)] showed no photoactivity and did not affect cell viability. Figure 4 presents a statistical summary of light-induced fluorescence images in the absence (blue bars, left) and presence of calcium (green bars, right). The bimodal distribution of Fig. 4 demonstrates a 10-fold greater photoinduced fluorescence intensity in the presence of external calcium. Further control experiments with inactive (chlorophyll present, but no photoinduced charge separation) PSI-proteoliposomes and the low-energy fluence used for these experiments exclude photodynamic therapy-like (PDT) effects as the origin of the light-induced Ca^{2+} signal. PDT treatments utilize photon fluences about 1000 times higher than that used in our experiments.[24]

In natural photosynthesis, photon-induced picosecond primary charge separation set in motion the much slower processes of biosynthesis and carbon dioxide assimilation.[25] Similarly, the kinetics of the PSI optically triggered increase in Ca^{2+} concentration in retinoblastoma cells is relatively slow. Yet, that is not surprising since Ca^{2+} is a universal signal messenger and multiple sequential reactions are likely to be set in motion by light-activated PSI. The calcium ion is a universal messenger in eukaryotic signal transducing pathways such as vision, muscle contraction, and many other physiological events.[1] The relatively slow kinetics of extracellular Ca^{2+} uptake (Ca^{2+} channel opening) may be secondary reactions of intracellular biochemistry that are triggered by PSI and the PSI-generated transmembrane voltage in the mammalian cell and not by voltage gating of ion channels directly. This interpretation is consistent with the description of lipid-gated transient receptor potential (TRP) channels.[26] The long-lasting Ca^{2+} influx may be due to the combined effect of TRP with analogs of *Drosophila* postsynaptic density proteins (PDZ), hypothesized to cause a long-persisting signal in retinal neuroepithelium.[27] This suggestion is supported by the findings that the TRP family can respond indirectly to a number of sensory stimuli.

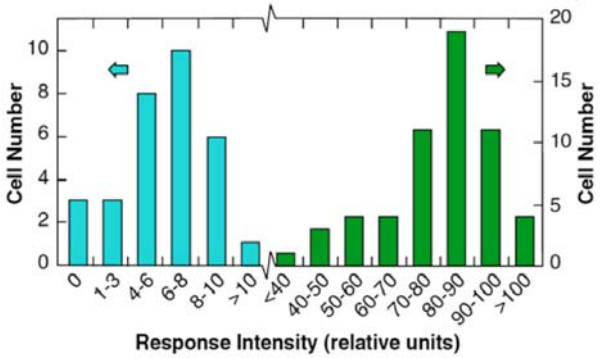


Fig. 4 Statistical distribution of fluorescence images in response to 670-nm illumination in cell populations fused with PSI-proteoliposomes. The bimodal distribution illustrates the effect of the absence (blue bars) and presence (green bars) of externally added calcium. The response was much higher (median intensity of 80–90 brightness units, n = 57) in the cells provided with external Ca^{2+} and lower in the cells devoid of external Ca^{2+} (median intensity of 6–8 brightness units, n = 31). The cells bore a similar load of proteoliposomes (6.34 ± 2.61 pg [P]/cell with external Ca^{2+} vs 7.21 ± 4.8 pg/cell without external Ca^{2+}).

Summary and Conclusions

1. In conclusion, and irrespective of the to-be-determined photobiophysics and photobiochemistry, two biological structures that are evolutionarily separated by at least 2 billion years have been integrated into a functional hybrid photosynthetic-mammalian cell light-induced calcium transport system.
2. The results of our project are of direct benefit to the programmatic mission of DOE's Medical Sciences Division and the NIH's National Eye Institute.
3. This is will be a new FWP in DOE's Medical Sciences Division. Dr. Dean Cole has informed E. Greenbaum that $400K in new funding will be made available to ORNL in FY 2004 to continue the promising results of this successful project.

References

[1]L. Stryer, *Biochemistry*, W. H. Freeman and Company, New York (1995). D. Baylor, *Proc. Natl. Acad. Sci. USA* **93**, 560–565 (1996).

[2]R. Blankenship, *Molecular Mechanisms of Photosynthesis*, Blackwell Science, Oxford, UK, (2002).

[3]I. Lee, B. L. Justus, J. W. Lee, and E. Greenbaum, submitted for publication (2003).

[4]I. Lee, J. W. Lee, A. Stubna, and E. Greenbaum, *J. Phys. Chem. B* **104**, 2439–2443 (2000).

[5]J. F. Millsaps, B. D. Bruce, J. W. Lee, and E. Greenbaum, *Photochem. Photobiol.* **73**, 630–635 (2001).

[6]I. Lee, J. W. Lee, and E. Greenbaum, *Phys. Rev. Letters* **79**, 3294–3297 (1997).

[7]M. S. Humayun, M. Prince, E. J. de Juan, Y. Barron, M. Moskowitz, I. B. Klock, and A. H. Milam, *Invest. Ophthalmol. Vis. Sci.* **40**, 143–148. (1999).

[8]S. Kim, S. Sadda, J. Pearlman, M. Humayun, E. J. de Juan, J. Melia, and W. Green, submitted to *Retina* (2002).

[9]J. A. Shiozawa, R. S. Alberte, and J. P. Thornber, *Arch. Biochem. Biophys.* **165**, 388–397 (1974).

[10]S. Hoshinah and S. Itoh, *Plant Cell Physiol*. **28**, 599–609 (1987); J. Cladera, J.-L. Rigaud, H. Bottin, M. Duñach, *J. Bioenerg. Biomembr.* 28, 503–515 (1996)].

[11]K. Gourovskaya, M. D. Mamedov, I. R. Vassiliev, J. H. Golbeck, and A. Yu. Semenov, *FEBS Lett.* **414**, 193–196 (1997).

[12]The liposomes were prepared by sonicating in buffer (25 mM Tris-HCl, pH 7.8, 10 mM MgCl2, 10 mM NaCl, 1 mM DTT) a mixture of hydrogenated soy phosphatidyl choline and cholesterol (10:1, molar ratio). The proteoliposomes were obtained following solubilization of the liposomes in 1% Triton X-100 according to the methodology of ref. 10 followed by addition of PSI complexes (100–200 times by weight) and reconstitution by gel-chromatography on a Sepharose 6B column equilibrated with the buffer. PSI to lipid ratios were determined by analysis of total phosphorus [C. H. Fiske and Y. Subbarow, *J. Biol. Chem.* **66**, 374–389, (1925)] and by chlorophyll *a* measurement according to MacKinney (*J. Biol. Chem.* **140**, 315–322, 1941) after proteoliposome solubilization in chloroform

[13]The apparent range of liposome diameters was 70–100 nm. The AFM-SSPM technique uses a slender cantilever probe with a slightly blunt apex. It provides accurate voltage measurements but exaggerates lateral dimensions. The theory for this technique has been developed by Jacobs et al. [*J. Appl. Phys.* **84**, 1168–1173 (1998)]. References 4 and 6 provide additional information on our techniques for working with single PSI reaction centers.

[14]J. Cladera, J. L. Rigaud, H. Bottin, and M. Dunach, *J. Bioenerg. Biomembr.* **28**, 503–515 (1996).

[15]J. Kruip, N. V. Karapetyan, I. V. Terekhova, and M. J. Rogner, *J. Biol. Chem*. **274**, 18181–18188 (1999).

[16]J.-L. Rigaud, B. Pitard, and D. Levy, *Biochim Biophys Acta* **1231**, 223–246 (1995).

[17]J. Knol, K. Sjollema, and B. Poolman, *Biochemistry* **37**, 16410–16415 (1998).

[18]R. Blumenthal, M. J. Clague, S. R. Durell, and R. M. Epand, *Chem. Rev.* **103**, 53–69 (2003); K. J. Harrington, K. N. Syrigos, and R. G. Vile, *J. Pharm. Pharmacol.* **54**, 2573–1600 (2002); K. Takiguchi, F. Nomura, T. Inaba, S. Takeda, A. Saitoh, and H. Hotani, *Chem Phys Chem*. **3**, 571–574 (2002). These references provide additional data on the fusion of liposomes with lipid bilayers.

[19]I. Ahmad and T.M. Allen *Cancer Res.* **52**, 4817–4820 (1992); D. E. Lopes de Menezes, L. M. Pilarski, and T. M. Allen, *Cancer Res.* **58**, 3320–3330 (1998).

[20]Retinoblastoma is a childhood intraocular tumor due to uncontrolled growth of undifferentiated retinal photoreceptor (rod) cells [R. C. McFall, T. W. Sery, and M. Makadon, *Cancer Res*. **37**, 1003–1010 [1977)]. M. Campbell and G. J. Chader *Ophth. Paed. Genetics* **9**, 171–199 (1988); G. J. Chader, *Cell Different.* **20**, 209–216 (1987); and A. Kyritsis, G. Joseph, and G. J. Chader, *J. Natl. Cancer Inst.* **73**, 649–654 (1984) reported the ability of retinoblastoma cells to differentiate into cells with characteristics of photoreceptors, conventional neurons, glia and pigment epithelium.

[21]Slide wells with Fluo-3-loaded WERI-Rb-1 cells were fixed in the dark on a thermostated stage (Bionomics BC-1, Technology 2020) and maintained at 37°C throughout the experiment. The dye was excited at 490 nm. Fluo-3 fluorescence emission was collected by an intensified charge coupled device (CCD) (Quantix, Photometrics). We chose a one-second interval for the data acquisition time for digital image collection and a one-second interval during which no data was collected before repeating the cycle. These time intervals allowed acquisition of high-quality images while avoiding excessively large files. Data were collected for up to 4000 sec and processed using Metamorph software (Universal Imaging, West Chester, Pennsylvania). The cells were counted before and after each experiment.

[22]N. K. Packham, P. L. Dutton, and P. Mueller, *Biophys. J.* **37**, 465–473, (1982).

[23]T. Kuritz, G. Qui, E. T. Owens, D.-Y. Wu, S. Sadda, M. Humayun, and E. Greenbaum, unpublished data (2003).

[24]M. L. Agarwal, H. E. Larkin, S. I. A. Zaidi, H. Mukhtar, and N. L. Oleinik, *Cancer Res*. **53**, 5897–5902 (1993).

[25]P. R. Chitnis, *Ann. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 593 (2001). Also reference 2.

[26]C. D. Benham, J. B. Davis, and A. D. Randall. *Neuropharmacol*. **42**, 873–888 (2002); C. Zitt, C. R. Halaszovich, and A. Luckhoff, *Progr Neurobiol* **66**, 243–364 (2002); M. C. Nowycky and A. P. Thomas, *J. Cell Sci.* **115**, 3715–3716 (2002).

[27]S. Tsunoda, J. Sierralta, Y. Sum, R. Bodner, E. Suzuki, A. Becker, M. Socolich, and C. S. Zuker, *Nature* **388**, 243–249 (1997).

Protein Microarray Interactions Readout Using Stepping Sampling Probe/ Electrospray Mass Spectrometry

G.. J. Van Berkel,[1] M. J. Doktycz,[2] G. E. Giles,[3] and S. J. Kennel[2]
[1]*Chemical Sciences Division*
[2]*Life Sciences Division*
[3]*Computational Sciences and Engineering Division*

A rapid, specific, and sensitive method for serial readout of interacting components on protein chip microarrays has been developed. This method makes use of a novel analytical detection concept that requires no extrinsic labeling, which we call stepping sampling probe/electrospray mass spectrometry (SSP/ES-MS). The project consisted of three separate subtasks, viz., development of the probe-readout device, development of affinity capture arrays for readout, and furthering of the "top-down" protein identification of proteins via tandem mass spectrometry. Protein array interactions are read out (i.e., identified) using a miniature sampling probe, connected to an electrospray ion source coupled with a mass spectrometer. The mass spectrometer was used to identify the interacting species on the basis of molecular mass alone or by means of tandem mass spectrometric analysis. Tandem mass spectrometry (MS/MS) of whole proteins in combination with high-resolution accurate mass determinations was shown to generate sequence tags or other diagnostic fragmentation allowing identification of unknown proteins via data-based searching ("top-down" identification). In addition, the developed surface sampling system has demonstrated an unanticipated analytical utility in the read out of thin-layer chromatography (TLC) plates, providing a capability long desired in the analytical chemistry community for on-line TLC/MS. With further experimentation it is anticipated that other important surfaces in biology research, such as protein blotting membranes, will be amenable to analysis with this device.

Introduction

Two-dimensional gel electrophoresis (2-DE) is the most common and powerful platform for monitoring protein expression. Protein "arrays" or "chips" are one potential alternative technology. However, protein microarrays cannot easily exploit the simple labeling and detection (readout) schemes used for DNA microarrays, and thus, detection methods that require no extrinsic labeling are being sought. Mass spectrometry offers a powerful detection option that requires no labeling. While the use of ES-MS for protein-chip-array readout has great potential, it is yet to be demonstrated because of the technical challenge of sampling analytes from a small spot on a sample surface with a liquid flow system in an automated way. The stepping sampling probe/electrospray-mass spectrometry (SSP/ES-MS) approach to reading out protein chip microarrays addresses this challenge. This unique ES-MS surface sampling system combined with our state-of-the-art mass spectrometry instruments is expected to ultimately provide fast, automated, sensitive, and molecular mass and structure-specific means of protein-array readout that cannot be matched by other technologies.

Technical Approach

This two-year project was subdivided into three subtasks that could be independently successful, viz., development of the probe-readout device, development of affinity capture arrays for readout with this probe sampling device, and the furthering of "top-down" protein identification via tandem mass spectrometry. The first six months of the project focused on development of a novel self-aspirating surface sampling probe device. The successful completion and implementation of this device for sampling of surface spotted proteins and thin-layer chromatography (TLC) plates[1] led to submission of a patent application.[2] For the remainder of the project period, emphasis was placed on the other two subtasks, along with application of the probe to the read out of a variety of different surfaces in addition to affinity arrays. The foray examining other surfaces was aimed at exploiting for follow-on funding alternative uses for the probe sampling device.[3]

The affinity capture portion of the project was studied largely off-line from the mass spectrometer using ^{125}I labeled proteins and the corresponding antibodies. In this way protein capture density on the arrays and protein elution efficiency from the arrays with ES-MS-compatible

solvents could be optimized using established detection methodologies. Successful readout of affinity captured proteins, the major milestone of this project, was demonstrated.[4]

Identification of proteins on the basis of the tandem mass spectrometry, top-down approach was studied using two dissociation methods—sustained off-resonance irradiation (SORI) and multipole storage-assisted dissociation (MSAD). Each method was evaluated for generating sequence tag data from intact proteins by FTICR mass spectrometry.

Results and Discussion

Protein Array Preparation. Selective capture of proteins on arrays amenable to readout by the probe sampling device was accomplished by covalently linking antibody to the desired surface and then adding the target protein for specific binding (capture) by the antibody. In the first year of support, FY 2002, this was accomplished for capture of several cytokines using antibodies attached to either glass microscope slides or silicon. Glass slides were masked with Teflon grids leaving 1-mm squares exposed with approximately 1-mm Teflon spacing. Experiments were done with radio-iodinated proteins to quantitate the attachment of the antibody and capture efficiency of the target protein and elution characteristics of the captured protein with solvents compatible with ES-MS. In the second year, approximately six major experiments were done largely in capturing murine gamma interferon, γIFN, with antibody bound to glass slides via a poly-l-Lysine gluteraldehyde bridge. The slides were prepared in parallel, one set with radioactive reagents for quantitative analysis and one for mass spectrometry. Specific capture and elution of radio-labeled γIFN was verified in every experiment. Quantitation of absolute amounts of captured and eluted γIFN varied among the experiments. By use of SDS-PAGE, it was determined that γIFN dissolved in water or normal saline solutions without carrier protein was not stable. Degradation occurred over days to weeks of storage at –20°C with several freeze/thaw cycles. In contrast, protein that had been radio-iodinated and diluted in buffer containing carrier protein (0.5% BSA) was stable and showed consistent integrity as analyzed by SDS-PAGE. In the last several weeks of the project, slides with two specialized ceramic surfaces were tested for their capacity to couple antibody and to capture and subsequently release γIFN. These results indicated that the ceramic slide surface bound about two-fold higher amounts of antibody per unit area than did the glass slide treated in an identical manner. Capture of radio-iodinated γIFN was also about two-fold higher, consistent with the increased amount of capture antibody present. Elution characteristics of the radiolabeled proteins from glass or ceramic were indistinguishable (Table 1). A commercially available aldehyde pre-derivatized ceramic slide performed about the same as the flat glass slide. Two explanations for this result were suggested: (1) the aldehyde derivatization was less dense as deposited by the commercial company than that applied here through the poly-l-Lysine gluteraldehyde bridge, or (2) some loss of reactivity of the aldehdye occurred over the time of shipping and storage of the slides before use.

Table 1. Comparison of glass, ceramic, and aldehyde pre-derivatized ceramic surfaces for antibody-based capture of γ IFN

	Glass	Ceramic	Ceramic prepared
Maximum amt Ab bound,ng/mm^2	33	65	46
Maximum amt γ IFN captured, ng/mm^2	0.35	0.76	0.49
Fraction eluted, %	97	98	92

For the expressed goal of ES-MS analyses, the captured γIFN must be rapidly eluted from the antibody with solvents compatible with ES-MS. In the first year of experiments, we determined that acetonitrile:water:formic acid at 70:30:0.1 volume ratio was the best compromise solvent for elution and spraying. The ability of this solvent to elute captured γIFN was evaluated in each experiment. When elutions were conducted within minutes after removal of the wash solution, greater than 90% of the radioactivity was eluted from all capture surfaces (Fig. 1). The elution occurred with a few second exposure to the solvent. More rapid kinetics could not be evaluated with the radioisotope method. However, if slides with captured γIFN were dried and stored for 18–24 h, the fraction of radioisotope that could be eluted dropped to 50–60%. This phenomenon is difficult to explain at the molecular level but has been observed in other capture systems as well (data not shown).

In summary, efficient capture of cytokines with antibody immobilized on flat surfaces has been demonstrated. The capture is specific, saturable, and

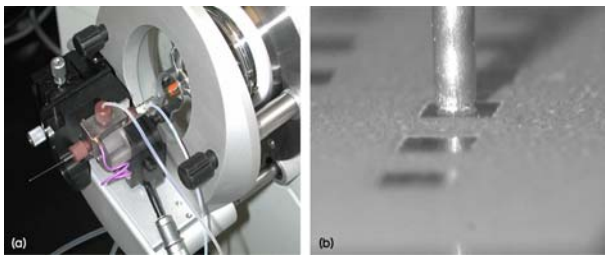


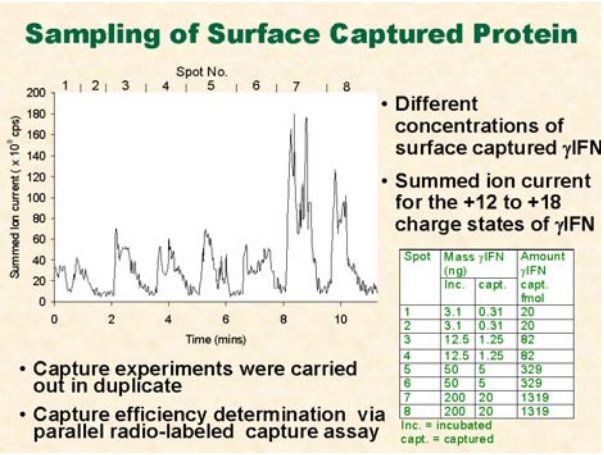
Fig. 1. (a) Surface sampling probe mounted at the atmospheric sampling inlet of a DecaXP Plus ion trap mass spectrometer. (b) View of sampling end of probe sampling from a spot on an affinity array arranged on a glass microscope slide.

related to antibody amounts bound to the capture surface. The captured cytokine can be eluted in nearly quantitative fashion under the right conditions and should be available for detection by ES-MS.

Surface Sampling Probe. Pictures of the surface sampling probe mounted on an ion trap mass spectrometer and the probe sampling from the well of a masked array slide are shown in Fig. 1. Using this probe system, detection capabilities were tested using lysozyme (15153 Da) deposited directly onto a masked glass slide using a peizioelectric spotting device. The amount of protein spotted was varied by increasing the number of droplets deposited in a single location. Each droplet from the peizioelectric spotter was 14 pL in volume and equivalent to 5 fmol of lysozyme per droplet. Two lysozyme spotting experiments were conducted; the first using 60.5, 302.5, 605, 1210, 6050 fmol of lysozyme as the amount spotted; the second using a more dilute series 1, 5, 10, 20, 50, 100 fmol spotted into the wells of the masked slide. Blank spots, with no spotted material, were sampled for both experiments. The first series of spots was sampled 7 times; the second series was sampled 3 times. Examination of the protein signals obtained (+8–+12 charge states) showed that spot to spot reproducibility was good (ca. ±10 %RSD) and linearity in signal with amount spotted up to about 1.2 pmol where signal roll-over occurred. Detection levels of 5 fmol of lysozyme deposited by a peizioelectric spotter were demonstrated.

Figure 2 shows a mass chromatogram acquired from the analysis of surface captured γIFN using the surface sampling ES-MS system. The data shown is the result of the duplicate analyses of 5 different amounts of surface captured protein. The chromatogram is representative of the +10–+19 charge states of this protein. The table in Fig. 2 shows the amounts of γIFN incubated in each masked off area of the slide and the respective amount captured by the antibody as determined by capture efficiency values generated for the array using a parallel radio-labeled assay. The sampling conditions used differed from those used for the surface spotted material in that the operators discretion played a more substantial role in whether or not a sampling event was complete or not. The ES mass spectrum and deconvoluted (zero charge) spectrum in Fig. 3 were obtained by sampling from spot 6 in Fig. 2. This array spot contained about 329 fmol of γIFN. The clearly visible multiply-charged peaks of the protein and the correct mass of the protein determined in the deconvoluted spectrum indicate the successful sampling of a surface captured protein. The controlled experiments did not show a discernable spectrum for γIFN, indicating that our positive result in Fig. 3 was the result of an antibody specific interaction. Spots 7 and 8 which contained more protein gave even better mass spectra. However, when sampling the array spots containing less than 300 fmol of protein, recognizable protein mass spectra were not observed.

Top-Down Protein Identification. Protein samples are often analyzed using a "bottom-up" approach, which typically involves one- or two-dimensional chromatography or electrophoresis, proteolytic digestion of the resulting fractions, and mass spectrometric fingerprinting of component peptides. Tandem mass spectrometry is sometimes performed to provide a more thorough analysis of particularly complex protein mixtures. However, the enzymatic digestion process central to the bottom-up approach not only complicates sample handling but also generates large quantities of data that can be difficult and time-consuming to process. In addition, valuable information about post-translational modifications to the sequence encoded by the genome can be lost because intact protein masses are not determined.



Spot	Mass γIFN (ng) Inc.	capt.	Amount γIFN capt. fmol
1	3.1	0.31	20
2	3.1	0.31	20
3	12.5	1.25	82
4	12.5	1.25	82
5	50	5	329
6	50	5	329
7	200	20	1319
8	200	20	1319

Fig. 2. Mass Chromatogram obtained from the surface sampling of four different amounts of surface captured γIFN.

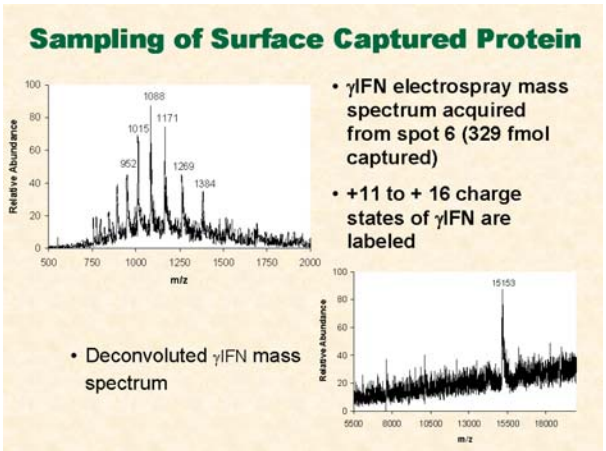


Fig.3. Electrospray mass spectrum and deconvoluted (zero charge) spectrum acquired from sampling of surface captured γIFN from spot 6 in Fig. 2.

The array-readout strategy aims to avoid these time consuming and labor intensive steps in the protein identification.

We explored a "top-down" protein identification strategy that eliminates the digestion process and instead involves direct MS/MS analysis of intact proteins. In this approach, primary sequence information derives exclusively from fragmentation patterns, so robust ion activation techniques are critical. The most effective methods of protein fragmentation include sustained off-resonance irradiation collisionally activated dissociation (SORI-CAD), nozzle-skimmer dissociation (NS), infrared multiphoton dissociation (IRMPD), and electron capture dissociation (ECD).

To date top-down protein analysis has largely been restricted to infusion experiments, in which analyte concentrations are constant. In a variety of applications, such as LCMS or high-throughput protein array analysis techniques under development in this project, analytical signals are transient in nature. Fourier-transform ion cyclotron resonance (FTICR) platforms offer excellent resolution and mass range capabilities in such applications. However, certain dissociation methods available in FTICR instruments, such as SORI and ECD, possess relatively low-duty cycles, and significant portions of a transient signal profile may be poorly sampled. IRMPD offers a relatively high-duty cycle, and was successfully incorporated by Li et al. in the LC-FTMS analysis of a mixture of intact proteins.

IRMPD is not universally available, however, so a role still exists for other methods that allow efficient acquisition of MS/MS data. In multipole storage-assisted dissociation (MSAD), space charge-mediated dissociation occurs in the linear RF-only multipoles commonly used to interface electrospray ionization with FTICR cells. Under certain conditions (extended accumulation times and/or larger dc offset voltages), ion activation and dissociation can be induced during the ion accumulation interval in an inherently high-pressure region of the instrument, so no pump-down periods are required and the overall duty cycle is quite high. Linear ion traps are often used to interface electrospray ionization with FTICR cells, so MSAD can be readily performed on many FTICR instruments with no additional components.

We undertook a systematic comparison of SORI and MSAD for a variety of 8–18 kDa proteins, with particular emphasis on the relative ability of these methods to provide fragmentation information that can be used in database search strategies for the positive identification of the parent proteins. MSAD spectra of intact proteins are influenced by several factors. Data for bovine ubiquitin suggests that because dissociation patterns are charge state dependent, the charge state distribution made available by the ionization conditions may dictate the range of fragment ions that can be generated during the experiment. In addition, conditions of high space charge within the hexapole impair transmission and/or trapping of high m/z species, which can result in loss of important precursor and product ions. And finally, the nonresonant nature of activation in MSAD can provide access to secondary dissociation processes that are not available by SORI. Because of these considerations, the two activation methods do not always provide the same fragmentation information, and MSAD is somewhat less reliable for generating sequence tag data. However, it appears that in general MSAD samples "preferred" cleavage processes (i.e., those occurring at D and P residues) just as well as SORI, which implies that MSAD data may be somewhat more compatible with search algorithms that utilize unprocessed fragment ion masses.

In these experiments the use of MSAD reduced overall analysis time by a factor of two to six over SORI. Other potential advantages of the technique include the fact that it is not necessary to pulse a collision gas into the cell, nor is it necessary to accomplish resonant isolation and activation. However, it remains to be seen whether these advantages can be exploited in the analysis of transient signals. While the nonresonant nature of MSAD could provide rapid, high-throughput analysis of multiple analytes, it also suggests that data interpretation for complex mixtures is potentially problematic. In addition, the charge density dependence of MSAD implies that the degree of fragmentation observed across a transient peak would vary with concentration (i.e., from low to high to low). Judicious choice of MSAD parameters would be necessary to maximize overall dynamic range and sensitivity, and multiplexed scan functions might prove beneficial. Additional experiments will therefore be required to fully characterize/optimize MSAD for applications such as surface sampling or possibly LCMS.

Conclusions

A self-aspirating electrospray probe/emitter coupled with an ion trap mass spectrometer was used to sample and mass analyze proteins deposited onto glass slides or affinity captured on glass slides. Detection levels of 5 fmol of lysozyme deposited by a peizioelectric spotter were demonstrated. From this amount of protein, the correct molecular mass was obtained from the mass spectrum using a commercial deconvolution algorithm. Proof-of-principle sampling of antibody captured proteins was demonstrated for gamma interferron (γ-IFN) using surface immobilized antibody. Results point to increased surface binding capacity and smaller volume probe to limit volumetric dilution and improve ionization efficiency by operating at sub-microliter per minute flow rates. Additional experimental and development work is needed to allow the exploitation of top-down protein identification

of proteins eluted from affinity arrays. While the transients signals proof-of-principle from this study and others in the literature indicate that this possibility may not be far off.

Our demonstrated implementation of this protein array readout technique will lead to an increase in the value of protein arrays in understanding protein-expression monitoring and will establish it as a suitable method to measure murine cytokine levels during a biological reaction. Success will lead to outside interest in development of commercial products that will transcend the current MALDI-MS approach to protein chip array read out. Our success will now also allow us to effectively compete for outside funding from federal agencies including the National Institutes of Health (NIH), the National Cancer Institute, and the Office of Biological and Environmental Research, as well as other sections at DOE, exploiting the current SSP/ES-MS technology for the study of protein expression. With the current data in hand we are in a position to submit a proposal for funding to NIH. Our success will also allow us to pursue funding to develop a refined device, possibly based on the mating of protein microarrays with soon to be commercialized microfabricated arrays of ES nozzles for faster, fully automated serial read out of chip arrays. This tool will meet the call in the DOE Genome To Life program document for revolutionary advance in protein chip arrays and mass spectrometry for the study of protein chemistry. Thus, ORNL's position as a leader in Complex Biological Systems and innovator in characterizing protein biochemistry as part of the DOE Genomes To Life Program will be strengthened.

References

[1]G. J. Van Berkel, A. D. Sanchez, J. M. E. Quirke "Thin-Layer Chromatography and Electrospray Mass Spectrometry Coupled using a Surface Sampling Probe," *Anal. Chem.* **2003***, 74,* 6216–6223.

[2]"Sampling Probe for Microarray Read Out Using Electrospray-Mass Spectrometry," U.S. Patent Application filed April 16, 2002. Inventor: G. J. Van Berkel.

[3]M. J. Ford and G. J. Van Berkel, "Thin-Layer Chromatography and Electrospray Ion Trap Mass Spectrometry Coupled using a Surface Sampling Probe," *Anal. Chem.*, in preparation.

[4]G. J. Van Berkel, M. J. Ford, S. J. Kennel, and M. J. Doktycz, "Electrospray Mass Spectrometry System for Analysis of Proteins Deposited or Captured on Surfaces," *Rapid Commun. Mass Spectrom.*, in preparation.

Reactive Membranes for Clean Coal Technologies

D. W. DePaoli,[1] S. Dai,[2] D. C. Duckworth,[2] J. M. Simonson,[2] C. V. Thompson,[2] M. R. Ally,[3] and R. E. Baltus[4]
[1]*Nuclear Science and Technology Division*
[2]*Chemical Sciences Division*
[3]*Engineering Science and Technology Division*
[4]*Clarkson University*

Coal is the dominant U.S. resource for producing electricity. However, emissions from coal combustion contribute to potential global warming and pose significant environmental threats. This project is aimed at new separations technologies based on ionic liquids that would allow sustainable use of coal resources. Because they have tunable chemical properties, a wide stable temperature range, relatively benign character, and essentially zero vapor pressure, ionic liquids show great promise as reaction and separations media. This project has explored the feasibility of ionic-liquid-based separations technologies for removal of carbon dioxide and mercury from flue gas through laboratory experimentation, thermodynamic modeling, and preliminary economic evaluation. Ionic liquids having significantly higher solubility of carbon dioxide than those previously reported in the literature were synthesized; preliminary analysis indicates that, with further development, absorption or membrane processes based on these liquids may be competitive with existing technologies. An ionic liquid with high affinity for elemental mercury was discovered and shown to be effective for removal of mercury from the gas phase; to enable practical application, further research is needed to determine effective means of regeneration. This project is expected to advance DOE's energy-security mission by providing initial data on ionic-liquid-based separations technologies for clean-energy processes.

Technical Approach

Currently, approximately 50% of the country's electricity is generated from coal, and it remains a vital raw material for the United States and world economy. However, emissions from coal power—including greenhouse gases, acid gases, and mercury—pose significant environmental threats.[1] Because carbon dioxide (CO_2) is generally recognized as the major greenhouse gas leading to global warming, there has been increasing interest in the development of economically viable technologies for the removal of CO_2 from flue gas streams. The President's Energy Plan identifies emission of mercury as a major obstacle to clean coal power. The Environmental Protection Agency is expected to release new standards for power-plant mercury emissions by December 2004. The regulations will aim for a 90% reduction in the 43 tons of mercury released annually by coal-fired power plants.[2] With the current threat of mercury contamination, and the pending EPA regulations, new, cost-effective technologies for mercury removal are sought.

This project was aimed at determining the technical feasibility of creating "designer" ionic liquids for clean-energy processes, such as the selective removal of carbon dioxide and mercury from gas streams produced by coal combustion. Ionic liquids are relatively new compounds that have received increased attention in recent years as "green" designer solvents that may potentially replace many conventional volatile organic solvents in reaction and separation processes.[3] These unique compounds are organic salts that are liquid over a wide range of temperatures near and at room temperature. Ionic liquids have no measurable vapor pressure; hence, there has been considerable interest in using them in place of volatile organic solvents that can emit problematic vapors.

The most commonly investigated room-temperature ionic liquids have an alkyl-substituted imidazolium salt as the cation. The imidazolium-based salts are relatively easy to synthesize and have physical properties that make them attractive for many chemical processes. The generally accepted acronym for the alkyl-substituted methyl imidazolium cations used in these studies is C_nmim^+, where *n* is the number of carbon atoms on the alkyl side chain. Common anions used in ionic liquids include bis[trifluoromethylsulfonyl] amide (Tf_2N^-), hexafluorophosphate (PF_6^-), tetrafluoroborate (BF_4^-), and ethyl sulfate ($EtSO_4^-$), although ionic liquids with a variety of other anions have also been synthesized.

This project employed ORNL capabilities in synthetic chemistry, thermodynamics, analytical chemistry, and engineering research to investigate the feasibility of gas separations using ionic liquids. Elements of the work included (1) synthesis and screening of ionic liquids for

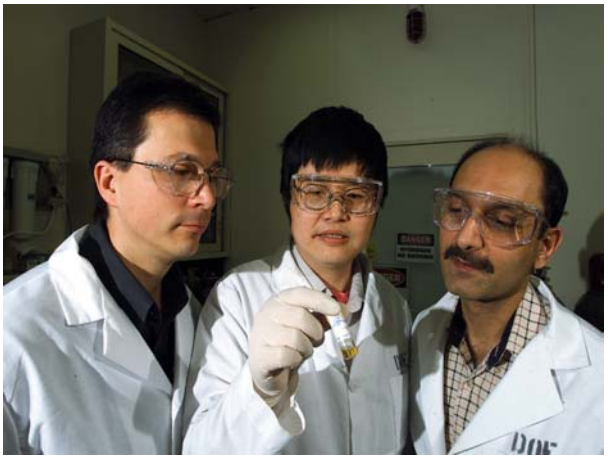


Fig. 1. Researchers examine a vial of ionic liquid. This ionic liquid was tested to determine its effectiveness in removing carbon dioxide from a mixture of gases.

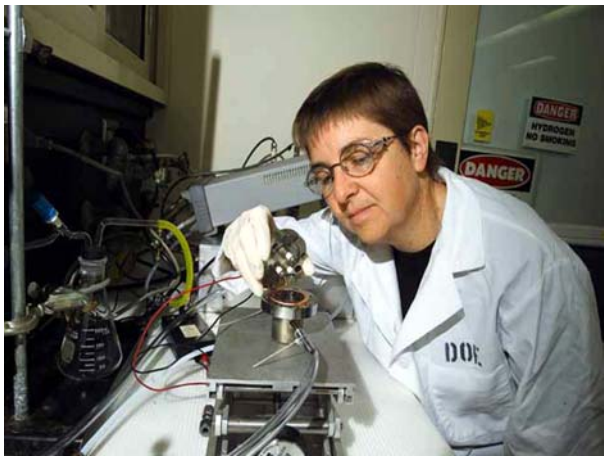


Fig. 2. A quartz-crystal microbalance test cell, which is used to screen different ionic liquids for their ability to absorb CO_2 from a gas mixture, is assembled. The ionic liquid is coated on a piezoelectric quartz crystal that is electronically excited into resonance. As the ionic-liquid-coated quartz is exposed to varying concentrations of CO_2, the resonance frequency changes because of the uptake and release of dissolved CO_2. These frequency changes are interpreted in terms of the effectiveness of the liquid in removing CO_2.

efficient separations, (2) predicting and verifying chemical and physical properties of ionic liquids, (3) development of ionic-liquid-based processes for removal of target compounds, and (4) preliminary technical and economic evaluation of novel processes.

Results and Accomplishments

Fourteen ionic liquids were synthesized, and their relative affinities for the target compounds (carbon dioxide and/or mercury) were assessed.

Carbon Dioxide Separation

The solubility of CO_2 in a series of imidazolium-based room-temperature ionic liquids was determined by a technique developed that employs a quartz crystal microbalance (QCM). Results were reported[4] in terms of Henry's Law constants, as shown in Table 1. A smaller value of Henry's Law constant indicates greater solubility of the gas in the liquid. For comparison, the Henry's Law constants for CO_2 are 1670 in water and 100 in toluene. The measured values of Henry's Law constant are in good agreement with published data for similar compounds that were obtained by other means,[5] supporting the viability of the QCM technique. A comparison of the results in Table 1 helps in understanding the role of chemical structure on the separation capabilities of these materials. Notable among the results is a significantly greater measured CO_2 solubility in an ionic liquid having a fluorine-substituted cation as compared to the corresponding ionic liquid with a non-fluorinated cation. The Henry's constant of 4.5 for that liquid (C_8F_{13}mimTf_2N) is several times lower than the smallest value reported in the literature for ionic liquids that do not undergo a chemical reaction with CO_2. This finding is important both for practical reasons in discovering better solvents for carbon dioxide and for fundamental science, since researchers at other leading institutions have reported that CO_2 solubility is primarily determined by the anion of the ionic liquid. The presence of water vapor was determined to have a minor effect on CO_2 solubility in hydrophobic ionic liquids, which is of practical importance given the water vapor present in flue gas.

Table 1. Henry's law constants for CO_2 in ionic liquids

Ionic liquid	H_{CO2} (bar)
C_3mim Tf_2N	37 ± 7
C_3mim PF_6	52 ± 5
C_4mim Tf_2N	37 ± 3
C_6mim Tf_2N	35 ± 5
C_8mim Tf_2N	30 ± 1
C_8mim Tf_2N with 20% relative humidity	30 ± 2
C_8mim Tf_2N with 40% relative humidity	27 ± 4
C_8F_{13}mim Tf_2N	4.5 ± 1
58 mol% C_8mim Tf_2N/42 mol% C_8F_{13}mim Tf_2N	15 ± 1
1,4-dibutyl-3-phenyl-imidazolium Tf_2N	63 ± 7
1-butyl-3-phenyl-imidazolium Tf_2N	180 ± 17

An irregular-ionic-lattice model (IILM) was developed to provide a means to predict the vapor pressures and solubility of carbon dioxide dissolved in ionic liquids throughout a range of pressures and

temperatures.There are only two parameters in this thermodynamic model and they were shown to be independent of an arbitrarily chosen reference state. Model predictions were compared with good agreement against published experimental data[5] for C_3mim PF_6 and C_8mim BF_4. The model may be useful in predicting CO_2 solubility at temperatures and pressures where experimental data is unknown. This is a significant finding; two papers on the subject were submitted to peer-reviewed journals.[6,7]

Laboratory-scale membrane tests systems were set up and operated to measure the flux and selectivity of gas transport through supported ionic-liquid membranes. The pores of anodic alumina membranes (Whatman, 47-mm diameter, 0.02-µm pore size and 60-µm thickness) were saturated with ionic liquids to create the membranes for the experiments. Transport rates of gases were measured for several ionic liquids using single- and mixed-gas feeds. Considerable selectivity of carbon dioxide over nitrogen transport through the membranes was measured, as shown in Table 2. The greater permeance of CO_2 through membranes containing C_4mimTf_2N compared to C_8F_{13}mimTf_2N—despite the significantly greater CO_2 solubility in the latter—can be explained by recognizing that the permeance of a supported liquid membrane to a particular species reflects both the solubility and the diffusivity of the species in the liquid. C_8F_{13}mimTf_2N was noticeably more viscous than C_4mimTf_2N. This higher viscosity is consistent with lower diffusivity, which apparently counters the favorable thermodynamics for this ionic liquid. These results point to the importance of addressing transport properties as well as thermodynamic equilibria in development of ionic liquids for separation applications.

Using the measured values of CO_2 solubility in ionic liquids and the measured membrane transport characteristics, a preliminary economic analysis of a separation process based on supported ionic liquid membranes was performed.[8] The cost of employing this membrane-based separation process for removal of carbon dioxide in the effluent of a coal-fired power plant was estimated and compared with a published analysis for a conventional amine scrubbing process.[9] The comparison indicates that, with continued technology development, an ionic liquid membrane process may potentially be economically competitive with the project cost of $33/mt CO_2 emission avoided for amine scrubbing.[9] In addition, an estimate of the cost of replacing the solvent in an amine scrubber with an ionic liquid also indicates that an ionic-liquid based process could potentially be competitive. This preliminary evaluation suggests that further investigation of ionic liquid processes for carbon dioxide capture is warranted.

Table 2. Measured CO_2 and N_2 permeance in supported ionic liquid membranes

Supported ionic liquid	Gas	Permeance (mol bar^{-1} cm^{-2} sec^{-1})	Selectivity
C_4mimTf_2N	CO_2	4.0×10^{-9}	127
	N_2	3.2×10^{-11}	
C_8F_{13}mimTf_2N	CO_2	1.5×10^{-9}	72
	N_2	2.1×10^{-11}	

Mercury Capture

Three ionic liquids were synthesized with the goal of imparting high affinity for mercury. These "task-specific" liquids were screened by conducting experiments in which gas streams containing trace levels of redox-generated metallic mercury vapors were passed by membranes coated with ionic liquid. A time-of-flight mass spectrometer was used to detect changes in the mercury concentration in the outlet stream from the membrane cell as mercury was removed by the ionic liquid. Two of the ionic liquids tested removed less than 30% of the mercury in the gas stream, while the third exhibited very high affinity for mercury. Greater than 99.9% mercury removal persisted for the duration of the tests, which lasted over 2 h.

Further experiments focused on the ionic liquid with high mercury affinity, and they were aimed at obtaining information needed for preliminary evaluation of the liquid for practical applications. The mercury capacity of the ionic liquid was measured to be 4.4 mg mercury per g of ionic liquid by analyzing the mercury content of a sample of ionic liquid that was equilibrated with mercury-saturated air at 26°C. Column testing was conducted to observe mercury removal performance in a flowing gas stream. Air containing a given concentration of elemental mercury vapor was passed through filters impregnated with a small amount ionic liquid, and samples of the inlet and outlet gases were analyzed. Excellent mercury removal performance was measured, as shown in the representative results of Fig. 3. In that test, a very small amount (12 mg) of ionic liquid was coated onto the surfaces of a porous filter (Millipore AP prefilter AP25-02500). Mercury-laden gas was passed through the filter at a relatively high flow rate (150 mL/min) that resulted in a contact time of the gas with the ionic liquid films of less than one second. Essentially complete mercury removal was achieved during the first 43 h of the experiment at an inlet concentration of 45 ng/L, and good removal persisted for many hours after the inlet concentration was increased by a factor of 23. The breakthrough from the column was consistent with the estimated mercury capacity of the ionic liquid. These positive results indicate rapid mass transfer performance of the ionic liquid and support its further investigation for gas-phase separations.

Further experiments were conducted to explore practical means for regeneration of the solvent for mercury

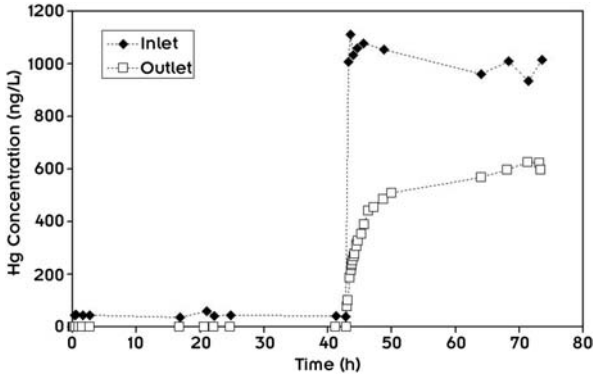


Fig. 3. Excellent results were obtained in column testing of mercury removal from a gas stream using an ionic liquid. Essentially complete mercury removal was achieved during the first 43 h of this test with 12 mg of ionic liquid, indicating significant mercury capacity and rapid mass transfer. Good removal persisted for many hours after the inlet concentration was increased by a factor of 23.

recovery and solvent reuse. Thermogravimetric analyses of both neat and mercury-loaded ionic liquid indicated that mercury was retained in the ionic liquid above the temperature at which the ionic liquid begins to thermally decompose. While those results indicate that thermal regeneration is not feasible, they underscore the strong binding of the mercury in the ionic liquid and indicate the capability for mercury capture at elevated temperatures.

Summary and Conclusions

This work has contributed to DOE's energy-security mission by providing positive information on the feasibility of ionic-liquid-based separations that could be used in future clean-energy processes. Ionic liquids were synthesized that were shown to have high affinity for carbon dioxide and mercury, target compounds for capture from exhaust gases from fossil-fuel combustion. Preliminary technical and economic evaluations indicate that further investigation of these ionic liquids is warranted. This project has also contributed to DOE's science mission by providing basic information on the chemical and physical properties of new ionic liquids and will help guide synthesis and application of these "solvents of the 21st century."

Products of this project included four research papers submitted to peer-reviewed journals, a master's thesis,[8] and three presentations at national technical meetings. An invention disclosure is in preparation on an ionic liquid for gas-phase mercury removal. In addition to ORNL staff from three divisions, the project supported efforts by a visiting professor on sabbatical (R. E. Baltus), a graduate student from the University of Tennessee (B. H. Culbertson), and an undergraduate summer student from Texas A&M University (C. Harris).

References:

[1]C. Krause, "Clean Coal Power Technologies," *Oak Ridge National Laboratory Review*, Vol. 35, No. 2, 2002. <http://www.ornl.gov/info/ornlreviewvol35/_2_02/clean_coal.shtml>.

[2]*Fact Sheet: EPA to Regulate Mercury and Other Air Toxics Emissions from Coal- and Oil-Fired Power Plants*, U.S. Environmental Protection Agency, Washington DC (2000). [Online]. Available: http://www.epa.gov/ttn/oarpg/t3/fact_sheets/fs_util.pdf.

[3]R. D. Rogers and K. R. Seddon, "Ionic liquids—Solvents of the future?" *Science* **302**, 792–793 (October 2003).

[4]R. E. Baltus, B. H. Culbertson, S. Dai, H. Luo, and D. W. DePaoli, "The low-pressure solubility of carbon dioxide in room-temperature ionic liquids measured with a quartz crystal microbalance," *J. Phys. Chem. B,* in press (2003).

[5]J. L. Anthony, E. J. Maginn, and J. F. Brennecke, "Solubilities and thermodynamic properties of gases in the ionic liquid 1-*n*-butyl-3-methylimidazolium hexafluorophosphate," *J. Phys. Chem. B* **2002**, 106, 7315.

[6]M. R. Ally, J. Braunstein, S. Dai, J. M. Simonson, and D. W. DePaoli, "Carbon dioxide gas solubility in ionic liquid 1-n-octyl-3-methylimidazolium hexafluorophosphate [C8mim][PF6] from the irregular ionic lattice model," submitted to *J. Chem. Eng. Data* (2003).

[7]M. R. Ally, J. Braunstein, R. E. Baltus, S. Dai, D. W. DePaoli, and J. M. Simonson, "Irregular ionic lattice model for gas solubilities in ionic liquids," submitted to *Ind. Eng. Chem. Res.* (2003).

[8]B. H. Culbertson, R. E. Baltus, H. Luo, S. Dai, and D. W. DePaoli, "Examination of the potential of room-temperature ionic liquids for gas separations," submitted to *Sep. Sci. Technol.* (2003).

[9]D. R. Simbeck. "CO_2 mitigation economics for existing coal-fired power plants," U.S. Department of Energy National Energy Technology Laboratory First National Conference on Carbon Sequestration, Washington, D.C., May 14–17, 2001.

[10]B. H. Culbertson "CO_2 Separation and Fuel Desulfurization Involving Room-Temperature Ionic Liquids," M.S. Thesis, University of Tennessee, Knoxville, Tennessee, May 2003.

Structure and Dynamics of Fluids in Confined Geometries

D. R. Cole,[1] J. M. Simonson,[1] M. Gruszkiewicz,[1] A. A. Chialvo,[1] G. D. Wignall,[2] Y. B. Melnichenko,[2] J. S. Lin,[2] G. W. Lynn,[2] B. Gu,[3] K. L. More,[4] T. D. Buchell,[5] P. T. Cummings,[5] Y. Leng,[5] K. E. Gubbins,[6] A. Striolo,[6] W. T. Cooper,[7] M. Schilling,[7] and A. Habenschuss[1]

[1] *Chemical Sciences Division*
[2] *Condensed Matter Sciences Division*
[3] *Environmental Sciences Division*
[4] *Metals and Ceramics Division*
[5] *Department of Chemical Engineering, Vanderbilt University, Nashville*
[6] *Department of Chemical Engineering, North Carolina State University, Raleigh*
[7] *Department of Chemistry and Biochemistry, Florida State University, Tallahassee*

Fluids including hydrocarbons, aqueous solutions, and gaseous species (e.g., CO_2, CH_4) can occupy the pores or fractures within solid matrices. The size, distribution, and connectivity of these confined geometries, the chemistry of the solid, and the chemistry of the fluids and their physical properties collectively dictate how fluids migrate into and through these micro- and nano-environments, wet and ultimately react with the solid surfaces. Our multidisciplinary approach combined scattering experiments, simulations, and thermodynamic measurements to quantitatively assess molecular properties of fluids confined to well-characterized porous media, subjected to a wide range of experimental conditions. Results were obtained in four distinct, interrelated areas. Scanning electron microscopy (SEM), transmission electron microscopy (TEM), X-ray diffraction (XRD), and small-angle X-ray and neutron scattering (SAXS, SANS) have been used to characterize a number of porous silicas, carbon fiber monoliths, zeolites, and clays, prior to interaction with fluids. Water adsorption/desoprtion isotherms have been determined on these materials from 105 to 200°C. Fourier transform infrared (FTIR), nuclear magnetic resonance (NMR), quasi-elastic neutron scattering (QENS), and SANS have been used to investigate structural and dynamic features of fluid-matrix interactions. Molecular dynamics (MD) simulations have focused on the behavior of water in carbon slit pores, carbon nanotubes, and mica. These studies are providing an understanding at the molecular level of how intrinsically different fluids behave in confined geometries, compared to bulk systems.

Introduction

The porosity (void) volumes within the solids span wide length scales including micro-, meso-, and macroporous regimes (d/Å < 20, 20 < d/Å < 500, and d/Å > 500, respectively, as defined by IUPAC). A number of factors dictate how fluids, and with them reactants and products of intra-pore transformations, migrate into and through these nano-environments, wet and ultimately adsorb and react with the solid surfaces. These include the size, shape, distribution, and interconnectivity of these confined geometries, the chemistry of the solid, and the chemistry of the fluids and their physical properties. The dynamical behavior of fluids and gases contained within solids is controlled by processes occurring at the interface between the various phases (e.g., water-water, water-solute, water-solid, solute-solid, etc.), as well as the rates of supply and removal of mobile constituents. The richness and complexity of fluid behavior (e.g., phase transitions, molecular orientation and relaxation, diffusion, adsorption, wetting, capillary condensation, etc.) in confined geometries only underscores the need to adopt a multidisciplinary approach when trying to quantify this behavior regardless of the fluid type or nature of the porous medium. If properly calibrated and scaled, an atomistic or molecular understanding of fluid-solid interaction may provide quantitative insight into the behavior of systems at the macroscopic level. The overall objective of our work was to show that the fundamental molecular properties and behavior of fluids confined to restricted geometries can be understood through the combination of detailed experimental studies with theory and molecular-level simulation, and that this understanding can open new avenues for future improvements in processes and systems in natural and technological environments relevant to DOE.

Technical Approach

In order to assess key features of the fluid-matrix interaction under confinement at the nanoscale, a multidisciplinary approach was taken that addressed the

following areas: (a) characterization of the physical nature (e.g., size, shape, fractal geometry, etc.) of the pores using various types of microscopy (SEM; TEM), scattering (SAXS; SANS), XRD, and gas sorption methods; (b) determination of the thermodynamic properties of water in confined geometries using the isopiestic method over a wide ranges of temperature and pressure; (c) determination of the local short-duration dynamics and longer-scale hindered motion of water, CO_2 and hydrocarbons through use of complementary techniques including neutron scattering, NMR, and FTIR spectroscopy, as a function of pressure and temperature; (d) application, refinement, and, where necessary, development of molecular-based simulations methods that describe the properties and structure of confined fluids. Selected aspects of these activities are highlighted below.

Considerable effort was expended to obtain a set of porous materials exhibiting varying porosity, pore sizes, shapes, distribution, and chemistries. Because many of these materials were used in more than one kind of study, multiple-gram quantities of each material were obtained so they could be shared among the investigators. Table 1 summaries the porous solids used in experiments and/or simulations, and some of their characteristics. The solids we selected for study had many of the following characteristics: (a) sparingly soluble at the pressure-temperature-fluid composition conditions of the experiments; (b) suitable for detailed characterization of pore features such as size, shape, distribution, roughness, etc.; (c) susceptible to chemical manipulation from hydrophobic to hydrophilic by chemical pretreatments; and (d) limited crystallinity to minimize the impact on small-angle scattering. In addition to the more conventional methods typically used to characterize porous matrices (SEM, TEM), we used a number of scattering and spectroscopic techniques. Many of these materials were either machined or synthesized into specific shapes and sizes in order to accommodate the various sample or environmental chambers used in our experiments. Vycor 7930 pore glass, machined into various shapes (e.g., elongate cylinders, plates) and polished, was cleaned initially by boiling for several hours in H_2O_2. Silica aerogels (densities of 0.01, 0.1 and 0.2 g/cm^3) were synthesized from different mixtures of TMOS and methanol, and supercritically dried. All silica materials were heat treated to drive off unwanted organics (e.g., SiO-CH_3 groups) by slowly heating over a several-hour period to 600°C in air, followed by slow cooling back to room temperature. Carbon fiber monoliths with varying porosities and surface areas were synthesized here at ORNL. Natural Ca-zeolite was hydrothermally pretreated and converted to Na or Na-K zeolites. The pillared aluminous-rich clay (montmorillonite) was obtained from Fluka (Sigma Aldrich; Lot 381121/1 13301). Prior to any characterization or experimentation, these materials were preheated to 100–110°C under vacuum to drive off adsorbed water.

Because of its ubiquitous presence in both natural and industrial settings, water in its bulk liquid or solid forms has been the subject of numerous theoretical and experimental investigations. As such, we elected to focus much of our effort in this project on the behavior of water in confined geometries. In addition to water, we also explored the influence of confinement in porous silica glass on the behavior of aqueous electrolytes (LiCl, $CaCl_2$, $NdCl_3$). A select number of non-aqueous systems (e.g., chloroform, aniline, and CO_2 were investigated using NMR, FTIR, and SANS, respectively).

Table 1. Porous solids used in experiments and/or modeling

Solid	Surface	Pore shapes	Pore widths (nm)
Porous glass	O, Si	Cylinder	2–8
Silica sol-gels	O,Si	Cylinder	2.5–20
Silica aerogel	O, Si	Irreg. cyl.	6–7
Activated C fibers	C (O, H)	Varied	0.4–1.0
C nanotubes[a]	C (O, H)	Cylinder	1.0–1.6
Na-K Zeolites	K, Na, Si, Al, O	Cylinder, cage	0.3–0.8
Montmorillonite	Na, Si, Al, O	Slit, pillars	0.5–0.8
Muscovite (2M1)[a]	K, Si, Al, O	Slit	0.6–1.0

[a]Solids investigated in modeling studies only.

Results and Accomplishments

Physical Characterization of solids is a crucial step that sets the stage for how variations in pore size, roughness, distribution, connectivity, composition (hydrophobic versus hydrophilic), surface charge, and structure (i.e., crystalline versus amorphous) influence molecular behavior of confined fluids. In addition to the more conventional methods typically used to characterize porous matrices (gas sorption isotherms, SEM, TEM), we used a number of scattering and spectroscopic techniques. While not yielding direct imaging information, scattering and diffraction experiments can provide global statistical information about interfaces (e.g., total porosity, pore surface area, pore length scale, skeletal density, fractality) over an enormous range of length scales (Å to mm), represented by reciprocal or Q-space (where $Q = 4\pi\lambda^{-1}\sin\theta$ is the momentum transfer, 2θ is the angle of scattering, and λ is the wavelength). The scattering intensity, $I(Q)$, is measured as a function of momentum transfer, and exploits the difference in scattering power (contrast) between the solid matrix, and air or fluid occupying the pore space. An example of results from small-angle neutron scattering (SANS) is shown in Fig. 1 for three different porous silica sol-gel glasses. This plot reveals that the correlation peaks (vertical arrows) shift to successively longer distances, which is

indicative of variations in the dehydration process designed to produce the different pore sizes. The glass with 75-Å pores yields the minimal excess scattering in the small-angle region, which makes it better suited for studying the influence of confinement on liquid-gas critical behavior. For the most part, the disordered porous-silica glasses have rather smooth pores that are uniformly distributed throughout the solid. The porous carbons have more wide-ranging pore sizes and exhibit varying degrees of crystallinity, depending on the temperature of burn-off and the duration used to activate the solid.

The Sorption Capacity of a porous or fractured solid matrix depends on the total exposed solid surface area, the shape and size distribution of voids in the matrix, and the nature of the interaction between fluid molecules and the solid surface, and between the molecules themselves. The competition between liquid-solid and liquid-liquid interactions leads to monolayer and cluster formation at low pressure and capillary condensation at high pressures. In order to understand these effects in a fluid with strong intermolecular interactions, adsorption/desorption isotherms of water were measured for 15 solid samples (e.g., porous silicas, carbon fiber monoliths, zeolites, pillared clay) at 105, 150, and 200°C from vacuum to saturated vapor pressure in an isopiestic apparatus. This apparatus is unique in its capability for precise weighing of samples at elevated temperature and pressure between vacuum and 4 MPa. This gravimetric method, in contrast to the more common volumetric techniques, allows for continuous monitoring of the changes of the adsorbent itself due to degassing, oxidation, or decomposition.

Figure 2 shows representative water sorption results at 105°C from vacuum to saturated vapor pressure (0.121 MPa) for several of the porous solids. At low pressure the isotherms reflect competition between liquid-solid and liquid-liquid interactions (forming the monolayer and subsequent clusters or layers; pore filling). The behavior at higher pressures reflects the impact pore size and structure have on capillary condensation and possible hysteresis. Both zeolites exhibit very steep isotherms at low pressure followed by relatively flat trends at moderate to high pressures. This behavior is indicative of strong water-solid intermolecular interaction dominated by mono-layering. The porous carbon and silica glass have rather flat slopes at low pressure followed by steep trajectories indicating weak water-solid interaction and capillary-like pore filling. The montmorillonite exhibits a steady increase in water uptake, suggestive of multi-layering characteristic of dipolar electrostatic interaction. In general, an increase in temperature produces a narrowing of the hysteresis loops, a decrease in the magnitude of water uptake, but essentially unaltered shapes of the sorption curves. One exception was the behavior of the clay, wherein the isotherms exhibited a different shape at each temperature due to a repeated sequence of desorption and adsorption of relatively weakly attached H_2O ligands and OH^- groups followed by physical adsorption. Results such as these can form the basis of testing various MD simulation schemes.

Vibrational Behavior and Molecular Motion of water and nonaqueous solutions (e.g., aniline, chloroform) in confined geometries have been assessed with FTIR and NMR spectroscopy, respectively. The FTIR data indicate that once water molecules are sorbed or "confined" within the pillared Al-montmorillonite, their structural arrangements became more ordered as demonstrated by the shifts in the symmetrical and asymmetrical stretching vibration and deformation of the water molecules. Free water molecules (for fully hydrated Al-clay) exhibited O-H stretching-deformation at ~1652 cm^{-1}, whereas water molecules in confined micropores showed crystalline-like structure with O-H stretching deformation shifted to lower wave number at ~1630 cm^{-1}. This blue shift in wave number (~22 cm^{-1}) is indicative of a relatively high energy required for the O-H stretch deformation because of confinement of water molecules.

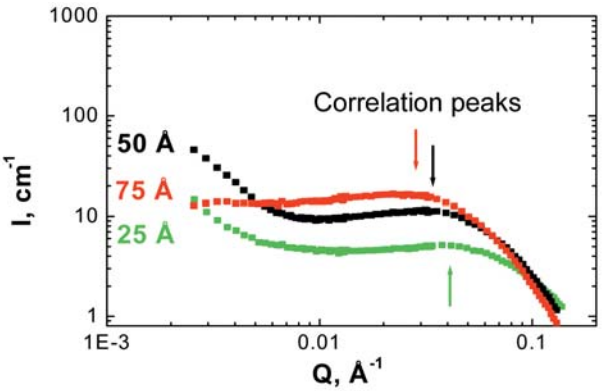


Fig. 1. SANS results from various silica sol-gel glasses having different average pore sizes.

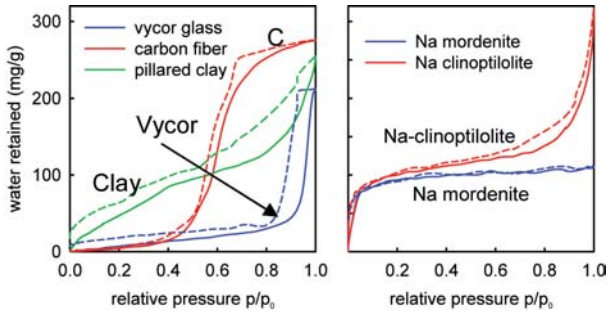


Fig. 2. Adsorption (solid) and desorption (dashed) isotherms of water measured for a variety of porous solids at 105°C.

Relaxation Parameters derived from NMR measurements were used to characterize the tertiary structure of soil organic matter (SOM) and the molecular motion of water (and chloroform) confined to pores in soil and Vycor silica glass, respectively. In the case of water, NMR data indicate that at pore filling levels below ~30%, water is highly mobile, probably existing as surface-bound vapor and not as a water cluster with liquid-like properties. Beyond 30% filling, water behaves an isotropic liquid and exhibits diffusion coefficients characteristic of bulk water. Chloroform, however, appears to behave differently, with restricted motion (i.e., decreasing diffusion coefficients) extending well beyond the 30% pore volume filling exhibited by water.

Phase Transitions can be significantly modified by confinement and disorder. The liquid-vapor behavior in porous systems with quench disorder, such as aerogels, is of particular interest because it may provide an experimental test of which theoretical model best describes critical phenomena in confined fluids. Currently, there are two competing theoretical models—the Random Field Ising model and the "single-pore" model. The experimental efforts to verify the validity of the models for confined binary liquid solutions have been obscured by experimental complications caused by the local composition gradients in pores as well as sluggish kinetics in the liquid-liquid region. SANS is particularly well suited to study the critical phenomena in small pores due to a short wavelength of neutron radiation (~ several Å), which allows for extracting information on the correlation length of the order of a nanometer and higher. In our experiments, silica aerogel (~96% porosity) was filled with CO_2 at different temperatures and pressures appropriate for bracketing the critical region. SANS measurements (Fig. 3) of the correlation length of the density fluctuations ξ indicate that quench disorder works to suppress density fluctuations in the critical region. The correlation length does not exceed the characteristic pore size which provides strong evidence in favor of the single-pore model.

Molecular-Based Simulation Techniques have been used to assess the behavior of water and water + CO_2 mixtures confined in a number of different nanoporous media (e.g., slit carbon, carbon nanotubes and mica). Experimental information determined from characterization of the porous media, adsorption-desorption isotherms, and scattering and spectroscopic assessments of fluids at different length and time scales provide crucial synergistic input to these simulations. The grand canonical Monte Carlo method (GCMC) was used to simulate water (SPC/E) in carbon-slit pores with widths of 0.6, 0.8, 1.0 and 1.6 nm and in (12:12; 20:20) single-walled carbon nanotubes with diameters of 1.36, 1.63, and 2.7 nm in the temperature range 298–600 K. A comparison of experimental sorption results (Fig. 2) and the

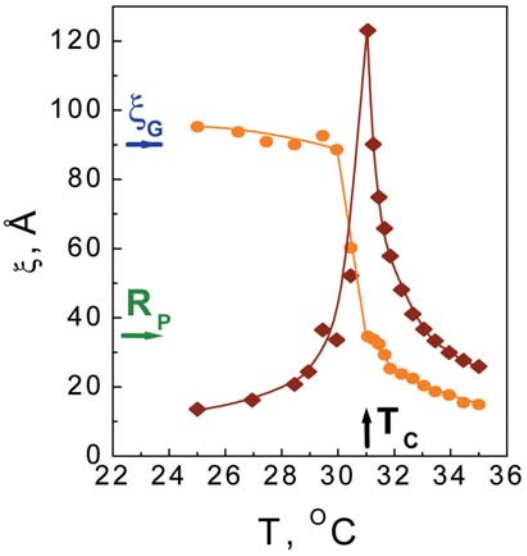


Fig. 3. Correlation length of density fluctuations of CO_2 for bulk (♦) and confinement (●). R_p = pore radius, ξ_G = aerogel correlation length, T_c = critical temperature.

simulations (Fig. 4) indicates that because of a competition between strong water-water and weak water-carbon interactions, adsorption isotherms are characterized by negligible uptake at low pressures, by pore filling following a capillary-like condensation mechanism, and by adsorption/desorption hysteresis loops. As temperature increases, the relative pressure at which pore filling occurs increases and the size of the hysteresis loops decreases. Differences between simulated and experimental results are primarily due to heterogeneous chemical composition (O, H in C fibers), pore connectivity, and non-uniform pore-size distribution, which are not accounted for in the simulation models.

Molecular dynamics (MD) simulations of water (TIP-4P) confined between two muscovite mica (2M1) surfaces focused on the hydration player structure and friction dynamics (Fig. 5). We investigated the hydration layer structure at the mica (001)-water interface. Under ambient conditions, for the "thick" water film (>1.0nm), we determined that the density oscillation of water oxygen as a function of distance *z* from the mean surface oxygen position of mica is only within 6 Å, roughly two water layers. When the confinement is below 1.0 nm, we also observed the adsorbed and hydration peaks of water molecules adjacent to the mica surface, similar to X-ray reflectivity experiments. Shearing of water films in MD regime showed that the hydrated potassium ions remains attached to the mica surfaces, and there were no significant stick-slips. The shear viscosity is found to be only 2–4 times that of the bulk values. This phenomenon suggests that the bound water molecules retain shear fluidity properties even at extreme confinement

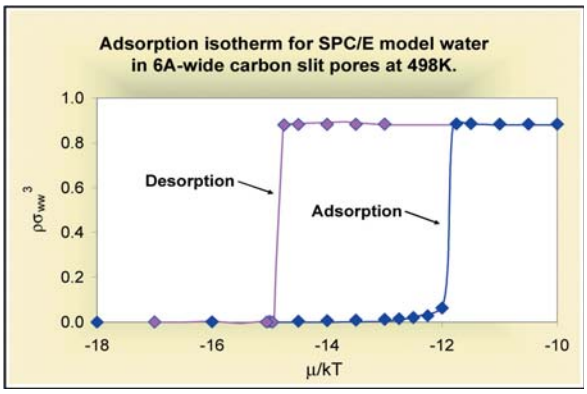


Fig. 4. GCMC results for water in slit-carbon pores. $\rho\sigma^3_{ww}$ is the reduced water density; μ/kT is the chemical potential $\cong p/p_o$.

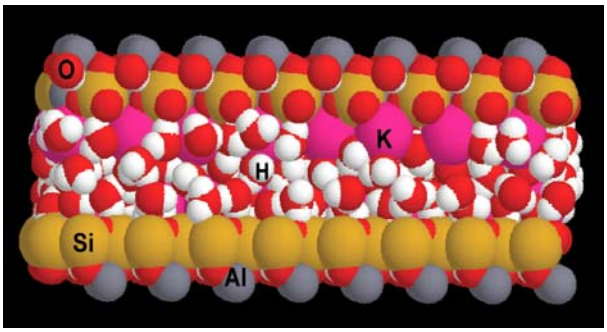


Fig. 5. 384 water molecules confined between two mica surfaces with a 0.96-nm separation.

Summary and Conclusions

A multidisciplinary, multi-institutional effort has addressed a number of key issues relevant to fluid-matrix interactions under confinement. Contributions have been made in the characterization of numerous complex nanoporous solid (e.g., porous silica, activated carbon fibers, zeolites) using a variety of techniques including TEM and small-angle X-ray and neutron scattering. Advanced sorption, scattering, and spectroscopic methods have documented fluid-fluid and fluid-matrix interactions of water, organic solvents such as chloroform, and CO_2 in confined geometries. MD simulations have provided the theoretical underpinning of the experimental results, particularly the water sorption isotherms.

Our new results on the structure and dynamics of fluids confined in porous materials at the nanoscale have broad relevance to DOE missions in Materials Sciences and Chemical Sciences, Geosciences and Biosciences (CSGB) within the Office of Basic Energy Sciences. An understanding of fluid behavior under confinement has broad application to problems related to catalysis, chemical separations, materials synthesis and durability, and environmental remediation. The effort also contributes to an understanding of fluid-solid interactions and corrosion relevant to the disposal of both civilian and defense-derived high level radioactive waste.

This project has already benefited the laboratory in several ways. Two BES renewal proposals to CSGB have been funded that redirected significant portions of their efforts to addressing issues of fluid confinement and associated reactivity in both experimental and natural mineralogical systems. Results obtained in this project were pivotal in helping establish the credibility needed to successfully defend the new directions taken in these proposals. Contributions realized from this project were instrumental in forming a vital part of the foundation upon which we developed a Research Focus Area (RFA) titled "Nanoconfined and Nanostructured Fluids" for the Center for Nanophase Materials Science (CNMS). Finally, results from this project along with the team we have assembled provide the necessary ingredients that will allow us future proposal opportunities in areas such as catalysis and hydrogen storage.

Publications Derived from This Project

D. R. Cole et al., "Influence of nanoscale porosity fluids behavior," *Proceedings of the 11th International Symposium on Water-Rock Interaction* (to be published).

Y. B. Melnichenko et al., "Density fluctuations near the liquid-gas critical point of a confined fluid," *Phys. Rev. Lett.* (submitted).

Y. B. Melnichenko et al., "Liquid-gas critical phenomena under confinement: small-angle neutron scattering studies of CO_2 in aerogel," *J. Molec. Liquids* (to be published).

J. L. Rivera et al., "Layering behavior and axial phase equilibria of pure water and water + carbon dioxide inside single carbon nanotubes," *Nano Letters* **2**, 1427 (2002).

M. Schilling and W. T. Cooper, "Effects of chemical treatments on the quality and quantitative reliability of solid state ^{13}C NMR spectroscopy of mineral soils," *Analytica Chimica Acta* (to be published).

M. Schilling and W. T. Cooper, "Identification of copper binding sites in soil organic matter through chemical modification and ^{13}C CP-MAS NMR spectroscopy," *Org. Geochem.* (submitted).

A. Striolo et al., "Water adsorption in carbon-slit nanopores," *Langmuir* **19**, 8583 (2003).

A. Striolo et al., "Simulated water adsorption isotherms in carbon nanopores," *Molec. Phys.* (submitted).

A. Striolo et al., "Temperature effect on water adsorption in porous carbons," *Fluid Phase Equilibria* (submitted).

Self-Organizing Polymers as Biomaterials

P. F. Britt,[1] J. W. Mays,[1,2] J. Pickel,[1] I. D. Chung,[2] and M. Liu[2]
[1]*Chemical Sciences Division*
[2]*Department of Chemistry, University of Tennessee, Knoxville*

This proposal seeks to create a new capability at ORNL in the design, synthesis, and characterization of polymeric materials, especially biomolecular materials, with novel properties. Because of the opportunities that exist for research into novel materials that mimic the properties and function of biological materials, this research project is focused on the synthesis, self-assembly, and characterization of amphiphilic block copolymers as biomaterials. Specifically, block copolymers that form micelles and vesicles are of interest because of their ability to compartmentalize additives into different domains and to be environmentally responsive. In this project, three different synthetic targets (block copolymers) were pursued utilizing living anionic and atom transfer radical polymerization techniques.

One of the greatest technical challenges facing chemistry, physics, and materials science today is the controlled synthesis of mesoscopic materials with well-defined properties, and assembly of these materials into macroscopic functional devices. One approach to address this grand challenge involves the synthesis and self-assembly of block copolymers to generate complex architectures with tailored mechanical, optical, electrical, or other physical properties. Block copolymers can be prepared by controlled/living polymerization methods with a high degree of control over the molecular architecture, which in turn controls molecular properties and function. This proposal seeks to create a new capability at ORNL in the design, synthesis, and characterization of polymeric materials, especially biomolecular materials, with novel properties. Because of the opportunities that exist for research into materials that mimic the properties and function of biological materials, this research project is focused on the synthesis, self-assembly, and characterization of amphiphilic block copolymers as biomaterials. Specifically, block copolymers that form micelles and vesicles are particularly interesting because of their ability to compartmentalize additives into different domains and to be environmentally responsive.

Three different synthetic targets were pursued in parallel using living anionic and controlled radical polymerization techniques. First, living anionic polymerization techniques were used to make linear block copolymers of styrene and ethylene oxide (PS-PEO) and mixed-arm star polymers of PS-$(PEO)_2$ to investigate the impact of molecular architecture (i.e., branching) on the self-assembly and biocompatibility of these materials. Attempts to prepare PS-$(PEO)_2$ by reacting living PS with excess CH_3SiCl_3, followed by removal of excess chlorosilane, and addition of living PEO were unsuccessful. Thus, a new synthetic approach was designed in which the hydroxyethyl terminated PS was coupled to 2,2-bis(phenyldioxymethyl) propionic acid. Deprotection by hydrogenolysis afforded the bis(hydroxymethyl) groups, which can serve as the initiation points for the polymerization of ethylene oxide to make PS-$(PEO)_2$ and potentially PS-$(PEO)_4$.

Second, block copolymers of substituted polystyrenes and acrylic acid were prepared by atom transfer radical polymerization methods (ATRP) to determine how the structure of the core influences the formation and stability of the polymer aggregates in solution, and how this impacts the encapsulation and release of hydrophobic substrates. Homopolymers of styrene (PS) and *t*-butyl acrylate (PtBA) were initially prepared by ATPR with narrow molecular weight distributions (M_w/M_n ca. 1.10) and controlled molecular weights (M_n = 5–10K). These macroinitiators were used to prepare block copolymers of PS-PtBA and polystyrene-polyacrylic acid (PS-PAA) after hydrolysis of the *t*-butyl group. In aqueous solution, PS-PAA formed spherical micelles and rods. The ATRP of substituted styrenes, including 4-*n*-butylstyrene and 4-*t*-butylstyrene, with *t*-butyl acrylate will be investigated next to determine the role of substituents in the aggregation of the polymers in aqueous solution (after hydrolysis of the *t*-butyl group) and the impact of the glass transition temperature of the micelle core on the release of hydrophobic substrates.

Finally, acryloyl amino acid derivatives were polymerized by ATRP in water to determine if the controllable properties of synthetic polymers could be combined with the exceptional functionality, activity, and

specificity of biological macromolecules. Monomers were synthesized by the reaction of β-alanine or phenylalanine with acryloyl chloride. Polymerization of acryloyl β-alanine (ABA) in water was surprising fast at room temperature, and polymers with M_n = 5, 8, and 12K were prepared in 70–87% yield. To provide better control of the block length of the poly(amino acids) and improve the copolymerization with other monomers, the impact of catalyst, inhibitors, initiators, solvent, temperature, and acid protecting groups are under investigation. Once polymerization conditions have been optimized, block copolymers will be prepared with poly(*N*-isopropylacrylamide), and the thermoresponsive behavior of the copolymers in aqueous solution will be investigated.

Advanced Ion Trap Mass Spectrometry for the Rapid and Confident Identification of Biological Agents

M. B. Wise,[1] K. J. Hart,[1] D. E. Goeringer,[1] L. J. Hauser,[2] I. F. Robbins,[3] and D. A. Clayton[4]
[1]*Chemical Sciences Division*
[2]*Life Sciences Division*
[3]*Computational Sciences and Engineering Division*
[4]*Engineering Science and Technology Division*

This proposal responds to the need for the rapid detection and positive identification of biological threat agents, including bacteria, toxins, and viruses, in complex environments. Because the protein component will most likely be differentiating for all types of biological agents, computational methods will identify and assess the efficacy of using biomarker proteins that are characteristic of and exclusive to the targeted biological agents. Employing an array of modern mass spectrometric techniques for their determination, the amino acid sequence(s) and molecular weight(s) then will be evaluated as an interrogation probe for the biomarker(s). Based on our findings, we will develop a concept for a biological agent mass spectrometer, produce a laboratory prototype instrument, and demonstrate its capabilities using a limited number of gamma-killed biological agents, toxins, and simulants.

The goals of this project are to develop novel mass spectrometry–based methodologies and portable instrumentation for rapidly detecting and confidently identifying low levels of biological agents including bacteria, toxins, and viruses.

The combination of analytical speed, sensitivity, and versatility offered by mass spectrometry (MS), especially when electrodynamic analyzers are employed, is virtually unmatched by any other single analytical technique. Consequently, a multi-step approach centered on MS is being pursued in developing an advanced strategy for the detection of biological agents. Computational simulations are being used to determine the applicability of protein biomarker analysis, within the context of MS, for differentiation and identification of specific biological threat agents. Experimentally, the actual detection and identification of a target organism within a complex mixture of other organisms is being assessed with simulations based on MS. In the instrumentation development area, our strategy is to incorporate proven MS technologies for biological agent detection while simultaneously considering novel scientific and engineering approaches. Thus, the proposed measurement approach is based on application of electrospray (ES) ionization for ion production, ion/ion reaction and dissociation of the mass-selected biomarker ions, and mass-to-charge analysis of the diagnostic fragment ions using electrodynamic devices.

Computational simulations compared a subset of 376 abundant genes from a target organism (*E. coli*) against a complex background mixture database containing ~80,000 unique entries from twelve microbial organisms. The purpose was to assess both the tryptic peptide level (bottom-up) and the whole protein level (top-down) ES-MS approaches for protein analysis as well as to determine the mass accuracy needed to make confident identifications. Results indicated that it is likely possible to identify a target organism within a complex background mixture using protein biomarkers while avoiding false positive identification. Furthermore, the simulations indicated that the top-down approach would be more viable for biothreat detection and identification as compared with the bottom-up method due to reduction in complexity of the database analysis, provided that ten fragments or more are produced per protein and that the mass measurement accuracy is at least ±10 Da.

Experimental simulations of the ES-MS detection and identification of a target organism (*E. coli*) within a complex mixture of four other organisms assessed the bottom-up technique under two scenarios: a "battlefield" scenario, in which threats are present at high concentration and rapid detection and identification is essential, and a "homeland security" scenario, in which threats are present at a considerably lower level and detection and identification is not as time sensitive. In the battlefield scenario, each simulation reported definitive detection of the target by returning multiple (>20) target peptides. The minimum ES-MS analysis time tested, 85 minutes, resulted in the identification of 21 target peptides, with only 7 false positive peptides. In the homeland security scenario, the target organism was present at varying dilution levels with respect to background: 1:1, 1:4, 1:40, 1:250, and 1:400.

The presence of the target was detected with relative certainty at the 1:1 (average of 267.3 peptide matches versus 7 false positive peptide identifications) and 1:4 (average of 83.3 peptide matches compared with 5 false positive peptide identifications) levels. For experimental simulations performed below these levels of the target organism, it could not be definitively stated that the target organism was present.

We have also completed the conceptual design of a mass spectrometer and associated operational method that promises to be particularly well suited for rapid and sensitive identification and analysis of protein biomarkers. The concept consists of a tandem configuration of a three-dimensional RF quadrupole ion trap, a linear RF quadrupole, and a time-of-flight mass spectrometer, along with the associated method of operation whereby the combination generates a complete three-dimensional mass spectrum of parent and product ions on a time scale that is significantly faster than current commercial mass spectrometers. Fabrication and evaluation of such a prototype instrument is currently in progress.

As indicated in the original proposal, we anticipate that this project will result in instrumentation and techniques that will give ORNL a technological advantage in the area of biological agent detection. More specifically, the MS-based detection and identification system will surpass the capabilities of other analytical technologies such as polymerase chain reaction and immunoassay with regard to the range of biological agents that may be detected. We also believe that by being able to detect changes in the basic structure of key proteins, this approach may be less vulnerable to being thwarted by genetic engineering of the biological agents. The DoD, DHS and other agencies should be very interested in this technology as the basis for their next-generation biodetectors/identifiers.

CHEMICAL SCIENCES AND TECHNOLOGY

Seed Money Fund

Carbonation of Serpentine for Long-Term CO_2 Sequestration

J. G. Blencoe,[1] L. M. Anovitz,[2] J. S. Beard,[3] and D. A. Palmer[1]
[1]*Chemical Sciences Division*
[2]*University of Tennessee, Knoxville*
[3]*Virginia Museum of Natural History*

Results of autoclave experiments show that serpentine [$Mg_3Si_2O_5(OH)_4$] can be quantitatively converted to magnesite ($MgCO_3$) + silica (SiO_2) + water in a multistep chemical process that is much more energy efficient than the "direct" method for achieving that conversion. In the new process, serpentine is reacted in a concentrated solution of caustic soda (NaOH), producing brucite [$Mg(OH)_2$] + solubilized silica. The brucite- and silica-rich liquid are then physically separated to allow (1) reaction of the liquid with CO_2 to form bicarbonate ion + silica gel + water and (2) reaction of the bicarbonate ion with the brucite formed previously to produce magnesite + aqueous NaOH + water. The number of moles of NaOH produced during magnesite crystallization is exactly equal to the number of moles of that base consumed during conversion of serpentine to brucite; therefore, there is no net consumption of NaOH in the process. In addition, because the process produces separate streams of high-purity silica and magnesite, both of these solids become valuable commercial by-products. Fully optimized for a reactor-based mineral-carbonation technology implemented at large point sources of CO_2, the new serpentine-carbonation process provides a safe and effective means for permanently sequestering huge masses of CO_2.

Introduction

Mineral carbonation is a promising technology for reducing CO_2 emissions from fossil fuel-fired power plants, cement factories, and steel mills.[1] It involves reacting CO_2 with non-carbonate (predominantly silicate) minerals to form one or more thermodynamically stable and environmentally benign carbonate compounds. The process mimics natural weathering of silicate minerals to form carbonate minerals such as calcite ($CaCO_3$), dolomite [$CaMg(CO_3)_2$], magnesite ($MgCO_3$), and siderite ($FeCO_3$). However, natural mineral carbonation reactions are very sluggish at ambient temperature and pressure; therefore, one of the principal technical challenges for CO_2 sequestration by mineral carbonation is to develop one or more commercially viable chemical processes that significantly accelerate the rates at which silicate minerals are transformed to stable, crystalline carbonates.[2] Physicochemical factors that affect mineral carbonation reaction rates (kinetics) include process temperature, pressure, CO_2 fugacity, and redox state; the mass proportions of the reactants, "additives," and catalysts; pretreatment and size reduction of the mineral feedstock; and the overall solution chemistry of the liquid phase. Several potential silicate feedstocks for industrial-scale mineral carbonation have been proposed and evaluated, with greatest attention being given to olivine (Mg_2SiO_4) and serpentine [$Mg_3Si_2O_5(OH)_4$].[3–5] The carbonate produced by olivine/serpentine carbonation, magnesite, is highly stable and environmentally neutral in nearly all terrestrial environments.

An early olivine/serpentine carbonation scheme involved production of magnesite by (1) acid (HCl) decomposition of olivine ± serpentine, producing brucite [$Mg(OH)_2$] + silica (SiO_2) and (2) reaction of the $Mg(OH)_2$ with CO_2 to form magnesite.[1,3] However, a detailed feasibility study of this approach has shown that it is too energy intensive for industrial-scale implementation.[2] Therefore, over the past five years, attention has been focused on CO_2 sequestration by "direct" carbonation of olivine and serpentine.[6,7] In this method, CO_2 is sequestered in magnesite with no acid pretreatment of the silicate feedstock, and the primary reactants are olivine and/or serpentine, and supercritical CO_2. For olivine, the overall reaction is

$$Mg_2SiO_4 + 2CO_2 \rightarrow 2MgCO_3 + SiO_2 \ ; \qquad (1)$$

the corresponding reaction for serpentine is

$$Mg_3Si_2O_5(OH)_4 + 3CO_2 \rightarrow 3MgCO_3 + 2SiO_2 + 2H_2O \ . \qquad (2)$$

Recent experiments performed to determine the kinetics of these reactions indicate that, at a given elevated temperature and pressure—and using essentially identical kinds and amounts of "additives" and catalysts ($NaHCO_3$ and NaCl)—olivine reacts with CO_2 much more readily than serpentine.[6] This is unfortunate because, while

significant masses of olivine-bearing rocks exist for sequestering CO_2, naturally occurring serpentine is approximately an order of magnitude more abundant.[4,5] To date, the only remedy for sluggish serpentine reaction is to heat it to 600–650°C prior to carbonation, which drives off structurally bound water (hydroxyl groups).[6] Tests of this altered (dehydroxylated) serpentine have shown that it is much more reactive than untreated (hydroxylated) serpentine. However, at a typical fossil fuel–fired power plant, heat treating serpentine at 600–650°C prior to carbonation would require ~200 kW·h of electricity per ton of serpentine feedstock.[6] With 1 ton of carbon in a fossil fuel producing ~3.7 tons of CO_2, and each ton of CO_2 consuming ~2 tons of serpentine during carbonation, the power requirements for serpentine dehydroxylation represent 20–30% of total power output. This large energy penalty threatens the economic viability of CO_2 sequestration by serpentine carbonation.

Technical Approach

In this study, olivine/serpentine carbonation was investigated at elevated temperatures and pressures using conventional methods of hydrothermal experimentation. Solid, liquid, and gaseous reactants were the following: finely ground natural serpentine (antigorite variety from the Cedar Hill Quarry, Lancaster County, PA) and olivine (from Twin Sisters Peak, Washington), fine-grained brucite formed by caustic decomposition of the natural serpentine, reagent-grade synthetic sodium silicate (Na_2SiO_3), a 50 wt % solution of caustic soda, and research-grade (99.999% pure) CO_2 dispensed from a standard gas cylinder fitted with a dip-tube. These starting materials were loaded into a 1.75-inch O.D., 1.40-inch I.D., 6.28-inch long, closed-one-end Teflon liner; inserted into a 300-mL, bolted-closure autoclave (Fig. 1); heated to experimental temperature in an AC-powered resistance furnace; and allowed to react for periods ranging from 2 h to 3 days. Experimental temperatures were 22 and 200°C, and experimental pressures ranged from ~10 to 64 atm. Experiments were concluded by switching off furnace power, allowing the autoclave to cool to a temperature below ~40°C, and when necessary, venting any high-pressure CO_2 that remained in the autoclave. [Note: numerous experiments were conducted with no high-pressure CO_2 in the autoclave.] After each experiment, reacted solids were removed from the autoclave, dried in a forced-air convection oven, and examined by optical microscopy, scanning electron microscopy (SEM, Fig. 2), powder X-ray diffractometry, and/or electron-probe microanalysis. Phase identification was sufficient to achieve the main aims of the project; therefore, no attempts were made to determine the minor- or trace-element contents of the reacted solids, or the relative proportions of solid products (when more than one solid compound was formed during an experiment).

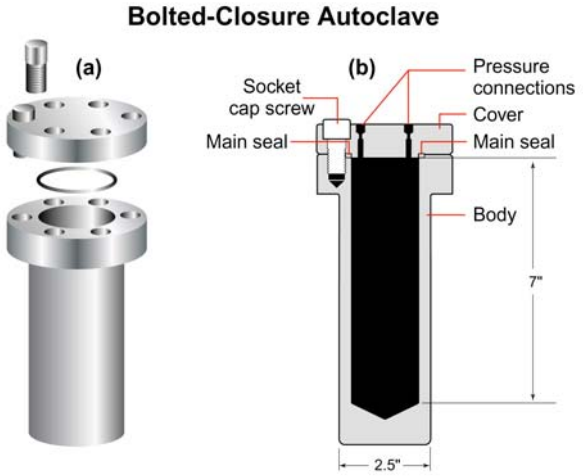


Fig. 1. Schematic illustrations of the bolted-closure autoclave used to perform hydrothermal experiments in this study: (a) perspective view, and (b) cross section through the middle of the vessel.

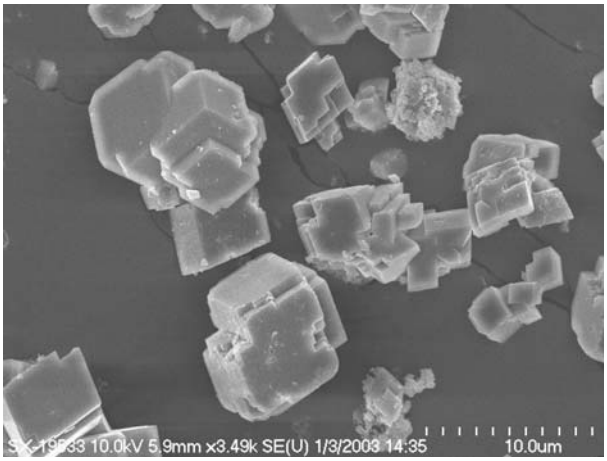


Fig. 2. Secondary electron SEM image of magnesite crystals grown by reacting brucite with CO_2 (Reactions 6 and 11).

Results and Accomplishments

The principal goal of the research was to develop a serpentine-carbonation process that is energetically superior to the direct method discussed above (Reaction 2). To achieve that objective, more than 40 autoclave experiments were performed. The resulting phase equilibrium data are definitive, collectively indicating that hydroxylated serpentine is converted to magnesite most efficiently by the following multistep process.[8] First, the serpentine feedstock is reacted with a concentrated solution of caustic soda (NaOH) to produce crystalline brucite and an aqueous fluid rich in dissolved silica. Written in ionic form, the reaction is

$$Mg_3Si_2O_5(OH)_4 + 2OH^- + 3H_2O \rightarrow 3Mg(OH)_2(\downarrow) + 2\,SiO(OH)_3^- \quad . \qquad (3)$$

[Note: Na^+ ions present in the caustic solution do not participate in serpentine → brucite conversion. Therefore, they have been omitted from Reaction 3.] Next, the solid and liquid products of caustic treatment are physically segregated for the second and third steps of the process, which are, respectively, (i) decomposition of $SiO(OH)_3^-$ by the reaction

$$2\,SiO(OH)_3^- + 2CO_2(aq) \rightarrow 2HCO_3^- + 2SiO_2(\downarrow) + 2H_2O \qquad (4)$$

and (ii) carbonation of brucite by the reactions

$$2Mg(OH)_2 + 2HCO_3^- \rightarrow 2MgCO_3(\downarrow) + 2OH^- + 2H_2O \qquad (5)$$

and

$$Mg(OH)_2 + CO_2 \rightarrow MgCO_3 + H_2O \quad , \qquad (6)$$

where the HCO_3^- used to carbonate brucite in Reaction 5 is formed in step 2 of the process (Reaction 4). The OH^- "regenerated" in Reaction 5 is then used to convert more serpentine to brucite (Reaction 3). Reactions 3–6 form the basis for a mineral carbonation technology that produces magnesite by the *net* reaction

$$Mg_3Si_2O_5(OH)_4 + 3CO_2 \rightarrow 3MgCO_3 + 2SiO_2 + 2H_2O \quad , \qquad (7)$$

which is identical to the direct serpentine-carbonation reaction discussed above (Reaction 2). Significantly, our experimental data also indicate that olivine is converted to magnesite by a multistep process that is very similar to the one developed for carbonation of serpentine (Reactions 3–6): viz.,

$$Mg_2SiO_4 + OH^- + 3H_2O \rightarrow 2Mg(OH)_2(\downarrow) + SiO(OH)_3^- \quad , \qquad (8)$$

$$SiO(OH)_3^- + CO_2(aq) \rightarrow HCO_3^- + SiO_2(\downarrow) + H_2O \quad , \qquad (9)$$

$$Mg(OH)_2 + HCO_3^- \rightarrow MgCO_3(\downarrow) + OH^- + H_2O \qquad (10)$$

and

$$Mg(OH)_2 + CO_2 \rightarrow MgCO_3 + H_2O \quad . \qquad (11)$$

Adding up Reactions 8–11 yields the net reaction

$$Mg_2SiO_4 + 2CO_2 \rightarrow 2MgCO_3 + SiO_2 \quad , \qquad (12)$$

which is identical to the direct olivine-carbonation reaction discussed previously (Reaction 1).

Considering that the solid phases and proportions of solids produced by the multistep carbonation method (Reactions 3–6 and 8–11) and the direct method (Reactions 1 and 2) are identical, it could be concluded that the two approaches are nothing more than different chemical pathways to the same overall result. However, several aspects of the multistep method give it important advantages over the direct method. First, because both olivine and (hydroxylated) serpentine are rapidly decomposed by caustic soda at very low temperatures (≤200°C) in the multistep method, no special pretreatment of the silicate feedstock is required. This contrasts sharply with the direct method for serpentine carbonation, which requires extreme grinding and/or high-temperature (600–650°C) heat pretreatment of the feedstock to ensure satisfactory reactivity.[6] Second, in the multistep method, the fluid pressure required to achieve rapid and efficient carbonation is <10% of that applied in the direct method (i.e., ~15 atm in the multistep method versus 185 atm in the direct method). Therefore, future mineral-carbonation reactors that implement the multistep method at the power-plant scale would require no expensive pressure-intensifying equipment. Third, in the multistep method, separate streams of silica gel and magnesite are produced, allowing each (nearly pure) substance to be used as a raw material for various commercial applications. In contrast, the direct method generates a single stream of solid material, consisting of intimately mixed silica gel and magnesite. The high costs of separating these materials effectively precludes any value-added commercial utility of the solid effluent. In addition, because gelatinous silica is readily mobilized by meteoric water, it is an inherently undesirable waste product for near-surface disposal. Thus, compared to separate masses of its individual components, mixtures of silica gel and magnesite are an economic liability.

The relative merits of the direct and multistep methods for olivine/serpentine carbonation could be assessed more fully by quantitatively comparing their overall kinetics and energetics; determining the capital and operating costs of the different kinds of reactors and processing protocols that would be required to implement each method in industrial settings; accurately calculating the costs of dealing with large masses of mixed silica and magnesite; and completing a detailed cost/benefit analysis of using silica gel and magnesite to create commercially viable byproducts. Unfortunately, these assessments are very difficult to perform at the present time due to a lack of detailed technical and economic information on the particular manner in which each carbonation method would be applied in reducing CO_2 emissions generated by industrial activity.

Summary and Conclusions

Multistep olivine/serpentine carbonation processes have been developed for sequestering CO_2 in industrial settings. The new method has the great advantage that it

avoids the following serious technical problems associated with the "direct" method for olivine/serpentine carbonation, viz., the need to dehydroxylate serpentine prior to carbonation, the necessity of pressurizing CO_2 to ~185 atm to achieve carbonation by reaction with supercritical CO_2, and production of a single solid effluent consisting of intimately mixed silica gel and magnesite. These problems are avoided in the new process by using caustic soda, rather than supercritical CO_2, as the rock "solvent," and by completely converting gaseous CO_2 to aqueous bicarbonate ion prior to forming magnesite. The main challenges for olivine/serpentine carbonation by the multistep method arise in separating reaction products formed in the first three steps of each process. Specifically, the multistep method relies on efficient and cost-effective separations of residual OH^- from $SiO(OH)_3^-$ in step 1 (Reactions 3 and 8), HCO_3^- from silica gel in step 2 (Reactions 4 and 9), and OH^- from residual HCO_3^- in step 3 (Reactions 5 and 10). Potential processing procedures for achieving these separations include the following: in step 1, precipitating solid NaOH by boiling off water; in step 2, gently centrifuging the aqueous solution to physically segregate the silica gel from the HCO_3^--rich aqueous fluid; and in step 3, adding "excess" $Mg(OH)_2$ (produced in step 1) to the reactants to ensure maximum consumption of bicarbonate ion during magnesite precipitation. [Note: In both olivine and serpentine carbonation by the multistep method, the solid effluents of step 3 (magnesite + residual brucite) are reacted with CO_2 in a fourth and final processing step to achieve the maximum possible production of magnesite (Reactions 6 and 11).]

By further demonstrating that olivine/serpentine carbonation is a potentially viable option for sequestering large masses of CO_2 produced by combustion of fossil fuels, the results of this study benefit governmental departments and agencies in both the United States and foreign countries, many of which are striving to reduce anthropogenic CO_2 emissions by either voluntary or regulatory means. However, it is clear that much work remains to be done to optimize olivine/serpentine carbonation for future applications in industrial settings. This reality has prompted the writers to submit proposals to potential funders in the U.S. DOE's Office of Fossil Energy outlining follow-on research that will further demonstrate the technical and economic viability of the new multistep method for achieving this carbonation, which has the potential to supplant all other proposed approaches for global-scale CO_2 sequestration by mineral carbonation.

References

[1]K. S. Lackner, C. H. Wendt, D. P. Butt, E. L. Joyce Jr., and D. H. Sharp, "Carbon dioxide disposal in carbonate minerals," *Energy* **20**, 1153 (1995).

[2]P. M. Goldberg, Z.-Y. Chen, W. K. O'Connor, R. P. Walters, and H. Ziock, "CO_2 Mineral Sequestration Studies in US," Paper presented at the First National Conference on Carbon Sequestration, May 14–17, 2001, Washington D.C.

[3]K. S. Lackner, D. P. Butt, and C. H. Wendt, "Magnesite Disposal of Carbon Dioxide," in B. Sakkestad (ed.), *Proceedings of the 22nd International Conference on Coal Utilization and Fuel Systems,* pp. 419–430, Coal and Slurry Technology Association, Washington, D.C., 1997.

[4]F. Goff, G. D. Guthrie, D. A. Counce, E. Kluk, D. Bergfeld, and M. Snow, *Preliminary Investigations on the Carbon Dioxide Sequestering Potential of Ultramafic Rocks*, Los Alamos National Laboratory Report LA-13328-MS, 1997.

[5]F. Goff and K. S. Lackner, "Carbon dioxide sequestering using ultramafic rocks," *Environ. Geosci.* **5**, 89 (1998).

[6]W. K. O'Connor, D. C. Dahlin, D. N. Nilsen, G. E. Rush, R. P. Walters, and P. C. Turner, "Carbon Dioxide Sequestration by Direct Mineral Carbonation: Results from Recent Studies and Current Status," paper presented at the First National Conference on Carbon Sequestration, May 14–17, 2001, Washington, D.C.

[7]G. D. Guthrie, J. W. Carey, D. Bergfeld, D. Byler, S. Chipera, and H. J. Ziock, "Geochemical Aspects of the Carbonation of Magnesium Silicates in an Aqueous Medium," paper presented at the First National Conference on Carbon Sequestration, May 14–17, 2001, Washington D.C.

[8]J. G. Blencoe, L. M. Anovitz, D. A. Palmer, and J. S. Beard, "Carbonation of Metal Silicates for Long-Term CO_2 Sequestration," U.S. patent application filed November 12, 2003.

Online Characterization of Individual Airborne Bacteria

W. B. Whitten[1] and S. Hoffmann[2,*]
[1]*Chemical Sciences Division, ORNL*
[2]*Institute for Spectrochemistry and Applied Spectroscopy, Dortmund, Germany*

An apparatus was developed to decouple the laser desorption and ionization processes in real-time mass spectrometry of individual airborne particles. Experiments were performed to explore atmospheric pressure chemical ionization of laser-desorbed species from airborne particles. An important discovery that we made was that certain biological molecules that were ionized in an atmospheric pressure discharge did not undergo the fragmentation that usually occurs when molecules are ionized in vacuum. Absence of fragmentation leads to much simpler mass spectra that make identification of the particle composition more positive.

Introduction

Timely detection and identification of airborne microorganisms is becoming increasingly important. Traditionally, bioaerosols are collected on culture media and identified by colony growth and response to various agents. There is considerable interest, however, in chemical techniques that would be potentially more rapid and that could also identify nonviable organisms. These techniques fall into two classes—those that use the highly specific attraction of the antibody-antigen reaction or the hybridization of gene probes to DNA or RNA and those based on conventional chemical analysis. The latter methods are usually faster and can respond to a wide range of species, at the expense of specificity. We have been developing an instrument to characterize airborne microparticles in real time by laser ablation mass spectrometry. Previous studies of laser ablation mass spectrometry of airborne particles showed that charge transfer from ions generated in the laser ablation process to neutral species with lower ionization potentials caused the measured mass spectrum to be dominated by the latter species. For the characterization of microorganisms, it would be desirable to detect ions of proteins of other biomarkers of high molecular weight that are specific to a particular organism. The mass spectrum of bacteria that we obtain is predominately from adducts of potassium with small fragments of unknown composition as well as ions of iron and other inorganic substances. We anticipate that performing the ionization after the laser ablation plume has expanded sufficiently to eliminate most ion-neutral collisions will yield a more representative mass spectrum.

[*]This project was part of the PhD thesis research of Sven Hoffmann who was visiting ORNL from Germany.

Technical Approach

A Varian Saturn II mass ion trap mass spectrometer was modified by the addition of an atmospheric pressure inlet and an ion guide based on two linear electrodynamic quadrupoles operating at approximately 4 MHz and 600 V peak amplitude. A photograph of the atmospheric-pressure inlet system and ion guide housing is shown in Fig. 1. By applying appropriate bias voltages to the quadrupole electrodes and apertures, it was possible to transfer ions entering the inlet to the quadrupole ion trap of the mass spectrometer for mass analysis. The ions were generated by atmospheric pressure chemical ionization in a corona discharge external to the inlet. Other experiments were carried out with a Finnigan LCQ Deca XP Plus mass spectrometer that has an atmospheric-pressure inlet system.

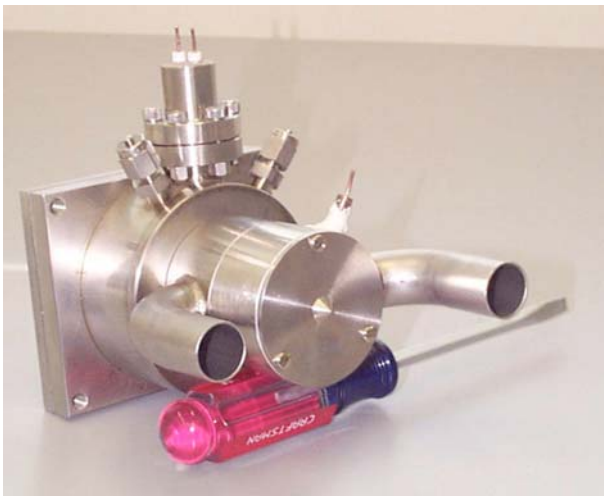


Fig. 1. Photograph of the atmospheric-pressure inlet system and ion guide housing.

Results and Accomplishments

For the first tests, samples were deposited on a metal target placed near the inlet, as shown in Fig. 2. The corona discharge was produced between the needle and the inlet housing. Sample molecules were desorbed from the target by a pulsed Nd:YAG laser and ionized by proton exchange with water ions produced in the discharge. Results with substances such as benzophenone were encouraging. Interesting results were also obtained for samples desorbed from the vicinity of the entrance orifice. However, the sensitivity of the instrument was apparently insufficient to obtain laser desorption mass spectra from airborne particles.

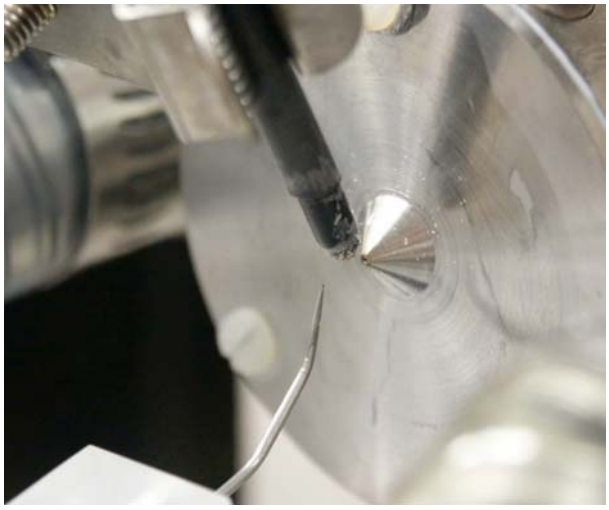


Fig. 2. Setup for preliminary tests.

To test the effects of instrumental sensitivity, the experiments were moved to the Finnigan Deca mass spectrometer. This instrument was designed for atmospheric pressure sampling of a liquid chromatograph effluent and contains a high-efficiency ion guide to direct the externally produced ions into the mass analyzer. A photograph of the experimental arrangement at the entrance of this mass spectrometer is shown in Fig. 3. Some exciting preliminary results were obtained from airborne tryptophan particles in the brief time available with this instrument. The mass spectrum obtained from a single tryptophan particle is shown in Fig. 4. The spectrum is almost entirely due to the protonated parent ion.

Summary and Conclusions

The important result from these studies is that the ionization resulting from the laser desorption atmospheric pressure chemical ionization produced little or no fragmentation of the parent molecule. If this should hold true as we expect for larger biological molecules, the mass spectrum from a complex mixture will be easier to interpret and a larger concentration of parent ions will be available for further characterization by tandem mass spectrometry.

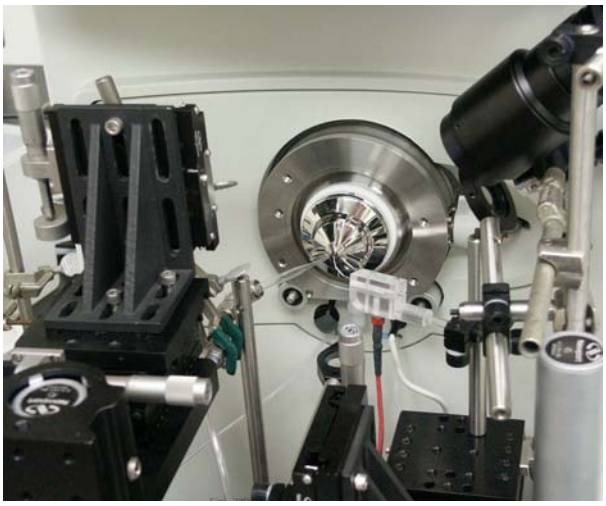


Fig. 3. Experimental arrangement with Finnigan LCQ Deca XP Plus.

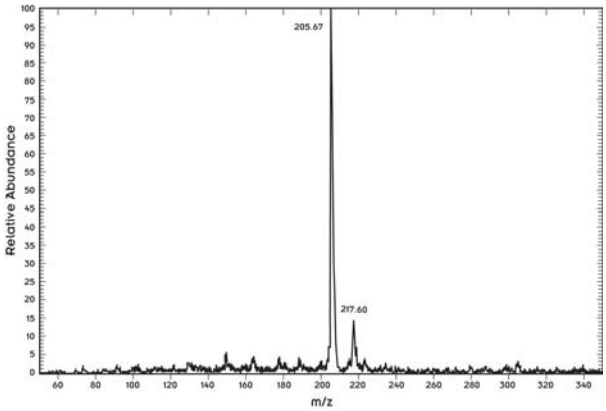


Fig. 4. Mass spectrum from a single tryptophan particle.

While at ORNL, Mr. Hoffmann designed and constructed several pieces of apparatus that will be used by our group in future work. These include an atmospheric pressure inlet for the Saturn mass spectrometer, discharge cell, quadrupole ion guide, and a high-voltage radio-frequency generator to power the guide. A task describing a laser desorption/atmospheric pressure chemical ionization front end for a next-generation biological mass spectrometer was included in a whitepaper submitted in response to a call from the Department of Homeland Security.

Photomolecular Comb

R. J. Warmack,[1] T. G. Thundat,[2] T. L. Ferrell,[2] and G. M. Brown[3]
[1]*Engineering Sciences and Technology Division*
[2]*Life Sciences Division*
[3]*Chemical Sciences Division*

A novel concept was explored that has the potential to accelerate and automate chromatographic analyses using photo-directed accumulation. Photoexcitation of a semiconductor surface in an electrochemical cell produces an electric field that can accumulate charged proteins. Motion of focused light beam can then be used to move the proteins in a directed, channel-less fashion. Molecular motion and chromatographic separation in a buffered gel depends upon the charge of the molecule and its size compared to the gel pores. This exploratory work has been used to demonstrate photo-accumulation, motion, and separation of protein mixtures for the first time. When the method is fully developed, this technology will provide direct control of chromatographic separation, hopefully with higher speed and resolution, which should greatly accelerate proteome analyses.

Introduction

The need for increased efficiency in drug discovery and disease diagnostics has driven a tremendous amount of research aimed at understanding the underlying mechanism of disease. With the sequence of the human genome nearly complete, there has been a significant shift to the post-genome sciences of gene expression and proteomics. Proteome analysis requires fast methods with high separation efficiencies in order to screen the various cell and tissue types for their proteome expression and monitor the effects of environmental conditions and tissue development. While many of the important tools for protein research have been in place for decades, current techniques lack the resolution, integration, and speed of analysis required for current-generation needs. Also, a perpetual problem with chromatographic separation is the need to attach hazardous fluorescent compounds or radionuclides as labels.

The most widely used tool for protein separation, two-dimensional gel electrophoresis, has been available for nearly 30 years. The technique resolves complex mixtures of proteins first by isoelectric point and then by size, generating a gel image pattern, or "fingerprint," of the proteome of a particular biological system. While significant knowledge has been gained with the technique, current systems lack the seamless integration and automation to allow for the reproducible processing of samples on a truly high-throughput basis. Other limitations include low speed, lack of reproducibility, limited range, and the need for attaching labels.

The work described below explores a new basis for chromatographic separations using a swept light beam to both direct the chromatography and monitor its results without using labels. The fundamental concepts are verified, and it is hoped that a fully developed method will provide direct control of chromatographic separation, with higher speed and resolution, as well as provide a built-in electronic readout. Automation, so essential to high-speed screening, is an intrinsic feature of the method.

Technical Approach

The photomolecular comb is a novel technological concept, invented at ORNL, for performing chemical separation and detection of biomolecules such as DNA or proteins using photoelectrochemistry. The idea, shown in Fig. 1, uses a photo-excited semiconductor to induce localized electrophoresis through a medium (buffer solution or gel) in contact with the semiconductor. For example, an electrochemical photo-anodic current using certain photoactive semiconductors causes a positive

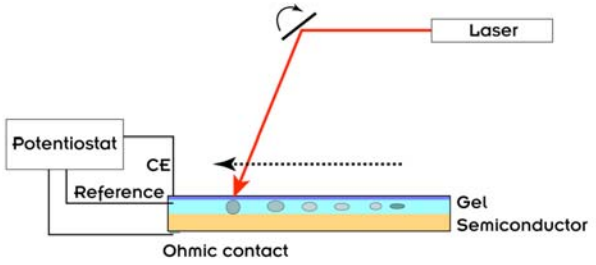


Fig. 1. Molecular comb concept uses a scanned light source to sweep a biosample through a gel to chromatographically separate the components. An electrochemical cell with reference electrode and counter electrode (CE) establishes the proper bias at the semiconductor surface.

charge to develop at the illuminated region. The resulting electric field attracts negatively charged species, such as proteins, and causes them to accumulate at the illuminated spot. An accumulation effect was previously demonstrated for polyanionic DNA using manganese-oxide-coated silicon.[1] Lateral motion of the light spot causes the charged species to be pulled along. When a sieving material such as polyacrylamide is present, chromatographic separation of the biomolecules is possible. In other words, the sieving gel exerts larger forces on high-molecular-weight protein molecules and less so on smaller species. Biomolecules with a low charge or high mass move more slowly and will be left behind if the beam moves with sufficient speed. Consequently, mixtures of biomolecules with different charges and sizes can be separated. This is a unique form of chromatography in which separations can be performed *without* channels or high voltages. In principle, the resolution is limited by the size of the focus so that protein separation can be achieved in a small cell, which could dramatically reduce the analysis time.

Semiconductors such as Si, Ge, GaAs, CdS, and certain metal oxides such as Mn_2O_3, TiO_2, and ZnO exhibit a change in surface charge upon irradiation with light of an appropriate wavelength. Absorption of photons of energies greater than the corresponding semiconductor band gap forms electron-hole pairs. Proper biasing of the solid-liquid interface to create a depletion layer at the surface of the semiconductor results in separation of the charges. The separated charge carriers reach the interface and create a localized photovoltage, whose magnitude is proportional to the light intensity and the extent of the band bending, controlled by adjusting the biasing voltage. The applied potential can be alternated to control the photovoltage, and thus the motion of the biomolecules as well, so that saturation of the biomolecules onto the substrate can be avoided. The illumination duty cycle of the laser can also be modulated to control the motion of the molecules. These experimental conditions can be optimized according to the diffusion and the effective viscosity of the molecules in the medium.

The experimental apparatus consisted of a special electrochemical cell that was open to illumination of a focused laser source. The cell was generally formed of the semiconductor, as the working electrode, upon which was cast agarose or polyacrylamide gel from 20 μm to 2 mm thick. The counter electrode was either a platinum wire loop or plate, or a slide coated with a transparent conducting film of indium-tin-oxide or gold. The buffers were Tris-glycine and others that were adjusted to provide reasonably low conductivity, allowing penetration of the photo-generated electric field across the gel. Ohmic contacts to the semiconductors were provided by epoxy-encapsulated gallium/indium eutectic or annealed aluminum film on the backside. A silver-wire reference electrode was used in some experiments but found to be unnecessary in general. Additionally, a pulsed power supply was designed to allow rapid control of the applied electrochemical bias. A low-power laser (e.g., 633-nm helium-neon, <0.5 mW) was focused onto the sample surface with a specially constructed apparatus designed to scan the laser spot across the working electrode. Two orthogonally arranged galvo mirrors (Model CT-6210, Cambridge Technologies) were interfaced under computer control to provide high-speed control and raster scanning of the light beam.

Results and Accomplishments

A number of experiments were performed using various semiconductors and electrochemical cell arrangements. Using germanium as the semiconductor substrate (working electrode), the photocurrent due to a low-power green or red laser was easily recorded. After depositing an agarose gel and injecting a stained protein, an accumulation was noticed after an exposure of a few minutes under the laser. By slowly moving the laser, the proteins could be moved within this gel. Figure 2 shows a few images of a stained protein (Kaleidoscope™, BioRad Laboratories) being accumulated and moved by two laser

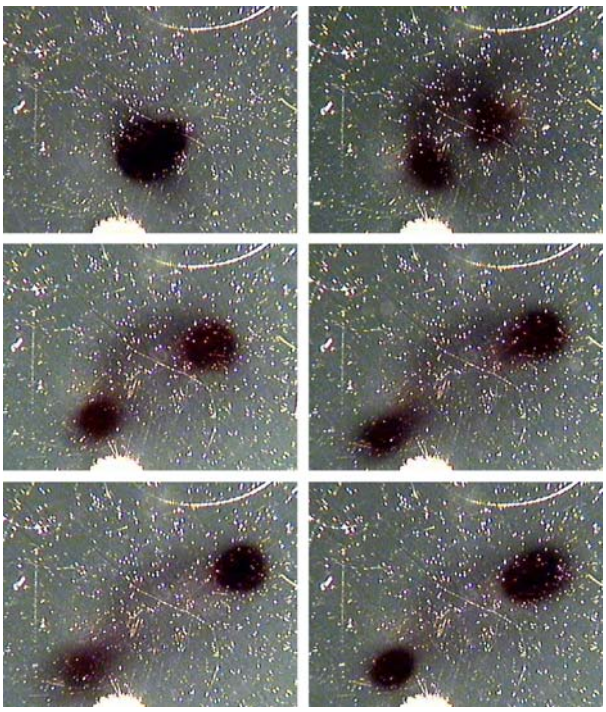


Fig. 2. A series of images of a stained protein (Kaleidoscope™, BioRad Laboratories) being accumulated and moved by two laser spots in an agarose gel located on a germanium working electrode. The light beams were formed by rapidly switching the galvo mirrors and increasing the separation between spots. The spots are moved roughly 5 mm from center to center. The scratches and blemishes in the surface do not affect photo-electrochemical activity.

spots. The size of the spots is roughly 1 mm in diameter and is significantly larger than laser-beam focus due to long charge-carrier diffusion in this single-crystal germanium.

There are a number of challenges that were faced in the electrochemical environment. The electric field depends upon the current density, so that the ensuing photochemical processes can alter the state of the electrodes and the electrolyte. Photocorrosion of the anode as a consequence of water hydrolysis causes oxidation of the surface or bubbles of oxygen to appear. Either effect can disturb the current density. Noble metals,[2] metal oxides,[3] and conducting polymers[4] have also been used to achieve varying levels of stabilization. Additionally, the electrolyte in the vicinity becomes more acidic, lowering the pH, which can alter the charge of proteins, which are amphoteric. In fact, in many cases, fluorescent-stained proteins first accumulate at the illuminated spot but then begin to be repelled as the local pH drops below the isoelectric point of the protein. The electric field is also limited by the conductivity of the electrolyte, so proper buffer preparation was found to be beneficial. Obviously, high contrast in photocurrent between light and dark regions is preferred. A variety of anodic and cathodic materials were tested.

Another goal of the project was to demonstrate the separation of proteins. This was achieved on the basis of charge difference. Figure 3 shows images of two proteins with opposite charge, which can be adjusted under pH control. A plug containing a mixture of the two proteins was located at the bottom of the cell next to the working electrode. The positively charged bovine serum albumin (BSA) is attracted to the photoanode, which repels the negatively charged cytochrome-C.

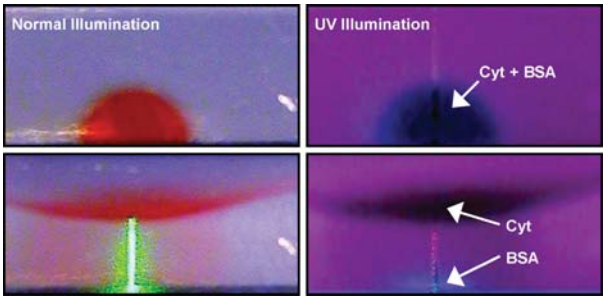


Fig. 3. Images showing the separation of two proteins, cytochrome-C (natural reddish color) and BSA (stained with Marina Blue, Molecular Probes). Top: *Initial cell distribution. The mixture is located near the working electrode.* Bottom: *The cytochrome-C moves to the cathode while the BSA is attracted to the photoanode. Images at the left are under normal illumination, while the images on the right are illuminated under ultraviolet light. The laser light appears as vertical beam. The fluorescent-stained BSA appears as a band near the bottom, while the cytochrome-C appears at the top.*

Summary and Conclusions

The basic premises of the photomolecular comb were established, including the *first known observation of protein photoaccumulation and separation*. When optimally implemented, this concept should have a very significant impact upon diagnostics related to genomics and proteomics. Gel electrophoresis is the mainstay of the burgeoning field of proteomics but is resolution and rate limited. Our work strengthens ORNL's position in biomedical technology and will help us establish long-term research programs with funding from private industry, DOE, and NIH. Already, a funds-in CRADA has been established with Qgenics Biosciences, Inc., which has licensed the intellectual property and is actively and aggressively developing the technology for biomedical diagnostics.

The authors gratefully acknowledge the assistance, advice and encouragement of Qgenics Biosciences personnel, especially Jay Harkins and Chuck Witkowski.

References

[1]I. C. Gurtner, C. Edman, R. Formosa, and M. Heller, "Photoelectrophoretic Transport and Hybridization of DNA Oligonucleotides on Unpatterned Silicon Substrates," *J. Amer. Chem. Soc*. **122**, 8589–94 (2000).

[2]S. Menezes, A. Heller, and B. Miller, "Metal Filmed Semiconductor Photoelectrochemical Cells," *J. Electrochem. Soc.* **127**, 1268–73 (1980).

[3]T. Osaka, K. Ejiri, and N. Hirota, "Photoelectrochemical Behavior of Iron-Oxide N-Si Heterojunction Electrodes with An Outer Pd Layer," *J. Electrochem. Soc.* **131**, 1571–4 (1984).

[4]R. Noufi, A. Frank, and A. Nozik, "Stabilization of N-Type Silicon Photoelectrodes to Surface Oxidation in Aqueous-Electrolyte Solution and Mediation of Oxidation Reaction by Surface-Attached Organic Conducting Polymer," *J. Am. Chem. Soc.* **103**, 1849 (1981).

Protein Surface Mapping by Chemical Oxidation and Mass Spectrometry

R. L. Hettich,[1] J. Sharp,[2] D. Xu,[3] and Y. Xu[3]
[1]*Chemical Sciences Division*
[2]*Graduate Student, ORNL-UTK GST Graduate School*
[3]*Life Sciences Division*

The solvent-accessible surface area of proteins is known to be important in biological function for many reasons, including protein-protein interactions, protein folding, and catalytic function. Because protein surfaces play a vital role in all facets of protein activity, any experimental technique capable of rapidly probing the surface of proteins would greatly assist the investigation of how protein structure is related to function. In this project, we have developed and demonstrated a chemical technique for protein surface mapping based on the oxidation of amino acid side chains in both model proteins such as apomyoglobin, lysozyme, and lactoglobulin, as well as a protein (yeast sml1p) with unknown structure. The elucidation of the effect of solvent accessibility on the sites of oxidation was determined by using both high-resolution electrospray Fourier transform ion cyclotron resonance mass spectrometry (ES-FTICR-MS) and liquid chromatography quadrupole ion trap tandem mass spectrometry (LC-QIT-MS/MS). Mass spectrometry is an ideal tool for analysis of sites of chemical modification due to its high accuracy, sensitivity, ability to be coupled to powerful separations techniques, and fragmentation techniques. Such experimental information can be combined with computational techniques for more comprehensive protein characterization.

Introduction

In modern biology, protein analysis is aided immensely by the determination of the three-dimensional structure of the protein, allowing for rational experimental design to answer specific questions about the function of the protein. Three-dimensional protein structure analysis is often used to study not only the intact structure of a protein, but also its ligand binding sites, protein-protein interactions, and induced conformational changes. Three-dimensional structures are also useful for rational design of enzyme inhibitors, essential in modern drug development. In all fields of biology, protein structure determination always advances the study of any protein to which it is successfully applied, often answering questions that otherwise would require years of tedious research to address.

The central hypothesis of this project is that chemical labeling methods with high-performance mass spectrometric characterization can be used for rapid protein surface mapping by identifying surface accessible amino acids residues in native-folded proteins, and this information can be utilized in an integrated fashion to enhance the accuracy and throughput of computational structural prediction methods. It is the unique combination of radical labeling of native protein structures, high-resolution mass spectrometric characterization, and high-throughput computational structural prediction techniques that provides the power for this approach. If successful, this method could enhance dramatically the ability to characterize the structures (at low resolution) of a wide variety of proteins, providing critical input into the refinement of computational prediction methods.

Technical Approach

The need clearly exists for a high-throughput method for accurate determination of protein structure, preferably based on actual biophysical data. Current structural techniques generally require prior purification of the protein of interest, inherently limiting their speed as well as limiting the range of proteins that can be analyzed to those that can be successfully purified in a native form. It has also been established that computational methods, which are currently the backbone of high-throughput structural analyses, often yield inaccurate structures but can be improved through the use of biophysical constraints.[1–3] The research proposed for this project attempted to address these needs through development of an experimental protocol for protein surface mapping by mass spectrometry.

A viable surface mapping methodology must be based on the specific modification properties, small size, and high hydrophilicity of the reagent species while avoiding the problems resulting from protein cleavage. One promising reagent is the hydroxyl radical, which can modify amino acid side chains via a radical-mediated oxidation. The sites of radical interaction are characterized

at the molecular level by mass spectrometry, a high-sensitivity, high-dynamic range analytical method that can readily handle complex mixtures. The hydroxyl radical appears to meet almost all of the desirable properties of a surface labeling reagent. This radical has been utilized to label a variety of model proteins, and it has been shown that the radical will not interact with side chains that are buried in the native structure, showing that the process itself does not modify the protein topography, and that the presence of the labeling compounds also does not modify the protein topography.[4–5] Moreover, the history of radiation biology provides an excellent collection of data on the reactivity of side chains with hydroxyl radicals.[6]

In this project, we sought to develop and demonstrate the ability to rapidly and accurately probe the surfaces of known as well as uncharacterized proteins in mixtures with hydroxyl radicals. Such an experimental approach could be used to greatly improve the accuracy of structural modeling, with each computational model supported by experimental biophysical data. The longer term goal of this research is to demonstrate how experimental surface mapping and computational prediction can be integrated into an approach that is more powerful than the sum of each technique separately.

Results and Accomplishments

There were three specific focus areas for this project. Each one resulted in a separate manuscript and is summarized below.

1. Fenton Chemistry-Generated Hydroxyl Radical Probes for Surface Mapping of Apomyoglobin

We have developed an experimental protocol to use *chemically* generated hydroxyl radicals to probe the surface areas of proteins by determination of the solvent-accessible amino acids which can be oxidized and then identified by mass spectrometry. This method has rapid reaction times and requires commonly available chemicals. Apomyoglobin was employed as the initial model protein because its secondary and tertiary structure has been extensively characterized by NMR;[7] it is commercially available at high purity, easily digested, and has no prosthetic groups to complicate chemistry or mass spectrometric analysis.

Aqueous buffered solutions of apomyoglobin were oxidized for varying amounts of time using chelated iron and hydrogen peroxide to generate solution-phase hydroxyl radicals, which then react with accessible amino acid side chains to oxidize the protein. Analysis of the oxidation products by high-resolution electrospray Fourier transform mass spectrometry (ES-FTMS) showed that oxidation of apomyoglobin occurs in a time-dependent manner, and several amino acid side chain oxidation events can be detected before significant backbone cleavage occurs. The increase in oxidation events per molecule was clearly time dependent, with cleavage of the protein backbone by the hydroxyl radicals occurring at extended times (several minutes).

In order to determine which amino acid side chains are solvent accessible, we interrogated the sites of oxidation by both ES-FTICR-MS and by reverse phase capillary LC-QIT-MS/MS. The use of high-resolution ES-FTICR-MS allows for high-accuracy mass measurements to ensure the identity of oxidized peptides, while LC-MS/MS is a robust, partially automated method for interrogating the complex mixture to assign oxidation sites. Oxidized peptides identified by their masses were compared to an *in silico* digestion of apomyoglobin. The site of oxidation then was determined by tandem mass spectrometry, looking for a neutral mass loss 16 Da heavier than the expected neutral mass loss. Oxidized peptides containing the two methionines were detected, as well as oxidized peptides containing either Phe151 or Trp7 and Leu11.

Under conditions of low protein oxidation, we have positively identified five oxidation sites by liquid chromatography–tandem mass spectrometry and high-resolution mass spectrometry. With the exception of methionines, all of the residues oxidized and identified in this study are highly solvent accessible according to the NMR structure of apomyoglobin. One interesting note is that this hydroxyl radical attachment technique was able to accurately identify amino acids located in the disordered regions of the apomyoglobin protein (as suggested by NMR measurements). This data provides evidence that the use of chemically generated hydroxyl radicals as probes to determine protein surface residues is feasible, with the exception of methionine oxidation. The short reaction time (30 seconds) along with relatively short digestion and analysis times suggest that this method would be useful as a rapid analysis of certain aspects of protein structure. The results from this proof-of-principle experiment are evidence that selective amino acid side chain oxidation by chemically generated hydroxyl radicals can be used for the rapid analysis of protein solvent accessible surface areas. This work was published in *Analytical Biochemistry* **313**, 216–225 (2003).

2. UV-Radiation-Generated Hydroxyl Radical Probes of Model Proteins

The purpose of this task was to demonstrate how a simple, rapid method for generating hydroxyl radicals from UV-irradiation of hydrogen peroxide solutions avoids the sample cleanup difficulties and high salt concentrations required for the previous approach utilizing Fenton chemistry and generates sufficient hydroxyl radicals to probe the solution-phase structure of two model proteins with well-characterized, stable NMR structural models:

hen egg white lysozyme and bovine β-lactoglobulin A. These two proteins have very different secondary and tertiary structures and serve nicely as test models for quantitating the surface mapping protocol. The rate of oxidation of any one side chain should be influenced by the solvent accessibility of the reactive atoms in the side chain, as well as the inherent chemical reactivity of the side chain. Hydroxyl radicals were generated by UV irradiation of hydrogen peroxide, and the model protein was allowed to react with the hydroxyl radicals. The resulting oxidized protein was analyzed by high-resolution, high-accuracy ES-FTMS to determine the extent of oxidation. The oxidized protein was then digested with trypsin; the resulting peptides were analyzed by LC-MS/MS in order to determine the residues that were oxidized and then analyzed by ES-FTMS in order to both confirm the identity of oxidized peptides and to gather more comprehensive information on oxidized peptides that were not identified by LC-MS/MS. Finally, the sites of oxidation were analyzed against the known solution-phase NMR structures of the model proteins in order to determine the extent to which solvent accessibility influenced the rate of oxidation. A manuscript on this work has been submitted to *Analytical Chemistry* and is in press.

Aqueous lysozyme oxidized for 5 min was denatured and reduced, and then digested with trypsin. The tryptic digest mixture from oxidized lysozyme was loaded onto a C18 column for LC-MS/MS analysis to determine the sites of oxidation. After the LC-MS/MS experiment, the measured peptides were analyzed computationally for differential oxidation of cysteine, tryptophan, methionine, phenylalanine, tyrosine, histidine, proline, leucine, and isoleucine. The non-methionine oxidized residues of lysozyme as detected by LC-MS/MS are shown as colored in Fig. 1. When plotted onto the X-ray crystal structure of lysozyme, PDB accession number 193L, with water removed for clarity,[8] the four oxidized residues can clearly be seen to be on the surface of the protein.

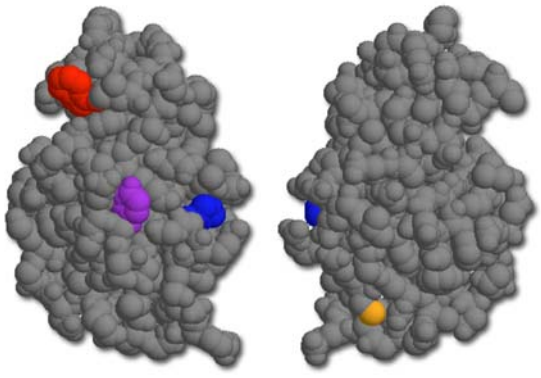


Fig. 1. Two views of the X-ray crystal structure hen egg white lysozyme with waters removed for clarity (PDB accession number 193L). Oxidized residues confirmed by LC-MS/MS are colored, with the sulfur of Cys6 is shown in orange, with the side chain of His15 (blue), Tyr20 (purple), and Trp62 (red) also indicated. Additional oxidized residues were detected by ES-FTMS but are not included here since the exact oxidation sites were not confirmed.

Photochemical oxidation has the advantages of being a simple protocol for labeling of a variety of residues, most of which are normally expected to be buried in protein structures. Since this method probes residues that are expected to be buried in the protein due to their hydrophobicities, even though the degree of surface labeling is limited, the information value of each oxidation event is much greater than if, for instance, the method labeled charged residues such as arginine and lysine that would be expected to be present on the protein surface. The ability of this technique to provide structural data for targeted proteins in a rapid, relatively simple fashion without the necessity for the X-ray synchrotron beamline could be of great use to structural biologists to test structural models or investigate conformational changes.

3. Integrating Experimental Constraints with Computational Structure Prediction for the Yeast DNA Checkpoint Protein Sml1

As the regulator of ribonuclease reductase (Rnr1p), the protein Sml1p is a highly important checkpoint protein, controlling intracellular levels of deoxynucleotide triphosphates and is thus important in maintaining genetic stability.[9] A significant amount of work in multiple labs has been performed in an attempt to elucidate the structure of Sml1p, although to date only limited structural details are known. Previous NMR work performed showed that Sml1p has two alpha helices comprising residues 6–16 and 63–82.[10]

Photochemically generated hydroxyl radicals were used to map solvent exposed regions in the functional as well as denatured form of the C16S mutant of Sml1p from *Saccharomyces cerevisiae*. The data on amino acid solvent accessibilities were then utilized to evaluate pseudo-*ab initio* computational models generated by the Rosetta algorithm, as shown in Figs. 2 and 3. All of the top five models returned by Rosetta disagreed with the solvent accessibility data, as well as with previously published NMR data. A model adjusted using the NMR data as well as tryptophan fluorescence data satisfied most of the solvent accessibility constraints. Through rotation of two amino acid side chains, a model was generated that satisfies previously reported NMR data, fluorescence data, and the solvent accessibility data presented here, as well as possessing a lower free energy than the model constrained solely by NMR and tryptophan fluorescence data. A manuscript on this work has been submitted to the *Journal of Biological Chemistry*.

While the approach taken with the analysis of Sml1p is a valid method for eliminating incorrect structures, better results would be achieved if the constraints were built into the modeling algorithm itself. Current efforts are being made to generate a constraining term for the protein

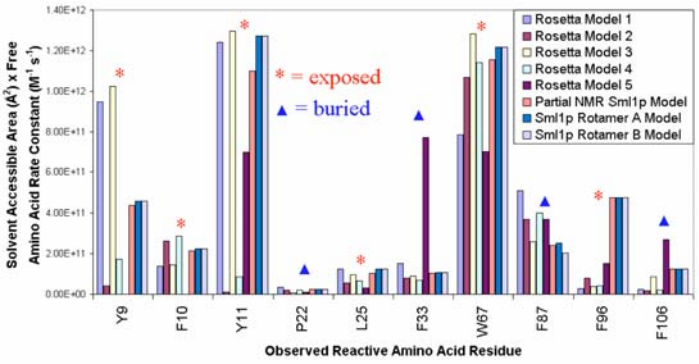


Fig. 2. The product of the solvent-accessible reactive area and the respective free amino acid rate constant for each of the ten reactive, non-methionine residues in C16S Sml1p. Residues that were oxidized in the native state are marked with a red asterisk (exposed), and residues that were oxidized only in the denatured state are marked with a blue triangle (buried).

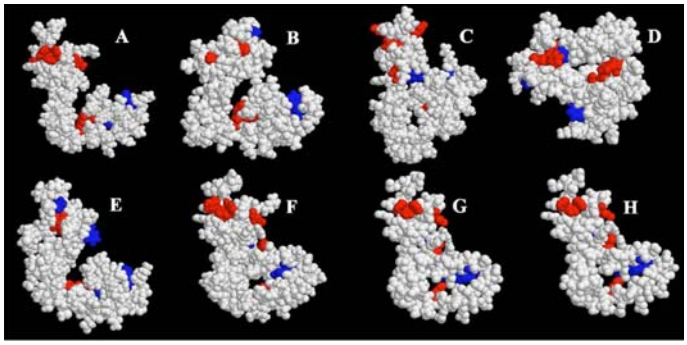


Fig. 3. Spacefill representations of each of the eight computational models discussed. A-Rosetta model 1; B-Rosetta model 2; C-Rosetta model 3; D-Rosetta model 4; E-Rosetta model 5; F-partial NMR Sml1p model; G-Sml1p rotamer A model; H-Sml1p rotamer B model.

threading algorithm PROSPECT, in an attempt to improve threading template selection and energy minimization.[11] Once the solvent accessibility data can be inserted within a threading algorithm, a direct pipeline between surface mapping experiments and constrained computational structures can be established for pseudo–high throughput constrained structure prediction, a tremendous goal for systems biology.

Summary and Conclusions

This project enabled research to demonstrate a viable approach to combine experimental protein structural information (obtained by surface mapping) with computational modeling methods. Although most of the research proposed was directed at demonstration and validation of the protein surface mapping technique with known proteins, the work with the Sml1 protein, whose structure had only previously been partially characterized, should demonstrate how such technology could become very valuable for a wide range of protein structural studies, ranging from bacterial protein characterization to rapid structural interrogation of proteins which have been altered as a function of disease conditions. Such experiments are likely to involve small quantities of proteins in simple or complex mixtures. As such, there presently are virtually no experimental techniques that can even begin to attempt to unravel structural details under these conditions.

One favorable outcome if the proposed experimental approach is successful is the large amount of structural data (albeit low resolution) that can be generated from protein mixtures. Note that such experiments are currently impossible with either x-ray crystallography or nuclear magnetic resonance techniques, which are the primary structural biology tools at present. Such data would be useful to rapidly screen and compare general or targeted proteins in prokaryotic and eukaryotic systems.

References

[1]O. Tcherkasskaya, E. A. Davidson, and V. N. Uversky, "Biophysical Constraints for Protein Structure Prediction," *J Proteome Res* **2**, 37–42 (2003).

[2]P. M. Bowers, C. E. Strauss, and D. Baker, "De Novo Protein Structure Determination Using Sparse NMR Data," *J Biomol NMR* **18**, 311–318 (2000).

[3]Y. Xu, D. Xu, O. H. Crawford, Einstein, F. Larimer, E. Uberbacher, M. A. Unseren, and G. Zhang, "Protein Threading by PROSPECT: a Prediction Experiment in CASP3," *Protein Eng* **12**, 899–907 (1999).

[4]J. G. Kiselar, S. D. Maleknia, M. Sullivan, K. M. Downard, and M. R. Chance, "Hydroxyl radical probe of protein surfaces using synchrotron X-ray radiolysis and mass spectrometry," *Int J Radiat Biol* **78**, 101–114 (2002).

[5]J. S. Sharp, J. M. Becker, and R. L. Hettich, "Protein Surface Mapping by Chemical Oxidation: Structural Analysis by Mass Spectrometry," *Anal Biochem* **313**, 216–225 (2003).

[6]W. M. Garrison, "Reaction-Mechanisms in the Radiolysis of Peptides, Polypeptides, and Proteins," *Chemical Reviews* **87**, 381–398 (1987).

[7]J. T. Lecomte, Y. H. Kao, and M. J. Cocco, "The Native State of Apomyoglobin Described by Proton NMR Spectroscopy: the A-B-G-H Interface of Wild-Type Sperm Whale Apomyoglobin," *Proteins* **25**, 267–85 (1996).

[8]A. C. Young, R. F. Tilton, and J. C. Dewan, *J Mol Biol* **235**, 302–317 (1994).

[9]X. Zhao, E. G. Muller, and R. Rothstein, *Mol Cell* **2**, 329–340 (1998).

[10]X. Zhao, B. Georgieva, A. Chabes, V. Domkin, J. H. Ippel, J. Schleucher, S. Wijmenga, L. Thelander, and R. Rothstein, *Mol Cell Biol* **20**, 9076–9083 (2000).

[11]Y. Xu and D. Xu, "Protein threading using PROSPECT: Design and evaluation," *Proteins* **40**, 343–354 (2000).

Gold Nanocrystal Sensors/Concentrators for Chemical Weapons Agents

T. G. Schaaff
Chemical Sciences Division

The goal of this project was to lay the scientific foundation for developing a program for using gold nanocrystals as selective chemical warfare agent sensors and collectors that are simple and capable of being field deployed with little effort. This proposal was designed to address the scientific proof of principle that gold nanocrystals may constitute a viable sensor and/or concentrator for organosulfur and organophosphorous compounds in civilian and military settings. Gold nanocrystals in various environments were exposed to compounds that are toxic (e.g., pesticides) or simulate chemical warfare agents. Two conditions were tested: (i) changes in optical properties of gold nanocrystal compounds due to exposure of toxic compounds and (ii) collection of toxic compounds by gold nanocrystal compounds. The gold nanocrystal compounds changed color upon exposure to both cyanide and organophosphorous compounds. In addition, it was shown that organophosphorous compounds could be collected and subsequently desorbed and analyzed by mass spectrometry.

Introduction

The ability to detect, concentrate, and identify chemical warfare agents (CWAs) are lacking in both military and civilian settings. Ideal detection systems would accomplish all three tasks, but currently few exist that can readily accomplish more than one task in a given setting and still remain field deployable. In order to accomplish more than one task, detection systems must be designed so that they are both highly sensitive and selective. In addition, the resulting system must be small, rugged, and of simple design, such that it can be deployed in any environmental setting. Ideally, the system would be man-portable. For example, mass spectrometers are highly sensitive and can identify chemical agents readily, but miniaturizing, simplifying, and field-deploying mass spectrometers has proven a daunting task. Simple systems that rely on color changes would ideally suit a battlefield environment.

Of special interest to counterterrorism and military CWA detection are nerve and blister agents. The most widely used nerve agents (e.g., Tabun, Sarin, Soman, GF, VX) are organophosphorous compounds, and blister agents (e.g., Mustards) are typically organosulfur compounds. There is also a need to monitor other environmental toxins, such as pesticides and so-called dual-use compounds (compounds that are used in the production of CWAs). These compounds share the same chemical functionality, typically consisting of organophosphorous and organosulfur (see Fig. 1). Much of the difference lies in the volatility of the organic compounds (e.g., CWAs have appreciable vapor pressure at RT, whereas pesticides are typically nonvolatile) and the reactivity of various leaving groups in the structure. As such, dual-use compounds and pesticides can be used as surrogates or simulants for testing detection systems for CWAs. In addition, the detection of traditional toxins (e.g., cyanide) is of interest to a wide ranging community: from environmental to military.

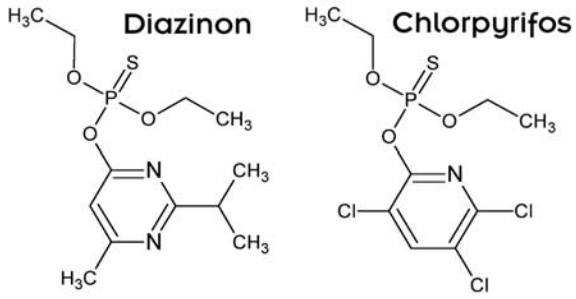


Fig. 1. Surrogates used for these studies.

Gold surfaces are considered "noble" because under most conditions, surface reactions (e.g., oxidation) do not occur readily. However, this classification is not always warranted. There are two classes of compounds for which gold surfaces have extremely high affinity—organophosphorous and organosulfur compounds. Other types of molecules may adsorb (physisorb) to gold surfaces, but with few exceptions, only phosphorous- and sulfur-containing molecules chemically bond to gold surfaces. While these types of molecules bond to gold surfaces, the interaction is still relatively weak, and

chemically bound organophosphorous and organosulfur compounds can be readily removed by gentle heating.[1]

Technical Approach

In many ways, gold nanocrystal compounds possess the same chemical properties of bulk gold but have added features that make them an attractive choice for use in sensors and collectors. In nanocrystalline form, the surface area is enhanced enormously. For example, using crystallographic data obtained on gold:thiolate cluster compounds with 145-atom cores,[2] a 1-cm × 1-cm × 1-μm film of nanocrystals has an equivalent gold surface area of ~1.5 m^2. In addition to the straightforward increase in surface area, nanocrystals have an increase in the number of edge and step atoms, which may also enhance the reactivity of the gold because surface reactions typically occur at faster rates along edge and step atoms. Gold nanocrystal compounds also possess special optical properties, which are inherent to not only the gold core-size, but the type of molecule attached to the surface of the nanocrystal (i.e., the environment of the cluster). The optical properties of gold nanocrystals with thio- and phospho-groups attached to identical gold cores have been studied previously and these show stark differences.[3,4] In addition, the ligands attached to gold nanocrystals can undergo facile exchange in solution with either organophosphorous or organosulfur compounds.[5,6] In these studies, the selective org-P and org-S chemistry of gold is combined with the enhanced surface area (and presumably reactivity) and optical properties inherent to nanocrystal compounds to provide a robust detection and collection technology of sufficient simplicity to be deployed.

This project is designed to provide the proof-of-principle experiments needed to expand programs in chemical warfare agent detection and nonproliferation. Relatively simple experiments will indicate if the gold nanocrystal can be used as a sensor/detector for org-P, org-S, and cyanide by measuring changes in optical properties upon exposure to surrogate molecules under various conditions. The gold nanocrystals will be investigated for detection of various toxins in different environments: (i) toxins in solid form, (ii) toxins in vapor form, and (iii) toxins dissolved in liquid media. To address the collection/concentration and sampling issue, gold nanocrystals will be exposed to CWA surrogates and then surface-bound molecules will be identified by thermal desorption and electron impact mass spectrometry.

Gold nanocrystal compounds with both hydrophilic and hydrophobic ligands were synthesized using established methods. For solution-phase detection and collection/concentration, the nanocrystal compounds were used as obtained from the synthesis. Reactions were performed in solutions native to the nanocrystals (e.g., hydrophilic nanocrystals were exposed to aqueous solutions containing CWA surrogates).

For vapor-phase detection and collection/concentration, it is likely that passivating thiolate molecule will hinder or even prevent the adsorption of other org-P and org-S molecules. For this reason, bare gold nanocrystals are desired in high density in porous material. To achieve such a structure, the gold was supported on titanium dioxide powders. Using methods developed for catalysts by Iwasawa and coworkers, gold glutathione nanocrystal compounds were co-precipitated with TiO_2/$Ti(OH)_4$ nanocrystals. Heating the precipitated powders under air to 400°C causes both dehydration of the TiO_2/$Ti(OH)_4$ crystallites and desorption and/or decomposition of the thiolate ligands attached to the gold nanocrystals. While preparation of these structures for catalytic purposes usually concentrates on the minimum loading possible for catalytic activity, in these studies, the maximum loading was desired so that changes in the optical properties upon exposure to surrogate molecules could be easily determined and the maximum amount of material concentrated on the gold surface. Co-precipitates were generated at different Au-loadings. Upon heating, it was easy to determine if the loading is too high because a color change from orange to red occurs, indicating sintering of the individual nanocrystals to form large colloids.

Results and Accomplishments

Solution-Phase Detection of Toxins

One of the goals of this project was to provide baseline detection limits for detection of various toxins that are found either in a battlefield environment or in the environment due to terrorist activity or other nefarious processes. Initial work concentrated on the detection of three compounds that are either found under these conditions or surrogates of those compounds: cyanide, diazinon, and chlorpyrofos.

In addition to the results for organosulfur and organophosphorous compounds (shown below), the sensitivity for cyanide in solution was found to be comparable to that of existing technologies. Figure 2a shows the change in optical absorption that occurs when solutions of gold nanocrystal compounds are in the presence of potassium cyanide. Spectroscopically, the absorbance in the visible region is depleted and absorbance in the ultraviolet region of the spectrum is enhanced. Visually, the color of the solution changes from orange/brown to colorless. The detection limit for cyanide using gold nanocrystal compounds under these conditions is approximated [see Fig. 2(b)] to be between 10–100 ppb in solution.

Diazinon and chlorpyrofos constitute simulates for nerve agents due to their chemical structures (see Fig. 1).

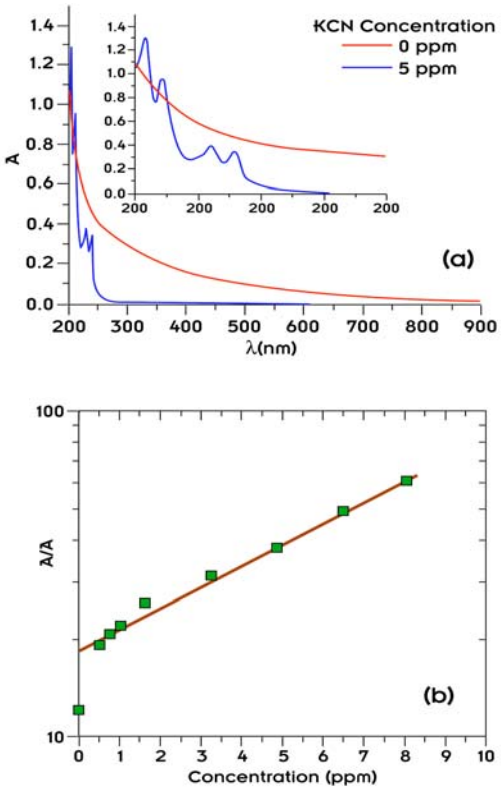


Fig. 2. (a) Optical absorption spectra of gold nanocrystals in water (dotted line) and solution after exposure to 5 ppm CN^- solution (solid line). (b) Ratio of absorbance measurements at 210 nm (UV) and 675 nm (visible) showing the dependence of optical properties on cyanide concentration.

It was found that detection of these types of compounds could not be achieved under strict solution conditions (i.e., using experiments similar to those for detecting cyanide). All solutions tested (formed in polar and nonpolar solvents) retained their visual appearance as well as spectroscopic properties even when these two compounds were added at high concentrations (0.01 M). It is likely the bonding associated with the gold-thiolate nanocrystal compounds is too strong for the displacement of the thiolate ligands by the diazinon or chlorpyrofos molecules, which would be required to measure a change in optical properties.

To overcome this barrier, the gold nanocrystals were deposited on a titanium dioxide support and heated to 250°C, which (i) drives off the existing thiolate ligands and (ii) does not cause an extreme amount of aggregation of the nanocrystals. This was evident from the color of the composite TiO_2/Au films, which retained its characteristic color even after heating. To determine if these molecules could induce a change in optical properties of the composite film, 1–3 μL of diazinon or chlorpyrofos solutions were delivered by a micropipet. Figure 3(a) shows a photograph of a film after delivering 2 μL of 0.001 M diazinon in methanol. The region to the right of the image corresponds to the area where the diazinon solution was delivered. It is clear that the color change occurred only at the region of solution delivery. Blank solutions were also delivered by similar methods, and no color change was measured in the films. Figure 3(b) shows the contrast provided between the area of film not affected by diazinon and the area of the film where the diazinon depleted the color (of the gold nanocrystals).

Vapor Detection/Concentration

A second goal was the concentrating and subsequent detection of organosulfur and organophosphorous compounds by adsorbing these species to gold nanocrystals and using thermal desorption electron impact mass spectrometry. For vapor-phase detection and collection/concentration, it is likely that passivating thiolate molecule will hinder or even prevent the adsorption of other org-P and org-S molecules. For this

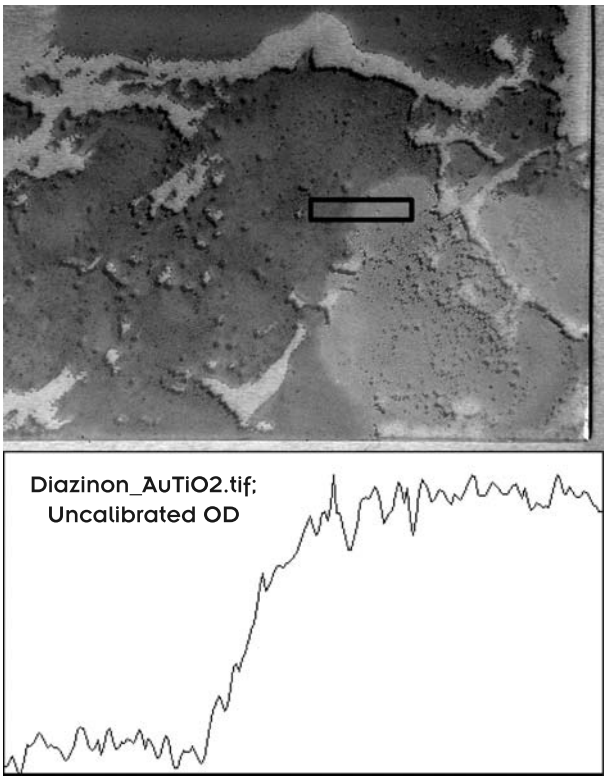


Fig. 3. Photograph of film generated by depositing gold nanocrystals on titanium dioxide and heating for 4 h at 400°C after exposure to 10 μL of 1 ppm diazinon in methylene chloride. The darker area corresponds to the Au/TiO_2 film that was not exposed to diazinon. Note the loss in color of the section of film exposed to diazinon. Total diazinon deposited on film was 12 ng. The optical density plot shows the change in optical density (not calibrated) averaged across the box indicated in the upper image.

reason, the same TiO_2/Au composite films were used for testing the ability to concentrate and detect the presence of toxins in vapors.

After heating the TiO_2/Au composite films on glass coverslips, they were placed into a thin-layer-chromatography developing chamber. Two experiments were performed on these composite films. In the first, one or two drops of volatile toxin was placed into the bottom of the chamber and allowed to evaporate. Thus, a dilute vapor simply reached equilibrium conditions within the chamber and was allowed to passively react with the composite films. Under these conditions, the presence of the toxins (diazinon and chlorpyrofos) was not detected by mass spectrometry.

A second approach was attempted following the inability to detect the molecules under passive conditions. A nebulizer was used to generate a more concentrated vapor form of the two toxins. This was only attempted once because the amount of toxin consumed under this type measurement precluded extensive studies due to cost, cleanup, and waste management considerations. However, under these conditions, the concentrated vapor allowed for detection of the toxins by the TiO_2/Au films. Figure 4 shows the chemical ionization mass spectrum from heating the films to 170°C. The ion corresponding to the $(M+H)^+$ of diazinon is at m/z 305.

Summary and Conclusions

As demonstrated through the detection of the selected toxins under different conditions by these experiments, a gold nanocrystal–based technology holds promise for a field-deployable detection, collection scheme. The goal of these experiments was to provide baseline information regarding the detection limits afforded by this type of technology. In the solution phase, cyanide can be detected at levels comparable to current technology by measuring the changes in optical properties of the nanocrystals. However, other toxins (e.g., other organosulfur and organophosphorous compounds) did not appear to displace the thiols or react in such a way to change the optical properties under solution conditions. Though when the thiols were removed from the nanocrystal cores and composite TiO_2/Au films were generated, the presence of diazinon did cause a change in the optical properties of the composite films. As for the vapor-phase detection, detection limits can only be said to lie between that of the passive vapor and condensation from a saturated vapor. The narrow scope of this proposal (to provide proof of principle for this technology) precludes the ability to make detailed measurements for this system. However, the main goal has been reached for this overarching theme—a simple detection technology that relies on color changes inherent to gold nanocrystals under different conditions.

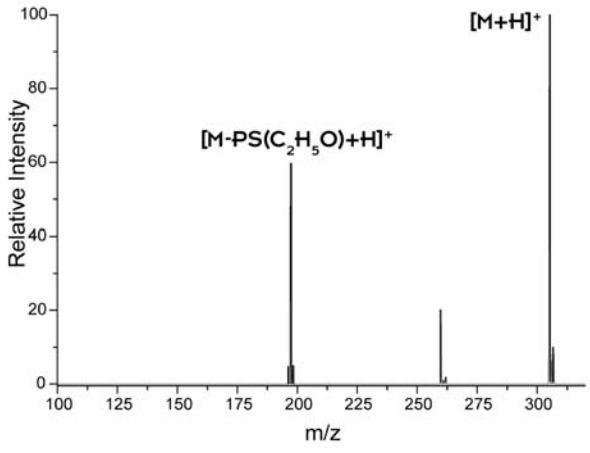


Fig. 4. Chemical ionization mass spectrum from heating the TiO_2/Au composite film to 170°C after exposure to a saturated diazinon vapor. The peak at m/z 305 corresponds to the $(M+H)^+$ ion of diazinon.

Currently, ORNL has elected to pursue a patent on this type of technology for sensing and collecting toxic chemical under typical environment and wartime settings. This technology, if fully developed, would greatly benefit both civilian and military detection schemes for toxins because it inherently targets the most toxic of chemical substances (organophosphorous, organosulfur, and cyanide).

Results from these proof-of-principle experiments indicate that gold nanocrystal compounds may be used in specific settings for the detection or monitoring of toxic chemicals. The detection and monitoring of toxic chemicals is not only of interest to defense-related agencies (e.g., Department of Defense) but is of increasing interest to a number of other agencies mandated to protect the environment and public health and safety (e.g., Environmental Protection Agency, Department of Health and Human Services, and Department of Homeland Security). Collaborating with Thomas Thundat, we plan to incorporate these nanocrystal compounds into cantilever sensor technologies to provide improved detection limits above what is currently capable for cantilever-based sensors. To further this nanocrystal/cantilever technology, we will approach DARPA, DOD, and DHS for follow-on funding, though we anticipate other agencies (e.g., EPA and DHHS) will have requests for proposals in areas where the nanocrystal/cantilever technology will have direct application.

References

[1]R. H. Terrill, T. A. Postlethwaite, C.-H. Chen, C.-D. Poon, A. Terzis, A. Chen, J. E. Hutchison, M. R. Clark, G. Wignall, J. D. Londono, R. Superfine, M. Falvo, Jr., E. T. Samulski, and R. W. Murray, *J. Am. Chem. Soc.***117**, 12537–12548 (1995).

[2]T. G. Schaaff, M. N. Shafigullin, J. T. Khoury, I. Vezmar, and R. L. Whetten, *J. Phys. Chem. B* **105**, 8785–8796 (2001).

[3]T. G. Schaaff, M. N. Shafigullin, J. T. Khoury, I. Vezmar, R. L. Whetten, W. G. Cullen, P. N. First, C. GutierrezWing, J. Ascensio, and M. J. JoseYacaman, *J. Phys. Chem. B* **101**, 7885–7891 (1997).

[4]E. Gutierrez, R. D. Powell, F. R. Furuya, J. F. Hainfeld, T. G. Schaaff, M. N. Shafigullin, P. W. Stephens, and R. L. Whetten, *Eur. Phys. J. D* **9**, 647–651 (1999).

[5]L. O. Brown and J. E. Hutchison, *J. Am. Chem. Soc.* **119**, 12384–12385 (1997).

[6]M. J. Hostetler, J. E. Wingate, C.-J. Zhong, J. E. Harris, R. W. Vachet, M. R. Clark, J. D. Londono, S. J. Green, J. J. Stokes, G. D. Wignall, G. L. Glish, M. D. Porter, N. D. Evens, and R. W. Murray, *Langmuir* **14**, 17–30 (1998).

Controlling Size and Function of Metal Oxide Nanoparticles: Coupling Micellar Nanoreactor Synthesis and Hydrothermal Processing

D. B. Beach, A. J. Rondinone, P. Benezeth-Gisquet, D. A. Palmer, and D. J. Wesolowski
Chemical Sciences Division

Oxide nanoparticles of well-defined sizes and properties are critical to applications such as nonlinear optics, communications, magnetics, and electronics. This project has produced complex, multicomponent oxide nanoparticles with high degrees of purity, crystallinity, and uniformity of particle size and shape, using a combination of micelle-confined synthesis and hydrothermal post-processing. Hydrothermal processing serves two purposes: it removes organic surfactants and provides a means for crystallization of the oxide nanoparticles without the undesirable agglomeration, growth, and sintering effects of dry-firing.

The predominant hindrance to conducting research and development of oxide nanoparticles is the lack of means through which to crystallize the particles. The properties of oxide materials are highly dependent on crystallinity; without proper crystallinity, the usefulness of oxide nanoparticles cannot be realized. However traditional means of crystallization, such as dry firing, are not available to nanoparticle research because unavoidable aggregation and sintering destroy the nanoparticle morphology. Hence we are attempting to develop a novel method for crystallization that could be a general replacement to dry firing, thereby allowing oxide nanoparticle research and development to move forward.

A small variety of metal oxide nanoparticles have been synthesized using our multistep hydrothermal approach, yielding highly crystalline particles at very small sizes (<10 nm). First, amorphous precursor nanoparticles were synthesized in micellar nanoreactors using sol-gel chemistry. The micellar nanoreactors define and control the nanoparticle size and size distribution. The amorphous particles were then transferred to a hydrothermal autoclave for crystallization. It is important to note that these two steps are discrete and represent a new approach to what has traditionally been attempted as a one-pot process. The technique has been successfully applied to zirconia, titania, and doped and undoped yttrium aluminum oxide nanoparticles.

A large chunk of the project time has been devoted to developing the necessary chemistry to achieve the goal of the project. Removing high-surface-area nanoparticles from a completely hydrophobic environment (micellar nanoreactor), while in the presence of a surfactant, into a completely aqueous environment (for hydrothermal treatment) is not trivial. The nanoparticles must remain completely stable in solution so that irreversible aggregation does not occur. Published chemistry for this step was not available and had to be developed.

While the project has demonstrated feasibility with three nanoparticle oxide systems, focus has been applied to the development of one particularly interesting system, yttrium aluminum oxide. The authors set out to develop a nanoscale version of yttrium aluminum garnet (YAG), a material used in numerous optical applications, but instead have repeatedly synthesized an oxide material with identical stoichiometry whose structure is yet to be determined. The properties of the YAG-like nanoparticle are similar to those of bulk YAG but are different in fundamental ways. For example, the fluorescence lifetime of the nanoparticles doped with fluorescent rare-earth ions is an order of magnitude shorter than the bulk YAG material. Furthermore, the lifetime is temperature dependent at room temperature rather than 500°C as with the bulk. This makes the nanoparticle useful as a remote temperature probe for biological and other systems.

Bulk YAG has a lattice length of 12 Å. These particles have a diameter of 50 Å (see Fig. 1). It is likely that the limit of four or so crystalline units across the diameter of a particle prevents the particle from having the same crystal structure as the bulk material. As the diffraction pattern from these particles cannot be indexed to any known material, it is likely that this is a crystal structure that is not accessible to the bulk material. The goal of synthesizing nanoscale YAG was not achieved; however, a novel material with novel and useful properties was synthesized. Ultimately, novel materials are the goal of nanotechnology research and this project has furthered that goal.

The yttrium aluminum oxide nanoparticles described above may be used as clandestine materials tags for

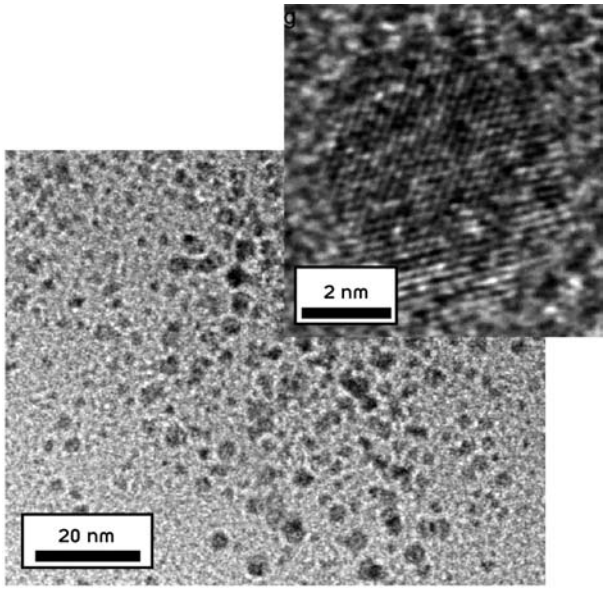


Fig. 1. Cerium-doped yttrium aluminum oxide nanoparticles.

national security and commercial applications. Other oxides such as zirconia are useful as catalysts or catalytic supports. As a result, our synthetic technique as well as the particles themselves have been elected for patent by ORNL.

The authors are also seeking further funding through the offices of the Department of Defense as there are several oxide nanoparticles that the military may find useful. The authors have also proposed to set up a hydrothermal user facility within the upcoming Center for Nanoscale Materials Science.

Metallic Nanofuels for Vehicles

B. G. Sumpter,[1] S. D. Labinov,[2] D. W. Noid,[1] and D. B. Beach[3]
[1]*Computer Science and Mathematics Division*
[2]*Engineering Science and Technology Division*
[3]*Chemical Sciences Division*

This proposal addresses the concept of novel, renewable, high-energy density, non-emission fuels for vehicles based on clusters of metallic nanoparticles and experimental proof of the outstanding characteristics of the combustion these fuels. It is expected that, due to the high surface-to-volume ratio typical of nano particles, complete and fast oxidation (combustion) will take place at temperatures under 1400–1600°C in the condensed phase. This will serve to keep the combustion products on a carrier and avoid pollution of the air by nitrogen oxides. Owing to the exceptionally high energy density, this novel fuel will be very efficient as an energy source for military vehicles, heavy trucks, locomotives, etc. The combustion of fuel clusters and their thermodynamic and thermochemical characteristics will be experimentally explored in order to prove the validity of the proposed concept. ORNL is able to develop this new direction by employing the joint efforts of specialists in chemistry, solid-state physics, energy, transportation, computation, and mathematics.

The problem addressed by this proposal is that of an alternative energy supply system for vehicles that is renewable, nonpolluting, and high in energy density. Limited reserves of liquid fossil fuel, security issues due to foreign oil dependence, and the air pollution caused by the burning of fossil fuels are the major drivers for the search for alternative fuels for transportation. We propose the use of engineered clusters of metallic nano particles as a fuel for vehicles. When compared to conventional liquid fuel, metallic fuel features greater energy capacity per volumetric unit, low flammability, and is virtually explosion proof. (The ignition temperature of bulk material is more than 2000°C in open air.) Its storage does not require any special conditions. Metals useful as fuels and their oxides are not toxic and do not pollute the environment. With the same size fuel tank, a vehicle using iron as a fuel will cover without refueling a distance two times greater than that covered by a car using gasoline. With aluminum or boron as fuels, the distances would be three and five times greater than gasoline! The combustion products of metals are solid oxides that can easily be stored and transported to reduction facilities.

The following problems need to be solved in order to demonstrate the capabilities of a nano fuel–based on metallic clusters: (1) the control of the metallic fuel combustion process; (2) the quick decrease or increase of released heat quantity depending on engine load, and (3) systems providing for fuel delivery to a combustion chamber and ash removal. Our proposed research addresses these problems by using a fuel cluster composed of metallic nano particles, delivered using a thermo-resistant thin-film carrier or through air flow.

In our approach, nano particles will be arranged in fuel clusters. Each cluster includes a number of nano particles that is appropriate for the elementary heat output in a combustion chamber. By arranging clusters on a carrier and activating them as necessary, heat release can be regulated to give the output required by engine loading. Proof-of-concept experiments will be carried out and be considered successful if it allows establishing the following:

1. A fuel cluster consisting of certain number of metallic nano particles and covered with thin metal oxide film can be manufactured with determined configuration and have mechanical strength sufficient to sustain transport and storage.
2. A fuel cluster can be ignited in air under given temperature and pressure with the help of a laser, electrical or electromagnetic short-time pulse, and the energy expenditure to create the pulse is less than 1 % of the energy released by burning the cluster.
3. Metal nano particles composing a cluster are oxidized by the atmosphere oxygen for the time no longer than 10^{-2} sec, with the maximum combustion temperature no higher than 1500°C and without forming volatile components.
4. After a fuel cluster is burnt, it preserves its configuration and strong bond with the carrier (internal combustion versions).

The project has just started (August 2003), but some progress has been achieved. To date, an experimental apparatus for examining the peak combustion temperature of an engineered cluster of metallic iron nano particles has been designed. Following combustion of the cluster (to be performed in FY 2004), the cluster can be removed and examined via microscopy to determine if the combustion caused significant amounts of volatile products instead of occurring mainly in the solid/condensed phase. The approximate fuel cluster size to achieve the necessary heat flow has been computed and a method for experimentally producing such a cluster composed of iron nano particles (the iron nano particles are synthesized using iron carbonyl in a solution of and then compressed into a small cavity followed by heating to produce a cluster with controllable size) has been developed.

The success of this work will lay the foundation for developing the technology for an alternative fuel production and its practical use. It has direct impact on both energy production and environmental quality, two of DOE's main missions. This type of nano fuel will provide an additional fuel source with a high energy density that combusts with near-zero emission. It can be expected that military, space, and specialized vehicles and ships would be the first to make use of such a metallic fuel. (For example, fast burning metal nano particles could be used to boost the output of a fighter plane's afterburner with significant reduction in fuel weight.) Passenger cars, trucks and locomotives converting to metallic fuel will occur gradually.

ENGINEERING SCIENCE AND TECHNOLOGY

Director's R&D Fund

Nanoscale, Explosive Energy-Burst Generators Using Controlled Nuclear-Mechanical Triggering of Pretensioned Liquids

R. P. Taleyarkhan,[1] C. D. West,[2] S. Hicks,[1] and J. S. Cho[3]
[1]*Engineering Science and Technology Division*
[2]*Physical Sciences Directorate (retd.)*
[3]*Oak Ridge Associated Universities*

This research focused on developing insights into a revolutionary class of explosive mechanical burst energy generators by tapping key physical attributes of pretensioned liquids which are fractured/triggered on the nanoscale (time and space) for explosive energy releases thereafter. Controlled triggering was achieved using neutron science technology in combination with pulsed lasers. The technology's potential was evaluated using theoretical analyses coupled with focused experimentation and was found to be promising. Estimates of energetics from experimental evidence indicate significantly larger quantities of energy-storage capability compared with conventional propellants. A table-top demonstration system was designed, fabricated, and tested.

Introduction

The goal of the project was to investigate the potential for energy storage in tensioned liquid metastable states followed by controlled release of the stored energy. The project workscope utilized neutron science, engineering expertise in acoustic studies for liquids and highly energetic materials, and life-science health effects assessment skills developed over the past several decades in various ORNL directorates.

It has been known for some time (although not widely recognized) that fluids, like solids, can be put under negative pressure (i.e., placed under tension). Upon reaching an appropriate state of instability, liquids can explode and release their stored energy. Such states can be reached via pretensioning or via suitable combination of pretension and thermal superheat. Experiments for determination of thresholds for fracture of pretensioned fluids under various states of treatment have been conducted[1] at ORNL and elsewhere.[2–7] The maximum possible extent of pretensioning prior to onset of cavitation for a properly prepared liquid such as water and mercury can be quite large and can reach close to -1400 bar (-20,000 psi) and -17,000 bar (-250,000 psi), respectively.[2–6] These tensile states are metastable and may be viewed for analogy as placing a spring under intense tension prior to snapping and explosive vaporization. The triggering of metastable fluids gives rise to explosive vaporization and release of stored energy which may be permitted to grow uncontrollably or induced to collapse to produce very high localized pressures and temperatures. This process, if conducted with suitable fluids, can possess the capability of producing enormous energy density (factors of 10^6 greater than that from CHNO compounds like TNT). Indeed, the controlled nuclear-based triggering of metastable fluids has recently been shown[8] by the authors to be capable of inducing nuclear fusion reactions between deuterium atoms during cavitation of deuterated acetone.

Technical Approach

The approach taken included the study and analysis of the literature on metastable fluid states induced via thermal superheat under positive pressure regimes and also by placing liquids under tension. The second phase of the project involved designing and fabricating a tensioned-metastability-inducing system to demonstrate nanoscale (space-time) triggering of the tensioned states in a variety of liquids with subsequent energy releases to propel an object.

Results and Accomplishments

A systematic study was undertaken to develop theoretical understanding of the various facets involved in energy storage during tensioned metastability and the nanoscale-to-macroscale triggering of energy bursts.

The nanoscale triggering of pretensioned liquids can be attained by subjecting individual nuclei of liquid molecules to knock-on collisions with energetic particles such as neutrons or the recoils from alpha particle decays. The collision of a high-energy (e.g., 14-MeV) neutron with an individual nucleus results in recoil-induced deposition of thermal energy in a nanometer-scale region of the targeted liquid. If this energy level is sufficient to cause a vapor nucleus to form with a radius larger than a certain

critical size of a vapor cavity, the cavity can then grow. Critical radii are usually in the nanometer range and are formed within nanoseconds. For example, for an organic liquid such as C_3H_6O, we calculate the critical radius to be about 28 nm with nucleation from dissolved alpha emitters or the knock-ons from energetic neutrons. Note that the extent (or quantity) of triggered formation of critical vapor pockets is dependent on the state of the fluid. The greater the pretensioned or metastable state, the lower is the initiation energy required. The closer the state is to that for homogeneous-explosive transformation, the larger will be the inherent, statistically induced fluctuations amenable to explosive vaporization. This will lower the initiation energy requirements. Figure 1 shows the experimentally determined variation of the normalized homogeneous or spontaneous-nucleation temperature with normalized pressure of the fluid field. Normalization is done relative to the critical point. Note that the data on which this plot is based involves extrapolation below the normalized pressure value of ~-1. This is based on the fact that the data above -l have been found experimentally[2,9] to follow a linear profile; therefore, the linear extrapolation into as yet uncharted territory is considered a reasonable initial engineering judgment. As will be noted, the greater the degree of pretension, the lower the stability limit and the easier it becomes to destabilize the entire system to explode. At a normalized negative pressure of -7 (i.e., -1400 bar) for water, the normalized temperature is only ~0.1 to 0.14 (i.e., ~30–50°C).

Triggering of explosive-burst generation can also be conducted via use of pulsed lasers. In this case, even if the metastable state is far from the stability limit, localized explosive-burst generation can be attained. Bulk triggering of metastable fluids can be attained via increasing the flux of incident neutrons, via use of laser (thermal) energy deposition, as also via use of acoustic energy that reduces the pressure to the limiting tensile state (as demonstrated for ether at ~120°C by Apfel[2] by reducing the pressure to ~–15 bar, at which point the conditions for system destabilization were met and the droplet exploded). As may be obvious by now, if the combination of pretension and temperature are at or close to the stability limits, bulk vaporization should be possible to initiate via intentional destabilization of a small region of the fluid space which propagates through the bulk fluid.

The means of storing energy in the fluid can be varied. The simplest device of storing and maintaining pretension is a centrifuge-type apparatus shown (for a specific configuration) in Fig. 2. The fluid to be explosively vaporized is placed in the central cavity, after which the centrifugal forces introduce the liquid to the desired degree of pretension. The degree of pretension will vary quadratically with the arm length of the centrifuge apparatus and rotation speed. As such, significant metastable states may be attained with a reasonable choice of parameters (e.g., -1000 bar with a length of 0.5 m and a spin rate of ~150 Hz). Other methods for introduction of metastable states in tension include acoustic means[2,7,8] and the Berthelot[3] method (with which the theoretical thresholds for pretension of water have been attained[6]). The use of acoustic means provides the capability to vary the frequency of tensile state attainment over a very wide range (e.g., up to tens of kHz). For our sonofusion studies,[8] we employed a chamber where ~15-bar (~250-psi) oscillating pressures were induced. We have also worked with multifrequency excitation[10] in which the transient pressure field can be composed of more than one pressure

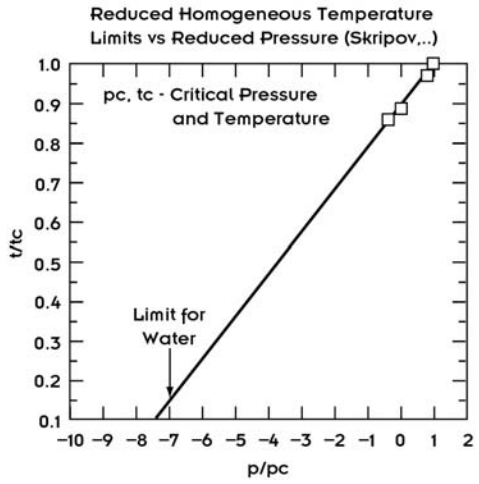


Fig. 1. Variation of normalized homogeneous nucleation temperature with normalized pressure (normalization done using critical state parameters).

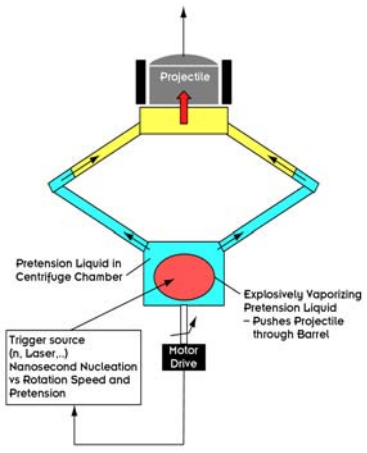


Fig. 2. Schematic arrangement of spinner experimental setup for determining energetics of Laser (nanosecond duration 0.3 mJ peak energy pulse) triggered explosive vaporization [arm length ~240 mm; liquid—acetone at 20°C; Radius of glass tubes ~2 mm (internal); Projectile mass ~0.5 g; Barrel length ~5 mm].

component. Such an acoustically excited system can possess the ability to store very large amounts of energy that can be released on demand (i.e., on triggering, which is something we have done successfully via pulsed-neutron-generator triggering). Transformation factors of output to input power have been observed experimentally by us to be in the several 10^3 range. The induction of extreme states of metastability requires careful preparation of the test liquid and its enclosure. Purity and degassing play an important role;[4,5] however, our studies indicate that prior precompression (to induce close-to-perfect wetting, and dissolve microbubbles) can play a far more effective role (as is clear when one analyzes the methods used and results obtained from tests by several experimenters[2,6]).

Energy is stored in a metastable fluid that, upon release in a controlled fashion, may then be used to deliver mechanical output to projectiles or surrounding structures. For pretensioned fluids, it is important to realize that the compressibility and specific heat capacity increase[11,12] as the degree of pretension increases, rising theoretically to infinity at the stability limit. Therefore, the amount of stored energy can become quite large, increasing nonlinearly with progressively enhanced states of metastability. An indirect method for estimating the work energy potential is to conduct experiments for bulk vaporization and localized explosive vaporization. Experiments[2] conducted with ether at ~120°C in an oscillating pressure field produced a relative volume change of ~10,000 to 20,000 (i.e., from liquid to vapor) over a millisecond of expansion. Conducting the calculation for *pdV*-type work, we estimated energy output of ~1–2 kJ/g of vaporizing material under positive pressure for conditions close to the limits of (thermal-superheat-induced) metastability. This is very significant when we realize that the work output from propellants such as nitrocellulose is only ~1.2 kJ/g.

An alternate experiment conducted involved introduction of metastable pretension in test liquids such as acetone, ethanol, etc., using a centrifugal system as shown in Fig. 2. As an example, for one case acetone was placed under pretension via spinning the apparatus to attain ~-5bar of negative pressure. Thereafter, a nanosecond timed laser pulse from a nitrogen laser caused explosive triggering at the focus point for a 1-mm-diameter region of the test liquid. The burst vaporization even at this relatively modest negative pressure (under room temperature conditions) resulted in the projectile being ejected at ~40 m/s (~150 fps) from which the *pdV* work output due to vaporization was estimated as ~0.6 to 0.7 kJ/g. While low compared to the ~1 kJ/g normally obtained from propellants such as nitrocellulose, it should be realized that this work output was obtained at only a small fraction of the theoretical pretension threshold. It is not known at present how the process will scale with various levels of metastability and other process parameters.

Summary and Conclusions

In summary, the combination of metastable states and appropriate triggering can offer the possibilities of nanoscale-to-macroscale triggering-induced explosive work output for a wide range of fluids. Various means ranging from centrifugal spinners to acoustic generators can be employed together with thermal management to induce desired states of metastability, which can then be destabilized explosively using nuclear to macroscale triggering techniques. The energy storage and work output potential from such releases can be significant and appear to be controllable. This promises to open up exciting possibilities in the fields of propulsion and possibly for nuclear fusion energy generation.

As a result of this project, the lead author was invited to present a keynote lecture entitled "Nano-to-Macroscale Triggering of Metastable Liquids"[13] at an international conference. We were also invited by the U.S. Army and several senior Pentagon officials (including the Vice-Chief of Staff Gen. J. Keane) to draft a white paper for planning the *Army of the Future* to potentially utilize this revolutionary technology. An invention disclosure was prepared and the technology was elected for filing with the U.S. Patent Office. An ORNL/TM report has been prepared.[14]

References

[1]R. P. Taleyarkhan and F. Moraga, *Nucl. Eng. Design* **27** (2001).

[2]R. Apfel, *Scientific American* **227**, 6 (1972).

[3]M. Berthelot, *Ann. De Chimie et de Physique*, Tome (XXX).

[4]L. Briggs, *Science* **109** (1949).

[5]M. Greenspan and C. Tscheig, *J. Res. Natl Bureau of Standards* **71C** (1967).

[6]Q. Zheng, D. J. Durben, and G. H. Wolf, *Science* **249** (1991).

[7]C. D. West and R. Howlett, *Acustica* **21** (1968).

[8]R. P. Taleyarkhan, C. D. West, J. S. Cho, R. T. Lahey, Jr., R. I. Nigmatulin, and R. C. Block, *Science* **295** (March 8, 2002).

[9]V. P. Skripov, *Metastable Liquids*, J. Wiley, New York (1974).

[10]F. Moraga, R. P. Taleyarkhan, R. T. Lahey, Jr., and F. J. Bonetto, *Phys. Rev. Lett.* **62** (2000).

[11]A. Hayward, *Nature* **202**(4931), 1964.

[12]R. J. Speedy, *Journal of Physical Chemistry* **86** (1982).

[13]R. P. Taleyarkhan, C. D. West, and J. S. Cho, *The Third International Conference on Transport Phenomena in Multiphase Systems (HEAT-2002), Kielce, Poland, June 24–27, 2002.*

[14]R. P. Taleyarkhan, C. D. West, and J. S. Cho, ORNL/TM-2022/233, November 2002.

Remote Emission Sensor Technology for Heavy-Duty Truck Emissions

R. N. McGill,[1] K. E. Lenox,[1] J. M. Storey,[1] A. Akerman,[1] C. W. Ayers,[1] H. D. Haynes,[1] T. Q. Dam,[1]
R. K. Richards,[1] S. Goedeke,[1] M. L. Simpson,[1] M.-D. Cheng,[2] E. A. Wachter,[3] and W. G. Fisher[3]
[1]*Engineering Science and Technology Division*
[2]*Environmental Sciences Division*
[3]*Galt Technologies, LLC*

The objective of this project is to develop low-cost, fast-responding, non-intrusive instrumentation for measuring certain emissions of heavy trucks as they pass on a highway. Prototype optical instrumentation has been developed for remotely measuring concentrations of nitrogen oxides (NO_x) and particulate matter (PM) in the exhaust plumes of heavy trucks. It is also important to acquire certain parameters about the truck's operating mode at the time emissions data are acquired. To this end, additional technologies were developed to analyze the acoustic signatures of trucks as they pass. From the acoustic signature, analysis data on the engine speed, turbine speed, and truck speed and acceleration can be developed. The three prototype instruments (including acoustic signature analysis software) have been tested against standard measures of exhaust species, and results compare favorably. They have also been validated in controlled, outdoor field tests using three different sizes of vehicles.

Introduction

Heavy-duty trucks account for large portions of the mobile-source NO_x and PM emissions in our urban environments, but truck emissions in the real world are not well characterized. It is difficult to attribute the emissions of a truck to its different modes of operation because truck engines, not the vehicles themselves, are certified for meeting emissions standards. In order to acquire large databases of in-use truck emissions, we must have access to emissions instruments that can acquire "snapshots" of a truck's emissions as it passes on a highway; that is, remote and non-intrusive instrumentation must be available. However, satisfactory instrumentation of that type has not been available to date. Therefore, to meet the need, this project has been focused on development of such instruments.

Technical Approach

We chose optical methods to achieve remote and non-intrusive measurements for NO_x and PM emissions. The NO_x instrument is based on an application of ultraviolet (UV) absorption, while the PM instrument is a greatly modified light distancing and ranging (LIDAR) approach. While these instruments are expected to yield concentrations of the pollutants in the exhaust plume of a truck, those measurements alone are not enough to characterize the emissions in terms of the truck's operation (e.g., grams per mile). To accomplish that goal we need a means to determine certain parameters of the truck's instantaneous operation, such as speed, acceleration, weight, engine speed, and turbine speed. It was thought that these parameters could be deciphered from the acoustic signature of a passing truck. Therefore, the third leg of the project is to develop acoustic signature analysis instrumentation that will provide data on the truck's operating parameters, including engine and turbine speeds, as the truck passes on the road. The technical approaches for each of these three technologies are discussed individually below. In all three cases, the development took similar courses. Bench-scale experiments were first used to test and improve the methodologies. After methodologies were confirmed, experiments moved to the engine laboratory scale, using real engines to generate the exhaust. At the last stage of development, all three instruments were tested in a chassis dynamometer laboratory with a vehicle generating the exhaust. Included in this last stage of development were outdoor tests with selected vehicles moving past the measurement station.

UV Absorption for NO_x Measurement

A double-pass UV instrument was developed for the measurement of nitric oxide (NO) concentrations. A deuterium lamp was used as the UV source to provide a broad continuum. The light from the lamp was directed with quartz optics into an approximately parallel beam from the instrument. This beam was directed through the region to a mirror—a UV-enhanced retro reflector—that returned the light beam back to the UV instrument. The returning light was focused into a spectrometer (HR2000, Ocean Optics) that measured its spectrum from 200 to

300 nm. A laptop computer controlled both the spectrometer and the data using a USB serial line.

Data from the spectrometer were collected and analyzed using software developed with Visual Basic version 6.0. The determination of the NO concentrations was made by first measuring the spectral intensity under conditions without NO present. This provided a reference intensity level. Measurements were then made with NO present, and the differences used to determine the spectral absorption. The NO concentration was determined using the strength of the absorption lines around 226 nm. The spectral background between 220 and 230 nm was used to estimate broadband absorptions and scattering losses due to water vapor, unburned fuel, rain, dust, etc. After correcting for theses other losses of light, the remaining absorption feature around 226 nm was fit to a multiple line shape, which then provided the absorption due to nitric oxide. The 226-nm wavelength was selected because it had no interference from sulfur dioxide or ammonia absorption.[1]

This instrument was calibrated using a 1.6-m cell set in the path of the light beam between the instrument and the retro reflector. A gas flow was maintained through the cell at 1 atmosphere of pressure using a combination of the chemical to be analyzed and nitrogen. A gas mixer was used to adjust the concentration of the chemical under study. The chemicals studied in this way included NO, sulfur dioxide, and ammonia.

Most of the calibration effort involved the study of NO. The concentrations of NO were varied between 0.5 to 1000 parts per million (ppm). The results were fit with a third-order polynomial to determine the concentration based on absorption strengths. The results were found to be reproducible to within approximately 2% with long measurement times (~10 s). Such a long integration time would not be practical for the measurement of effluents from passing vehicles, so the accuracy as a function of integration was examined. For short time periods, the inaccuracies grew as the inverse square root of the integration time. For an integration time of 50 ms, the results were found to be reproducible to within approximately 10%.

FM-CBR LIDAR for PM Measurement

The challenges of measuring a diffuse target, such as truck PM exhaust, require the ability to monitor rapidly varying particulate concentration profiles that can be very small compared to the ambient concentration profiles. Using the superior noise suppression provided by phase-sensitive detection enables the use of low-power, eye-safe, inexpensive light sources such as diode lasers. To achieve these goals, we have developed a hybrid modulation scheme called frequency-modulated coherent-burst ranging (FM-CBR). This combines key aspects of traditional LIDAR systems to achieve sensitive, precise measurement of the distance and concentration of diffuse targets at data bandwidths exceeding 1 kHz. This measurement technique is a variation of CBR, which was first developed at ORNL.

FM-CBR LIDAR uses an interrupted series of high-frequency amplitude modulation pulses encoded on an optical carrier. The use of high modulation frequencies, combined with high data acquisition rates, allows several positions in a passing exhaust plume to be monitored in quick succession. The system is monostatic (i.e. the laser and light collector are all on the same axis), which greatly simplifies alignment without a significant sacrifice of signal (Fig. 1). The laser used is a widely available red laser diode (632 nm) combined with a frequency matched photomultiplier tube for light collection.[2,3,4]

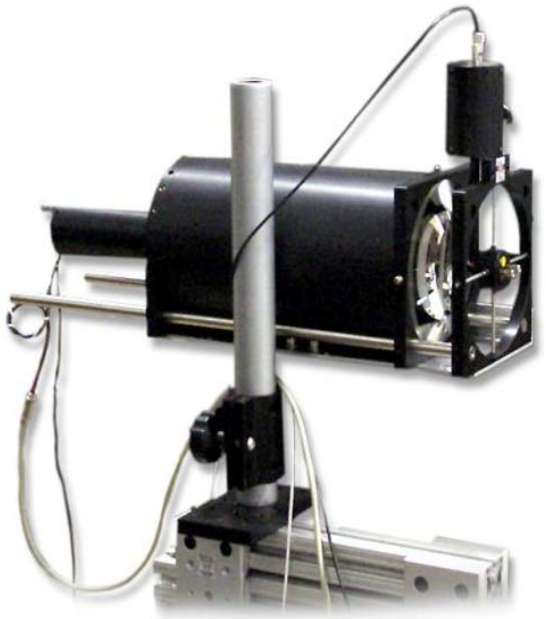


Fig. 1. Monostatic LIDAR system for measuring PM emission concentrations.

Acoustic Signature Analysis for Acquiring Engine Operating Parameters

The challenge in deciphering a sound recording of a passing truck is to efficiently extract the desired information from the sound. In theory it should be possible to determine the truck's engine and turbine speeds from the acoustic signature. These data, together with knowledge of the typical size of a heavy-duty truck engine, will allow the estimation of the exhaust flow rate at the time of the measurement. Knowing the exhaust flow rate, the emissions concentrations from the instruments developed in this project, and the speed and acceleration of the truck, we should be able to closely estimate the emissions of NO_x and PM on a grams-per-mile basis.

Results and Accomplishments

UV Results for NO_x

The goals of this aspect of the project were met with a demonstration of the instrument's capability for measuring several chemicals (NO, sulfur dioxide, and ammonia) and the measurement of NO concentration from the exhaust from a diesel engine. The diesel exhaust measurements were performed with both a stationary vehicle on a dynamometer (see Fig. 2 for results) and with several passing vehicles. Note that the measurement results are in ppm · m since the light beam integrates through the cloud of effluent. For the investigation of the passing vehicle, three vehicles were tested; a Mercedes A Class, a Ford F-250, and a Peterbilt heavy-duty truck tractor. The Peterbilt was also tested on a street location in Chattanooga, Tennessee. Example results from these tests are illustrated in Fig. 3 showing the NO concentration from the Ford F-250.

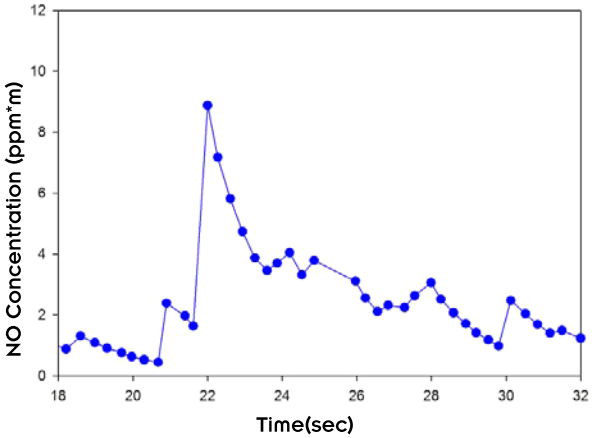


Fig. 3. Measurements of NO concentration from a Ford F-250 diesel passing through the optical beam at approximately 22 seconds. Although the data integration time was 50 ms, the data analysis required several hundred milliseconds yielding only a few data points per second.

FM-CBR LIDAR Results for PM

To confirm that the range and resolution of the LIDAR system were sufficient to meet the measurement goals, a series of experiments were performed using "hard" targets, such as clear sheets of plastic and Delrin®, a highly reflective material. This initial work provided the information needed to refine the hardware and software in order to proceed onto more challenging measurements. After successful completion of these experiments, the focus of the experiments became the investigation of the ability of the LIDAR system to detect particle sizes and size distributions. Thus we moved from hard targets to laboratory-generated aerosols.

In general, aerosols interact with laser light in three ways: they absorb the light, they scatter the light forward, and they scatter the light backward. By knowing the aerosol particle-size distribution, the wavelength of the laser light, and the effective refractive index of the particles, one can derive a theoretical volume backscatter coefficient, β, analogous to the reflectivity of a surface, using Mie theory.[5] The amount of light scattered backward from an aerosol plume can then be calculated from this coefficient. Thus β is the common link between LIDAR measurements and microphysics models of aerosols. As a first step in validating our instrument, experiments were performed with known aerosols, and the measured value of β from the LIDAR measurements was compared to that from Mie theory.

We had mixed results from our experiments with laboratory aerosols. In all cases, the data clearly show a difference between two aerosols and the location of the aerosol streams, the signal to noise ratio is good, and the results are repeatable. Typical results with the "hard" targets and with a highly reflective aerosol are illustrated in Fig. 4.

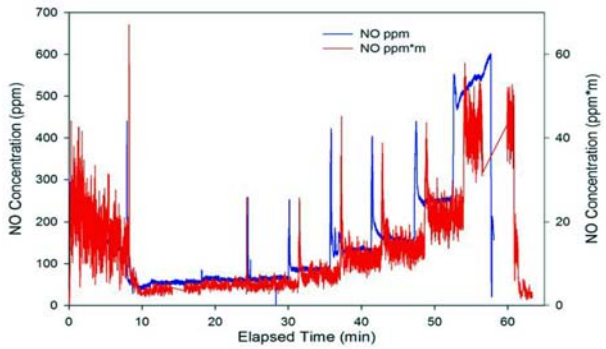


Fig. 2. Comparison of NO measurement techniques. The blue curve is from the heated chemiluminescent detector taking samples in the exhaust tail pipe and the red curve from the UV absorption instrument measuring the concentrations 40 cm after the tail pipe.

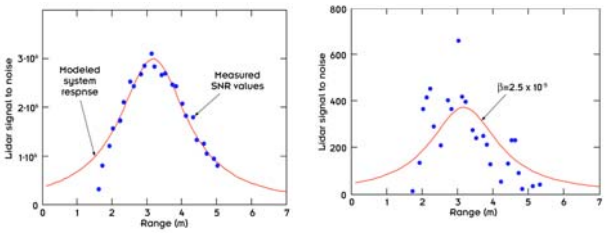


Fig. 4. The measured and modeled FM-CBR LIDAR system response to a hard target made of Delrin (on the left) and a laboratory generated aerosol of ammonium sulfate. Note that the scatter has increased dramatically for the aerosol. However, in the aerosol case the calculated beta value of 2.5E–6 does agree favorably with the theoretical value of 3.23E–6.

The experimental focus moved to real exhaust measurements—first in a chassis lab and then in an outdoors setting. FM-CBR LIDAR and scanning mobility particle sizer (SMPS) measurements were taken of the exhaust of a light-duty diesel vehicle on a chassis dynamometer operating at a range of steady-state conditions. The FM-CBR measurements were repeated with the same vehicle driving by in a parking lot at speeds up to 40 mph. Similar measurements were taken using a full-size diesel truck and a heavy-duty diesel tractor. For the outdoor experiments, no data existed to make a β value comparison, so the signal-to-noise ratio was used to compare the data. The high signal-to-noise ratios, as shown in Fig. 5, indicate that the location and presence of the exhaust plume are being detected by the FM-CBR system.

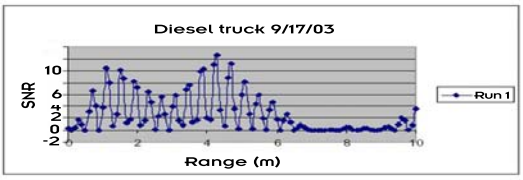


Fig. 5. The FM-CBR system response to exhaust from a moving vehicle.

It is the correlation between backscatter and particle size and distribution that needs refinement.

Acoustic Signature Analysis for Acquiring Parameters of Engine Operation

A typical fast Fourier transform (FFT) shows little or no signal at the fundamental engine speed, but as one would expect, significant harmonics, including strong 3rd and 6th harmonics are present. From this there is sufficient information to determine the engine speed from the acoustic signature. In addition we have identified the signature sound from the turbocharger in the acoustic data. A part of the turbocharger spectrum is shown in Fig. 6. This is a waterfall plot to demonstrate the change in frequency of the turbocharger as the truck goes up a slight grade maintaining a constant engine speed (rpm).

Many techniques have been developed for analysis of time-series data such as a sound signature. We have employed linear techniques to date. Using a method that looks for discrete spectral components, we have written programs in LabView that automatically analyze a data set for engine speed and show the turbocharger frequency as a function of time.

One version of this is shown in Fig. 7. On the left side are two data sets showing signals from trucks going uphill on I-40 at Lovell Road, Knoxville, Tennessee. Subsequent traces are superimposed so that the change in frequency of the turbocharger signal may be clearly seen. Unless the truck is using engine braking, the turbocharger signal is absent when the truck is going downhill. A third program called Truck Acoustic Data Analyzer (TADA) combines these with additional analysis tools that provide an indication of engine load in nearly automatic fashion.

Another consideration in the development of the acoustic signature analysis techniques is the need for the most efficient methods for collecting data along the highway unattended. A two-channel broadband microphone and amplifier was constructed and used for much of our collection, indicating a clear pathway to an unattended system.

Each of the three technology developments included in this project has advanced the body of knowledge in their respective fields. Invention disclosures are being prepared for both the FM-CBR LIDAR technology and the acoustic signature analysis technology. The researchers have produced a number of technical papers and presentations thus far, and we anticipate more of such in the near future. The high-frequency amplitude modulation of the FM-CBR LIDAR along with the phase-sensitive lock-in detection of signal has produced a unique system with large signal-to-noise ratio and capable of rendering the particulate concentration in the beam path in a manner

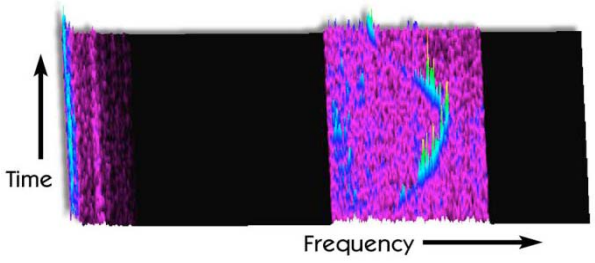


Fig. 6. A "waterfall" plot of engine and turbine speeds as a truck goes up a hill. Engine speed (to the left) remains relatively constant as the truck maintains a constant speed. But, turbine speed (to the right) first increases, as the truck requires more power to go up the hill at constant speed, then decreases as the truck crests the hill.

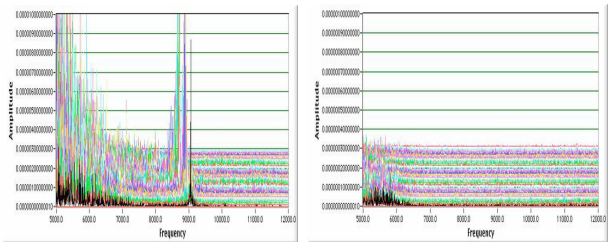


Fig. 7. The change in frequency shown in subsequent traces on the uphill (left side) charts is due to the Doppler shift as the truck passes the microphone. The downhill trucks (right side) do not show the turbocharger peaks.

that has not heretofore been accomplished. This technology promises to be applicable in a number of other scenarios including the issues of homeland security. The advances in the field of acoustic signature analysis represented in this project have a number of potential commercial applications relative to vehicle/engine diagnosis and maintenance as well as the intended use in this project in remote sensing of truck operations.

Summary and Conclusions

Working prototypes have been produced for the three different instruments to measure remotely the NO_x and PM emissions and engine operating parameters of passing trucks. Novel technologies have been employed, and patent disclosures have been filed for two of the technologies. Our goal of low-cost instruments has been achieved; all three technologies can be fabricated at much lower cost than typical instruments of a similar type. Validations of the operations of the instruments have been performed in real-world settings. More work is needed to "harden" the instruments for use alongside a road and to automate their processes to produce real-time results of measurements.

These developments will benefit DOE, EPA, and the Federal Highway Administration (FHWA) as they attempt to gain a better understanding and characterization of in-use emissions from heavy-duty trucks. Data are sorely needed to characterize the heavy trucking sector's role in urban air quality problems, and these instruments will help to clarify the effects.

This project has been part of a larger vision for an ORNL role in studying truck emissions and their impact on local environment. That vision is captured in the Watt Road Environmental Laboratory Initiative (WRELI). Knox County, Tennessee, is funding the University of Tennessee and ORNL to monitor emissions at one of the large truck travel centers at the Watt Road interchange, and discussions continue with the FHWA to fund a similar but larger project to study the emissions inventory along the roadbed of I-40. Efforts to further develop the WRELI vision, and the technologies of the project will continue.

References

[1]T. E. Barber, J. Castro, S. Hnatyshyn, N. L. Ayala, J. M. E. Storey, and W. P. Partridge, "Analysis of Diesel Engine Exhaust by Ultraviolet Absorption Spectroscopy," *Analytical Letters* **34**(14), 2493–2506 (2001).

[2]E. A. Wachter and W. G. Fisher, "Coherent Burst Laser Ranging: Decoupling Resolution and Unambiguous Range," *Opt. Lett*. **22**, 570 (1997) .

[3]E. A. Wachter and W. G. Fisher, "Method and Apparatus for Coherent Burst Ranging," U.S. patent 5,745,437.

[4]E. A. Wachter and W. G. Fisher, "Method and apparatus for improved ranging," U.S. Patent 5,889,490.

[5]H. C. Van de Hulst, *Light Scattering by Small Particles*, Wiley New York, 1957.

Nanoscale Photonic Crystal Laser

J. T. Simpson,[1] M. L. Simpson,[1] S. Rajic,[1] P. G. Datskos,[1] G. E. Jellison,[2] C. W. White,[2] and S. P. Withrow[2]
[1]*Engineering Science and Technology Division*
[2]*Condensed Matter Sciences Division*

The goal of this project was to demonstrate the first silicon nanocrystal waveguide–based laser (i.e., to show stimulated emission from a high-"Q" resonant cavity using silicon nanocrystals as the cavity's optical gain medium). Before we achieved our main goal of showing stimulated emission, a number of smaller goals had to be met. We demonstrated the ability to form silicon nanocrystals in SiO_2 by ion implantation and for ultraviolet light to generate silicon-nanocrystal-based photoluminescence. We then demonstrated the ability to form a matrix of high-Q distributive-feedback structures, each of which contained a large concentration of silicon nanocrystals. And finally we demonstrated the ability to optically pump these structures in order to generate stimulated emission. This project has been a multifaceted research effort involving not only ORNL but also Lawrence Livermore National Laboratory, the University of Arizona, and Digital Optics Corporation.

Introduction

Integration of optical components within semiconductor microchips has been a goal for many years. Such integration could create new and improved devices. The main reason why this integration has not occurred is the lack of any small CMOS compatible laser sources. Current solid-state lasers generally use gain media of non-standard III-V (or II-VI) materials, such as GaAlAs formed in a multiple quantum-well configuration. Such non-standard materials are difficult to fabricate and are highly incompatible with standard semiconductor microchip processes, which are generally silicon based. A solid-state laser suitable for integration with standard semiconductor microchip processes would be constructed from silicon-based materials, or at least be CMOS compatible, and would include a semiconductor process compatible optical waveguide material to facilitate energy transport. However, several challenges, including the indirect bandgap characteristic of bulk silicon, have generally prevented fabrication of silicon-based lasers. In 2000 Pavesi et al.[1] showed that silicon nanocrystals (embedded in a silicon dioxide matrix) produced an intense photoluminescence when optically pumped with ultraviolet light. Further investigation by Pavesi's group and others[2–6] showed the ability of such nanocrystals to act as an optical gain media with gains reported as high as 100 cm^{-1}. This project combines the recently discovered optical gain of silicon nanocrystals in the near IR region of the spectrum with an ORNL-designed, high-Q, distributive feedback (DFB) optical-waveguide structure. When optically pumped, using a high-intensity pulsed ultraviolet laser, this structure showed stimulated emission at 835 nm, thus demonstrating the world's first silicon nanocrystal waveguide–based laser. This groundbreaking research project enables the integration of micro-optics and opto-electronics with semiconductor microchips. Such integration is enabling to a host of optical technologies including optical interconnects, new types of chemical and biological sensors, new types of optical waveguide-based modulators, and is a vital component for optical computing.

Technical Approach

The goals of this research were to model, design, simulate, fabricate, and characterize the world's first silicon-nanocrystal-waveguide laser. To accomplish these goals, it quickly became apparent that multiple research and development paths were required. We began with two parallel paths of effort. The first path involved the creation of photoluminescent silicon nanocrystals, while the second path was to model, design, and computer simulate a high-Q DFB waveguide resonant cavity structure. While the ion implantation team worked on perfecting the implantation of the silicon and the formation of silicon nanocrystals, the resonant cavity design team worked on the design, modeling, and simulation of the DFB resonant structure. The ion implantation team used ORNL's ion

implantation facility to establish the best recipe for silicon nanocrystal creation. Once the ion implantation group had the right recipe for making silicon nanocrystals and the resonant cavity design team had a good cavity design, the two paths merged into the integration and fabrication phase of the project. During this phase, both groups worked together and began collaborating with other research and fabrication groups at the Lawrence Livermore National Laboratory (LLNL), the University of Arizona's Optical Sciences Center, and Digital Optics Corporation in Charlotte, North Carolina.

Initially we planned to do all the waveguide cavity fabrication and optical measurements at ORNL, but, based on our simulations and subsequent design choices, it soon became apparent that we required state-of-the-art photolithography equipment and expertise. This necessitated the photolithography work going outside the Laboratory. LLNL researchers assisted us in both high-quality optical thin film deposition and in e-beam photolithography. Some of the modeling and simulation problems and results were discussed with faculty and graduate students from the University of Arizona. We also subcontracted with Digital Optics Corporation to fabricate a two-dimensional array of DFB structures using state-of-the-art deep ultraviolet photolithography using I-line steppers. We also took advantage of their optical processing expertise with silicon dioxide to etch and dice some of the fabricated devices.The testing and characterization phase was primarily performed at ORNL.

Results and Accomplishments

Silicon nanocrystals: The ion implantation steps were performed at ORNL. A number of different ion implantation recipes, consisting of differing concentrations of silicon atoms, annealing temperatures, annealing durations, implantation energies, and substrates were tried and evaluated. The recipe that produced the largest photoluminescent characteristics (i.e., the brightest emission when pumped with ultraviolet light) is the one we used with our high-Q DFB cavity design and is described as follows: Multiple ion implantation energies were used to obtain a relatively uniform (flat) concentration profile of 5×10^{21} atoms/cm^3 throughout the top 500 nm of the substrate. The matrix substrate material used was fused silica. The samples were annealed for 1 h at a temperature of 1100°C. Annealing at this temperature caused significant migration of the excess silicon atoms. This migration had three distinct advantages. This large silicon migration allowed surface damage from the ion implantation to be repaired. Migration also caused the coalescence of silicon atoms into nanocrystals and increased the uniformity of the resulting effective index of refraction profile. The resulting nanocrystals have an estimated mean diameter of approximately 3 nm.

Planar waveguide: The index of refraction of fused silica is known to be 1.45 for near-infrared wavelengths. When fused silica contains silicon nanocrystals with a concentration of 5×10^{21} atoms/cm^3, the resulting index of refraction is increased to at least 1.55. We took advantage of this increase in refractive index and the uniformity of its profile to form a planar waveguide within the top portion of the fused silica substrate. Next, a thin film (approximately 1.0 μm thick) of SiO_2 was deposited (by chemical vapor deposition) on the planar waveguide's top surface. This allowed the planar waveguide to have a semi-symmetric structure which significantly enhanced the waveguide's modal characteristics. Figure 1 shows our fabricated planar waveguide and the photoluminescence observed when the sample was illuminated by a 1-W CW laser having a wavelength of 413 nm. Note how intense the photoluminescence appears in the image. Figure 2 shows the measured photoluminescence spectra from the sample seen in Fig. 1. Note that most of the emission from this sample is in the near-infrared regime, while only a small amount emits in the visible. Thus most of the emission can't be seen by eye or CCD camera. This makes the CCD generated image in Fig. 1 that much more impressive.

Modeling, Design, and Simulation: Modeling and simulation was performed using two commercially available sets of modeling software. The first software package we used, GSOLVER,[7] allowed us to establish an initial design for the cavity. Unfortunately, GSOLVER ran very fast by making a lot of simplifying, and sometimes unrealistic, assumptions. The algorithm GSOLGER uses

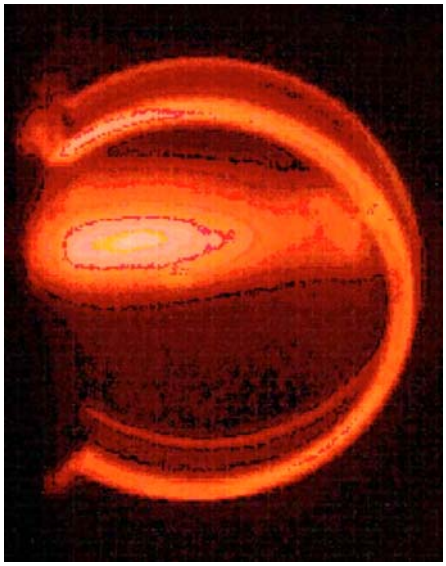


Fig. 1. CCD Image of the silicon nanocrystal photoluminescence from a 0.5-μm layer of silicon nanoparticles in a 2-cm-diameter SiO_2 wafer. Excited with 413-nm, 1-W CW laser.

what is known as "rigorous coupled wave equation" analysis. The second software package, FIMMWAVE[8] ran much slower, using algorithms based on much more complicated but more realistic assumptions. FIMMWAVE uses algorithms based on "modal analysis." We based our final design on simulations using this modal analysis package. A series of different waveguide cavity designs were modeled and simulated. The one chosen to be

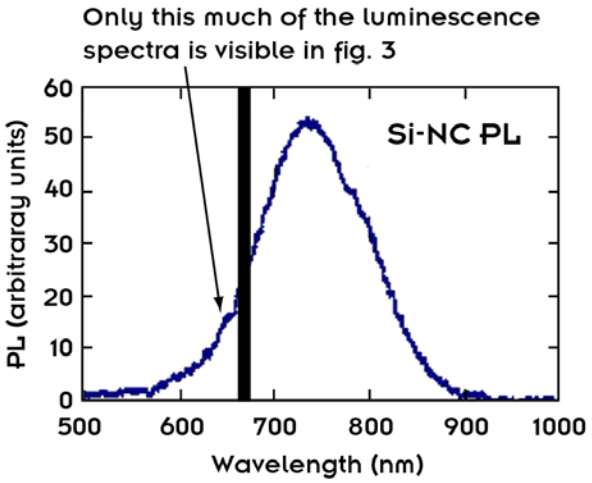


Fig. 2. Spectral response of the photoluminescence from a 0.5-μm layer of silicon nanoparticles in a 2-cm-diameter SiO_2 wafer. Excited with 413-nm, 1-W CW laser. Measured with an Ocean Optics spectrum analyzer.

implemented was a channel waveguide DFB. This design was chosen for its fabrication simplicity, superior performance characteristics, and its ability to be fabricated from the planar waveguide provided by our ion implantation team members. Because all the etching was on the overcoat layer and didn't directly affect the channel waveguide core region, this design was very insensitive to small processing defects. Any periodicity defects simply shifted the resonant wavelength slightly while any defects in the etch depth or uniformity simply reduced the resonant Q slightly. The design called for a 1-mm^2 channel to be etched into the silicon nanocrystal planar waveguide. Within this channel, 4000 shallow (500-nm-deep) grating periods were to be etched. Each period was to be one-half of an optical wavelength in extent. Assuming the waveguide's effective index of refraction was 1.55, the required periodicity turned out to be 250 nm. The theoretical Q of this design was roughly 100,000. Such a high Q means that virtually all of the stimulated emission generated within the structure will leak out of plane before coming out the ends. This extremely high Q was much greater than needed and made greatly relaxed the fabrication processing tolerances required. Figure 3 shows a simplified (much smaller number of periods) channel waveguide DFB structure schematic, and Fig. 4 shows its associated simulation results when formed into a resonant cavity having a quarter wave offset in its center. Since the silicon nanocrystals showed maximum photoluminescence at a wavelength of 750 nm, this was chosen to be the design wavelength for our cavity.

Fabrication: Our design called for forming a channel waveguide and distributive-feedback (DFB) structure via photolithography and chemical etching. Even though our DFB design was substantially easier to fabricate than most other designs investigated, it is still very difficult due to the small dimensions (250-nm periodicity) and the associated tight tolerances required (±5 nm). These fabrication issues led us to collaborate with Digital Optics Corporation in Charlotte, North Carolina. With their help we were able to come up with a fabrication strategy to successfully form such a channel waveguide having a DFB structure with 4000 periods. They also suggested we create a two-dimensional array of such structures using their state-of-the-art I-line photolithography steppers. Each channel

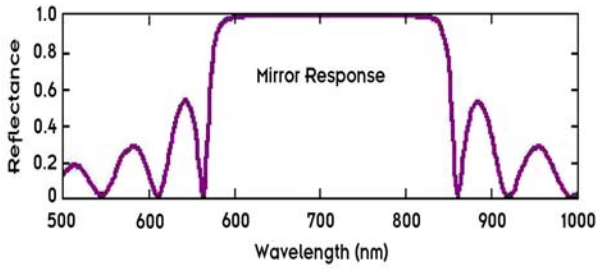


Fig. 3. Simulated spectral response of a distributed feedback (mirror) structure having a periodicity of 250.

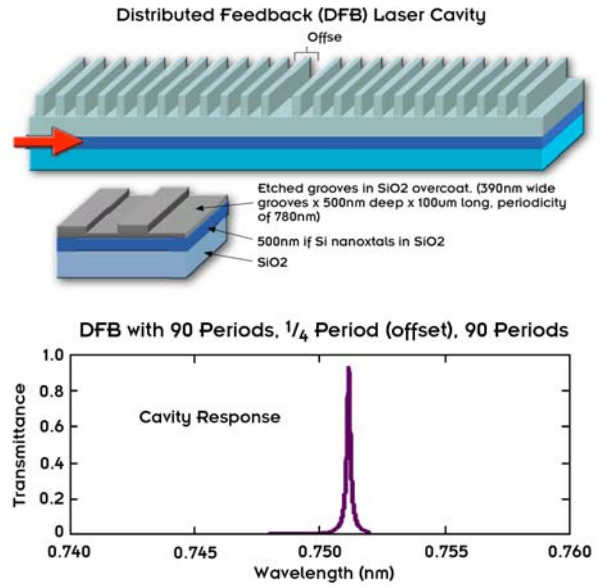


Fig. 4. Schematic and simulated spectral response of a distributed feedback (cavity) structure having a periodicity of 250 nm.

waveguide DFB structure was fabricated to be slightly different (i.e., slight differences in grating periodicities and line widths). This was accomplished by varying the optical focus and or the exposure dosage as the stepper imaged different channel waveguide DFB structures on different areas of the wafer. The result is known as a focus-exposure matrix, which is a fabrication development technique to help optimize fabrication processing parameters. Figure 5 shows an optical interference image of a fabricated DFB offset region and alignment mark. The distortion seen in the image is due to distortion in the imaging system and not the fabricated part. The etching line widths and parallelism design tolerances were all met according to the SEM and AFM measurements. Figure 6 shows an atomic force microscope (AFM) image of part of the DFB structure created. Figure 7 shows an image of many channel waveguide DFB regions making up a focus-exposure matrix of structures.

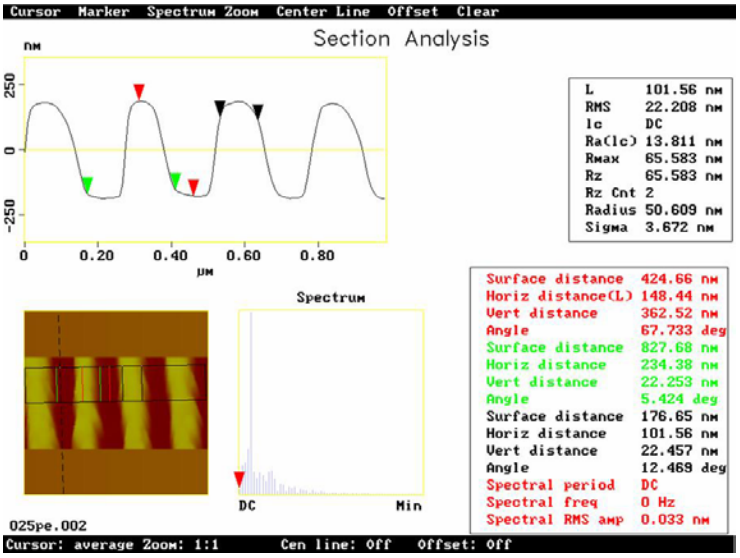


Fig. 6. Atomic force microscope image of a distributed feedback structure with a periodicity of 250 nm.

Testing and Characterizations

Once we completed fabrication of the focus-exposure matrix of channel waveguide DFB structures, we began the testing and characterization phase of the project. We used a pulsed 337-nm wavelength nitrogen laser having pulse durations of 800 ps and pulse outputs of 2.8 µJ to optically pump our structures. The nitrogen laser emitted a rectangular beam approximately 3 mm × 10 mm in extent. This beam was focused down to approximately 1 mm × 3 mm. This allowed us to illuminate a single channel waveguide structure at a time. Again, since the Q of the structures was so high, we expected most of the emission to leak out the top rather than come out the ends. As expected, we saw virtually no emission out the sides. But, as expected, we saw a large leakage emission coming out the top surface. Figure 8 shows the spectral emission of this leakage. While he had expected emission to be at 750 nm, you can see from Fig. 8 that it actually emitted at 835 nm. Most likely this discrepancy is due to an underestimation of the effective index of refraction of the channel waveguide. The 845-nm emission suggests that the actual channel waveguide index of refraction is 1.75 instead of 1.55.

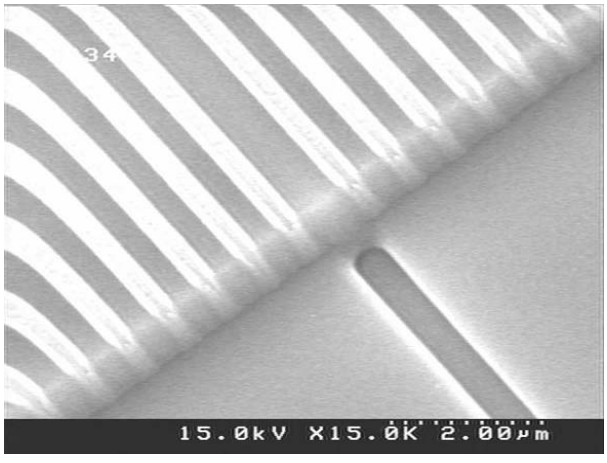


Fig. 5. Optical interference image showing the fabricated distributed feedback offset region and alignment mark. The apparent distortion is due to limitations in the optical imaging system.

Summary and Conclusions

In order to achieve the ultimate goal of this project of demonstrating silicon based stimulated emission, many intermediate goals and accomplishments had to be met. First and foremost, we had to form photoluminescent silicon nanocrystals. We then had to use these photoluminescent nanocrystals to form a planar waveguide within a SiO_2 substrate. Finally we measured the spectral emission observed when the channel-waveguide distributive feedback (DFB) structure was optically pumped using a pulsed nitrogen laser. A broad spectral photoluminescent emission is produced by the silicon nanocrystals being pumped by ultraviolet laser light, while a narrow spectral emission is observed which is caused by stimulated emission generated within the channel-waveguide DFB structure. This observed stimulated emission was our research goal and as such demonstrates

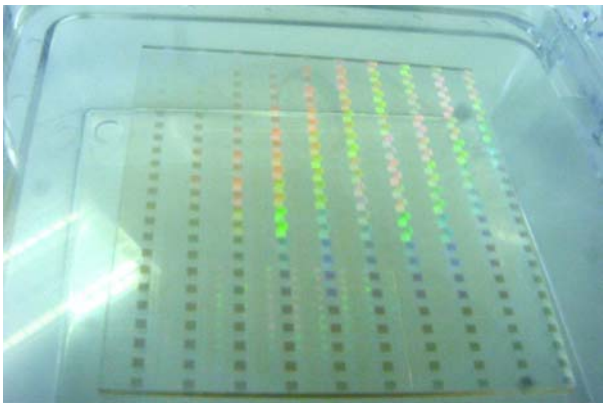


Fig. 7. Image of a glass slide patterned with a distributive feedback focus-exposure matrix.

the world's first silicon nanocrystals–based waveguide laser. Another major accomplishment of this research has been the collaborative effort involving two national labs (ORNL and LLNL), a major university (the University of Arizona's Optical Sciences Center), and an industrial partner (Digital Optics Corporation). This research has demonstrated the ability of silicon nanocrystals to be used as a gain media in conjunction with a high-Q distributive feedback structure. This groundbreaking research project enables the integration of micro-optics and opto-electronics with semiconductor microchips. Such integration is enabling to a host of optical technologies including optical interconnects, new types of chemical and biological sensors, and new types of optical waveguide-based modulators and is a vital component for optical computing. We are in the process of submitting these results for peer review and publication. Now that stimulated emission has been demonstrated using silicon nanocrystals, we are pursuing sponsorship to expand the scope of this research, which would include the possibility of electronically pumping the silicon nanocrystals and combining the silicon nanocrystals with erbium. We already have some follow-on DARPA funding to explore the silicon nanocrystals with erbium combination.

References

[1]Pavesi et al., "Optical gain in silicon nanocrystals," *Nature* **408**, 440 (2000).

[2]Khriachtchev et al., "Optical Gain in Si/SiO2 Lattice: Experimental Evidence with Nanosecond Pulses," *Appl. Phys. Lett*. **79**, 1249 (2001).

[3]Ivanda and C. W. White, "Stimulated Emission and Optical Gain in Silicon Nanocrystals," *Applied Physics Letters,* 2031–33 (2000).

[4]Degoli et al., "Electronic and Optical Properties of Silicon Nanocrystals: Structural Effects," *Mat. Res. Soc. Sym. Proc.* **770**, I1.9 (2003).

[5]M. A. Green et al., "Exceptionally High Radiative Efficiencies in Silicon," *Mat. Res. Soc. Sym. Proc.* **770**, I2.7.1 (2003).

[6]L. Dal Negro et al., "Time-resolved Gain Dynamics in Silicon Nanocrystals," *Mat. Res. Soc. Sym. Proc.* **770**, I3.4.1 (2003).

[7]*GSOLVE*R from the Grading design software from Grating Solver development Company of Allen, Texas.

[8]*FIMMWAVE and FIMMPROP-3D* custom waveguide design software from Photon Design, Inc., of Oxford, United Kingdom.

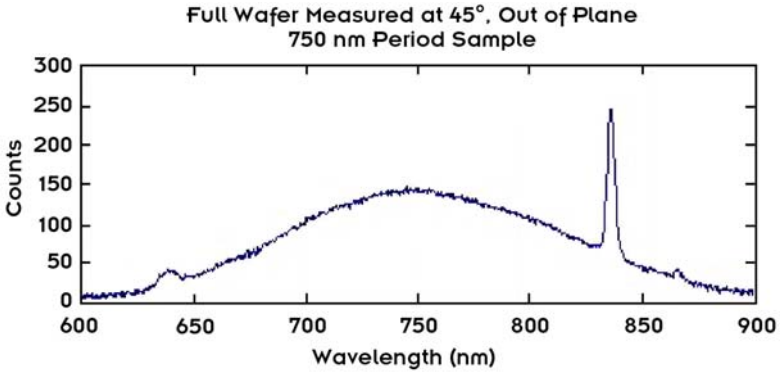


Fig. 8. Spectral emission showing a broad photoluminescent emission and a narrow distributive feedback structure emission at 835 nm.

Zero-Power, Low-Cost Sensor Platform

J. E. Hardy,[1] P. D. Ewing,[1] J. E. Christian,[1] R. J. Warmack,[1] P. G. Datskos,[1] C. L. Britton, Jr.,[1]
M. Z. Martin,[2] and S. F. Smith[1]
[1]*Engineering Science and Technology Division*
[2]*Environmental Sciences Division*

Numerous national studies and working groups have identified very low-power, low-cost sensors as a critical technology for increasing energy efficiency, reducing waste, and optimizing processes. This research will address that need by developing a zero-power, low-cost sensor platform based on microcantilever (MC) arrays that includes MC sensors, very low-power electronics, signal processing, and two-way data communications, all integrated into a robust package. MCs will be developed to measure carbon dioxide, temperature, and room occupancy. Research areas are coatings for the various gases, developing an electronic readout for the infrared occupancy sensor, and developing very low power readout and signal transmission (both over power lines and wirelessly).

The objectives of this research effort is the development a near zero-power, low-cost platform based on a microelectromechanical-system (MEMS) sensing array. This sensor platform will include microcantilever (MC) sensors, very low-power electronics, signal processing and filtering, and data transceivers all integrated into a robust package. The vision is to develop a flexible multisensor array (measuring several items concurrently) that will use little to no net power, be very low cost in volume (<$100/unit), and communicate bidirectionally either over existing power lines or wirelessly. The focus of this project is to develop MEMS sensors to be able to eventually monitor a variety of gases, temperature, and personnel occupancy that are important in many applications. The R&D required to achieve the proposed vision includes the development of coatings for the cantilever array and the development of low-power signal processing and communications electronics. State-of-the-art improvements will be needed in multisensor array/electronics integration, conversion of infrared sensing to a very low-power electronic readout (instead of optical), and low-power data transmission/coupling to existing power lines. Success in this proof-of-principle effort will enable us to obtain a leadership role in DOE programs related to energy efficiency/waste reduction [e.g., zero-energy buildings (ZEB), fossil energy, fuel cells, distributed energy resources].

The microsensor array prototype will be based on ORNL inventions in MCs, low-power electronics, and sensor networks. Coatings will be developed based on literature review and our polymer experience, the infrared (IR) occupancy sensor will be an extension of previous work in IR imaging with MCs, the electronics will be based on several years of experience in low-power design and integration, and the sensor network will be based on our inventions in hybrid spread-spectrum communications.

During the first year of the project, we developed a chemical coating that is sensitive to carbon dioxide, has reasonable response times, and is relatively stable with minor influences from humidity. We also successfully tested a ten-channel capacitive readout and signal conditioning electronic module. The electronics are very low power (1 mW in continuous use and microwatts in intermittent use) and have very low noise levels (equivalent to a nanometer of movement in the MC). We made good progress in the design and fabrication of infrared MCs that can be read out electronically (first of its kind) and completed conceptual design for a robust communication scheme that uses existing building wiring and can handle three-phase noise. An invention disclosure was filed on this robust communication scheme.

This project is relevant to DOE in several mission areas including building technologies, industrial technologies—sensors and controls, and fossil and fuel cell energy technologies. The zero-power, low-cost sensor platform will allow more extensive monitoring of building processes and environmental conditions for personnel health and comfort. The research into low-power electronics and multisensor arrays will advance measurement science—an enabling technology for most DOE programs. This project has applications to Department of Defense programs and Homeland Security programs. The low cost of our platform should allow more and better monitoring of ships, tanks, and planes, which will increase in importance as manpower is reduced in the armed forces. The low power and low cost allows more monitoring and detection for security needs. In addition, the zero net power enables sensing in remote locations, giving unattended monitoring—a must for security and homeland defense.

NEUTROMEGAS: A Pixel Detector for Neutron Imaging

C. L. Britton, Jr.,[1] W. L. Bryan,[1] A. Wintenberg,[1] R. J. Warmack,[1] T. E. McKnight,[1] S. Frank,[1]
R. G. Cooper,[2] N. Dudney,[3] and G. Veith[3]
[1]*Engineering Science & Technology Division*
[2]*Spallation Neutron Source (SNS)*
[3]*Condensed Matter Sciences Division*

A bright neutron source such as the Spallation Neutron Source (SNS) places extreme requirements on detectors and especially impacts a detector's imaging and dynamic range. Present imaging detectors have either shown position resolutions that are less than acceptable or they exhibit excessive paralyzing dead times due to the brightness of the source. A new detector concept known as MICROMEGAS (MICROMEsh GAseous Structure) has been developed at CERN in Geneva for high-energy charged-particle tracking and has shown great promise for handling high data rates with a rather low-cost structure. If properly developed, this detector platform has features that could revolutionize neutron imaging. We are presently developing a novel detector using the MICROMEGAS concept by creating NEUTROMEGAS, the first pixelated, high-rate detector for 2-D neutron imaging. This has not been successfully accomplished anywhere else, although it has been a goal for researchers. This would be an outstanding contribution to the overall detector instrumentation effort for the neutron community and would establish ORNL as a leader in the development of next-generation neutron detectors.

Objective

The goal of this project is to develop a prototype high-(position) resolution, low-dead-time neutron detector that we call NeutroMegas. In order to accomplish our goals two key projects have been undertaken. One focuses on the development of the detector structure and readout electronics, while the other is directed towards converter foil research. The detector structure portion of this project will entail building and testing both one-dimensional (1-D) strip and two-dimensional (2-D) pixel detectors. Initially a 1-D strip prototype will be developed since it allows easy electronics readouts, while our long-term goals will be to produce a true pixelated 2-D detector

During the first year of the project, we fabricated a detector prototype whose anode structure was formed on a quartz substrate using micro-fabrication techniques. A regular array of separator columns 73 μm tall was formed by chemically etching the photopatterned surface. A shadow mask was then aligned to permit 250-μm-wide chromium anode strips to be evaporated between the columns. Adjacent strips were spaced 1 mm apart and were alternately connected to vacuum feedthroughs in the aluminum chamber. A 1000-line-per-inch, 7-μm-thick nickel grid was placed atop the anode structure. A 3-mm insulator was then used to space a boron- or lithium-coated aluminum converter plate away from the grid.

The main objective of the neutron converter work has been to develop the synthetic techniques required to fabricate the converter foils as well as understanding and manipulating the chemical or electrochemical stability of these foils. This is essential to successfully building a stable, high-efficiency detector. To this end, we have been very successful in meeting our goals.

Converter films consisting of naturally abundant elements have been prepared using a variety of physical deposition techniques. Due to the air and moisture sensitivity of these materials, all film growths were carried out in a deposition system contained within an argon-filled dry box. Thin films of boron (1–2 μm) and gadolinium (1–2 μm) have been prepared using RF and DC sputtering techniques, respectively, while lithium films (3–100 μm) were synthesized by thermally evaporating elemental lithium. These thickness ranges were chosen from the literature for reasonable detector efficiency. The deposition conditions were optimized (i.e., argon pressure, power, distances between the targets and substrates, rate and choice of substrates) in order to minimize the film impurities as well as reduce any potential mixing of the reactants due to the energetic nature of the deposition process. By optimizing the deposition conditions, we have effectively reduced the impurity content to below 5 atomic %.

In summary, we have established the synthetic conditions required to prepare these foils, investigated a series of Li-substrate interactions, initiated a series of long-term studies to address the interdiffusion between the converter constituents, and determined the stability of the converter to the detector gases.

Because the detector has worked so well in a short period of time, we have had opportunity to do extensive testing. Our testing includes the following:

1. *Critical Voltage (E_{crit}) testing*—The critical voltage E_{crit} is the voltage when electron multiplication begins for a given gas and detector geometry. Calculations for our geometry at 1 atm of P-10 gas predicted a value of E_{crit} = 350 V for our multiplying grid voltage. Our measurements showed a value of approximately 330 V, demonstrating excellent agreement with the design value.
2. *Detector position mapping with ^{241}Am*—With the data acquisition system (DAQ) we developed, we can simultaneously read out each of the six instrumented channels of strips in two different modes. The observed position-sensitive detector spectrum is exactly what we expected to see with this detector and data system.
3. *^{252}Cf Testing (neutron testing)*—Mode 2 of the DAQ allows us to stop and read all channels any time one of the strips has an event. This allows position determination of an incident neutron in 1-D, the very purpose of the detector. We have been able to observe individual neutron events with little crosstalk between strips.

At the end of the second year of this project, we expect to have a working 2-D pixel detector. This should attract interest from instrument scientists at the SNS for use in beam and transmission monitoring applications in some of the proposed experiments. In addition, a variant of this detector is being considered by another experiment at SNS.

ENGINEERING SCIENCE AND TECHNOLOGY

Seed Money Fund

Investigation of a Novel Technique for Infrared Energy Detection and Imaging

K. Korsah,[1] R. A. Kisner,[1] L. A. Boatner,[2] and H. M. Christen[2]
[1]*Engineering Science and Technology Division*
[2]*Condensed Matter Science Division*

This project focused on the investigation of a new piezoelectric material for its suitability as a thermal detector using surface acoustic wave (SAW) technology. The long-term goal is to use the technology for infrared (IR) detection and imaging.

Thermal (IR) imaging remains a cornerstone technology for many military applications, including small unit operations, ground, air and sea target acquisition, missile seekers, and threat warning. However, the cost of present-day high-performance cameras can range between $70K and $170K. Besides cost, another disadvantage of these systems is that they typically are required to be cryogenically cooled. Therefore, there is a need for the development of more sensitive, uncooled, and less costly thermal imaging devices for these as well as other (civilian) applications. The continued need for improved IR imaging systems is evidenced by the statement in June 2001 by Dr. Tony Tether, Director of the Defense Advanced Research Projects Agency (DARPA), before the subcommittee on Military Research and Development, that the program began funding the investigation of new concepts for thermally sensitive microstructures in FY 2000.

The immediate, expected result of this project was a single-pixel surface acoustic wave (SAW) infrared detector. The long-term goal of the project is to use the results obtained under this project as a basis for further research and development of an uncooled IR detector camera, based on SAW technology that can be purchased for around $1000 or less. This would be a unique instrument and would revolutionize the IR camera market in much the same way as the development of the uncooled microbolometer camera has over the last several years. Low cost is the key that opens the door to hundreds of markets.

Introduction

The traditional application of SAW technology has been in communications and, more recently, in sensor applications. SAW application in IR imaging has not been explored to any extent. In almost all traditional applications of SAW sensors, the objective is to have negligible changes in the substrate parameters as a function of temperature. For application as an IR sensor, however, this is the means of transduction, so changes in substrate parameters as a function of temperature must be as high as possible. In fact, many of the requirements for a SAW IR (thermal) sensor are of less significant consideration in the more conventional applications of SAW devices. These requirements include the following: (1) high-temperature coefficient; (2) sensitivity to ~10 mK; (3) temperature changes; (4) small pixel size; (5) low interference from non-SAW modes; (6) robust design; and (6) simple instrumentation.

A SAW device is characterized by the electrical excitation of a surface acoustic wave in a piezoelectric single-crystal plate or substrate (e.g., quartz, lithium niobate).[1,2] The electrical excitation is produced by means of a metallic (e.g., aluminum) interdigital transducer (IDT) structure. When the SAW is used as a sensor in an oscillator configuration, the influence of the measurand (e.g., gas, temperature/infrared energy) on the acoustic wave phase propagation speed v is used for the sensor effect. In the oscillator configuration, this velocity perturbation gives rise to a frequency shift Δf given by

$$\Delta f = (\partial f/\partial m)\cdot\Delta m + (\partial f/\partial c)\cdot\Delta c + (\partial f/\partial \varepsilon)\cdot\Delta\varepsilon + (\partial f/\partial \sigma)\cdot\Delta\sigma + (\partial f/\partial T)\cdot\Delta T + (\partial f/\partial P)\cdot\Delta P \quad , \qquad (1)$$

where $\Delta\sigma$ and $\Delta\varepsilon$ signify conductivity and dielectric changes, respectively (the changes in σ and ε are due to *acoustoelectric interactions*); Δm and Δc are mass and viscoelastic constant changes, respectively, and are due to *mechanical interactions*; and $\Delta T + \Delta P$ are temperature and pressure changes, respectively. Equation (1) shows that the SAW frequency is sensitive to a number of parameters; however, depending on the type of sensor being developed, it is possible to reduce the effect of other parameters on the frequency. For example, for use as a gas sensor, the surface of the substrate is covered with a chemically active, selective thin film. This film is capable of adsorbing specific molecules. This adsorption gives an increased mass loading of the Raleigh wave. In this case, temperature effects are reduced by using a piezoelectric cut that has a zero temperature coefficient.

If the measurand is infrared energy (temperature), any of the parameters in Eq. (1) that change significantly with temperature will contribute to the output frequency shift. The fifth term $[(\partial f/\partial T)\cdot\Delta T]$ signifies a direct temperature effect of the substrate via the temperature coefficient of delay (TCD) or, as is more popularly referred to, the temperature coefficient of frequency (TCF), given by the following equation

$$TCD = (1/L)*(\partial L/\partial T) - (1/v)*(\partial v/\partial T) \ , \qquad (2)$$

where L is the acoustic path length.[2] However the temperature effect through the TCD is typically small, usually of the order of parts per million. To be a good candidate for IR sensing, we either require a material with a much higher TCD, or one in which other parametric changes have a much greater effect. In our study we investigated potassium tantalum niobate ($KTa_{1-x}Nb_xO_3$), or KTN, as a potential candidate because it exhibits very large changes in dielectric constant as a function of temperature. In particular, KTN exhibits significant permittivity and electro-optic effects at temperatures near the ferroelectric curie point T_c. By adjusting the Ta/Nb ratio, T_c can be varied continuously from –273°C ($x \le 0.008$) to 435°C (i.e., to the value of T_c for pure $KnbO_3$). For $x \ge 0.05$, T_c has been shown to follow the simple linear relationship given by

$$T_c[°C] = 676°C \cdot x - 241°C \ . \qquad (3)$$

The dielectric constants as a function of temperature for films grown with different Ta/Nb ratios is shown in Fig. 1. In the figure, curve (a) corresponds to $KNbO_3$, curve (b) corresponds to $KTa_{0.19}Nb_{0.81}O_3$, and curve (c) corresponds to $KTa_{0.25}Nb_{0.75}O_3$.

Technical Approach

The approach taken in this study was to first design and fabricate thin films of KTN of the desired composition, design and fabricate SAW structures using the KTN material as well as lithium niobate ($LiNbO_3$) and compare the temperature response of the two materials. $LiNbO_3$ was used for this comparison because it is one of the more common SAW device materials that possess a relatively large temperature coefficient of frequency (TCF). For example, a SAW device with propagation in the Z direction on Y-cut $LiNbO_3$ (YZ-$LiNbO_3$) has a temperature coefficient of approximately 90 ppm/°C. The temperature coefficient of 128° lithium niobate has a lower temperature coefficient (~75 ppm/°C), but the lower bulk wave interference tends to improve the response (lower distortion) in broad bandwidth devices. If KTN is going to show any promise as a thermal detector in a SAW configuration, it would be important to show that it's much more sensitive to temperature than $LiNbO_3$.

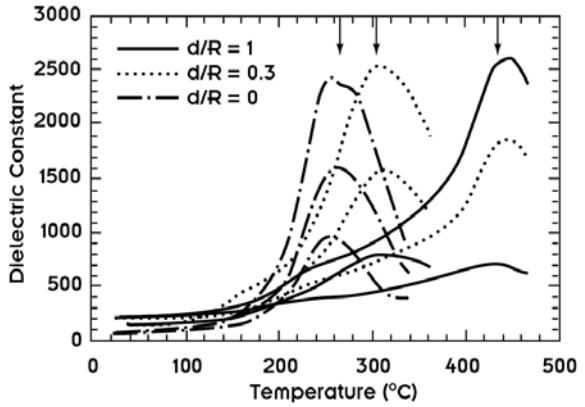


Fig. 1. Dielectric constant as a function of temperature for films grown with different Ta/Nb ratios. By changing the Ta/Nb ratio, it is possible to shift the (steep) slope of the curve to room temperature.

A major challenge was to be able to fabricate sufficiently thick films (several microns) of the KTN material. Eventually, after overcoming several challenges, 0.6 μm thick of $K(Ta_{1-x}Nb_x)O_3$ [$x = 0.5$] epitaxial films were successfully grown and polished on $KTaO_3$ (001) substrates approximately 1 mm thick. The film was grown after the method of Christen et. al[3] by pulsed laser deposition using a four-target mixing approach (repeated sequential deposition of sub-monolayer amounts of $KTaO_3$ – KNO_3 – $KNbO_3$ – KNO_3 – etc.). KNO_3 is added as an additional K-source to compensate for the volatility of this element. Target firing sequence is in 4-pulse bursts at 100 Hz (248 nm, 200 mJ per pulse), with the target sequence completed once per second.

The perovskite targets were prepared from commercial $KTaO_3$ and $KNbO_3$ powders (Alfa Aesar) by hot-pressing in vacuum at ~1300°C ($KTaO_3$) and 1100°C ($KNbO_3$) and from commercial KNO_3 powder (Alfa Aesar) by cold-pressing at 6000 lb in air after carefully grinding the power to small grain size.

Film characterization by profilometry on a reference sample indicated a thickness of 600 nm for a total of 40,000 laser pulses (i.e., deposition time of 2500 s). X-ray analysis of the crystal indicated single phase within resolution. Rocking curves (ω-scans) were broad but identical for film and substrate within experimental resolution, indicating that the film is following the crystalline template of the substrate. θ-2θ scans through the film and substrate lines indicate a $2\theta_{substrate}$-$2\theta_{film} = 0.24°$ (see Fig. 2). With $a^*(KTaO_3) = 3.989$ Å, we calculated $c^\dagger(\text{film}) = 4.009$ Å. With a unit cell volume for $KTaO_3$ of $V(KTaO_3) = 63.47$ Å^3 and for $KNbO_3$ of $V(KNbO_3) = 64.74$ Å^3, as well as the assumption that $V(\text{film}) \le [c(\text{film})]^3 = 64.433$ Å^3, we found that $x \le 0.75$. This assumption is based on the fact that the in-plane film lattice parameter (grown on a substrate with a smaller lattice parameter) will not be larger than the out-

*In-plane lattice parameter.
†Normal lattice parameter.

of-plane value. However, the opposite extreme estimate is to propose that the in-plane lattice parameter of the film is identical to that of the substrate (which is an unlikely case for a film of this thickness). This would lead to V(film) = $a(KTaO_3)^2 \times c$(film) = 63.79 Å^2, leading to $x \geq 0.25$. From optical growth monitoring and previous experience, we assume that the growth rate for $KNbO_3$ is only slightly larger than that of $KTaO_3$ (this depends on target quality), and therefore that $x \approx 0.5$. Using T_c[°C] = 676°C · x –241°C, we find that –72°C ≤ T_c (film) ≤ 266°C from X-ray data, with the mean value being T_c = 97°C. Using data from Christen et al. [*Thin Solid Films* **312**, 156 (1998)], we know that ∂ε/∂T is highest for $T = T_c - 50$°C; therefore, the film appears to have the appropriate composition. (Extensive characterization via an outside sub-contract would have to be performed to determine this number accurately.) Optical microscopy revealed defects at the surface of the film on portions of the substrate ("milky" regions), with most of the substrate being coated with a nice-quality epitaxial smooth film.

After thin films of $KTa_{1-x}Nb_xO_3$ were successfully fabricated, both $KTa_{1-x}Nb_xO_3$ and $LiNbO_3$ SAW delay lines were designed assuming a foundry with a 2-μm technology. In the following summary of the design, $LiNbO_3$ has been used as the example; however, the design principles are the same in both cases.

The 2-μm line width constraint allows sufficiently thick aluminum (120 nm) to eliminate the need for secondary metallization on the bond pads while staying below the manufacturing constraint of a 10:1 width-to-thickness ratio. An objective was to minimize the size of the active area of the SAW temperature sensor. This dictated that the transducer period be minimized. The only remaining parameter that dictates the size of the active device area is the number of electrodes per transducer period. Two-electrode-per-period (2/λ) transducers exhibit serious distortions from internal reflections. Four electrode-per-wavelength (4/λ) transducers are preferred for delay lines since they are mechanically reflectionless. Three-electrode-per-wavelength (3/λ) transducers are also reflectionless and this was the style selected for the design. The transducer period was thus 12 μm with three lines and three spaces of 2 μm per electrode (i.e., with equal electrode spacing, each line or space is 2 μm).

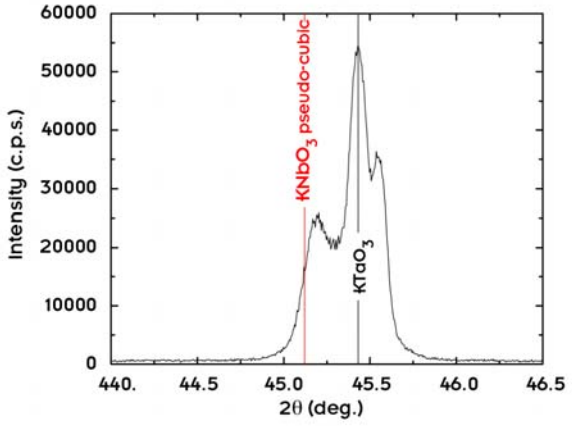


Fig. 2. Film and substrate X-ray diffraction (θ-2θ) peaks for the $KTa_{1-x}Nb_xO_3$ film on the $KTaO_3$ substrate.

The transducer bandwidth is the center frequency, F_c, divided by the effective number of transducer periods, N. Thus, any bandwidth specification would determine the number of periods. In practice, bandwidth specification is of less practical importance for the application under consideration. Therefore, the number of electrode periods was iteratively adjusted to optimize insertion loss and other electrical properties. A 12-period transducer was found to be optimal. The center-to-center transducer separation (L_{cc}) in wavelengths determines the slope of the phase shift at center frequency and, along with the number of transducer periods, N, the number of 2π phase shifts in the transducer pass band. A value of L_{cc} = 112 was chosen to balance device size with phase resolution. Thus the total active device width in microns was

$(12 \times 12 \times 2 + 112 \times 12) = 1632$ μm, or ~1.6 mm .

Using an aspect ratio (length to width) of 6:1 (typical in delay line design), the active device area was therefore ~0.27 mm × 1.6 mm.

The nominal phase shift is given by

$$\textit{Nominal Phase Shift} = 360° * L_{CC} = 40{,}320° \quad . \qquad (4)$$

Thus, a 0.1° phase resolution would correspond to ~2.5 parts per million (ppm) of the nominal phase shift. With a temperature coefficient of phase of 75 ppm/K, the sensor should have a nominal temperature resolution of 0.033 K (33 mK).

Results and Accomplishments

The phase response as a function of temperature for KTN as well as $LiNbO_3$ were measured with a network analyzer by heating the device in one-degree increments and measuring the phase at each temperature. The response is shown in Fig. 3. The slope of the phase (i.e., the change in phase per degree change in temperature) as a function of temperature is shown in Fig. 4. Comparison of KTN and $LiNbO_3$ responses from these two figures show the following:

1. The largest phase change exhibited by the $LiNbO_3$ is about –4.7°/°C, whereas the largest phase change exhibited by the KTN is about twice as much (11°/°C). Assuming that the network analyzer phase resolution is at least 0.5°, this corresponds to a temperature resolution of 0.1°C for the $LiNbO_3$ and better than 0.05°C for the KTN.

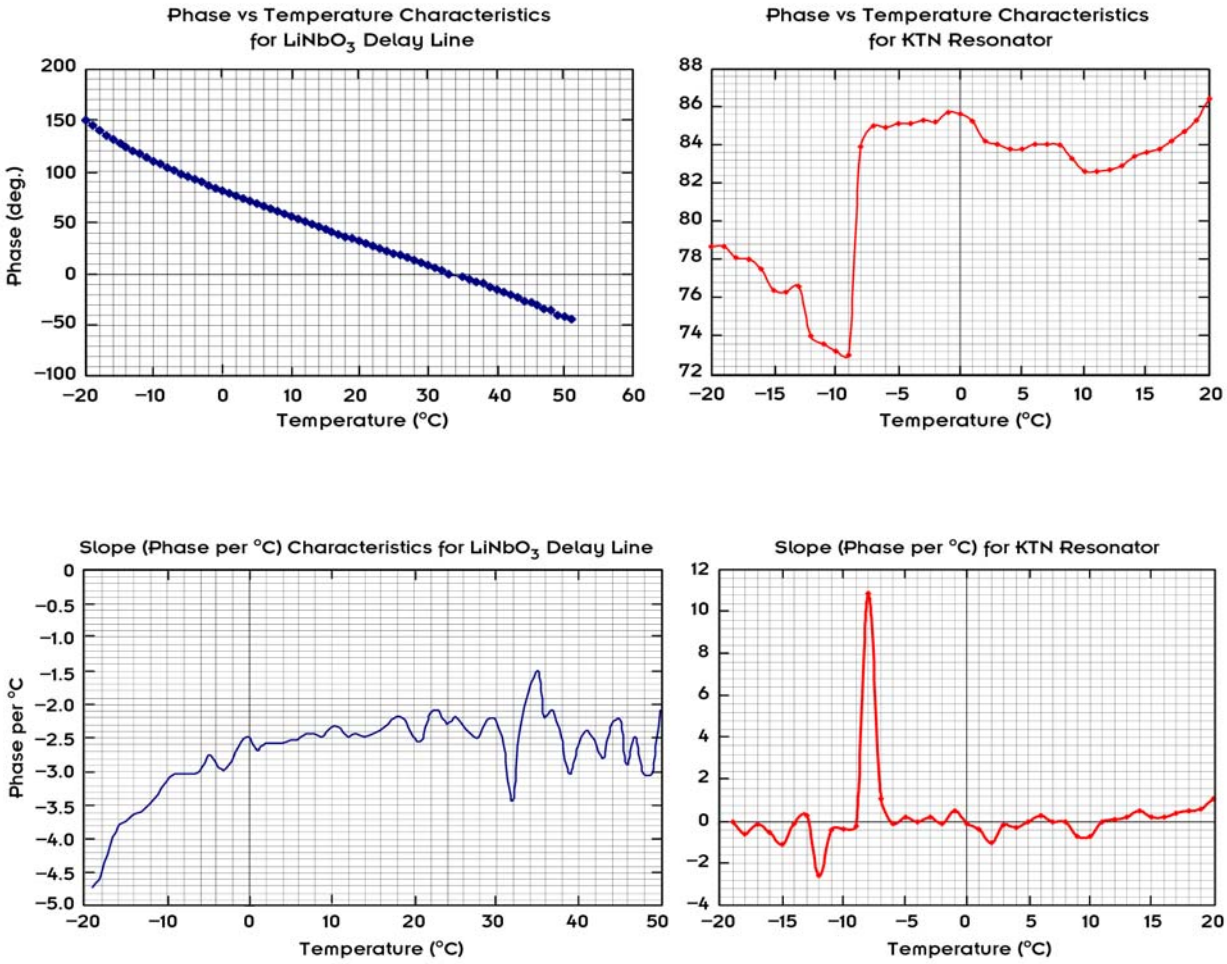


Fig. 4. Slope of phase as a function of temperature for $LiNbO_3$ and KTN SAW structures.

2. The temperature response of KTN does exhibit the characteristics the investigators were hoping for. However, the investigators believe that optimum response was not achieved due to fabrication difficulties (e.g., poor photolithography by the company that fabricated the SAW devices, learning curve in fabricating KTN SAW structures, etc.).

Summary and Conclusions

This study was successful in identifying KTN as a potential candidate for IR detection and imaging. With improved fabrication and signal processing, it will be possible to achieve a temperature resolution of better than 0.05°C. One observation that needs further study is that theoretical predictions placed the transition temperature within a wide temperature range (between –72°C and 266°C), the arithmetic mean of which gives a value of 97°C. We therefore expected the maximum phase changes to occur at a higher temperature than what the results indicate (–8°C). This apparent discrepancy will be a subject of subsequent studies.

Typical pixel size of current IR cameras is in the neighborhood of 50 µm. The "pixel" size of the sensor designed for this study was considerably larger. However, size was not a constraint for this initial study. We have identified a path for reducing the pixel size and using a different signal processing approach (to increase data acquisition speed in an array configuration) in subsequent studies.

References

[1]H. Mathews, *Surface Acoustic Wave Filters: Design, Construction, and Use*, John Wiley and Sons, New York, 1977.

[2]D. S. Ballantine et al., *Acoustic Wave Sensors: Theory, Design, and Physico-Chemical Applications*, Academic Press, 1997.

[3]H.-M. Christen, D. P. Norton, L. A. Gea, and L. A. Boatner, "Pulsed Laser Deposition of Soli-Solution Films Using Segmented Targets," *Thin Solid Films* **312**, 156–159 (1998).

High-Resolution, Real-Time, Three-Dimensional Imaging Using Two-Wavelength Spatially Heterodyned Interferometry

G. R. Hanson, P. R. Bingham, T. P. Karnowski, and J. T. Simpson
Engineering Science and Technology Division

The ability to make high-resolution, three-dimensional surface topology measurements in real time for inspection and metrology is a growing need in manufacturing, scientific research, stockpile stewardship, and medicine. The problem with existing topography methods is that they suffer from low-resolution, slow inspection speeds, and/or the inability to image large height changes with high spatial resolution. ORNL researchers have previously developed the Direct-to-Digital Holography (DDH) technique for imaging semiconductor wafers during the manufacturing process. DDH records the complete surface profile in a single digital image by using "spatial-heterodyning" to record the phase of the laser beam reflected off of the surface. This technique is ideal for surface topography measurements on wafers because the transverse (x, y) resolution requirements are of order 100–300 nm (of order 1/2 the laser wavelength), and the longitudinal (z) resolution requirements are of order 1–5 nm (much less than the laser wavelength). However, many potential applications involve longitudinal and transverse dimensions that are 10 to 1000 times greater than the probing laser wavelength. We have demonstrated a novel method of using two separate laser wavelengths to image the object surface. This method will allow high-speed or real-time inspection of surfaces with topological features in the 1- to 1000-μm range.

Introduction

The ORNL patented Direct to Digital Holography (DDH) technology uses holographic techniques along with digital image processing to measure and display surface profiles in real-time (more than 30 images per second is possible). The technology has been licensed to a start-up company and is currently being used in the development of a tool to image semiconductor wafers during the manufacturing process.

DDH has had limited impact on other industrial applications due to a significant limitation: surface profile features (both surface roughness and discontinuities) that exceed half of the laser wavelength are difficult to measure due to "phase wrapping." Skateboarders, figure skaters, and other extreme sport enthusiasts are familiar with phase wrapping, although they do not refer to it as such. Essentially, a 360-degree turn is equivalent to a 720-degree turn or even a 3600-degree turn in terms of the starting and finishing position.

In DDH, we are measuring the phase of a reflected laser beam whose wavefront encodes the surface depths. The surfaces must be relatively flat and smooth relative to the laser wavelength, with only a limited distribution of three-dimensional surface features. When the surface-feature depths and/or surface roughness exceeds half the wavelength of the measurement laser over a distance less than 2 pixels in the digital image, we find that the true depth becomes impossible to measure exactly (the reflective nature of our system imposes the half-wavelength criteria). Because the phase "wraps around," that is, goes through a 360-degree or 2π radian change, we cannot measure depths greater than a half-wavelength accurately, since the number of potential wraps past 2π is unknown. We are able to get a high-resolution measurement to within a half-wavelength, but we cannot measure multiple wavelengths with certainty.

The problem of phase unwrapping has been addressed in literature;[1,2] most techniques rely on a priori knowledge of the general characteristics of the surface under measurement and attempt to track phase wraps in an incremental fashion. However, these techniques require that the phase wraps occur over a large number of image pixels; a gradual slope can be tracked, but a sudden jump or too sharp a slope cannot be unwrapped (see Fig. 1).

To eliminate this problem, it would be desirable to measure the height variations at a much longer wavelength while still maintaining the longitudinal resolution of the shorter laser wavelength. This goal is accomplished in other forms of interferometry by acquiring the phase data at two or more wavelengths using multiple images and then looking at the difference of the phases measured by each wavelength.

We propose a new method for performing two-wavelength differential-phase interferometric imaging in a single digital image. This is not possible with existing types of digital holography or interferometry systems.

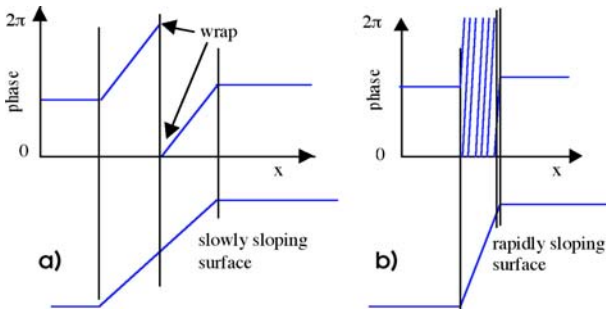


Fig. 1. In the gently sloping surface of (a), the phase can be unwrapped easily; in the steep sloping surface of (b), the problem is more difficult. Surface roughness can increase the difficulties in phase unwrapping tremendously. In addition, the limited number of pixels in an imaging sensor can also prevent unwrapping.

The technique of using two wavelengths to measure large objects is well known in digital holography, holographic contouring, and holographic interferometry.[3] In these techniques, phase information is obtained independently at two separate wavelengths. These two sets of phase data are then subtracted to transform the phase data to a scale length equal to the beat wavelength λ_b, where $\lambda_b = \lambda_1 * \lambda_2/(\lambda_1 - \lambda_2)$. Thus the phase is measured at an effective wavelength much longer than either of the two probing wavelengths. In this way, height variations many times greater than the original laser wavelengths used can be measured. This approach to two-wavelength measurements can be performed using DDH by sequentially (separate digital images) recording phase images at two different wavelengths, and then obtaining their difference. However, this sequential technique suffers from several limitations. First, noise in each individual image will be uncorrelated. The phase-image noise results in height errors proportional to the laser wavelength. When the difference between the two images is taken, the noise will be carried into the difference-phase image, but now it will be proportional to the beat wavelength λ_b. Thus the image quality and accuracy will be severely degraded. Second, any motion in the object between the acquisition of each image will result in pixel-to-pixel differences between the images. When the two phase images are subtracted, new errors will be introduced due to the pixel-to-pixel changes. Finally, acquiring each phase image sequentially increases the imaging time and limits the real-time ability of a potential system.

If the two phase-images could be obtained simultaneously using the same optics and digital camera, high-resolution differential-phase images could be generated. Image noise will be reduced because common-mode noise, vibrations etc. will be correlated and cancel out when the difference image is obtained. Since the two phase-images are obtained simultaneously in the same optics system with one digital camera, there is always exact pixel-to-pixel alignment and correlation. Finally, real-time topography is still achievable because only one image is required at each inspection location.

Technical Approach

This project will first demonstrate the ability to perform spatial-heterodyne interferometric imaging utilizing two laser wavelengths. By recording the two individual spatial-heterodyne interferograms and then taking the difference of the two phase images, object heights much greater than either individual wavelength can be measured. After demonstrating this novel technique, we will explore its capabilities and limitations.

Results and Accomplishments

We have successfully demonstrated the ability to record two spatial-heterodyne interferograms at different laser wavelengths. The recording of both wavelengths permits the imaging and inspection of objects with height variations much greater than either individual wavelength in real time. Utilizing two different laser wavelengths, we have accurately measured step height changes from a few hundred nanometers to over 6 μm. Figure 2 shows an example of a measurement made with the two-wavelength system on a sample with areas 2.86 μm high. The two individual wavelengths (left and center images) were unable to measure this height properly due to the many wraps; however, by beating the two wavelengths together, the third image shows proper measurement of the height. We have also imaged curved surfaces having tens-of-micrometers-of-height change over the image area. An example of this is the images of a convex mirror with a 52.8-mm radius of curvature shown in Fig. 3. The left and center images show the phase information obtained by the two individual wavelengths, while the right shows the beat wavelength image. This example quickly demonstrates the advantage of the beat wavelength. For this mirror surface, only one wrap appears in the beat wavelength image, while over 30 wraps occur in the individual wavelength images. The shape of the mirror

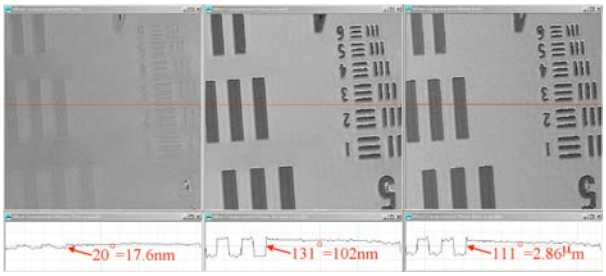


Fig. 2. Two-wavelength DDH imagery of 2.8-μm-deep resolution target. Left: 632.8-nm wavelength phase image; center: 611.9-nm wavelength phase image; right: 18.5-μm beat wavelength image.

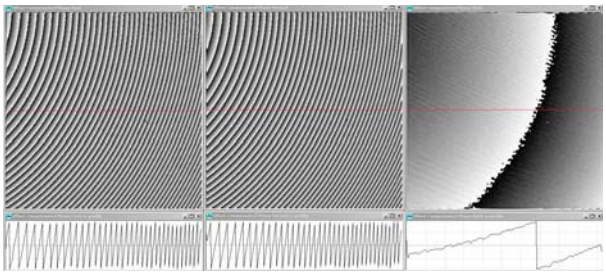


Fig. 3. Two-wavelength DDH imagery of concave mirror with 52.8-mm radius of curvature. Left: 632.8-nm wavelength phase image; center: 611.9-nm wavelength phase image; right: 18.5-μm beat wavelength image.

surface could be determined by unwrapping the individual wavelength images; however, phase unwrapping is a complicated problem that is greatly affected by noise in the image. A key limitation we have discovered is that the object surface roughness must be less than half the laser wavelength to successfully obtain a spatial-heterodyne interferogram. This limitation may reduce the range of potential applications.

We have utilized our experimental system to support the Spallation Neutron Source (SNS) Target Development program. The stainless steel vessel containing the mercury target material is damaged by cavitation in the mercury. Other inspection techniques available to the target group could not clearly resolve the depth of the pits caused by the cavitation. We used our system to measure the depth of these pits, which were from 10 to 50 μm across and from 1 to 10 μ/μm deep.

Summary and Conclusions

The successful demonstration of this technology has led to a joint proposal submitted to the National Institute of Standards and Technology—Advanced Technology Program (NIST-ATP) entitled "Spatial-Heterodyne Differential Digital Interferometry for MEMS Inspection." We are partnering with MicroSystem Technologies, Inc., to submit this proposal. Announcement of winners is expected by June 2003.

This project has also allowed us to build up a flexible laboratory capability to perform the spatial-heterodyne imaging. This has allowed us to begin a project with International SEMATECH to demonstrate some novel inspection capabilities of interest to the semiconductor industry.

References

[1]J. Strand and T. Taxt, "Performance Evaluation of Two-Dimensional Phase Unwrapping Algorithms," *Applied Optics* **38**(20), 4333–4344 (July 1999).

[2]D. Bone, "Fourier Fringe Analysis: The Two-Dimensional Phase Unwrapping Problem," *Applied Optics* **30**(25), 3627–3632 (September 1991).

[3]Wagner et al., "Direct shape measurement by digital wavefront reconstruction and multi-wavelength contouring," *Opt. Eng.* **39**(1), 79–85 (January 2000).

Signal Processing Architectures for Maskless Lithography

T. P. Karnowski,[1] D. Joy,[2] L. F. Allard,[2] and L. Clonts[1]
[1]*Engineering Science and Technology Division*
[2]*Metals and Ceramics Division*

In this work we analyze a redundant scanning array (RSA) geometry for low-energy, electron-beam maskless lithography. We then show estimates of the relationship between emitter spacing, gray levels, redundancy, and the printing accuracy as simulated by an image processing–based simple resist exposure and development model. We propose data compression architecture and algorithms to efficiently transfer control and printing data to the emitter array. Finally we present some experimental results determined with simulation software.

Introduction

Moore's law (i.e., the number of transistors on a microchip will double every 18 months) has become a driver for the semiconductor industry. Lithographic processes for printing device structures on integrated circuits (ICs) are the fundamental technology that has allowed this trend to persist. As device features shrink towards 130 nm, next-generation technologies for lithography are under frenzied development to replace optical lithography by the 100-nm technology node in 2005 and achieve 35-nm structures by 2014.[1] Electron-beam direct-write (EBDW) technologies provide the highest potential for maskless lithography (ML2). The International Technology Roadmap for Semiconductors (ITRS) states that "breakthroughs in direct-write technologies that achieve high throughput would be a significant paradigm shift [that] would eliminate the need for masks, offering inherent cost and cycle-time reductions."

ML2 methods such as EBDW have a unique advantage over the competing technologies: the long, tedious, and expensive process of fabricating a unique mask for the manufactured chip is not necessary. The costs of fabricating, inspecting, maintaining, and cataloging masks for alternate lithographic processes, along with the potential for yield loss associated with mask errors, are significant forces driving towards a maskless technology. While the increased mask cost can be amortized over the life of a particular IC product, such as a memory or microprocessor device, the ability to recover these costs in an application-specific integrated circuit (ASIC) foundry is highly problematic. In this case, a foundry rarely produces large quantities of devices for any given customer and the mask cost will be difficult to recover for sub-100-nm features.

As the ITRS states, the semiconductor community is extremely excited about the possibility for ML2. However, there are some rather daunting problems with establishing ML2 as a viable commercial technology. The data rate necessary for ML2 to be competitive in manufacturing is estimated to be in the 1- to 10-terabit/s range[2], which is not feasible in the near future. There is also doubt that the competing technologies for the writing mechanisms and corresponding photoresist (or analogous medium) will be able to accurately produce the desired patterns necessary to produce multilayer semiconductor devices.[3] Despite these issues, the payoff of a ML2 system is so great that there are still many ongoing efforts in developing technology for this exciting field. Thus, ML2 has three critical elements that must be addressed if a commercially competitive technology is to evolve: (1) the development of a robust array of signal emitters or alternate maskless lithographic device (MLD); (2) a resist medium that is responsive to the signal radiation of (1); and (3) a signal processing strategy and architecture that will enable competitive throughputs. All of the known institutions that are addressing this problem today are focusing primarily on one of these three aspects of maskless lithography. ORNL has been addressing part (1) through the development of an EBDW system composed of vertically aligned array of carbon nanofibers (VACNFs) under the auspices of a Defense Advanced Research Projects Agency (DARPA) funded program in Advanced Lithography. Other researchers investigating issues of device throughput are concerned largely with data compression.[2] These efforts explore the throughput as an image compression problem and do not explore issues of maskless printing accuracy, different maskless devices, or on-chip electronic processing. The crux of the issue is that the compression technique can depend not only on the mask pattern itself

but also on the scanning technique and writer/resist physics—an approach that the current EBDW data path estimates have failed to consider. Our specific goal with this effort is to develop compression techniques for candidate ML2 systems that include both algorithms and hardware implementations. If successful, we envisioned that this project and our complementary VACNF effort could potentially establish ORNL as a leader in maskless lithography.

Technical Approach

Our technical approach consisted of three main tasks.

Model Maskless Lithography Printing

We first modeled the maskless lithography printing process through two subtasks: analysis of the redundant scanning array (RSA) geometry as described in ref. 4 and modeling the impulse response of low-energy electron beams in resist material. These subtasks are designed to produce accurate models for simulations to help assess printing quality under a variety of different array geometries, which is directly related to the quantity of data needed to produce lithographic patterns.

Determine Required Control Signals through Simulation

We next investigated techniques for accurately determining what control signals were required. These methods were needed because the true data rates must be determined, as well as the actual data needed to print the patterns, before an effective compression scheme could be adopted or developed. Our approach first applied algorithms and methods for determining the best discrete "weights" for a particular array geometry and lithographic pattern, then assigning these weights to the redundant discrete emitters to optimize potential compression.

Design and Conceptualize Algorithms and Hardware for High Throughput

Finally, using results of the previous tasks, we sought to find compression techniques and hardware implementations that would exploit the array and data properties to produce high throughput solutions. This task was admittedly somewhat speculative and possibly the most risky endeavor of the project.

Results and Accomplishments

Model Maskless Lithography Printing

We first explored the geometry issues associated with a tilted RSA. The RSA is a practical technique for using massively parallel arrays of binary energy emitters to perform lithography. An array is passed over the resist medium at an angle so that the discrete emitters fall coincident to one another. Combined with clocked writings, a much finer pitch can be achieved than the actual array spacing. In addition, the use of an RSA can alleviate problems associated with nonuniform emitters and discrete firing locations. The basic concept is illustrated in Fig. 1. From our analysis, we determined that the desired tilt angle is given by

$$tan(\Theta) = -k\frac{M}{N} \quad ,$$

where k is the ratio of the x and y emitter spacing (generally unity) and M and N are positive-only integer values that represent the row and column difference between coincident elements. Consequently, there are only discrete angles that satisfy the co-incidence requirements.

A more significant result of the RSA analysis was the expected number of steps (i.e., discrete emissions) required per second S:

$$S = \frac{R\left(\frac{d_y}{d_r}\right)^2}{T} \quad ,$$

where R = redundancy (i.e., the number of times that an emitter passes over a specific site on the wafer), T is the writing time (60 seconds), d_y the spacing between emitters,

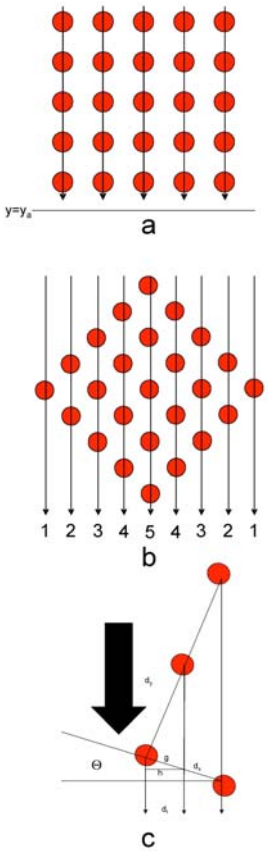


Fig. 1. The RSA concept. (a) An array with redundancy of 5. (b) Smaller array spacing produced with a tilted array. (c) Geometries and variables used to analyze the RSA.

and d_r the desired emission site spacing. We see that S is proportional to the square of the ratio of the actual spacing to the desired spacing. This makes sense since we are trying to create a fine discrete grid with more coarse elements in two dimensions, but such a result was not immediately obvious from the cleverness of the tilted array.

Our second subtask was the lithography system impulse response. We took the results of electron-beam profile experiments performed by David Joy and Larry Allard to create an approximate "system impulse response." The experiments were performed with Monte-Carlo simulations in PMMA, a resist material chosen for its general qualities. The results were not terribly encouraging, as the low energy used by candidate ML2 devices requires extremely thin layers of PMMA. There is considerable evidence that using a different type of resist—particularly chemically amplified resists—would be much more effective. We were not able to obtain material characteristics for any of these other candidates that would have allowed us to perform additional Monte Carlo experiments. A plot of this data is shown in Fig. 2. Nevertheless, this data was useful in creating a basic function and in examining how we may use it in performing maskless lithography. We were able to explore how scaling would be done on the function in association with different emitter spacings and redundancy factors or gray levels. We used an exponential curve fit to the data to determine a response modeled by

$$h(r) = \begin{cases} 1.8\frac{G}{R}, r \le 1 \\ \frac{G}{R} exp\left(-0.0283r^2 + 0.0613r - 1.0546\right) elsewhere \end{cases},$$

where G is the single-dose gain, R is the redundancy, and r is the distance from the center of the emitter. We show this response in Fig. 3.

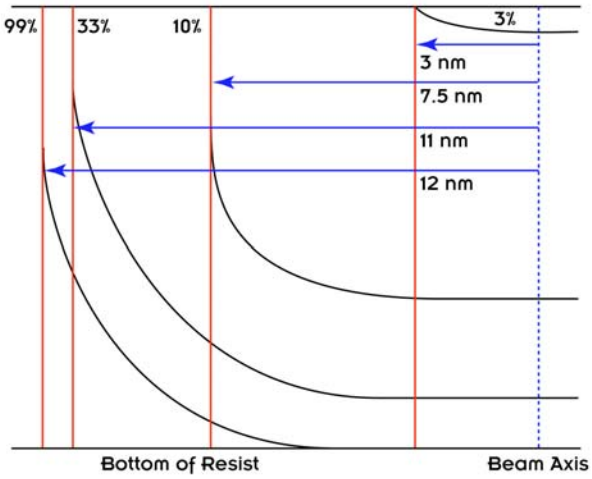


Fig. 2. The Monte Carlo simulations of the resist response.

Determine Required Control Signals through Simulation

The RSA geometry produces a number of discrete, approximately regularly spaced emission sites. We can model the printing process by weighting these sites with the number of "ON" emissions at each site, and convolving with the impulse response. Finally a threshold is applied to simulate the development process to produce a binary image that should match the desired pattern as closely as possible. These steps are illustrated in Fig. 4.

We used several different methods to try and determine optimal printing weights and examined the error produced by these methods. A few of our methods did not bear fruit, although they could stand to be revisited in the future. The most effective method was also the most time-consuming: an algorithm we call Direct Quantized Search (DQS) modeled after the direct binary search methods of ref. 5. We found that all our approaches produced errors, but of course more gray levels decreased errors. Determining what levels are appropriate is dependent on many factors, but we believe 16 gray levels seemed to work rather effectively with an effective (tilted RSA) spacing of 10 pixels, or 4 gray levels with a spacing of 5 pixels. Some examples of printed patterns with different spacings and redundancies are shown in Fig. 5. We were also able to use our simulation software to determine errors due to random emitter displacements and random emitter failures. The latter are shown in Fig. 6.

Design and Conceptualize Algorithms and Hardware for High Throughput

These results suggested a compression technique we called programmable basis functions (PBFs), which are geared toward very simple circuits that are located at each emitter and decode primitive code words for good compression rates. A block diagram of the circuitry for a decoder is shown in Fig. 7. When coupled with Huffman coding, they seem to be effective at reducing the data rate, but the rates we produced are still rather high because

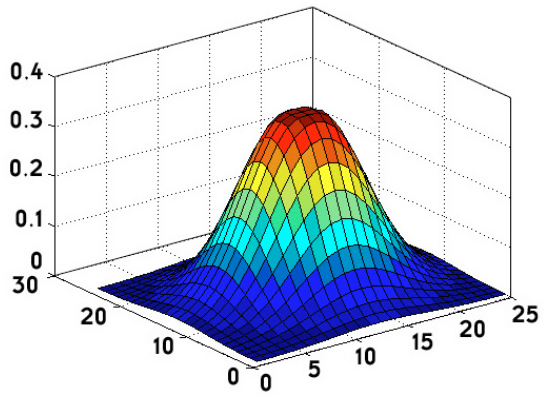


Fig. 3. 2-D plot of the modeled impulse response.

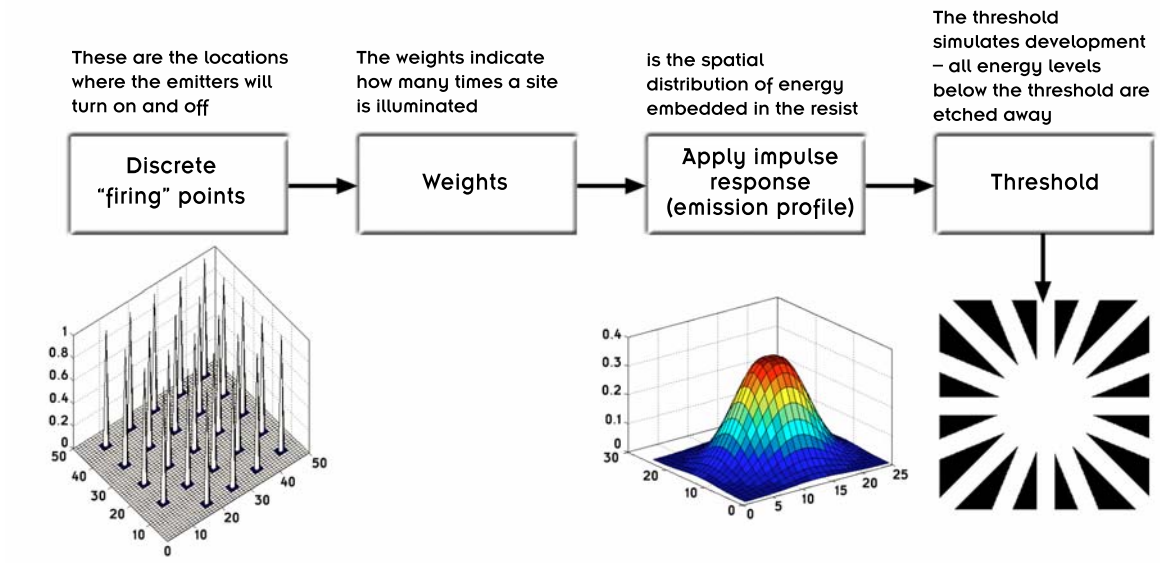


Fig. 4. System diagram of ML2 printing process simulations.

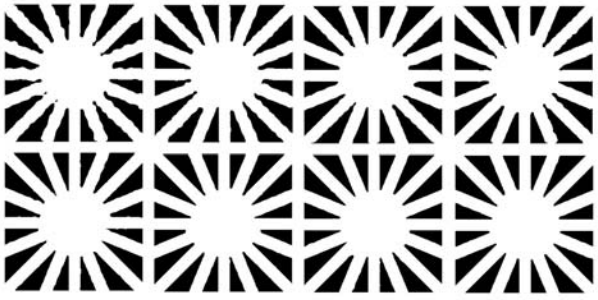


Fig. 5. Results with cross pattern and spacing of 10 (top) and 5 (bottom). R = 3,7,15,31.

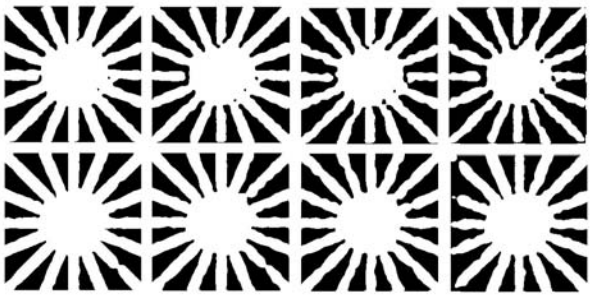


Fig. 6. Cross test images showing effects of random dead elements for 10%, 20%, 30%, and 40% bad elements. Top row: spacing = 10 pixels, R = 15; bottom row: spacing = 5 pixels, R = 3.

each decoder must be addressed at different times. In addition, for simplicity, we actually expand the data prior to transmission; for example, a redundancy of 15 can be represented as a 4-bit word, but because we seek simple circuitry, we actually issue 16 bits (prior to PBF and Huffman compression), which creates higher uncompressed data rates.

More encouraging, however, are the concepts of limited library compression (LLC) and a massively parallel optical data path. The LLC method was suggested by our printing simulation work. Researchers in this field have started with the concept that lossless compression is necessary, but the truth is that no ML2 device will produce a printed pattern that is exactly like the CAD files. The goal of LLC is to use a small library of control signals that produce a pattern that is "close enough" based on simulated print quality measurements. Similarly, we realized from our circuitry design and compression studies that an alternate data path would be beneficial. By using spatial light modulators or SLMs and using large arrays of photodetectors as the input devices, a more effective optical path could be created with relatively inexpensive components. This approach has some similarity to those used by some ML2 technologies, with the exception that our method transmits code words that can represent compressed data streams.

Summary and Conclusions

Our analysis revealed several problems we see with the creation of the ML2 system (besides creating the maskless writing device itself). These problems can be summarized as follows:

1. *The large data rates required*. Solutions seem to be compression methods (which take up a lot of chip real estate and require extremely large on-chip data rates as well) or a massively parallel optical solution (which requires either impossibly fast spatial light modulators, or photographic film storage mechanisms which may prove to be impractical).
2. *The ratio between the desired spacing and actual array spacing.* The tilted array is an ingenious idea,

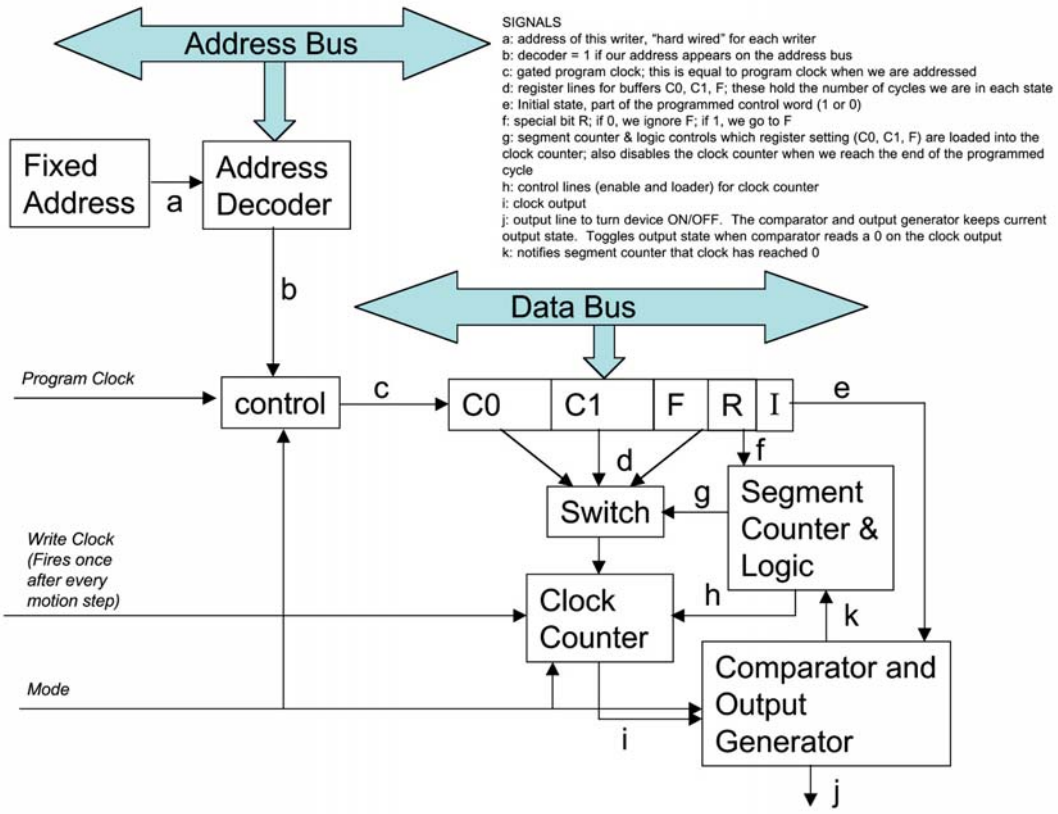


Fig. 7. Block diagram of PBF decoder circuit.

but it cannot get around the fact that many writing cycles are required to write a tight pattern with relatively sparsely spaced emitters. Due to the two-dimensional nature of the problem, the number of steps scales as the square of the ratio between the actual spacing and the desired spacing.

3. *The redundant scanning.* Again this is also an ingenious idea and very necessary to overcome the shortcomings of binary emitters, but the number of steps required scales linearly with the number of redundant elements desired.
4. *Creating the optimal weighting factors will be time-consuming.* This may be a fertile field for some clever solutions, however. The problem should be somewhat easy to scale, making it well suited for networked computer solutions.

We believe we have good solutions and directions for the following problems:

1. The LLC technique can help improve compression dramatically using very simple circuits for decoding.
2. Introducing signals through a massively parallel input stream, such as an optical path, would be an ideal way to input the data to the emitters.
3. Based on error measurements, we believe (if our simple thresholding is a reasonable practice) that lower gray levels than cited in ref. 2 can achieve effective results.
4. Missing elements can be a problem, as can motion problems, but these both are not show-stoppers if the impulse response of the system is considered in the weighting design.

Our next plans for this work involve three main efforts. We would like to determine if the LLC technique can succeed for larger circuits than our test patterns in these experiments. If that effort has merit, the compression technique could be useful for either an optical massively parallel method suggested here or even a more conventional electronic data path method. Finally, we explored some iterative techniques based on linear least-squares estimation and projections onto convex sets for the selection of the optimal weights. We believe these are important problems that could also bear fruitful research results.

This work has proceeded toward firmly establishing ORNL as a leader in maskless lithography systems engineering. We have submitted a white paper to selected government and nonprofit agencies (such as the Defense Advanced Research Projects Agency's Microelectronic Program), and we are optimistic that we will be able to

continue this research. In particular, our invention disclosures that are the direct result of this work represent radical new ideas for enabling technologies for maskless lithography which we believe will directly impact the future of this field.

References

[1]International Roadmap for Semiconductors, Semiconductor Industry Association, 1997.

[2]V. Dai and A. Zakhor "Lossless Layout Compression for Maskless Lithography Systems," *Proceedings of the SPIE: Emerging Lithography Techniques IV* **3997**, 467 (2000).

[3]D. Lammers, "EUV gains as venture ends e-beam litho work," EE Times, http://www.eetimes.com/story/OEG20010105S0023.

[4]L. R. Baylor, D. H. Lowndes, M. L. Simpson, C. E. Thomas, M. A. Guillorn, V. I. Merkulov, J. H. Whealton, E. D. Ellis, D. K. Hensley, and A. V. Melechko, "Digital electrostatic electron-beam array lithography," *J. Vacuum Science & Technology B* (Microelectronics and Nanometer Structures) **20,** 2646 (2002).

[5]J. P. Allebach, "DBS: retrospective and future directions [direct binary search]," p. 358 in *Proceedings of the SPIE, Color Imaging: Device-Independent Color, Color Hardcopy, and Graphic Arts VI*, Vol. 4300, 2001.

Fluorescence-Based Coatings Diagnostics

S. W. Allison and D. L. Beshears
Engineering Science and Technology Division

The protection of turbine blades and vanes by means of thermal barrier coatings (TBCs) is one of the key techniques that enable the envelope of turbine-engine operating temperature to be increased. Ensuring TBC integrity is vital for preventing hot section component failure. The role of the TBC is critical because failure of the TBC to be durable and to perform properly leads to severe engine damage. For example, if the TBC spalls off, it leads to temperature rise of the underlying substrate and exceeding of operational limits with possible subsequent failure. Ensuring TBC integrity is of concern to more than just the propulsion engine community, as evidenced by an assessment of the needs for sensors and controls for land-based advanced gas turbines that was performed for the Department of Energy.[1] This report indicated that, out of several critical instrumentation and sensor areas of concern, the requirement for a TBC failure and health monitoring capability was the number one need. We therefore conclude from this and other communications that a means for detecting TBC spallation on-line is of considerable value. The goal of this project was to develop approaches for long-term monitoring of TBC integrity via fluorescence techniques. The primary application is for operating turbine engines, but similar surfaces in other types of engines should also be viable.

Objective and Technical Approach

The objective of this project was to explore luminescence-based methods for diagnostics of TBC coatings. Layering of fluorescent materials can be the basis for heat-flux measurement, wear monitoring, temperature measurement, and overall coating health determination. Such a coating may consist of a thermal barrier coat, typically ytrria-stabilized zirconia (YSZ), and a thermographic phosphor on top and/or underneath the YSZ. As it turns out, YSZ can also be activated to fluoresce by doping with a rare earth so that it may function as a thermographic phosphor as well. The vision is to attain coatings whose layers are of controlled, known thickness and whose emissions are distinct. By monitoring the ratios of the characteristic emissions, the thickness and overall coating health may be determined. In principle, if the environment is clean, only two coatings may be required. In the event that soot or other material builds up on the coating, a third layer can provide added benefit by enabling the degree of attenuation of the soot coating to be inferred. By using an imaging system in conjunction with two or more fluorescent coatings, cracks may be more easily discernible. In such an application, incipient problems could be detected before catastrophic failure occurs.

The inset of Fig. 1 depicts three representative coatings for purposes of illustration. YSZ:Dy fluoresces in the yellow-orange and blue, YSZ:Tb fluoresces in the green and blue, and YSZ:Eu fluoresces mainly in the red. However, YSZ absorbs well in the near UV and blue so that appropriate visible wavelengths should be used for realistic thicknesses of the material. The physical properties of the coating are not expected to be significantly affected by the small doping level.

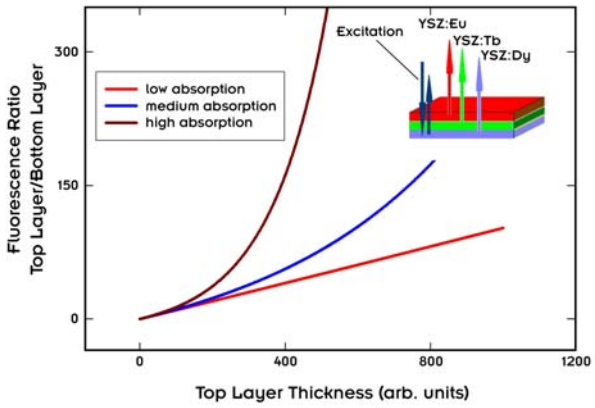


Fig. 1. Plot of fluorescence ratio versus top layer thickness.

A simple analysis provides instructive illustration of our approach. It begins with a derivation in ref. 2, which developed equations for emission from films of fluorescent dye mixtures. Consider a differential element of area in a film that is illuminated with excitation light. The fluorescence from such an element is

$$dI_f = I_e \cdot \alpha \cdot Q \cdot dx \ , \qquad (1)$$

where I_e is incident intensity, α is an effective absorption coefficient which includes scattering losses, Q is the quantum efficiency of the fluorescence, and dx is the film thickness. The well-known Lambert-Beer relation teaches that I_e will decrease as the excitation light penetrates the material according to

$$I_e(x) = I_0 e^{-\alpha x} \quad . \tag{2}$$

The differential fluorescing element becomes

$$dI_f = (I_0 e^{-\alpha x}) \cdot \alpha \cdot Q \cdot dx \quad . \tag{3}$$

The total fluorescence received by a detector for a single layer of thickness d is the integral of this expression times an appropriate collection efficiency, η

$$I_f = \eta \cdot \int_0^d \alpha \cdot Q \cdot I_0 \cdot e^{-\alpha x} dx = \eta \cdot I_0 Q \cdot (1 - e^{-\alpha d}) \tag{4}$$

$$I_f = \eta \cdot I_0 \cdot Q \cdot (1 - e^{-\alpha d}) \tag{5}$$

If we generalize and apply the above analysis successively to layers of two different materials of thicknesses d_1 and d_2, absorption coefficients of α_1 and α_2, for layers attached to a surface of reflectivity R, we may obtain the following two expressions for the received fluorescence at each fluorescence wavelength.

$$I_1 = I_0 \cdot \eta \cdot Q_1 \cdot \{(1 - e^{-\alpha_1 \cdot d_1}) + R \cdot (e^{-\alpha_1 \cdot d_1 - 2 \cdot \alpha_2 \cdot d_2}) \cdot (1 - e^{-\alpha_1 \cdot d_1})\} \tag{6}$$

$$I_2 = I_0 \cdot \eta \cdot Q_2 \cdot \{(e^{-\alpha_1 \cdot d_1}) \cdot (1 - e^{-\alpha_2 \cdot d_2}) + R \cdot (e^{-\alpha_1 \cdot d_1 - \alpha_2 \cdot d_2}) \cdot (1 - e^{-\alpha_2 \cdot d_2})\} \tag{7}$$

The ratio of I_1/I_2 is plotted in Fig. 1 as a function of the thickness, d_1, of the top layer. As expected, it is a well-defined function of the thickness of the top layer. If the absorption of the excitation light is low, then the dependence is approximately linear. Modeling reveals that the value of the substrate reflectivity does not change the position of the curves significantly. A more detailed model could be developed (for example, taking into account absorption of fluorescence in the layers which we now know can be significant for some of the shorter wavelengths), but what is presented here will suffice for the present purpose of illustrating our goal.

Results and Accomplishments

Given the above analysis, a number of steps are necessary to move forward toward demonstrating useful coatings diagnostics:

1. Determine the transmission of YSZ material.
2. Verify that top and bottom coatings can be viewed simultaneously throughout a useful temperature range.
3. Verify that rare-earth-doped YSZ material can fluoresce.
4. Place test coatings in operating degrading environments for in situ wear.
5. Obtain and test layered materials of known thicknesses.

In this project, we have collaborated with researchers at the NASA Glenn Research Center to address the first three steps, the results of which have been published.[3] Briefly, in regard to step 1, it was determined at NASA that YSZ is highly absorbing in the ultraviolet, near-ultraviolet, and into the blue region of the spectrum. This observation led to the use of visible excitation sources for further investigations. With respect to step 2, it was demonstrated both at ORNL and NASA that several phosphor materials coated underneath a YSZ could be excited from the top of the coating with green light at 532 nm. Emission could be distinguished spectrally as expected and also temporally. For example, YAG:Ce emission has a characteristically short decay time, around 60 ns, which is easily separable from typical longer-lived fluorescence that is many microseconds in duration. Measurements of top and bottom coatings were taken to 1100°C. With regard to step 3, several doped-YSZ materials were procured and emission spectra measured. This is represented in Fig. 2, which shows the characteristic emission spectra of two such materials, YSZ:Dy and YSZ:Eu. These materials exhibited the expected relatively narrow band emission that is desired. Future tests will characterize these materials versus temperature. For step 4, follow-on work is in progress regarding testing coatings in an operating environment. With regard to step 5, we did not obtained samples for detailed testing before the project completion date.

Overall, excellent progress in the development of fluorescent coatings techniques was made. During the course of work with NASA, an additional unexpected benefit appeared. Careful observation at NASA of a Y_2O_3:Eu coating measured the temperature dependence of the risetime of this material in this situation. We continued a study[4] of this emission feature and concluded that measurement of risetime promises to be another useful tool for thermometry and coatings health monitoring. The decay time of Y_2O_3:Eu is a sensitive function of temperature above about 600°C, as is well known, and it has been used effectively for thermometry above that value in the past. The risetime has the advantage that it is temperature dependent from at least ambient to well above 600°C. We have made a detailed study of the temperature dependence from a crystalline sample of Y_2O_3:Eu with the dopant at 0.5% and have compared it with the NASA data. The results are presented in Fig. 3. It is seen that our crystalline sample has a greater temperature dependence as well as longer overall risetimes. It appears that dopant concentration affects the risetime value and its temperature dependence. It is seen from the figure that temperature sensitivity is increased by decreasing the Eu concentration.

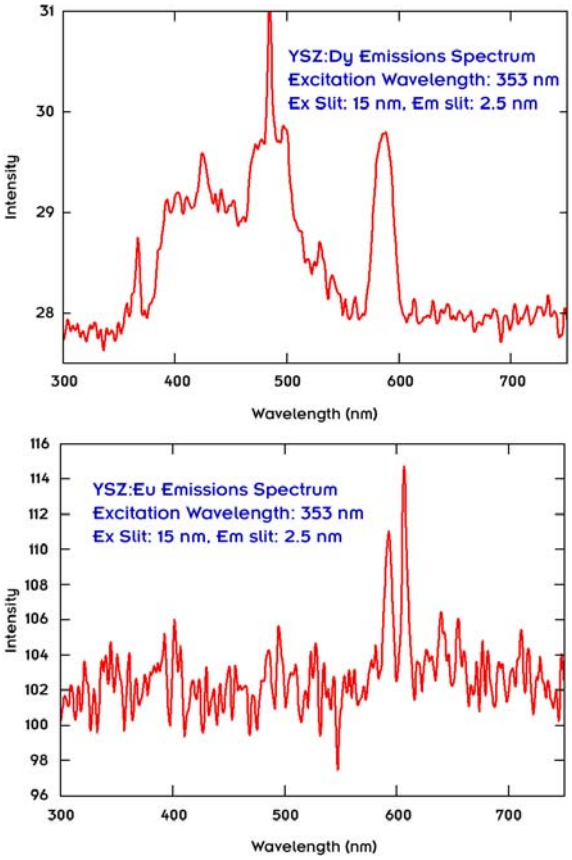


Fig. 2. Emission spectra of YSZ:Dy and YSZ:Eu.

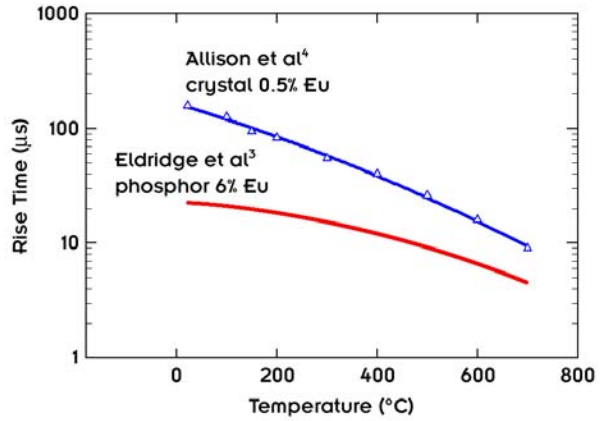


Fig. 3. Comparison of fluorescence risetime versus temperature for different Y_2O_3:Eu samples.

Summary

DOE, the Air Force, and NASA all have applications that can benefit from this research. In particular, we have participated with the Propulsion Instrumentation Working Group on a proposal for a specific Air Force program aimed at developing a thermal barrier coat health monitoring system for very-long-distance fighter jet turbine applications. The first such proposal was not successful, but we envision other opportunities. That proposal involved Rolls Royce, Pratt Whitney, Honeywell, Rolls Royce, and NASA Glenn. We are presently directing our effort at participating in a new joint US-UK Energy Partnership which DOE has promised to fund in FY 2005. A white paper has been submitted for turbine surface health monitoring.

References

[1]R. L. Anderson et al., *Advanced Turbine Systems Sensors and Controls Needs Assessment Study Final Report*, Oak Ridge National Laboratory report ORNL/TM-13335, February 1997.

[2]C. H. Hidrovo and D. P. Hart, "Emission reabsorption laser induced fluorescence (ERLIF) film thickness measurement," *Meas. Sci. Technol.* **12**, 467–477 (2001).

[3]J. I. Eldridge, T. J. Bencic, S. W. Allison, and D. L. Beshears, "Depth-Penetrating Temperature Measurements of Thermal Barrier Coatings Incorporating Thermographic Phosphors," accepted for publication in the *Journal of Thermal Spray Technology*.

[4]S. W. Allison, S. M. Goedeke, M. R. Cates, W. A. Hollerman, and T. J. Bencic, "Advances in High Temperature Fluorescence-Based Thermometry," CD-ROM, Huntsville, AL USA, Advanced Sensor Development Consortium Meeting, NASA MSFC, Huntsville, AL USA, 05/13/2003–05/15/2003.

Ultraviolet Electroluminescent Device Development

S. W. Allison[1], A. Akerman[1], and P. Rack[2]
[1]*Engineering Science and Technology Division*
[2]*Materials Science Department, the University of Tennessee.*

The development of solid-state light sources in the ultraviolet has been a long-term goal of defense and commercial organizations. Many millions of dollars have been directed at extending the III-V nitride blue lasers and light emitting diodes into the ultraviolet region. The present endeavor aimed at pursuing a different approach, that of electroluminescence (EL) of rare-earth doped materials. The advantages include inherent ruggedness, narrow spectral bandwidth, and the fact that such light sources may be made arbitrarily large in area. Sensing, illumination, and security applications are anticipated.

Introduction

The goal was to develop and demonstrate an ultraviolet-emitting, alternating current, thin-film electroluminescent (ACTFEL) device. The approach utilized gadolinium-based phosphor materials. These are known to produce emission lines in the ultraviolet as a result of transitions of the Gd^{3+}activator. The two most prominent transitions are at 312 and 275 nm. The purpose in building and testing ultraviolet emitting devices was to demonstrate the feasibility for a variety of applications. An additional motivation was to acquire valuable experience in designing and fabricating electroluminescent devices. Miniature UV sources have applications such as biological agent detection, communications, lighting, sensors, and security applications. The attributes of ruggedness and all solid-state construction are of obvious advantage to these and other applications.

Technical Approach

Two approaches were pursued in demonstrating ultraviolet electroluminescence. One aimed at learning how to properly sputter thin films of the relevant materials for subsequent ACTFEL devices. The other explored the powder electroluminescence approach. It should be noted that the most efficient means for producing electroluminescence is from thin films though powder EL, an inexpensive and easy means for producing light output. During the course of this project, a magnetron sputtering device was acquired by the University of Tennessee and its first use was for depositing thin films of the desired materials.

Results and Accomplishments

Materials Development and Characterization

The first step was to characterize the ultraviolet-induced emission from candidate phosphor materials. Three types of Gd-doped phosphors were ordered and received. They all photoluminesced with reasonable efficiency. Interestingly, the emission of the ZnF_2:Gd phosphor is shifted slightly, about 2 nm, to higher energy with respect to the Y_2O_3:Gd and YAG:Gd phosphors. This difference may be useful spectroscopically. A comparison of their emission spectra with respect to an SO_2 absorption spectrum shows a different response to the respective wavelengths and may indicate a useful approach to SO_2 detection. Tests with ultraviolet excitation revealed that that photoluminescence from the Y_2O_3:Gd and YAG:Gd phosphors were comparable and about ten times greater

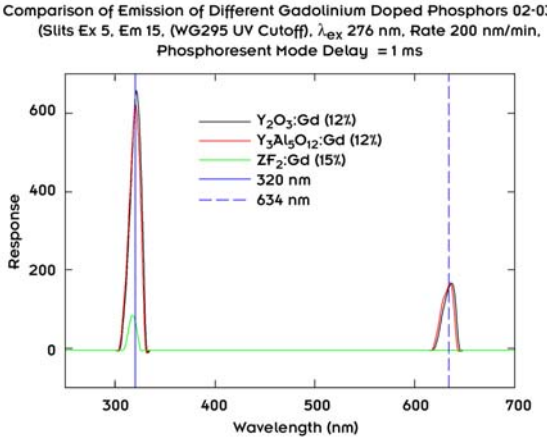
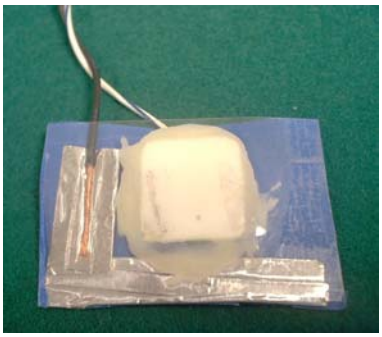


Fig. 1. Plot of ultraviolet photoluminescence of Y_2O_3:Gd, YAG:Gd, and ZnF_2:Gd.

than that for ZnF_2:Gd. This is depicted in Fig. 1. We recently acquired some thulium phosphors for other applications. It turns out that the Tm activator also exhibited useful, narrow-band near-UV light. Also, other standard visible electroluminescent phosphors, ZnS:Cu and ZnS:Mn, were characterized and exhibited their standard characteristics.

The next step in gaining experience and facility with electroluminescence was to construct a powder electroluminescence device. Commercially available night- lights are common examples of powder electroluminescence. Infrared, visible, and ultraviolet electroluminescence was achieved with some lab-made powder EL devices. Figure 2(a) is a photograph of one of a number of lamp assemblies that were made and tested for UV emission using many of the same potential UV candidate phosphors. A commercial night-light was cut and layer thicknesses measured. Following the layer structure and thickness measurements and literature accounts, electroluminescent structures were assembled using plastic, glass, or quartz windows coated with indium tin oxide (ITO) and tested using DC and AC excitation. Conditions were found under both types of excitation for electroluminescent operation in the visible using ZnS:Mn and ZnS:Cu. These glowed orange-yellow and blue-green, respectively, as expected and also exhibited a charging effect described in some literature accounts. Additional structures were built on the same lines, however, with the substitution of YAG:Tm, YAG:Gd, ZnF:Gd (various percentages and sources), Y_2O_3:Nd, and ZnS:Ag. Electrical excitation was provided in the same way as for the visible EL lamp structures, that is, using a high-voltage transformer.

The powder EL lamp consists of a phosphor layer sandwiched between two thin, clear-plastic insulators that in turn are sandwiched by brass screens. As an example, for ZnS:Mn, for transformer input at 450 Hz and 2000–4000 V, the center layer glowed orange-red through the brass screen; the higher the voltage, the brighter the light. With a stack thickness of 0.009 cm, the electric field was 2×10^5 V/cm to 4×10^5 V/cm. For YAG:Gd, using the same structure, but substituting phosphor layer, we observed UV emission at 312 nm as well as another UV wavelength, as was observed both with a monochromator and by secondary excitation of fluorescence from ZnS:Cu. Using the same structure, but substituting Y_2O_3:Nd, we observed IR with filters and a night vision scope, but not with the monochromator, possibly due to sensing choices. In fact this aspect of the project ended when serious mechanical problems with the monochromator developed. In addition to the brass screen, combinations of aluminum and partially transparent electronic ink–coated transparent insulators were used for electrodes with good results.



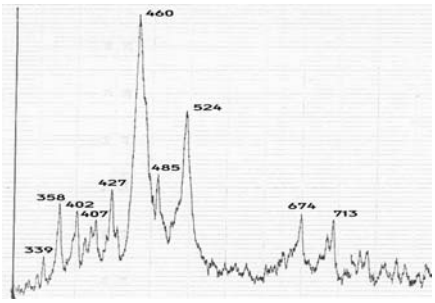


Fig. 2. (a) Photo of powder electroluminescence lamp and (b) emission spectrum of YAG:Tm powder luminescence lamp.

For some of the powder EL devices we detected some near ultraviolet emission that is characteristic of nitrogen. Other researchers see this in some instances also. This is not necessarily a disadvantage, just another source of ultraviolet light.

Representative results are presented in Fig. 2(b) for which a YAG:Tm powder EL lamp was tested using the monochromator scanning from 300 to 600 nm. The result was a spectrum similar to several seen in the literature with a peak near the expected emission, but with stronger peaks at 350 and 374 nm, also seen in the same papers. Using a Japanese ZnF:Gd formulation in a structure with pulsed excitation, a strong signal was observed with photomultiplier through a 313-nm narrow-band filter. Another filter centered at 360 nm allowed 1/5 the amount of signal under the same conditions. Turning the room lights on would "recharge" the lamp and allow a few minutes operation, typical of many of the visible structures fabricated. Subsequent attempts to measure the emission with a monochromator and photomultiplier were not successful due to the aforementioned mechanical problem.

Thin Film Electroluminescence

Gadolinium-doped thin films in yttrium oxide and yttrium aluminum garnet ($Y_3Al_5O_{12}$) were deposited on silicon substrates by radio-frequency magnetron sputtering.They exhibited the desired cathodo-luminescence (CL). This was typically on the order of

0.1%, depending on deposition parameters: substrate temperature, total chamber pressure, and the oxygen partial pressure. Post-process annealing also plays a role in final efficiency attained. Luminescence efficiency improved with increasing film crystallinity for films deposited at room temperature and subsequently annealed at 1000°C. Figure 3(a) shows a scanning electron microscope image of an Y_2O_3:Gd multilayer film, deposited on (001) Si, cleaved and shown in cross section. The film was deposited at room temperature and consists of five layers of Y_2O_3 and four layers of Gd. Figure 3(b) shows the emission spectrum for YAG:Gd and the emission intensity as a function of the dopant concentration. This curve peaks close to 6%. The corresponding figure for Y_2O_3:Gd is between 9 and 10%. These values are less than the maximal dopant concentrations of the neat phosphor powders. It is likely that this is due to thickness considerations. The details of the thin film development and characterization are provided in reports resulting from this effort.[1–4]

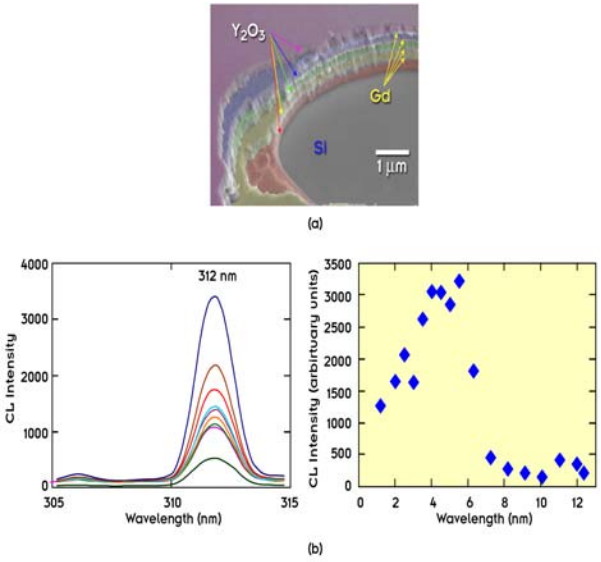


Fig. 3. (a) Y_2O_3 layer deposited by reactively sputtering yttrium in PO_2 = 0.212 mTorr and P_{Ar} = 2.788 mTorr with a forward yttrium gun power of 200 W. The Gd layer was deposited at 50 W in P_{Ar} = 3 mTorr and (b) cathodoluminescence intensity versus Gd concentration in the YAG films.

The one test of a thin film as part of an EL device on a substrate provided gratis by a commercial firm did not produce detectable luminescence. More testing would surely be successful.

Summary and Benefits

Ultraviolet electroluminescence was demonstrated using the powder EL approach. Progress in sputtering and optimizing high-quality thin films for future use in thin film alternating current electroluminescent devices was made. The cathodoluminescence of the YAG and Y_2O_3 materials, doped with Gd, was significant. This may have positive implications for lighting since YAG materials are highly desired for white-light light-emitting diodes. We will further explore this application. Also, the powder luminescence approach may be pursued as a spectroscopic light source that is of simple construction and may be of use in high- temperature and hazardous environments as a robust spectroscopic ultraviolet light source. We will commence investigating opportunities within the Defense Advanced Research Projects Agency where this technology may be best suited.

References

[1]J. D. Fowlkes, P. D. Rack, R. Bansal, and J. M. Fitz-Gerald, "Cathodoluminescence Study of Gadolinium—Doped Yttrium Oxide Thin Films Deposited By Radio—Frequency Magnetron Sputtering," *Proceedings of the New Applications for Wide Bandgap Semiconductors Symposium of the Materials Research Society Spring Meeting*, **764**, C7.12 (2003).

[2]Y. Deng, J. D. Fowlkes, J. M. Fitz-Gerald, and P. D. Rack, "Combinatorial Thin Film Synthesis of Gd-doped $Y_3Al_5O_{12}$ Ultraviolet Emitting Materials," *Applied Physics D* (in press).

[3]J. D. Fowlkes and P. D. Rack, "Structure and Cathodoluminescence Relationships in Y_2O_3:Gd Thin Films Prepared by rf Magnetron Sputtering," *Applied Physics A* (manuscript in preparation).

[4]Y. Deng, J. D. Fowlkes, and P. D. Rack, "Structural and Cathodoluminescence Properties of $Y_3Al_5O_{12}$:Gd Films Prepared by a Combinatorial rf Magnetron Sputtering Technique," *Journal of Vacuum Science and Technology B* (manuscript in preparation).

Highly Reflective Subwavelength Structures for Homeland Security Initiatives

S. L. Jaiswal and S. R. McNeany
Engineering Science and Technology Division

In synchrony with many initiatives currently being undertaken in the interests of homeland security, the development of highly reflective subwavelength structures (SWSs) has application in areas of tracking, tagging, locating, and identification. These subwavelength structures are essentially customized optical surfaces, which become highly reflective when illuminated with light of a specific wavelength. Unlike other retro-reflectors, mirrors, and reflective paints, these subwavelength structures, if designed correctly, can offer high reflectance without limited wavelength selectivity, polarization selectivity, and angular sensitivity. The structures are essentially located in microscopic optical waveguides, with imprinted features, having dimensions smaller than the wavelength of the incident light. This offers an additional advantage over other highly reflective sources, which tend to be larger in extent and therefore difficult to conceal from visual inspection. This project investigated subwavelength optical structures with reflectivities in the infrared where long-range propagation (>5 km) of light in the atmosphere is possible. This is of particular interest due to the low absorption of infrared light with wavelengths in the range of 9–11 μm in the presence of fog, smoke, or dust. Thus applications of tags include remote identification, tracking, and locating at distances of 5 km or greater under a variety of environmental conditions.

Introduction

Examples of subwavelength optical structures, which provide customized reflectivities, occur in nature. One of the more publicized examples, discovered in the 1960s, is the moth eye effect. In this case, the moth's eye has been cleverly engineered to contain subwavelength-sized cones (cones with dimensions smaller than the visible light) that provide an antireflection coating[1,2] to conceal the presence of the moth. The SWS cones form a graded index layer in which Fresnel reflection losses are minimized at the surface. Previous work to customize surface reflectivities using physically etched structures has concentrated on coupling the incident light into an embedded waveguide via a surface relief grating. This configuration is termed a "guided mode resonance filter" and results in a multiple- layer structure that responds to a very narrow range of wavelengths.[3] In this proposal, a simpler configuration is employed, in which the surface relief grating acts as the waveguide itself. This structure is advantageous over existing technology since it offers high reflectance, without limited wavelength selectivity, polarization selectivity, and angular sensitivity, and can be employed for remote tracking (>5 km) with reasonable discretion.

Technical Approach

The use of periodic structures within a waveguide medium enables a form of optical control. In the commonest example, a diffraction grating systematically forces a spatial redistribution of uniformly illuminating light into a series of dark and bright regions, termed "diffraction orders." It is the design of the spatial frequency of the diffraction grating, relative to the frequency of the incident light, that determines the diffraction pattern. While diffraction patterns occur when the separation of ruled lines (or periodic structures) is just greater than the wavelength of incident light, it is interesting to explore the consequences of having the spatial separation smaller than the wavelength of the incident light. This is the subwavelength regime. By decreasing the separation, the resulting diffracted orders become more orthogonal to the incident light, until eventually a resonant condition is achieved at a specific separation where light coherently scatters from the interface between high and low refractive index, in such a way as to form a highly reflective beam and no transmitted orders. By depositing a thin film onto a lower refractive index substrate and patterning and etching the surface, periodic subwavelength relief structures are created that allow the surface itself to become an effective waveguide. These structures have a feature size of about 5 μm for the infrared regime under consideration (9–11 μm). Creating many of these structures on the order of magnitude of a "dust particle" and randomly orienting them can produce reflective beams from a wide range of incident angles, eliminating any angular selectivity concerns. In addition, these SWSs are consistent with the properties of photonic crystals, so that the feature size scales with wavelength, to produce high reflectance for any desired wavelength.

Design: The key factors to be considered are the waveguide materials and the design of the subwavelength structures in the wavelength of interest (9–11 μm in this case). Commercially available simulation software called Gsolver, based on rigorously coupled wave analysis, can easily model these structures.

Fabrication: The first goal is to create a thin-film waveguide. The likely candidate for the waveguide material is germanium, due to its high refractive index. Candidates for the substrate are $BaFl_2$, ZnSe, and ZnS with lower refractive indices, but this will be determined by the modeling results. The coating material will be evaporated onto the substrate using an electron-beam deposition tool. After determining the periodicity and feature size of the SWS, a photolithography mask will be designed and prepared. The SWS features are then patterned onto the surface using photolithography with sufficient resolution to perform the task. In order to etch the pattern into the waveguide, a reactive ion etcher will be employed. This will require the development of an etch recipe for the material of choice. Finally testing will be realized through a tunable (9–11 μm) CO_2 laser, centered at 10.6 μm, by illuminating the structure at normal incidence and using an appropriate infrared detector to measure reflectance.

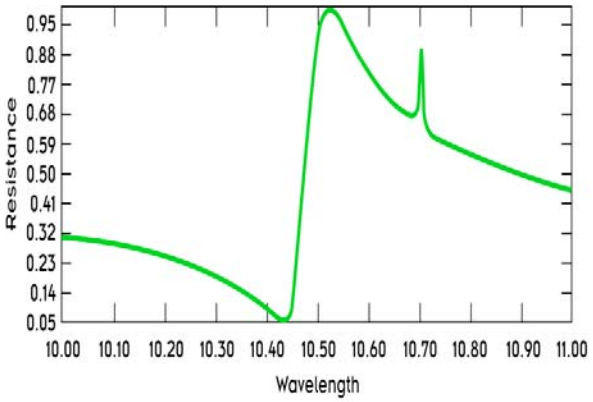


Fig. 1. Reflectance of a subwavelength structure in a germanium waveguide.

Results and Accomplishments

As noted above, the Gsolver computer program was used to model the optical performance of proposed subwavelength structures. Figure 1 gives an example of the calculated reflectance from SWS (an array of holes) in a 1-μm thin film of germanium on $BaFl_2$. Note the expected resonance peak in the reflectance curve near 10.6-μm wavelengths. Nearly total reflectance is expected at the resonance peak.

In an attempt to verify this type of optical behavior, efforts were undertaken to fabricate such structures at a micro-optics fabrication facility at Penn State University. Only one sample was produced that was tested for reflectance using a tunable CO_2 laser at ORNL. The sample was scanned over a wavelength range of approximately 9.5–11.5 μm. Unfortunately, the test results showed very little reflectance from the sample at any wavelength and no evidence of resonance structures.

Summary and Conclusions

Calculations of the optical performance of thin-film-based subwavelength structures clearly indicate that very high reflectance is possible. Unfortunately, the efforts to produce such structures as part of this project were unsuccessful. Nevertheless, because of the numerous potential applications and the high level of interest by potential sponsors for materials such as those investigated in this project, future research in technology will continue to be pursued.

References

[1]C. Bernhard, *Endeavor*, Vol. 26, p. 79, 196.

[2]P. B. Clapham, *Nature* **244** (5414), 281 (1973).

[3]R. R. Boye, "Physical Optics Approach to Guided Mode Resonance Filters," Ph.D. dissertation, University of Arizona, 2000.

A New Approach to Chemical Detection: Infrared Coherent Anti-Stokes Raman Scattering

R. K. Richards,[1] J. B. Wilgen,[1] M. R. Ally,[1] and M. Z. Martin[2]
[1]*Engineering Science and Technology Division*
[2]*Environmental Sciences Division*

We propose to revolutionize current thinking on making ultraperformance, small detectors for homeland defense, national security, and civilian uses by taking advantage of responses of chemicals in the relatively long wavelength electromagnetic (EM) spectrum for unambiguous identification of chemical threats. This is a completely unexplored area of investigation in sensor and detector development. This idea takes advantage of carbon dioxide (CO_2) lasers for a proof-of-principle demonstration. The technique is Coherent Anti-Stokes Raman Spectroscopy (CARS) using CO_2 lasers.

A proof-of-principle test was conducted using an innocuous chemical that has a nonlinear response in the microwave (long-wavelength) region. This idea has the potential to qualitatively change sensor and detector technology to exceed goals of the Pentagon, U.S. Department of State, and civil defense agencies for reliable, real-time information in case of chemical attacks or threats. We aim to provide early warning of a chemical warfare agent and eventually biological warfare agents released intentionally, for example, in a populated area to enable civil defense agencies to respond quickly. We should also be able to keep the first-generation unit cost under $10,000. We are motivated by our analysis to seek funding from DOE's Building Technology Program, Homeland Security Office, DoD, or DARPA to develop and demonstrate this potentially rewarding technology.

Introduction

Longer-wavelength CARS will have advantages over conventional CARS at shorter wavelengths, generally in the visible. The photon energy at 10-μm wavelength is only 0.1 eV (compared to 1.2 eV at 1 μm and 2.5 eV for visible CARS). Therefore, fewer resonances can be driven energetically by the lower-energy pump beam and the resulting spectra should be simpler. Appreciation of this idea is exemplified by considering the different responses of water molecules to visible and microwave energies. Heating water with visible light is inefficient even though visible light has higher energy per photon than microwaves. The reason why long-wavelength (such as microwave) heating of water is efficient (and convenient) is that water, like other chemical substances, has long-wavelength resonant frequencies, which respond well to this type of excitation. We intend to exploit the rotational and lower frequency modes of target compounds for positive identification. In addition, the frequency resolution provided by the heterodyne receiver will easily surpass even the highest-resolution visible spectrometers available today. Even simple molecules such as ammonia have extremely complicated infrared spectra. Simpler spectra with the ultimate in frequency resolution are the motivation for developing this technology to overcome a distinct limitation of existing instruments. Our goal is to build the next generation of high-performance, low-cost, extremely reliable, low-maintenance instruments for counterterrorism and homeland security.

Technical Approach

Ultrahigh-resolution CARS can be much more easily achievable at 10 μm than in the visible or near infrared. For CO_2 CARS, both pump beams would be produced by pulsed CO_2 lasers. The coherent output beam resulting from the CARS process would also be in the CO_2 wavelength range. A CO_2 heterodyne receiver would be used to down-convert the scattered signal to an intermediate frequency (IF) range of 1–30 GHz. By tuning the CO_2 local oscillator for the heterodyne receiver to make use of different lines in the 10 P series, a frequency range of 500 GHz, or more, can be covered. That allows access to >100 molecular lines from even a light molecule like methanol. Only this down-converted IF signal would be in the microwave/millimeter frequency range. Frequency resolution would then be determined by the resolution of the microwave/millimeter–wave spectrum analyzer, where frequencies can easily be measured to a fraction of a MHz. A significant part of the challenge is in achieving the necessary response time, as the frequency spectrum must be measured during a pulse time of <100 ns. At ORNL we have been using such heterodyne receivers for a fusion-

related diagnostic instrument and have developed leading-edge technology in pushing the frequency bandwidth and sensitivity of these CO_2 detectors/down converters for almost 20 years.

Our existing pulsed CO_2 lasers operate best at pulse power levels >10 MW, with pulse lengths of <100 ns, and are extremely reliable when operated in this low-power regime. At 10 μm the minimum spot size in the sample region will be 10 times bigger than it is for infrared CARS at 1μm. Except for a simple size scale-up by a factor of 10×, the scattering geometry would be the same as that developed for infrared CARS, consisting of two Gaussian beams intersecting at a small angle. Therefore, in order to achieve the same electric field in a cross-sectional area that is 100 times greater, the laser power must be 100 times greater (>10 MW versus >100 kW). That is possible, because it is difficult to degrade the power of pulsed CO_2 lasers below 10 MW. With the power level of the existing CO_2 lasers, instead of working hard to get the beam spot size down to a small dimension, we can comfortably accommodate sample volumes of less than 1 cc. Normally the CO_2 laser lines in the 10 P series (those around 10 μm) are spaced by about 53 GHz, so we could easily have the two pump frequencies separated by 53 GHz to probe for molecular resonances at that specific frequency. Continuous tuning of CO_2 lasers is done by raising the pressure in the laser cavity to 10 atmospheres, which can be accomplished in later developments of this diagnostic.

Results and Accomplishments

The core of the CO_2 CARS instrument consisted of five CO_2 lasers. Two high-power pulsed lasers were constructed as the source lasers, and these were driven by two seed lasers. The seed lasers set the wavelengths for the pulsed lasers. A fifth CO_2 laser was used as the local oscillator in the measurement of the resulting beam. The instrument is illustrated in Fig. 1 and shows the location of these lasers, the beam from the lasers, and the optical components. The detection of the resulting beam is with the heterodyne receiver, as illustrated in Fig.2. The beams from the two pulsed lasers were focused and directed to cross in a sample chamber containing a chemical to be studied. This chamber was constructed of one section of a laser cavity; it was 66 cm long and 10 cm in diameter. High-power laser windows (5-cm diam) were mounted to the ends of the sample chamber. With the two laser beams separated at the windows, the interaction length in the sample chamber was 50 cm. A photograph of the system showing the setup is provided in Fig. 3.

The first chemical selected for study was D_2O. This molecule was selected because it had a resonance at 54 GHz,[1] close to the primary frequency difference of the CO_2 lasers at 53 GHz and within the line width of several GHz. The chemical was added as a liquid in the bottom of the sample chamber (which was purged with dry air) and allowed to evaporate at room temperature, thus producing a partial vapor pressure of 17 Torr.

The two pulsed lasers were set to different laser wavelengths through the seed lasers. The higher-power

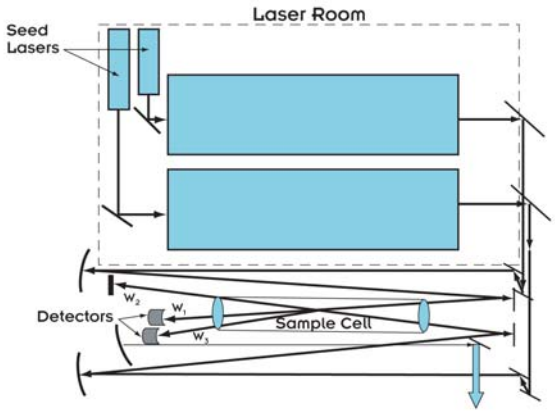


Fig. 1. Schematic of the CO_2 CARS instrument showing the location of the lasers, optical components, and sample chamber in the laser room.

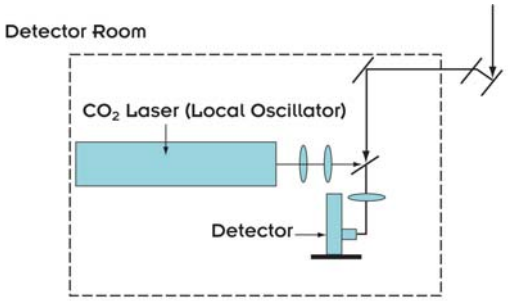


Fig. 2. Schematic of the CO_2 CARS instrument showing the location of the of the heterodyne detector in the detector room.

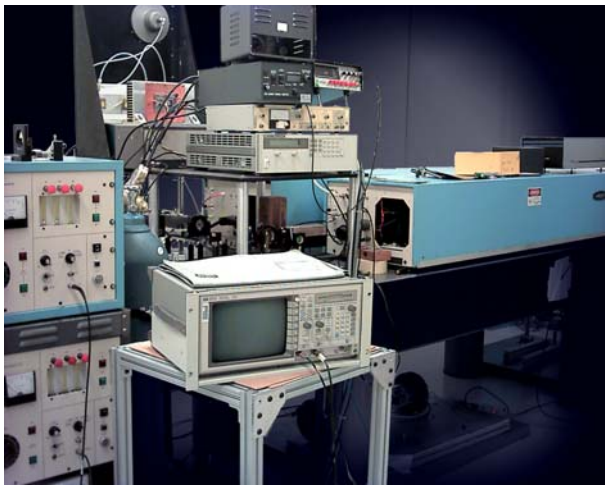


Fig. 3. CO_2 CARS instrument showing the location of the two pulsed lasers on the optic table and the optical path below the table with the sample cell.

laser of the two pulsed lasers was set to the 10P(18) laser line (at 10.57105 μm), and the lower-power pulsed laser was set to the 10P(20) line (at 10.59104 μm). These wavelengths were selected because they differ by 53GHz. The lasers energies were measured with a slow pyroelectric detector. The higher-power laser was adjusted to produce approximately 10 J, and the lower power laser was adjusted to 1 J. The lasers' pulse length was measured with a single-crystal germanium photon drag detector; both were found to produce a 100-ns-wide (FWHM) pulse. Therefore, the lasers produced pulses with 100 MW and 10 MW, respectively.

The nonlinear mixing was expected to produce a third beam at 10.5511 μm near the 10P(16) laser line which could be selected using the local oscillator laser. Unfortunately, the power supply for this local oscillator laser failed and prevented this measurement. To complete this measurement without a working local oscillator, a high- speed HgCdZnTe detector was set in the path approximately 1 m from the sample cell to detect the expected third beam. This detector was able to detect a signal coincident in time with the power from the pulsed lasers. The signal was not present when either pulsed laser was blocked, or when one was tuned to a different wavelength. With a calibration of the HgCdZnTe detector supplied by the manufacturer, and an estimate of the beam size at the detectors location, the power of this third beam was estimated at 3 W. To further check the probability that this third beam was produced by the nonlinear mixing of the two laser beams with D_2O, the sample was changed to H_2O and no signal from the third beam was observed.

Summary and Conclusions

The CO_2 CARS proof-of-principle instrument was constructed and demonstrated its capability in monitoring the presence of specific chemicals through the nonlinear four-wave mixing. With the addition of heterodyne detection, this technique holds the promise of extremely sensitive real-time sensing of complex chemicals. For the example with D_2O, the heterodyne sensitivity would allow detection at extremely low concentrations. In this proof-of-principle test, the D_2O concentration was 0.022 generating 3 W. With a heterodyne detection capability of less than 10^{-9} W,[2] D_2O can be monitored to concentrations below 10 parts per trillion.

- We have developed and demonstrated a new class of sensor for homeland security with proof-of-principle tests
- Unambiguous detection of chemicals in real time
- Detection is at 10 parts per trillion
- Instrument can built with off-the-shelf components
- Does not require the concentration of samples; ideal for direct sampling of airflow into buildings

Future efforts will involve producing a prototype sensor. This will be more compact than the sensor produced for this proof-of-principle test, possibly portable, and more versatile for wavelength tuning. Also, more complex chemicals will be investigated to verify the assumption that they will have distinct resonances accessible to the CO_2 CARS sensor with far greater sensitivity than 10 parts per trillion.

Future efforts may also enter into biological sampling. With the high-power lasers directly in the sample volume, biological specimens can be fragmented into chemical components for direct study. Identification of a specific biological agent would be through identification of several complex chemicals associated with this agent and the ratio of these chemicals.

For follow-on funding, a proposal has been submitted to Department of Homeland Security, and a proposal has is being drafted with the University of Tennessee for submission to the National Science Foundation.

References

[1]F. J. Lovas, J. S. Coursey, S. A. Kotochigova, J. Chang, K. Olson, and R. A. Dragoset, "Triatomic Spectral Database," National Institute of Standards and Technology on-line database <physics.nist.gov/PhysRefData/MolSpec/Triatomic/index.html>.

[2]R. K. Richards, D. P. Hutchinson, C. A. Bennett, H. T Hunter, and C. H. Ma, "Measurement of CO_2 Laser Small-Angle Thomson Scattering on a Magnetically Confined Plasma," *Appl. Phys. Lett.* **62**(1) 28–30 (1993).

Novel Platium Support for Proton-Exchange Membrane Fuel Cell Cathode and Anode Active Layer

T. J. Toops[1] and D.A. Blom[2]
[1]*Engineering Science and Technology Division*
[2]*Metals and Ceramics Division*

Fuel cells show great potential as a clean energy source for both mobile and stationary power, but considerable improvements are still necessary to reach this potential, primarily with respect to the cost/power ratio. This proposal details an alternative membrane construction for the proton-exchange membrane fuel cell (PEMFC) that may enable a 10-fold or greater reduction in the required platinum (Pt) mass, the most costly single component in the PEMFC. The key to this improvement comes from using a continuous array of aligned carbon nanotubules for the Pt support in the electrode. The aligned tubes constitute the active layer of the membrane assembly and will be tested against the current technology to quantify the improvement.

There is a growing interest in the development of alternative power sources in all areas of energy production. A field that has shown remarkable technological progress over the past 30 years is the fuel cell, particularly the proton-exchange membrane fuel cell (PEMFC).[1] Hurdles still exist in each step of this fuel-cell system, but one area that requires particular attention is the high cost of using Pt in the active layer of the anode and cathode. The chemical reactions necessary for fuel cells to generate power do not occur at the temperatures of interest (80–120°C) without a catalyst, and Pt is the most efficient catalyst for both the hydrogen oxidation reaction in the anode and the oxygen reduction reaction in the cathode; there are currently no known cost-effective alternatives to this noble metal. The current target for a competitive PEMFC is $45/kW by 2010, with an additional goal of $30/kW set for 2015. According to an Arthur D. Little study, the current cost of precious metals alone in the present day fuel cell is $65/kW, an estimated one-fifth of the total cost, which illustrates the necessity of marked improvement in this step.[2]

This project outlines a technique that will reduce the amount of platinum used in PEMFCs by at least an order of magnitude and significantly lower the cost of the fuel cell stack. This improvement is realized by minimizing the distance that H_2 and O_2 travel to the Pt catalyst and through the proton-exchange membrane. The current technology supports Pt on nanoporous carbon black in the electrode and requires the gas molecules to follow a long and tortuous path through the membrane electrode assembly (MEA) to reach the various reaction sites. The proposed Pt support is made of carbon nanochannels that enable the gas molecules to follow a linear path through the MEA to reach the various reaction sites. The inexpensive carbon nanochannels are easily formed in the pores of an aluminum oxide template, which are currently manufactured by several companies and are typically used in filtration processes. The nanochannels form an electrode that is essentially made of carbon nanotubes but without the high costs typically associated with nanotubes since the cost-prohibitive separation process is not necessary. These nanochannel-based electrodes could be used on both the anode and cathode of the MEA and would provide an optimum pathway for the reactants. This optimizes the use of Pt in the fuel cell and improves its efficiency and performance.

During FY 2003, we fully mapped the deposition rate of carbon on the walls of the alumina for both the small and large alumina templates and verified the rates with high-resolution transmission electron microscopy (HR-TEM) at ORNL's High Temperature Materials Laboratory (HTML). The HR-TEM also verified that there is no mass transfer limitation in the nominal 20-nm channels of the alumina template, as deposition was constant regardless of pore size (20–300 nm), that is, the channel thickness is constant throughout the membrane. Template removal has been executed for a wide array of carbon supports, and the method has undergone modifications to ensure the integrity of the nanotubular array. Platinum has been added to the nanotubular array after removal of the template, and particle size and deposition have been verified by HR-TEM. The channels before and after deposition are shown in Figs. 1 and 2, respectively. The tools and the materials to test both the current technology and the new electrode formulation have been acquired, with tests to commence shortly. A patent is pending based on the technology in

this project entitled "Integrated Membrane Electrode Assembly Using Aligned Carbon Nanotubules."

Improvements in the PEMFC is a key thrust to DOE's Office of Hydrogen, and this project fits directly into this area, which is the likely prime mover for FreedomCar. It is also relevant to the energy resources and the environmental quality missions of DOE. Many other federal agencies are investigating the usefulness of PEMFC, the primary barrier of which is high cost. If the research is successful in reducing the cost, multiple other federal agencies (Department of Defense, Department of Transportation, etc.) could benefit.

References

[1]P. Costamagna and S. Srinivasan, *J. Power Sources* **102**, 242 (2001) 242; P. Costamagna and S. Srinivasan, *J. Power Sources* **102**, 253 (2001).

[2]E. J. Carlson, J. Thijssen, S. Sriramulu, G. Stevens, and S. Lasher, *Cost Analysis o fFuel Cell Ssytem for Transportation*, final report, Ref 49739, SFAA No. DE-SCO2-98EE50526, Washington, D.C., Department of Energy (1998); http://www.fuelscells.org/adlittle2001.pdf

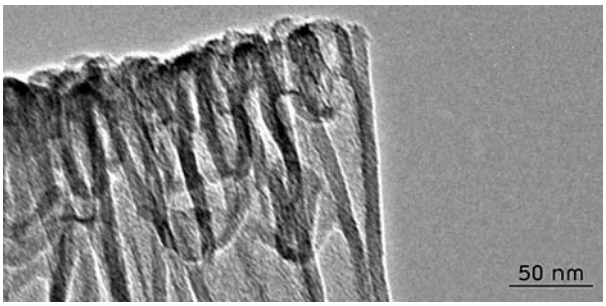


Fig. 1. HRTEM image of deposited carbon after removing the alumina template. Image illustrates aligned nanochannels of carbon with uniform wall thickness. Continuous array not possible to image at this scale due to thickness constraints of the microscope.

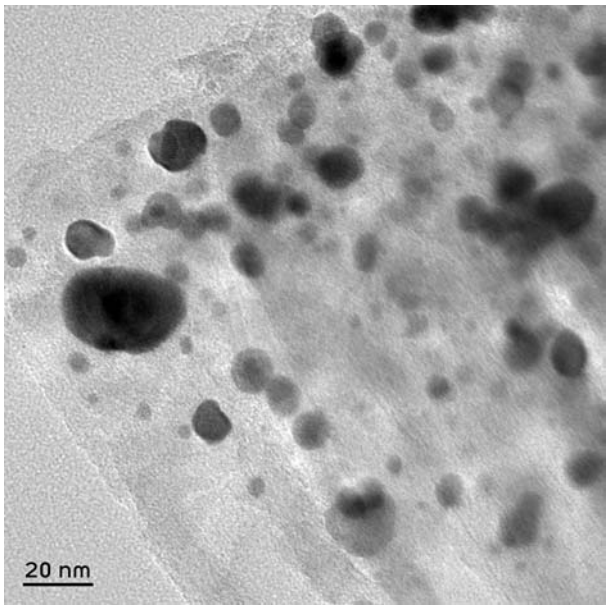


Fig. 2. HRTEM image of deposited carbon after removing the alumina template and depositing Pt.

Zero-Loss Fiber Optic Splitter

S. Rajic, P. G. Datskos, and W. R. Lawrence
Engineering Science and Technology Division

Development of a zero-loss fiber optic splitter would be an important technological advancement for the telecommunications industry. For instance, it would allow telecommunication providers to enable ancillary services to existing or additional customers at low cost. Such a development has proven to be elusive for previous researchers due to concerns about conservation of energy principles and associated signal amplification requirements. In fact, this view seems to be supported by numerous patents and open literature publications. All known prior approaches to signal splitting divert a small portion of the signal photons, usually by evanescent wave tunneling, to provide an additional information path for the original optical signal. When signal photons are removed, costly additional electronic amplifying repeater stations are needed to compensate for the diverted signal loss instead of simpler and less expensive optical repeaters. Thus, when considered for use in expanding service to new users on an existing transmission line, these techniques require expensive system modifications.

Our approach to splitting is a new concept for cascaded optical amplification involving only the inherent properties of the optical fiber. In this project, we intend to demonstrate proof-of-principle for the new technology while limiting the demonstration to only one channel of the 40-nm-wide telecom "C" band. To date, we have set up an experimental fiber-optic test system and collected initial coupling data. Additional work is under way to analyze the data and to optimize the system. Project plans are proceeding on schedule and within budget.

High-Speed Decay Lifetime Analysis Using Field-Programmable Gate Arrays

D. D. Earl
Engineering Science and Technology Division

Computing the lifetime of an exponentially rising or decaying signal is a common calculation associated with various noncontact measurement techniques. For a diverse range of sensing techniques (such as phosphor thermometry, molecular fluorescence lifetime spectroscopy, etc.), the speed at which this mathematical operation can be performed is often the key factor limiting a sensor's achievable measurement rate. Unfortunately, it is this limitation that prohibits decay-rate sensors from investigating and monitoring highly dynamic processes. The investigation of some of these dynamic processes (i.e., ultrafast surface heating/cooling, lattice energy exchange mapping, etc.) represents some of today's most exciting and promising research areas. Because of the varied analysis methods associated with decay-rate measurements and the low-volume demands of the average research consumer, a generalized user-configurable lifetime measurement digital signal processor (DSP) circuit suitable for laboratory use does not exist. The objective of this project is to develop the first DSP, written in VLSI Hardware Description Language, to realize high-speed, user-configurable, single-pulse lifetime measurements using field-programmable gate arrays.

Computing the lifetime of an exponentially rising or decaying signal is a common calculation associated with various measurement techniques. Phosphor thermometry, for example, accurately measures surface temperature by analyzing the fluorescent decay lifetime of certain temperature-sensitive phosphor coatings. Unfortunately, the mathematics associated with an exponential fit to real data can be computationally challenging, especially when signals exhibit a superposition of differing decay lifetimes. The speed at which this mathematical operation can be performed is often the key factor limiting a sensor's achievable real-time measurement rate. To provide the necessary computational power, without limiting data analysis options, a personal computer (PC)–based data acquisition system is typically the only feasible option for laboratory investigations. Unfortunately, the sluggish measurement rates of a PC-based system (<10 measurements/second) limits the decay-rate sensor to monitoring only slowly varying or static processes. With much of today's material research and sensor development involving highly dynamic processes, an improved data analysis platform is needed to enhance the applicability and potential of these unique sensing techniques.

The objective of this project is to develop a parameterized DSP algorithm, written in VLSI Hardware Description Language (VHDL), that can realize high-speed, user-configurable lifetime measurement circuitry in field-programmable gate arrays (FPGAs). FPGAs are re-configurable microcircuits that can be programmed (and reprogrammed multiple times) to realize complex digital circuit designs. Although FPGAs are generally slower than application-specific integrated circuits (ASICs), they can still provide significant hardware acceleration (1000–10000×) over PC-based systems while also permitting design modifications. The high computational speed and inherent design flexibility of an FPGA-based data analysis system makes it the ideal platform for enabling decay-rate sensors to investigate dynamic processes.

To date, we have written and simulated VHDL code to compute the lifetime of an exponentially decaying signal using three separate measurement algorithms. In tandem, an FPGA-based hardware prototype is being designed which allows signals from a photomultiplier tube (used to measure a decaying phosphor signal) to be converted into a digital signal accessible to an FPGA made by Altera Corporation and processed by this VHDL code. The prototype provides a hardware-user interface that allows various measurement values to be easily configured by the user, as well as a software interface that allows more sophisticated re-configuration of the hardware. The prototype is shown in Fig. 1. The prototype is still under construction and has not yet been tested on real-world signals. This testing is expected to start in late December 2003.

When completed, the FPGA-based data analysis prototype will enable ultrafast physical measurements previously considered impossible using decay-rate measurement techniques. This technology will have immediate applications to high-speed and two-dimensional phosphor thermometry research. ORNL researchers working on DOE's Freedom Car project have already expressed an immediate interest in using this device for

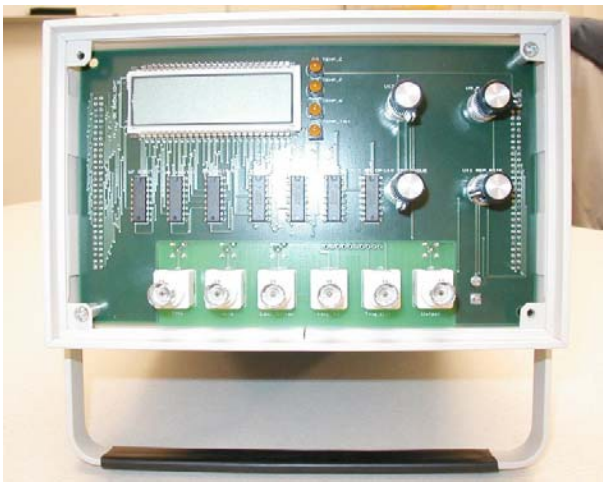


Fig. 1. FPGA-Based Decay Lifetime Measurement Prototype.

an upcoming field-test of phosphor thermometry sensors embedded in fuel cells.

The completed high-speed decay lifetime measurement system will be immediately applicable to thermographic phosphor research ongoing throughout the country. Specifically, it will be of immediate assistance to DOE projects related to the testing and development of fuel cells. Maximum efficiency operation of a fuel cell requires accurate knowledge of internal temperatures. Because of environmental considerations, these measurements are best made with a fiber-optic, phosphor thermometry sensor. Research at ORNL is already ongoing to develop these sensors for use in experimental fuel cell tests. Following the conclusion of this effort, the high-speed decay lifetime measurement system will be immediately applied to this research area as part of ORNL's contribution to DOE's Freedom Car program.

High Effective Hydrogen Storage Density

B. S. Richardson[1] and J. F. Birdwell[2]
[1]*Engineering Science and Technology Division*
[2]*Nuclear Science and Technology Division*

A significant barrier to the use of fuel cells for power generation in mobile and/or portable applications, particularly transportation, is the ability to store hydrogen at high density. A concept based on the release of chemically stored hydrogen by the catalytic hydrolysis of sodium borohydride using recycled water from the hydrolysis reaction excess and the oxidation of hydrogen in a proton exchange membrane (PEM) fuel cell (coupled to the reactor system) has the potential to exceed the DOE hydrogen storage density target. The objective of this research is to study and to demonstrate that sufficient water recovery and recycle can be achieved to sustain a continuous hydrolysis reaction with a significant reduction in the amount of required reagent water, thereby enabling a significant increase in storage density.

Mobile and portable power systems require high-density energy sources to operate efficiently. Increased interest in fuel cells as power sources for these types of applications, particularly automobiles, has resulted in the identification of high-density hydrogen storage technology as a critical need in the shift from hydrocarbons to hydrogen for transportation fuel. DOE has set target goals for hydrogen storage of 6 wt %, 1100 Wh/L, and 2 kWh/kg. While these targets have yet to be achieved, the automotive industry has stated a need for hydrogen storage at a minimum density of 12wt % as criteria for producing commercially viable hydrogen-powered vehicles.

One option for meeting the storage density target is to store hydrogen in a hydrolysable compound, couple the hydrolysis reaction with a fuel cell, and recycle excess hydrolysis water and fuel cell product water to the hydrolysis reaction. Sodium borohydride has several key characteristics that make it amenable for this approach. These include exothermic hydrolysis (to maximize net energy production), fast kinetics (facilitate hydrogen generation responsive to changing power demands), and hydrolysis only in the presence of a catalyst (for safety and controllability). In addition, the catalyzed hydrolysis is sufficiently exothermic to vaporize most of the excess water in the feed solution, facilitating separation of the water vapor from the $NaBO_2$ waste exiting the reaction chamber. Coupled with the fact that a fuel cell produces water as a by-product, it is theoretically possible to continuously segregate and recycle the water in a $NaBH_4$/fuel-cell system. In fact, since for every mole of water consumed in the hydrolysis reaction the reaction at the fuel cell produces two moles of water, the system is a potential net generator of water.

Water recycle eliminates the need for on-board storage of any significant volume of water, since solid $NaBH_4$ reactant can be dissolved on demand into a small volume of water that is continuously replenished during system operation. Consequently, the effective hydrogen storage density of the fuel to the system is 21 wt %: 10.5% from $NaBH_4$, which is doubled as a result of the H_2 contribution from recycled water.

This research effort was initiated in September of FY 2003, with this interim report covering startup activities performed in the last month of the fiscal year. We have conducted a more extensive literature search with regards to the properties of sodium borohydride and its hydrolysis. Of particular interest has been properties which would allow accurate real-time measurement of solution concentration. Potential catalysts have also been reviewed, with ruthenium shown to be an extremely effective catalyst for this reaction. Several ruthenium catalysts, on different substrate materials, shapes, and loadings, have been obtained for testing.

The current DOE hydrogen storage objectives are 6 wt %, with the automobile industry advocating a change in the target to 12 wt %. Thus there is the potential for significantly exceeding the DOE automotive industry targets with our approach. Upon demonstration of a proof-of-principle system, opportunities for follow-on funding within DOE and the commercial automotive industry will be pursued. In addition, this technology has potential application in a broad range of military applications. As this technology could be scaled, applications from unmanned air and ground systems, soldier-borne systems, and other portable systems could benefit from this research.

ENVIRONMENTAL SCIENCE AND TECHNOLOGY

Director's R&D Fund

Community-Wide Analysis of Unique Sequences and Functions from Uncultured Microorganisms

M. W. Fields,[1*] F. W. Larimer,[2] and J. Zhou[1]
[1]*Environmental Sciences Division*
[2]*Life Sciences Division*
**Present address: Department of Microbiology, Miami University, Oxford, Ohio*

Microorganisms play an integral and often unique role in ecosystem functions; however, the majority of microorganisms remain uncultivated. Due to the limitations of conventional detection methods, little is known about microbial ecosystems and metabolic capacity. Therefore, we propose to develop microarray methodology that would identify the novel genomic sequences from microbial communities. With this technology, the community DNA can be screened for unique sequences when compared to a background site and the novel sequences represented by decreased hybridization signals on a microarray. Our research will further strengthen ORNL as a leading institution for integrating genomics and associated technologies to understand the biocomplexity of microbial ecosystems. The use of microarray technology would greatly enhance the identification of unique sequences from any given environment and would be an attractive deliverable.

Introduction

The objective of this project was to develop technology for the identification of unique and novel genomic sequences from microbial communities. The detection of novel community sequences will impart a more holistic understanding of the structure, functional stability, and metabolic diversity, as well as aid in the discovery of uncultured microorganisms. As proposed our technology would detect unique environmental sequences and would not be based on previous molecular- and culture-based techniques. The metagenomic microarrays (MGAs) would identify clone library subsets; hence, sequence determination only occurs for clones of interest, and not the entire library. Moreover, the identification of functional genes is not directly reliant upon PCR-based methods. The BAC-libraries allow for the discovery of linked novel sequences and pathways, in particular the linked sequences of uncultivable microorganisms. The importance of system-wide ecology is becoming more evident, and our techniques would be an instrumental step for linking uncultivable microorganisms with specific functional and ecological properties. Another important issue for any environmental genomics study is the extraction and purification of high-quality, high-molecular-weight DNA (HMW-DNA). We have also been working to develop and improve protocols for the extraction of HMW-DNA from different environmental samples. We can achieve DNA fragments over 200 kb, and the DNA is suitable for general molecular manipulations. Larger DNA fragments result in a more informative and useful BAC or fosmid library and are an issue that has not been thoroughly investigated or evaluated in the literature. We are currently exploring different techniques for the efficient cloning of our extracted HMW-DNA.

Technical Approach

Environmental samples pose difficulties in laboratory experiments, including low biomass, diminished nucleic acid recovery, high species diversity, and community complexity. Recent work has sampled the metagenomes of a soil and a seawater microbial community, and thousands of clones were generated representing up to 1000 Mbp. The identification of unique and novel sequences in such large libraries is problematic at best and requires a vast amount of time and effort even for a limited subset of clones. Microarray-based genomic technologies represent a potential revolution in the biological sciences. However, the usefulness and performance of these technologies in studies on environmental microorganisms and microbial ecosystems are unproven. Because of the heterogeneity and diversity of environmental samples, enrichment cultures were used as a proof of principle in order to develop methodologies.

We addressed the issue of hybridization specificity of large DNA fragments on glass slides, which has not been previously addressed. Because the construction of BAC and fosmid libraries can be laborious and tedious, we tested specificity and sensitivity using HMW-DNA from bacterial cultures. The relationships between signal hybridizations and organismal relatedness were

characterized via multiple approaches and compared to previous research. Another important issue for any environmental genomics study is the extraction and purification of HMW-DNA. Methods were developed and improved to extract quality, large DNA fragments from microbial communities from groundwater, sediment, and enrichment cultures. The isolation of HMW-DNA that is suitable for cloning is a crucial step to the construction of non-biased libraries that represent entire communities. Another issue to be resolved is the purification of cloned DNA for the printing process. These issues have not been previously addressed with large-insert cloning vectors and the deposition of HMW-DNA onto glass slides.

Results and Accomplishments

Early data suggested that bacterial species could be differentiated at the species level. Further work confirmed that large, genomic fragments can be differentiated with microarray hybridization conditions, and that decreased signal intensities correspond to decreased DNA:DNA homology. When a mixture of DNA from different microorganisms was hybridized to the array, hybridization specificity and signal intensity was maintained. Also, it was determined that the signal intensities can correlate to actual cell numbers when at least 10^4 cells/mL are present, and hybridization of large genomic fragments could be quantitative between 0.2–1000 ng. These results indicated that hybridization specificity can be achieved with mixed populations of large, genomic fragments and that it will be possible to differentiate environmentally relevant microorganisms over 2 to 3 orders of magnitude (site of interest has 10^5 to 10^7 cells/mL). Because the technique appeared quantitative over a large range, and signal intensity correlated to DNA:DNA homology, BAC- and/or fosmid-DNA should be easily differentiated between environments.

We have also developed and improved protocols for the extraction of HMW-DNA from different environmental samples. Effective lysis of microorganisms and recovery of HMW-DNA of sufficient purity from environmental samples is critical for comprehensive characterization of microbial communities using metagenomics approach. The extraction of HMW-DNA can be limited by biomass density and the presence of contaminants in many groundwater and subsurface sediment samples. In this study, we developed protocols that can be used to achieve large quantities of HMW-DNA of soil or groundwater samples. Compared with similar procedures, the approach uses chemical flocculation to replace density centrifugation for bacterial cell isolation from soil matrix and thus can be used in recovering bacterial cells from large quantities of soil samples of low biomass. Combing in-well-lysis and electrophoresis separation, DNA fragments sized more than 300 kb were obtained from the samples. We also modified a protocol developed in our laboratory previously and widely used in direct DNA extraction from soil samples. By application of appropriate amounts of chemical extraction buffers, the humic substances recovered in DNA extracts were decreased significantly

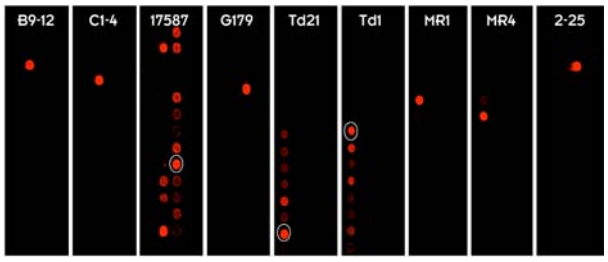


Fig. 1. Experiments with different bacteria in pure culture are shown above. Each spot is genomic DNA from a different bacterium, and a probe for a single organism is hybridized to the array. When conditions are modified (i.e., temperature and buffers), only the spot corresponding to the correct organism is hybridized.

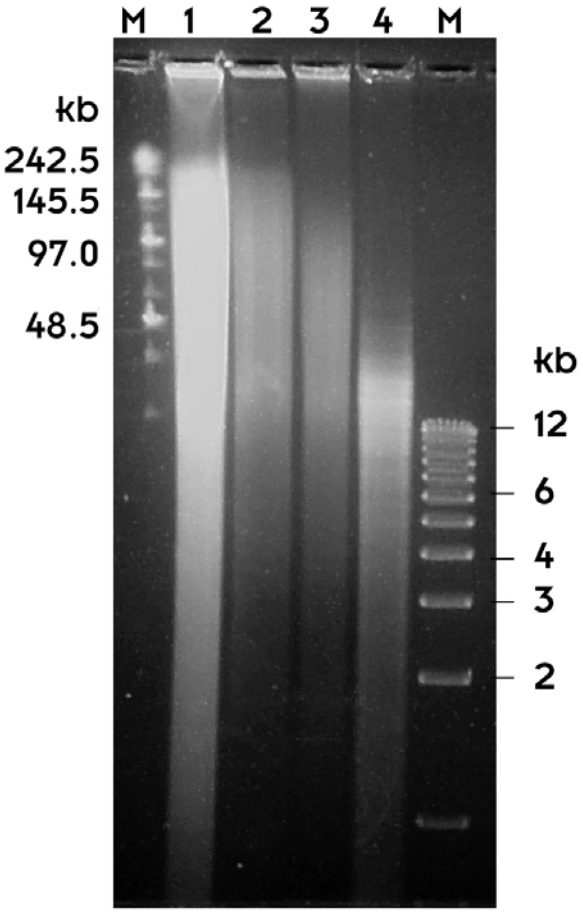


Fig. 2. Purification of HMW-DNA from sediments with improved grinding methods and extraction buffers.

and thus greatly facilitated the follow-up purification procedures. By eliminating Proteinase K treatment, the protocol can directly isolate DNA of 200 kb fragments. The purified HMW-DNA was pure enough for restriction enzyme digestions and PCR amplification. Larger DNA fragments result in a more informative and useful BAC library and is an issue that has not been thoroughly investigated or evaluated in the literature. We then set out to identify different techniques for the efficient cloning of our extracted HMW-DNA.

Samples were collected from different FRC sediments, and enrichment cultures were initiated. Community structure and diversity were analyzed via SSU rDNA sequence libraries. Cultures with 5 to 20 species were selected for library construction.

Communities of interest were used to construct fosmid libraries. BAC libraries have proved difficult to construct from mixed microbial communities, and similar results have been observed by other research groups. The fosmid approach was selected because the efficiency of cloning increases significantly with fosmids, and the cloned insert size is still relatively large (approximately 35 kb).

In order to clone HMW-DNA into fosmid vectors, 5 to 10 µg of DNA was end repaired using the Copy Control Fosmid Library production kit (Epicentre). End-repaired DNA was separated on a 1% LMP agarose gel and the gel slice excised that contained the DNA of interest. The gel slice was melted at 70°C for 10 minutes and transferred to 45°C water bath. Pre-warmed gelase buffer and gelase (1 Unit) were added to every 200 mg of gel and incubated for 2 h. The reaction was inactivated by incubating at 70°C for 10 min. The tube containing DNA was incubated in an ice bath for 30 min and centrifuged at 10,000 rpm for 20 min. The supernatant was precipitated by adding 1/10 volume of 3 M sodium acetate (pH 7.0) and 2.5 volume of ethanol and incubated at room temperature for 10 min. The precipitated DNA was centrifuged for 20 min and washed with 70% ethanol. The DNA pellet was resuspended in 20 µL of water. DNA concentration was measured with a nanodrop spectrophotometer. Extracted DNA was ligated to 0.5 µg of Copy Control pCC1FOS Vector in 10 µL of total reaction volume.

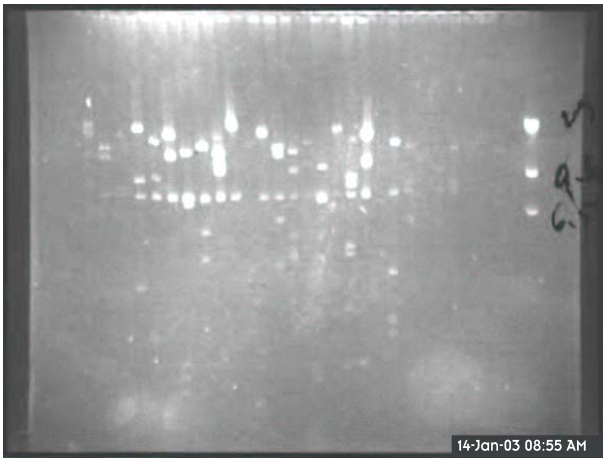


Fig. 3. Our results with HMW-DNA from nitrate-reducing enrichments indicate that sizable libraries can be produced (>8,000 clones) with an average insert size between 35 to 40 kb. In addition, over 800 clones have been generated from a sulfate-reducing enrichment.

The ligation mixture was heat inactivated by incubating at 70°C for 10 min and then combined with 25 µL of the MaxPlax Lambda Packaging Extracts and incubated at 30°C for 90 min. Subsequently another 25 µL of the MaxPlax Lambda Packaging Extracts was added to this mixture and incubated further for 90 min. These packaged clones were diluted with 1 mL of phage dilution buffer (10 mM Tris-HCl [pH 8.3], 100 mM NaCl, 10 mM $MgCl_2$). Packaged phage particles were mixed with 100 µL of host cells, EPI300-T1^R, grown in LB with 10 mM $MgSO_4$ and with 2% maltose to the cell density of $O.D._{600}$ 0.8–1.0. The mixture was incubated at 37°C for 20 min and plated out on LB plates containing 12.5 µg of chloramphenicol. The plates were incubated at 37°C over night to select for the Copy Control Fosmid Clones.

The constructed clonal libraries are being used to test methodology for purification of vector DNA for printing onto glass slides. The fosmid clonal DNA must be purified away from genomic DNA of the host cells and be of sufficient quality and quantity for printing onto glass slides. Current protocols do not exist for the purification of large cloning vectors for deposition onto glass slides for microarray analyses. Because of limitations in current protocols for high-throughput purifications, we had to develop new methodology.

Rolling circle amplification (RCA) was pursued for the random, nonbiased amplification of fosmid clonal DNA. Fosmid DNAs were purified using the Millipore bacDNA purification kit. DNA from each well was eluted with 50 µL of elution buffer. Then, 10 µL of eluted DNA was amplified using 1 µL of phi29 DNA polymerase (10 units), 1 µg of T4 gene 32 protein, 3 µg of random primer (modified at the 3' end), 4 µL of 10 × buffer, 4 µL of 10 × BSA, 2 µL of 10mM dNTP in 40 µL of total volume. The reaction was incubated at 30°C overnight. The concentration cannot be measured unless DNA is purified. The modified random primer worked much better than the previous primers (N7, two additional nitroindole residues at 5' end and a phosphorothioate linkage at the 3' end) for RCA. Once an ample amount of quality fosmid clone DNA was obtained, the amplified DNA was purified and prepared for deposition onto glass slides. After this is achieved, the determined protocols for hybridization specificity as determined earlier in the study were tested with the fosmid clone microarrays.

Conclusions

As a result of the project, we have developed protocols for the extraction of quality HMW-DNA from mixed microbial communities and environmental samples. This was an important and vital methods development for the project. A manuscript that describes the protocol is being prepared, and the work has been submitted for presentation at the national meeting for the American Society of Microbiology. A manuscript that describes the work concerned with the hybridization specificity and sensitivity of genomic fragments in a microarray format has been submitted and is currently being reviewed. The construction of fosmid libraries from microbial enrichment cultures has proven to be an efficient way to develop the metagenomic approaches, and this work was recently presented at the International Microbial Genomes Conference.

When working with mixed microbial communities or environmental samples that involve iron or sulfate reduction (which many microbial DOE projects do), the extraction of quality HMW-DNA with enough quantity for molecular approaches will always be problematic. We have developed protocols for the purification of HMW-DNA from these types of samples, and then used the HMW-DNA for the construction of fosmid libraries. Once obtained, the cloned DNA must be prepared in sufficient quantity for deposition onto glass slides. To circumvent identified problems, RCA has been proven beneficial for the non-biased amplification of cloned DNA. Work is under way to prepare this DNA for printing onto glass slides and to test the developed protocols to achieve hybridization specificity between large fragments of genomic sequences in a high-throughput fashion.

Simulation of Subsurface Environmental Processes

J. C. Parker
Environmental Sciences Division

This project addresses the development of practical methods for modeling field-scale processes that control the fate and transport of volatile organic contaminants that may occur as nonaqueous-phase liquids (NAPLs) and as dissolved or vapor-phase constituents. Models were developed to predict volatilization rates for dissolved contaminants in groundwater and for NAPL in the unsaturated zone to the atmosphere or into buildings under natural conditions. Models were also developed to predict contaminant removal for vacuum extraction, bioventing, and air sparging systems. Formulations were proposed to describe field-scale liquid-liquid and liquid-vapor mass transfer kinetics for steady-state and transient (e.g., pulsed) flow conditions. High-resolution numerical experiments were performed to evaluate the functional form and parameter values for field-scale mass transfer coefficients for the case of NAPL dissolution in groundwater.

Introduction

Volatile organic chemicals comprise an important class of subsurface environmental contaminants, which include solvents, fuel hydrocarbons, and other chemicals that commonly occur as nonaqueous-phase liquids (NAPLs). NAPLs that are denser than water (DNAPLs) are particularly problematic due to their ability to penetrate deep into aquifers and act as a long-term source of groundwater contamination and are very common at industrial sites and throughout the DOE and DoD complexes. A quantitative understanding of processes that affect contaminant attenuation is critical to assess long-term risk and to evaluate remediation alternatives.

Volatile contaminant transport in the vadose zone is often attributed to vapor-phase molecular diffusion. Measurements of vapor fluxes from contaminated groundwater, however, have been reported to exceed those attributable to diffusion alone. Additional processes that may induce vapor-phase transport include barometric pressure changes, water table fluctuations, air displacement due to water infiltration, and vapor-density variations. While volatilization from soil and groundwater may be beneficial from the standpoint of attenuating soil and groundwater contamination, vapor emissions from the ground, especially into buildings, can pose significant health and safety risks.

The rate of contaminant dissolution to groundwater over time is a crucial factor governing the feasibility and effectiveness of engineered remediation or natural attenuation at DNAPL-contaminated sites. Two issues are of paramount importance. First, how long will it take for a DNAPL source to become depleted, and second, to what degree will the contaminant flux from the source zone decrease with time as the DNAPL mass diminishes? Quantification of DNAPL dissolution rates is commonly formulated as a mass transfer kinetics problem. Predictions of DNAPL dissolution kinetics based on laboratory studies indicate near-equilibrium concentrations should occur within a very short travel distance in DNAPL-contaminated regions. However, such high concentrations are rarely observed at DNAPL-contaminated sites.

It is apparent that DNAPL dissolution in the field is controlled to a great degree by heterogeneity in the distribution of DNAPL and groundwater velocities within the subsurface, which are extremely difficult to delineate at a sufficiently fine scale to predict field-scale behavior. High-resolution numerical experiments offer a much more efficient means of studying scale-dependent mass transfer relations and developing practical large-scale mass transfer functions.

The goal of this project was to develop practical field-scale models for contaminant mass transfer from NAPL to air and to water in order to evaluate remediation feasibility by natural attenuation and/or engineered systems.

Technical Approach

Volatilization from Groundwater

A model was developed to predict contaminant fluxes due to volatilization from groundwater that considers (1) vapor-phase molecular diffusion, (2) a dispersive vapor flux due to periodic (sinusoidal) barometric pressure fluctuations, (3) dispersive transport due to water table fluctuations, (4) aqueous-phase advection in the vadose zone associated with upward or downward unsaturated water flow, and (5) vertical aqueous dispersion in the saturated zone. An analytical solution to the governing

equations for steady-state conditions was derived, yielding an apparent first-order volatilization coefficient with respect to the dissolved groundwater concentration. Model sensitivity analyses were performed for a range of parameter values to evaluate the magnitude of volatilization losses and effects of various processes and factors.

Volatilization from NAPL

A model was also derived to predict NAPL mass loss versus time due to volatilization of NAPL present in the unsaturated zone considering the same processes discussed above. Solutions are obtained by treating the contaminated zone as a moving boundary or as a stirred reactor to bracket effects of diffusive-dispersive mixing. Equilibrium partitioning is assumed, which is expected to be valid provided unsaturated hydraulic fluxes are not large.

Indoor Air Intrusion

A model for vapor intrusion into buildings overlying contaminated soil was developed considering the processes described above and additionally (1) multicomponent NAPL mixtures with vapor partitioning controlled by Raoult's Law based on the time-dependent mole fraction composition of the NAPL, (2) aerobic contaminant biodecay in the soil at rates controlled by oxygen transport into the soil and by species-dependent oxygen utilization coefficients, (3) advective-diffusive vapor transport through foundation cracks, (4) air cross-flow through the soil and foundation subbase, (5) building dilution due to ventilation rate, (6) health risk associated with indoor air inhalation, and (7) consideration of uncertainty in model parameters using a first-order error analysis. Sensitivity analyses were performed, and previously published data from field sites were analyzed.

Soil Vapor Extraction

Models for remediation using active or passive soil vapor extraction, bioventing, and air sparging were implemented that consider non-equilibrium NAPL-vapor partitioning. The model is unique in the use of a mass-transfer function that accounts for diffusive and velocity-dependent dispersive mass transfer that can accommodate pulsed-flow conditions. The models predict remediation time and life-cycle remediation cost based on defined unit capital and operating costs. A technique was developed to monetize the cost attributable to uncertainty in site properties due to deviations between computer-optimized and actual optimum remediation designs. Total design cost penalty is defined as the root mean square sum of the cost differentials between the model-optimized design performance and the actual system as estimated from sensitivity analyses.

DNAPL Dissolution Kinetics

Field-scale DNAPL dissolution kinetics was studied using a high-resolution numerical model for DNAPL percolation into a heterogeneous aquifer followed by dissolution in moving groundwater. Parameters in an empirical field-scale mass transfer function were determined by fitting the model to predicted average fluxes from the DNAPL source zone.

Results and Accomplishments

Volatilization from Groundwater

Simulations were performed to assess the contribution of barometric and water table fluctuations on vapor transport. The results indicated that dispersive fluxes are increasingly important as the frequency of fluctuations increases. Thus, seasonal water table fluctuations have little impact, while high-frequency (e.g., tidal) fluctuations are more significant, especially if air-filled porosity is low and/or groundwater is relatively deep. The ratio of dispersive flux due to barometric pressure fluctuations to diffusive flux is predicted to increase sharply with groundwater depth. For systems with low air-filled porosity, barometric pumping is predicted to be the dominant transport mechanism, while diffusion dominates for soils with high air-filled porosity and/or for shallow soils.

Simulations of perchloroethylene (PCE) volatilization from groundwater were performed for a range of site conditions. The results (Fig. 1) indicate that volatilization decreases sharply with groundwater depth up to a point and may exhibit slight increases with depth in certain circumstances due to barometric pumping effects. Volatilization coefficients are greatest for soils with high air-filled porosity. Net volatilization rate is decreased by a downward hydraulic flux in the vadose zone and increased by an upward hydraulic flux. Considering that apparent first-order biodecay coefficients for PCE in groundwater are less than the volatilization coefficients computed for many of the scenarios here, volatilization losses may represent a more important attenuation mechanism in many cases than has previously been thought.

Volatilization from NAPL

A simulation was performed for DNAPL volatilization from a soil initially contaminated with PCE at 1000 $mg \cdot kg^{-1}$ from the ground surface to the water table at a depth of 10 m. The soil is assumed to have an air-filled porosity of 0.2 and an air permeability of 10^{-10} m^2 (1 darcy). Results for unsaturated zone hydraulic fluxes of 0, 0.003, and 0.01 $m \cdot d^{-1}$ are presented in Fig. 2. Computed times for complete NAPL depletion from the soil are 123, 69, and 31 years, respectively. These durations would

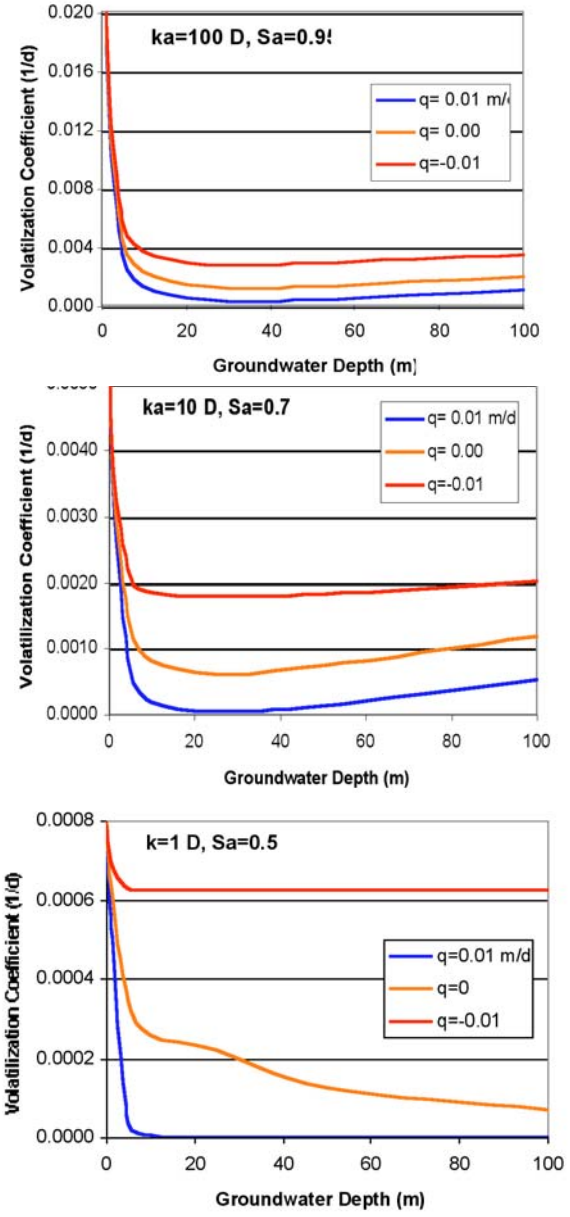


Fig. 1. Apparent first-order volatilization coefficients versus groundwater depth for selected air permeability, air saturation, unsaturated zone water fluxes and other parameters described in text.

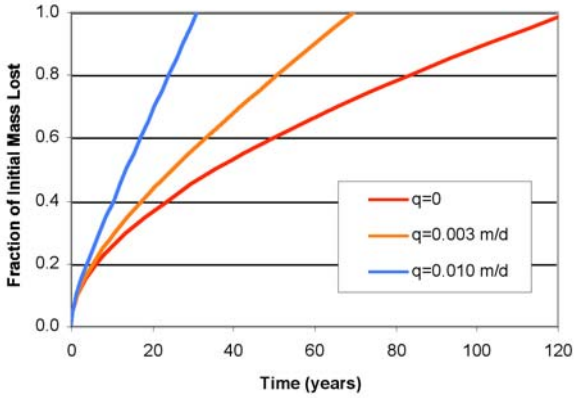


Fig. 2. Soil NAPL depletion versus time due to volatilization and leaching with different hydraulic fluxes for example problem.

decrease or increase in direct proportion to the initial DNAPL mass. Shallow contaminant sources would deplete much faster. The results indicate volatilization may be a significant loss mechanism but that leaching losses may be the dominant mechanism for deep NAPL sources.

Indoor Air Intrusion

The vapor intrusion model was verified by comparison to available field data for two sites. One site was a sandy soil in New Jersey with groundwater ~6 m below ground surface, and the second site was a fine-grained soil in southern California with groundwater at a depth of ~2 m. Both sites were contaminated with fuel hydrocarbons. Measured indoor air concentrations fell within the confidence limits of predicted concentrations for all monitored species considering uncertainty in model parameters. Sensitivity analyses were performed to assess potential effects of model parameters on indoor air concentrations, health risk in the absence of further remedial action (baseline risk), and groundwater cleanup level required to meet a baseline risk of 10^{-6} for a hypothetical problem involving a sandy soil with dissolved benzene in the groundwater below a building with a shallow basement. The results (Table 1) indicate that indoor air concentration and baseline risk are most sensitive to soil type, which strongly affects diffusive transport from groundwater to the building and advective air intrusion rates. An otherwise similar site with fine-grained soil or with slab construction rather than a foundation would meet the risk criteria without further remediation. The effects of biodecay are particularly interesting. Although the indoor air concentration and baseline risk for the base case, which accounts for biodecay, is only ~1/2 of that predicted if biodecay is disregarded, the risk-based cleanup

Table 1. Effects of various factors on indoor air concentration, baseline health risk and risk-based groundwater cleanup level

Case	Mean indoor air concentration ($\mu g\ m^3$)	Baseline risk	Groundwater cleanup level ($\mu g\ L^{-1}$)
Base Case	6.8	1.7×10^{-5}	2,300
1—higher source concentration	79.4	2.0×10^{-4}	2,300
2—finer grained soil	0.001	3.0×10^{-9}	34,000
3—slab on grade foundation	0.16	4.1×10^{-7}	10,000
4—no foundation cross-flow	41.7	1.1×10^{-4}	900
5—no source attenuation over time	46.5	1.2×10^{-4}	800
6—no barometric pumping	4.8	1.2×10^{-5}	2,350
7—no biodecay	19.0	4.9×10^{-5}	100

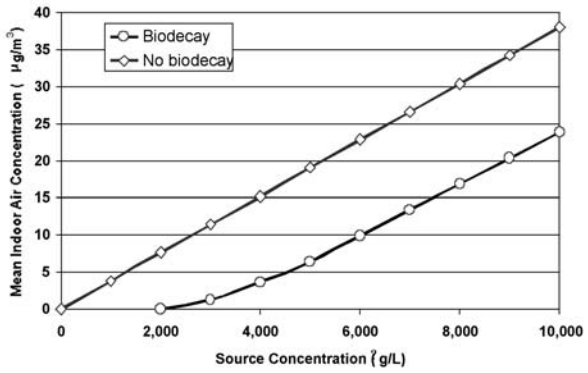


Fig. 3. Predicted indoor air concentrations with and without biodecay.

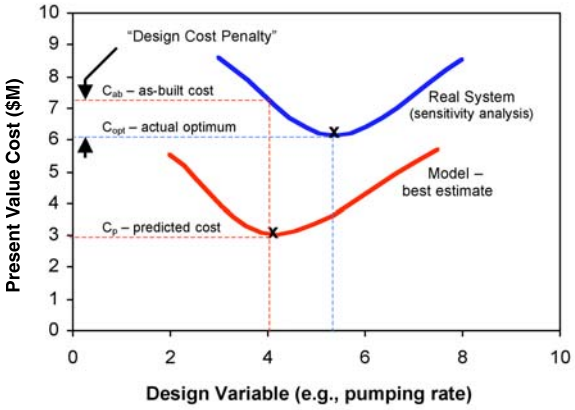


Fig. 4. Conceptual representation of design cost penalty with respect to a single variable.

level is 23 times higher. This effect arises because biodecay induces a shift in the relation between indoor air concentration and groundwater concentration (Fig. 3). The magnitude of the shift represents the biodecay capacity of the soil. The results indicate that disregarding soil bioattenuation capacity can result in risk-based cleanup levels that are greatly overpredicted and hence more costly than necessary to meet risk criteria.

Soil Vapor Extraction. Computer models are most often utilized to evaluate the effectiveness of alternative remediation system options and to optimize system design. Models are rarely utilized in parallel with site characterization efforts to guide site investigations and minimize total costs. To assess the potential benefits of models to guide characterization efforts, a hypothetical problem was analyzed involving a hydrocarbon-contaminated site. The model was used to optimize design variables for a bioventing remediation system using estimates of mean site parameters. The total remediation cost in net present value (NPV) was computed to have an expected value of $432,000 with confidence limits from $372,000 to $520,000 considering uncertainty in site parameters. An analysis of variance of the cost uncertainty with respect to individual site properties (Table 2) indicated that 68% of the cost uncertainty was due to uncertainty in the NAPL mass at the site. The design cost penalty (Fig. 4) associated with uncertainty in NAPL mass, however, was negligible. The largest design cost penalties were due to uncertainty in air-filled porosity and air permeability, which were minor contributors to the cost uncertainty. The magnitude of the design cost penalty for porosity and air permeability represent the maximum expenditure that would be cost justified in terms of reducing actual remediation costs. The large effect of NAPL mass on uncertainty in estimated cost reflects the fact that remediation time and hence total operating costs are nearly proportional to the NAPL mass. However, the remediation system design is largely controlled by system hydraulics that govern mass removal efficiency. Large expenditures to reduce uncertainty in NAPL mass would enable more accurate predictions of remediation time and cost to be made, but the effort would not pay off in terms of actually reducing the cost.

Table 2. Model-predicted uncertainty in remediation cost as a percent of total variance and design cost penalty as a function of various uncertain site parameters

Parameter	Cost uncertainty % of variance	Design cost penalty, $K
Total contaminant mass	68	<1
Biopreference coefficient	10	<1
Air permeability	10	21
Air-filled porosity	8	23
Other parameters	4	3

DNAPL Dissolution Kinetics

The distribution of residual trichlorethylene (TCE) was simulated for a DNAPL release into a 10-m × 10-m × 10-m heterogeneous aquifer region discretized into 1,000,000 computational cells (Fig. 5). TCE dissolution and dissolved-phase transport was simulated until essentially all DNAPL was removed from the domain. Field-scale mass transfer coefficients were computed as a function of time from average simulated contaminant fluxes leaving the model domain. Field-scale coefficients were found to be a linear function of mean groundwater velocity, in contrast to lab studies that indicate proportionality with velocity to a power of ~0.7. This finding is significant because linear dependence implies that reductions in mean groundwater flow will result in greater reductions in contaminant fluxes than would be predicted based on lab correlations. Engineering measures aimed at reducing flow through DNAPL source zones could significantly improve the effectiveness of natural attenuation processes.

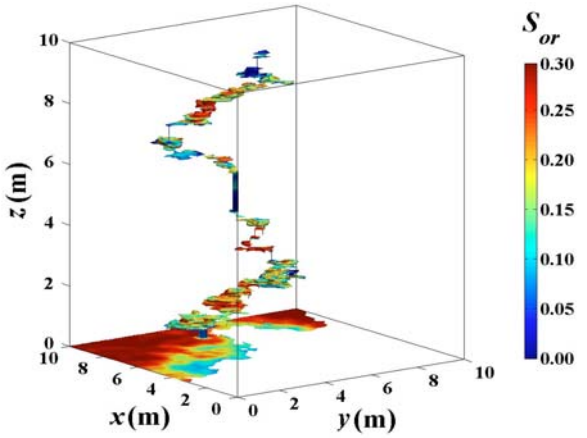


Fig. 5. Simulated initial residual DNAPL distribution in model domain.

Computed field-scale ("lumped") mass transfer coefficients vary approximately in proportion to relative DNAPL mass raised to an empirical "depletion exponent," which was observed to be less than one for DNAPL zones with laterally extensive DNAPL pools or lenses and greater than one for regions with randomly distributed DNAPL "fingers" regions. The field-scale numerical model results contrast sharply with numerous published lab-scale studies, which yield depletion exponents that are consistently less than one and net mass transfer coefficients that are much larger than those predicted here to occur at the field scale. A closed-form analytical solution for source concentration versus time was derived using the lumped field-scale mass transfer function subject to mass balance conditions. The model predicts that source zones with depletion exponents less than one will exhibit small reductions in effluent concentrations over time until most of the DNAPL is depleted when concentrations decrease sharply. Source zones with exponents greater than one are predicted to exhibit more gradual concentration decreases with time as DNAPL mass diminishes. The analytical model results are confirmed by the high-resolution numerical simulation results, which show DNAPL lens dominated zones (with small depletion exponents) exhibit slow reductions in source concentration with time as mass depletion proceeds, whereas finger-dominated regions (with high depletion exponents) exhibit steady declines in source concentration with time as depletion occurs (Fig. 6). These findings indicate that DNAPL source concentrations and fluxes may decrease much more substantially over time than has been commonly presumed, especially for finger-dominated DNAPL zones. The simple analytical model provided reasonably accurate predictions of source depletion versus time (Fig. 6). Since this model is amenable to calibration from reasonably attainable field

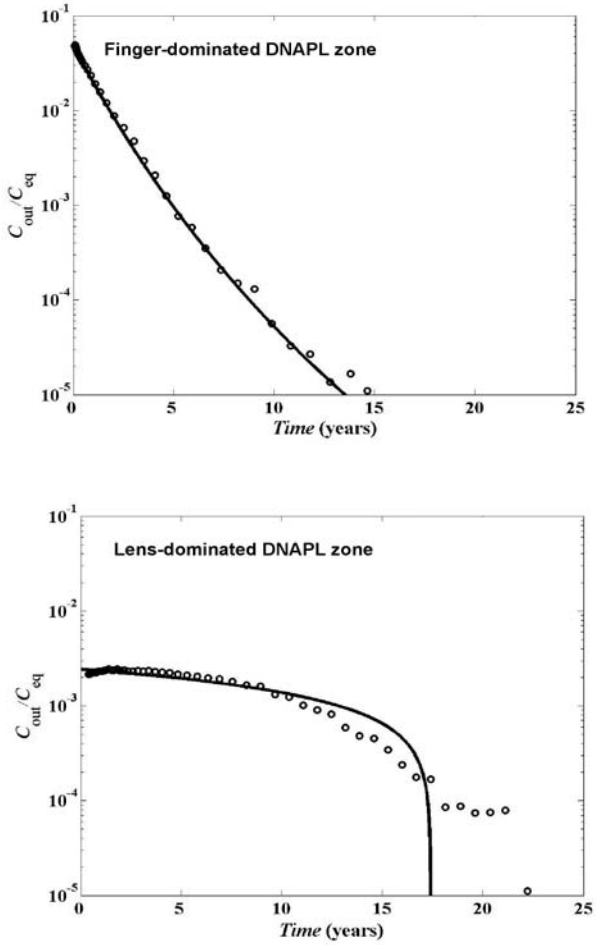


Fig. 6. Relative concentration versus time for DNAPL source zones dominated by fingers (top) or lenses (bottom). Points are high resolution numerical model results and lines are analytical model.

data, practical projections of DNAPL source attenuation, which is critical for assessment of long-term natural attenuation effectiveness, should be possible.

Summary and Conclusions

This project has resulted in significant advances in our understanding and ability to predict the behavior of organic contaminants that occur as NAPLs and as aqueous- and vapor-phase constituents in soils and groundwater. The knowledge and modeling approaches will enable remediation of contaminated sites to be performed more efficiently, with less uncertainty and lower costs.

Funding has been secured from DoD-SERDP to develop models to assess effects of physical and biological processes on natural attenuation at DNAPL sites ($636K total ORNL budget). DOE has just released an RFI for work on natural attenuation at DNAPL sites, which is being pursued. Funding has also been obtained that focuses on development of models for metal-contaminated sites from SERDP (~$110K for modeling tasks) and from DOE-NABIR (~$400K for modeling tasks).

Ecosystem Genomics—An Emerging Opportunity for Environmental Research

S. P. DiFazio,[1] S. Jawdy,[1] L. Gunter,[1] B. A. Wilson,[2] and A. Brunner[3]
[1]*Environmental Sciences Division*
[2]*Department of Biology, Jackson State University*
[3]*Forest Science Department, Oregon State University*

We have undertaken a research program that capitalizes on the poplar genome sequence as an entry point for studying the molecular bases of ecologically significant processes such as flowering, drought adaptation, and response to perturbations such as elevated carbon dioxide. Analytical techniques for studying molecular underpinnings of phenotypes are currently most developed for association studies involving qualitative traits such as disease incidence. Therefore, as a proof of principle, we analyzed gender in poplar because this is an easily assayed qualitative trait that is under complex genetic control.[1] Determination of gender is most likely under the control of floral meristem identity genes, which in turn are regulated by transcription factors that control the transition of a vegetative meristem to a floral meristem, and transduction of environmental signals such as day length, light quality, and cold. We identified a total of 98 polymorphisms in 4.1 kb of mostly noncoding sequence in putative regulatory regions of four floral homeotic genes. There were substantial levels of linkage disequilibrium between polymorphic sites for three of the sequenced regions, suggesting that these polymorphisms could be under selection. There were no significant associations with gender for any of the sequenced polymorphisms. This is the first study of nucleotide polymorphism and linkage disequilibrium for wild poplar trees and provides important baseline data for future association studies in this model organism.

Introduction

Ecosystem genomics has vast potential for enhancing understanding of ecosystem functioning as genomic information, methods, and analysis tools continue to develop. However, at present these tools are in an early stage of development, and much exploratory and proof-of-principle work is needed to develop and validate approaches for extending genomic information to ecosystem scales. We have begun this research program by focusing on a single model species that plays a dominant role in many ecosystems: the poplar tree (*Populus spp.*). There are many reasons that poplar serves as an excellent bridge from current, single-species genome studies to an ecosystem scale. First, poplar shares many of the desirable characteristics of other model species: it has a small genome size (550 Mb, 4 times larger than *Arabidopsis* and comparable to rice), it is readily transformed and clonally propagated, genetic linkage maps and large pedigrees exist, controlled crosses are readily performed within and between species, and a great deal is known about poplar physiology and genetics.[2] In recognition of its utility as a model species and relevance for agency objectives, the Department of Energy has sequenced the entire poplar genome, thus placing poplar in the same class as the handful of other model eukaryotes whose genomes have been sequenced.[3]

In addition to its characteristics as a model species, poplar provides an excellent point of departure for ecosystem research. The poplar genus consists of 29 species distributed throughout the northern hemisphere across a wide range of ecological amplitude.[4] Poplars often play a keystone role in riparian ecosystems, where they are pioneers on newly formed sediments and may be the dominant tree species on the landscape (e.g., in the high desert and intermountain west).[5] Individual poplar populations harbor a tremendous amount of genetic diversity in adaptive traits due to an outcrossing breeding system and extensive potential for gene flow among populations.[6] Also, closely related, sympatric poplar species form natural hybrid zones, further increasing the range of genetic variation in wild populations and creating excellent opportunities for studying the genetic mechanisms controlling species distributions.[7] Therefore, poplar offers an immediate opportunity for tractable studies of the molecular bases of adaptation with ecosystem-scale implications.

Association Studies for Detecting Molecular Bases of Adaptation

The study of the genetic basis of adaptation has long been of interest to evolutionary biologists and applied plant and animal breeders, and a rich array of analytical

techniques have been developed.[8,9] Early attempts at quantifying the effects of natural selection focused on spatial and temporal changes in phenotypes and, later, on changes in gene frequencies in response to environmental gradients or stimuli.[10–12] With the advent of high-density genetic maps, it became possible to identify quantitative trait loci (QTL) for traits with adaptive and/or commercial value, based on linkage with neutral molecular markers.[13,14] QTL analysis allows the identification of chromosomal segments containing one or more genes with quantifiable effects on adaptation, and a number of groups have successfully cloned genes of adaptive and commercial significance using this technique.[15]

Despite its widespread adoption and application, QTL analysis has some substantial limitations for the discovery of adaptive molecular polymorphisms. First, the size of the region identified by QTL analysis depends on the amount of linkage disequilibrium in the genome segment, which in turn is determined by the effective population size, selection, and local recombination rates.[16,17] Large amounts of linkage disequilibrium can result in identification of extensive genome segments containing hundreds of candidate genes and thousands of polymorphisms, making it nearly impossible to identify the genes and polymorphisms of adaptive significance.[15,18] Alternatively, low levels of linkage disequilibrium can make initial detection of adaptive QTL extremely difficult, requiring a very high density of mapped molecular markers.[19] Also, the range of inference of QTL analysis can be unacceptably narrow. Robust identification of QTL requires large pedigrees containing several hundred progeny, so these analyses are usually performed in one or two families only, especially in the case of forest trees for which progeny tests are large and expensive. QTL effects may not be conserved across families because of epistatic effects of the different genetic backgrounds.[8,20] Finally, adaptive polymorphisms identified in the context of controlled crosses and field trials may have little significance in wild ecosystems because of different selection pressures and genotype by environment interactions.[21,22]

An alternative approach to identifying adaptive polymorphisms is to take advantage of naturally occurring variation in a species and search for statistical associations with molecular polymorphisms.[19,23] This approach has been most commonly applied to qualitative traits such as case-control studies of human disease using individual single nucleotide polymorphism (SNP) markers or population haplotypes consisting of SNP's in linkage disequilibrium.[24,25] This is an extremely active area of research, with a number of notable successes in identifying disease genes.[26] However, skeptics have questioned the value of haplotype-based association studies because of uncertainties about the level of disequilibrium and hence the number of makers required for full genome coverage,[16,27] the confounding effects of population structure, epistasis, allelic heterogeneity, and the difficulties in establishing a causative link between polymorphisms and phenotypes.[19,28,29] A solution that addresses many of these concerns is the use of candidate genes to narrow the search for adaptive polymorphisms and thus reduce the number of markers that must be screened.[15,20,23] A candidate gene approach is particularly appropriate for a species like poplar, which has an outcrossing breeding system, high levels of genetic variation within wild populations, extensive gene flow among populations, and continuously large population sizes for thousands of generations, all of which suggests that linkage disequilibrium will be quite low.[16,30] Therefore, a tremendous number of markers would likely be required for a whole-genome scan for associations with adaptive phenotypes, a prospect that would be prohibitively expensive with current technology. However, information from model organisms such as *Arabidopsis* can guide the identification of candidate genes in poplar based on sequence homology, and these can be used in hypothesis-driven searches for polymorphisms with adaptive significance.

Pilot Study: Molecular Control of Gender Determination in Poplar

Analytical techniques are currently most developed for association studies involving qualitative traits such as disease incidence. Therefore, as a proof of principle, we analyzed gender in poplar because this is an easily assayed qualitative trait that is under complex genetic control.[1,31] Sex expression has been intensively studied in poplar because of its central importance in breeding and safety of genetically engineered plantations.[32] Also, sex expression has attracted considerable attention in model species such as *Arabidopsis*, in part because of the potential for direct improvements to food production through alteration of floral meristems, and also because of the central evolutionary importance of flowering. Determination of gender is most likely under the control of floral meristem identity genes, which in turn are regulated by transcription factors that control the transition of a vegetative meristem to a floral meristem, and transduction of environmental signals such as day length, light quality, and cold.[33] A large number of these genes have been identified in *Arabidopsis*, and a number of poplar homologs have been sequenced as well. The floral meristem identity genes are of particular interest, because they specify the formation of floral organs, and are likely to be involved in sex determination. In fact, overexpression of versions of poplar homologs of the transcription factors AGAMOUS and LEAFY resulted in production of female floral structures by a male poplar clone.[34] This suggests

that gender could be determined by a small number of sequence differences in regulatory elements and/or structural genes. This supposition is supported by results from an intensive effort aimed at finding neutral markers linked to sex determination in poplar[35] and willow (a member of the same family as poplar), which resulted in identification of only a single locus under complex genetic control.[31] By focusing on candidate genes known to affect floral morphology and gender, we should greatly enhance the prospects for elucidating the mechanism of gender formation.

Candidate Genes for Gender Determination

The ABC model of floral organ determination is a generally accepted framework for the actions of floral homeotic genes. Briefly, A-function genes specify sepal and petal formation, B-function genes specify petals and stamens, and C-function genes specify carpels. Petals require both A and B genes, and stamens require both B and C.[36] Genes from each of these functional classes have been isolated and sequenced in poplar, and we have reason to believe that each may be involved in gender determination.

LEAFY: This gene controls transition of vegetative meristems to inflorescence meristems, so it is upstream of the "ABC" genes. As a bud is forming, a decision point determines whether it will become a branch or a flower. LEAFY expression is intimately involved in that decision, and LEAFY probably controls expression of the other floral genes involved in the ABC model. The poplar gene is called PTLF:[37] Overexpression of this gene in poplar has caused a male tree to produce hermaphroditic and female flowers.[34] There is a Sort Interspersed Nuclear Element (SINE) insertion in the promoter region of this gene, and the region around the insertion is strongly differentiated between male and female trees (A. Brunner, unpublished data).

APETALA1: This is an A-function gene that specifies petal and sepal formation. Its possible role in gender is unknown, but it regulates expression of some of the genes involved in specifying male and female structures. The poplar genes are called PTAP1-1 and PTAP1-2 (A. Brunner, unpublished). .

AGAMOUS: This is a C-function gene, and knockouts result in completely sterile flowers. The poplar versions are called PTAG1 and PTAG2.[1] Overexpression of PTAG2 has caused male poplar trees to produce female flowers.[34]

Objectives

Objectives for this study were to explore molecular mechanisms of gender determination in poplar trees using a SNP association approach and assay allelic diversity and patterns of variation in major floral organ identity genes in poplar populations.

Technical Approach

Plant Collections

To accomplish the main objectives of this study, it was necessary to assay male and female trees that represented a cross section of the genetic variability that is present in *P. trichocarpa* populations. We therefore sampled trees across an east-west transect that represents the major axis of environmental variation in the range of *Populus trichocarpa*. We sampled trees from 7 populations, 3 of which were on the xeric east side of the Cascade mountains, and 4 from the more mesic west side of the Cascades (Table 1). Collections were made in November 2002. Many of the trees were selected because gender and microsatellite profiles had been determined for a previous study.[38] The remaining trees were haphazardly selected, taking care to avoid sampling multiple ramets from the same clone. We determined gender for trees with accessible floral buds. In total we analyzed 28 males, 26 females, and one cosexual tree that produced male, female, and hermaphroditic flowers, a fairly common occurrence in poplar.[39]

Table 1. Locations of sample collections
F, female; M, male; MF, cosexual.
Location is relative to the Cascade mountain range, which casts a substantial rain shadow to the east

Site	F	M	MF	Location
Hood River	0	3	0	West
Marchel	11	12	0	West
River Ranch	6	6	0	West
Scappoose	0	0	1	West
John Day	2	1	0	East
Snake River	1	3	0	East
Umatilla	6	3	0	East
Total	26	28	1	

Neutral Variation

We assessed variation for eight microsatellite loci derived from random genomic sequencing and an enrichment[40] (Table 2). We calculated genetic distance as the sum of squared differences between loci[41] and determined relationships among genotypes using the Unweighted Pair-Group Method using Arithmetic means in the PHYLIP software package.[42] We also calculated differentiation based on provenance (east versus west) and gender using *Rst* and exact tests in the Arlequin program.[43]

Selection of Genotyping Targets

It appears that poplars of each gender retain the ability to produce floral organs of the alternate gender, based on

Table 2. Characteristics of microsatellite loci used for assessment of neutral variation in poplar samples

LG, chromosomal linkage group on which marker is located;[56] N, number of trees; He, expected heterozygosity; Ho, observed heterozygosity; Fis, fixation index

Marker	LG	N	Alleles	He	Ho	Fis
O15	XVII	33	25	0.95	0.94	0.01
O349	IV	35	10	0.77	0.63	0.18
P2515	XIV	36	15	0.87	0.78	0.11
P2571	X	36	10	0.84	0.69	0.17
P2585	XV	29	13	0.83	0.59	0.30
P2610	VIII	33	10	0.59	0.45	0.22
P2885	XII	30	11	0.83	0.67	0.19
P649	XIII	32	37	0.97	0.75	0.23

the occurrence of cosexual genets and the conversion of gender in transgenics. We therefore hypothesized that gender was determined primarily by the patterns of expression of regulatory genes. We focused on known and suspected regulatory regions of some of the key regulatory genes involved in floral organ identity: PTAG1, PTAG2, PTLF, and PTAP1. We further refined our sequencing targets by comparing the genome sequence of the female clone, Nisqually-1, which has recently been sequenced by DOE's Joint Genome Institute, to sequences previously determined for a male clone.[1,37,44]

Polymorphism Detection

We tested two main approaches for discovering polymorphisms in our candidate gene regions: cloning and sequencing of PCR fragments, and direct sequencing of PCR products. In the first approach we PCR-amplified 2-kb fragments of each gene and cloned these using the Topo-TA cloning kit (Invitrogen). We then sequenced 3 to 10 clones per individual for each fragment, using universal primers for the initial forward and reverse sequences, and then designing specific sequencing primers for internal sequence. For direct PCR sequencing we amplified 500- to 700-bp fragments from the regions of interest, purified the products with EXO-SAP kits, and then sequenced the products directly. All amplifications were accomplished with Pfu DNA polymerase (Stratagene), and all sequencing was done with Big Dye Terminator kits (ABI). All sequencing and fragment analysis was accomplished on an ABI 3700 capillary sequencer.

SNP Analysis

We created alignments for each gene fragment using the Phred/Phrap/Consed suite of base calling and assembly programs.[45,46] We identified polymorphisms using Polyphred, an add-on to the above package that calls heterozygous bases for diploid sequences.[47] We processed the Polyphred output with a series of PERL scripts to identify haplotypes and format data for input into the programs for further analysis. For haploid data, derived by sequencing cloned fragments, haplotype diversity could be assessed directly using the Arlequin program, and comparison among haplotypes was accomplished using Jukes and Cantor genetic distance measure using the DNAdist program in the PHYLIP package, followed by construction of a UPGMA dendrogram. For diploid data derived by sequencing PCR products directly, haplotypes were inferred from allele frequencies using the maximum likelihood routine in the Arlequin package.[43]

We tested for linkage disequilibrium between all pairs of loci using exact tests as implemented by the Arlequin package for both haploid and diploid data. We used a Bonferroni criterion for significant linkage: the critical value for each gene fragment was calculated as 0.05/N, where N is the number of pairwise comparisons.

Results

Neutral Variation

The eight microsatellite loci were highly polymorphic in these populations, with 10 to 28 alleles observed for 37 to 44 individuals. Expected heterozygosities ranged from 0.59 to 0.97, but observed hetorozygosities were considerably lower, ranging from 0.45 to 0.94 and with FIS values between 0.09 and 0.30. These low-to-moderate FIS values may indicate that pooling of the samples has not caused a substantial Wahlund effect.[48] Furthermore, there was no evidence of differentiation between East and West trees ($R_{ST} = –0.021$, $P = 0.68 \pm 0.04$) and male and female trees ($R_{ST} = 0.083$, $P = 0.41 \pm 0.05$). Sampling was inadequate to assess pairwise population differentiation, so it is impossible to definitively assess the severity of the Wahlund effect caused by pooling these samples. However, a cluster analysis did not indicate substructure among these populations (Fig. 1), and the extremely long gene flow distances characteristic of *P. trichocarpa* have probably inhibited population differentiation for neutral loci in this species.[30,38]

Polymorphisms and Patterns of Linkage Disequilibrium

Overall Patterns

In total we produced over 2600 sequences for this project, and 832 of these were of sufficient length and quality to be included in the final alignments. We successfully surveyed 3638 bases of noncoding DNA and 498 bases of coding DNA in the vicinity of four floral homeotic genes (Table 3). We identified 98 polymorphisms, including 86 single nucleotide polymorphisms (SNPs) and 12 insertion/deletion (indel) polymorphisms. Ninety-two of the polymorphisms were

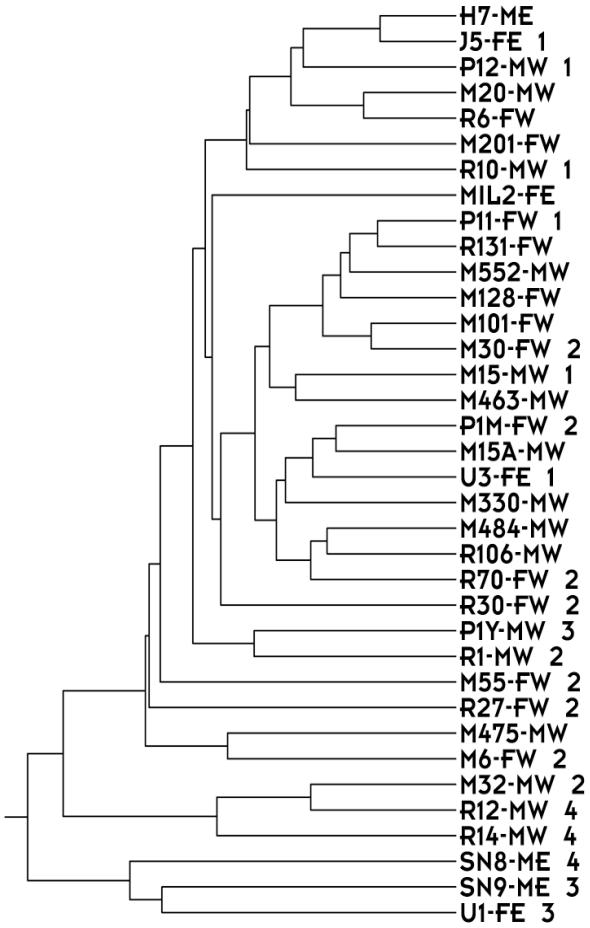


Fig. 1. UPGMA dendrogram depicting relationships among multilocus microsatellite genotypes. Distance measurement was the sum of squared pairwise differences in allele sizes (Slatkin 1995).

located in noncoding DNA, with 79 polymorphisms in upstream regions and 13 in introns. The only polymorphisms identified in coding regions were in the 5' untranslated region (UTR) of the PTAG1 gene. No polymorphisms were observed in 229 bp of exon sequence. In total, 169 sites showed significant LD, the vast majority of which were in the 5' region of PTAG1 (Table 4). There was no apparent relationship between LD and distance between sites (Fig. 2).

PTAG1

We sequenced portions of a cloned 2-kb fragment that began 1750 bp of upstream of the transcription start site, and extended 269 bases into the 5' UTR region. The first 1800 bp of sequence corresponds to a previously undescribed Ty1-copia class retrotransposon,[49] designated PtOP1. We observed a total of 51 SNPs in this region and no indels (Fig. 3). There were 36 different haplotypes observed among the 62 sequenced chromatids, and these ranged in frequency from 1.6% to 20%. There was

Table 3. Frequency of polymorphisms identified in coding and noncoding regions
P, number of polymorphisms observed

	Length	P	Rate (snp/kb)
	Noncoding		
Upstream	2605	79	30.3
Intron	1033	13	12.6
Total	3638	92	25.3
	Coding		
5' UTR	269	6	22.3
Exon	229	0	0.00
Total	498	6	12.1
Total	4136	98	23.7

significant linkage disequilibrium between 166 pairs of loci in this region (Fig. 4; Table 4). There were no significant differences in patterns of polymorphisms between males and females in this region, and patterns of linkage disequilibrium were similar between the two genders (Fig. 4). There were also no apparent patterns of relationships among haplotypes to suggest an association between these polymorphisms and gender (Fig. 5).

We also sequenced the 292 bases immediately upstream of the transcriptional start site, and 269 bases of the 5' untranslated region (UTR) (Table 4). In total we observed 13 polymorphisms, including two indels, and 22 haplotypes among the 56 assayed chromatids. Six of the SNPs were in the 5' UTR. No linkage disequilibrium was detected between the 156 pairs of polymorphic loci in the 561 bp region flanking the start site (Table 4).

Finally, we sequenced portions of intron 5 and exon 5 of the PTAG 1 gene, directly from a PCR product spanning positions 7060 to 7419. We observed 23 polymorphisms in this region and one indel, with a total of 23 haplotypes among the 25 PCR products examined. All of the polymorphisms were in the intron, and there were no fixed differences between males and females. We observed no significant LD between any of these polymorphisms.

PTAG2

For the PTAG2 gene we focused on a region 5' of the transcriptional start site that contained a large (426 bp) insertion/deletion polymorphism that differed between the initially sequenced male and female clones. This indel was particularly intriguing due to the presence of a 30 bp tandem repeat that could be indicative of a protein binding site, suggesting a possible regulatory role. A PCR assay using primers flanking the site of the insertion revealed that 22 out of 24 trees possessed at least one copy of the insertion, and four trees were heterozygous for the insertion polymorphism. The polymorphism was approximately equally divided between males and females, with one homozygote of each gender, and 2 male and one female

Table 4. Polymorphisms, linkage disequilibrium, and differentiation between males and females for sequenced fragments
T, type of fragment: H, haploid, D, diploid; N, number of trees sequenced; F, females; M, Males; A, All trees, B, polymorphism shared by both; M-F Dist, genetic distance between male and female clones; P Exact, value from Exact test

Gene	T	N	F	M	Start	End	Haplotypes			Polymorphisms			Indels			LD				M-F Dist	P Exact	Location
							A	F	M	A	F	M	A	F	M	A	F	M	B			
PTAG1	H	31	15	14	–1745	–611	36	18	25	51	36	27	0	0	0	166	83	64	35	–0.061	0.12	5'
PTAG1	H	28	13	14	–292	269	22	8	15	13	6	8	2	2	2	0	0	0	0	–0.978	0.06	5'/UTR
PTAG1	D	25	14	11	7060	7419	23	14	12	8	8	8	1	1	1	0	0	0	0	–0.006	0.79	Intron/Exon5
PTAG2	D	17	11	6	–908	–686	10	9	3	12	8	0	5	5	1	0	0	0	0	0.139	1.00	5'
PTAP1-1	D	29	15	14	657	1560	16	14	11	5	5	5	0	0	0	3	4	3	3	–0.048	0.07	Intron/Exon6
PTLF	D	21	12	9	–1630	–1119	36	19	16	9	8	7	4	4	3	0	4	3	3	–0.112	1.00	5'

heterozygotes. In addition, the cosexual tree was heterozygous for the polymorphism.

We also sequenced this PCR product for 17 trees. There were 12 polymorphisms in all, but nine of these occurred in just one female individual, and 5 were indels. The sequence of the large indel and flanking region was invariant for the 6 male trees we sequenced. No significant linkage disequilibrium was detected among the three polymorphisms with sufficiently high variation to allow testing.

PTAP1-1

There has been some uncertainty about the number of AP1 homologs in poplar due to the detection of multiple cDNAs corresponding to PTAP1-1, during initial cloning (A. Brunner, unpublished). One of the cDNA's had the entire 3' portion of the gene deleted following exon 5. It was unclear if this was due to alternate splicing or presence of a third gene. Furthermore, Southern blots suggested that there might be two copies of PTAP1-1 in females but not in males (A. Brunner, unpublished data). We therefore amplified the region between exons 5 and 6 to determine if a length polymorphism existed that differentiated male and female trees. Amplification of 30 trees, 15 males, 14 females, and one cosexual, failed to reveal any such polymorphisms, suggesting that alternative splicing accounts for the heterogeneous cDNAs rather than

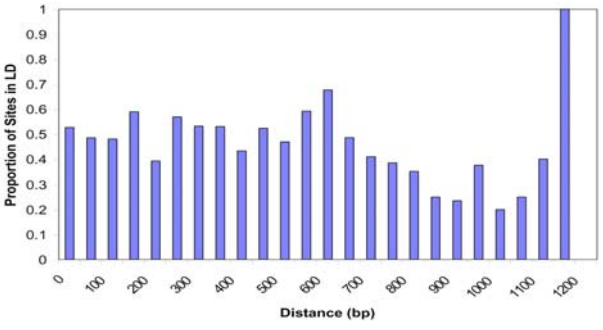


Fig. 2. Relationship between pairwise distance between polymorphisms and the occurrence of linkage disequilibrium for all sequenced fragments.

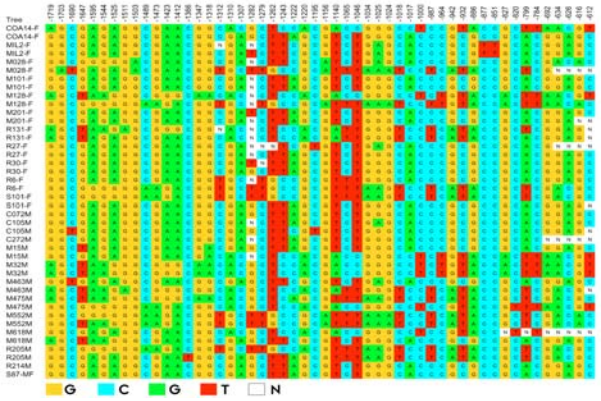


Fig. 3. Polymorphisms in a portion of a cloned fragment of the upstream region of the PTAG1 gene, which corresponds to a portion of the retrotransposon PtOP1. Individual trees are in rows, with a trailing F or M indicating female or male, respectively, and MF indicating a cosexual tree. Position relative to the start site is indicated in the first row.

duplicated loci. We also sequenced these PCR products and discovered five polymorphisms that were present in both male and female trees. All polymorphisms were located in introns. There was significant LD between 3 pairs of these polymorphic sites (Table 4).

PTLF

We focused on the 5' region of the PTLF gene that contained a SINE (short interspersed nuclear element) which also showed substantial similarity to a SINE in the 5' region of the PTAP1-1 gene. We identified nine polymorphisms in this region, four of which were indels. Three of these polymorphisms showed significant LD in males and four in female trees. There was no evidence of differentiation between males and females, however, and three of four pairs showed LD in both genders.

Discussion

Rates of polymorphism were moderately high in the regions we surveyed, more than an order of magnitude greater than polymorphisms observed in Arabidopsis (1 per kb,[50] about five times greater than Fugu (4 per kb[51]),

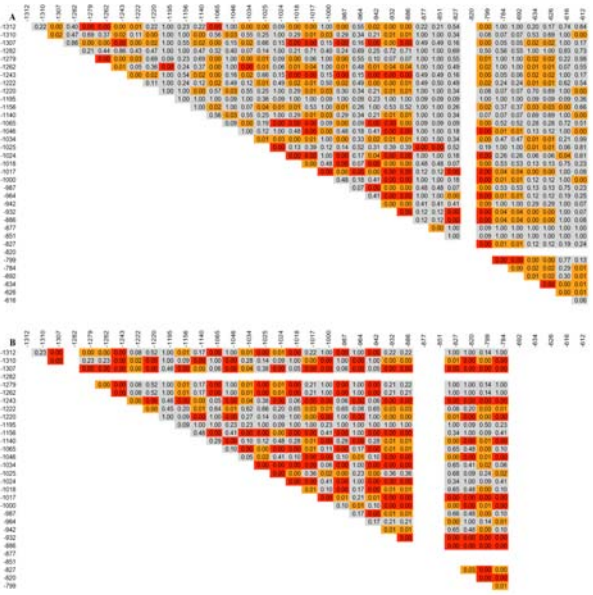


Fig. 4. Linkage disequilibrium between pairs of polymorphic sites in the retrotransposon 5' to the PTAG1 transcriptional start site. A. Female trees, upstream fragment. B. Male trees, upstream fragment. Blank rows and columns are loci that are fixed in male or female trees. Orange indicates significance at the 0.05 level. Red indicates significance with a Bonferroni correction ($P < 0.0001$ in this case).

comparable to sea squirt (12 per kb in a single individual[52]), and maize (32 per kb[53]). The patterns in this limited sample matched expectations: noncoding regions were greater than introns which in turn were greater than 5' UTR's. No polymorphisms were detected in exons, but the sample was extremely small. Much more robust estimates will soon be available for the individual clone that is being sequenced by the U.S. DOE.

These relatively high rates of polymorphism will provide abundant variation that can be exploited for association studies such as was attempted for the present study. However, LD was relatively low for most of the surveyed regions. If this pattern holds true on a genomic scale, this will make genome-wide association studies virtually impossible because an unrealistically large number of markers will be required to survey the genome.[16,27] At the same time, low LD will enhance the confidence of associations that are discovered between phenotypes and candidate gene polymorphisms, because low LD decreases the possibility of spurious associations.[19,28,29] Low LD also reduces the practical utility of QTL detected in experimental crosses, because marker-trait associations will be unpredictably disrupted outside of the mapped families.[54] Therefore, the present study provisionally supports the use of candidate gene association studies as a means of functional characterization in poplar.

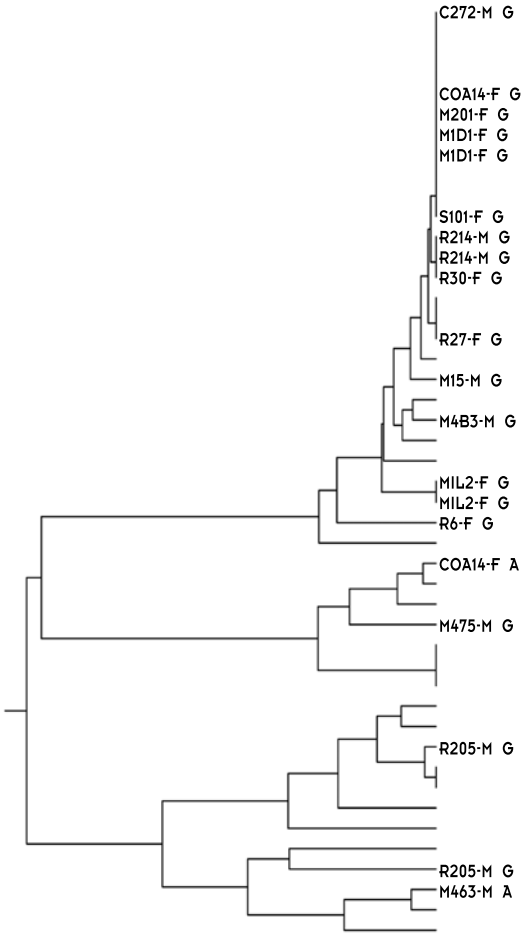


Fig. 5. UPGMA dendrogram for haplotypes for the same portions of the PTAG1 gene that are pictured in Figs. 3 and 4. All haplotypes of sequenced individuals are represented, and those connected by vertical lines are identical.

There are two important caveats about the LD estimates reported in this study. First, power was relatively low for most of the fragments due to the diploid nature of the data and relatively small number of individuals. Most LD was detected for the PTAG1 5' retrotransposon region, and this was also the data set for which the most power existed for detecting LD because linkage phase was known unequivocally, and the largest number of sequences were from this fragment. It is possible that LD would have been detected in several of the other fragments if more individuals had been sampled and if linkage phase had been determined. The other caveat is that LD was only tested over relatively short stretches, and there was no indication that rates of significant LD were declining with distance between tested markers, and LD was scattered throughout the fragments in which it occurred. This suggests that the LD is not due to repressed recombination

in these regions, because this would result in contiguous blocks of linkage.[55] Rather, this pattern may be indicative of coordinate selection on the paired polymorphisms. Such LD could occur over very large distances and illustrates that generalizations about LD will be difficult in the absence of direct tests of much larger numbers of loci.

The "proof-of-principle" portion of this study failed to uncover the hypothesized gender-associated polymorphisms. The observed polymorphisms occurred in both males and females, and LD generally occurred between the same pairs of loci in males and females (except in cases where spurious associations occurred due to splitting the data set and thereby reducing the frequency of individual polymorphisms). It should not be surprising that no associations were uncovered with this limited survey, since we only examined a small fraction of the possible regions of differentiation between the genders, and previous studies had indicated that gender determination was likely complex in the Salicaceae.[31,35] Furthermore, it is quite possible that epigenetic differences could account for much of the differential gene regulation that results in production of alternate sex organs in poplar, and these would not be detected with simple sequence surveys. Nevertheless, this approach is a valid one and warrants further effort, particularly as the finished poplar genome sequence emerges, and this can be readily compared on a large scale to ESTs and genomic sequence from male clones.

Finally, this study provided some insight into methodological issues surrounding SNP genotyping. The relatively high rates of polymorphism and low LD argue against SNP genotyping techniques that target individual polymorphisms. In fact, sequencing is currently one of the most cost-effective ways to survey polymorphisms in poplar because of the high density of polymorphisms. Furthermore, direct sequencing of PCR products is a viable strategy for identifying many poplar SNPs, but this technique suffers from significant shortcomings: (1) linkage phase cannot be determined, so power for detecting LD is greatly reduced; (2) the occurrence of indels renders diploid sequences uninterpretable; (3) identification of heterozygotes is tenuous except for the highest quality sequence regions; and (4) it is nearly impossible to sequence duplicated genome regions this way because mixtures of PCR products provide very poor sequence. Therefore, if SNP discovery and assessment of LD are the goals, cloning of mid-sized PCR fragments (2–4 Kb) is a viable but expensive strategy. Ultimately, shotgun sequencing of tiled BACs from multiple individuals may be the best way to determine SNP diversity and LD over large regions.

Summary and Conclusions

This proof-of-principle study has provided valuable initial data on SNP variation in poplar, a species with tremendous economic and ecological importance. The knowledge and techniques that we have gained from this work will be directly applied in future studies of adaptation of plant populations, and will therefore add to our understanding of ecosystem composition and aid in predicting ecosystem responses to climatic change. This will have direct benefits for multiple U.S. government agencies that are concerned with the effects of climatic change, including DOE, NASA, EPA, and DOD.

This work has already led to collaborations that will enhance the competitiveness of ORNL in plant genetic research. We have been involved in two proposals that build directly off of this work. One was for the NSF Frontiers in Integrative Biological Research. This multimillion dollar project is a collaboration between Northern Arizona University, ORNL, and the University of Wisconsin. The preproposal recently passed competitive review, and we have been invited to submit a full proposal this year. In addition, we have been invited to participate in another proposal for the NSF Systematic and Population Biology Program, which would bring several hundred thousand dollars to the lab if funded. Finally, this work has positioned us to pursue future opportunities with DOE programs such as the Program for Ecosystem Research and the Basic Energy Sciences program.

References

[1]A. M. Brunner, W. H. Rottmann, L. A. Sheppard, K. Krutovskii, S. P. DiFazio, S. Leonardi, and S. H. Strauss, "Structure and expression of duplicate AGAMOUS orthologues in poplar," *Plant Mol Biol.* **44**, 634 (2000).

[2]H. D. J. Bradshaw, R. Ceulemans, J. Davis, and R. Stettler, "Emerging model systems in plant biology: poplar (*Populus*) as a model forest tree," *J Plant Growth Reg* 306–313 (2000).

[3]C. C. Mann, and M. L. Plummer, "Biotechnology: Forest Biotech Edges Out of the Lab," *Science* **295**, 1626–1629 (2002).

[4]J. E. Eckenwalder, "Systematics and evolution of *Populus*." In *Biology of Populus and Its Implications for Management and Conservation*, pp. 7–32. Editors: R. F. Stettler, H. D. Bradshaw, Jr., P. E. Heilman, and T. M. Hinckley. NRC Research Press, Ottawa, Canada, 1996.

[5]J. H. Braatne, S. B. Rood, and P.E. Heilman, "Life history, ecology, and reproduction of riparian cottonwoods in North America." In *Biology of Populus and Its Implications for Management and Conservation*, pp. 57–85. Editors: R. F. Stettler, H. D. Bradshaw, Jr., P. E. Heilman, and T. M. Hinckley. NRC Research Press, Ottawa, Canada, 1996.

[6]R. E. Farmer, Jr. (1996) The genecology of *Populus*. In: Biology

of Populus and its implications for management and conservation, pp. 33–55. Editors: R. F. Stettler, H. D. Bradshaw, Jr., P. E. Heilman, and T. M. Hinckley. NRC Research Press, Ottawa, Canada.

[7]L. H. Rieseberg and S. E. Carney, "Tansley Review no. 102 plant hybridization," *New Phytol.* **140**, 599–624 (1998).

[8]N. H. Barton and P. D. Keightley, "Understanding quantitative genetic variation," *Nature Reviews Genetics* **3**, 11–21 (2002).

[9]P. W. Hedrick, M. E. Ginevan, and E. P. Ewing, "Genetic polymorphism in heterogeneous environments," *Annual Review of Ecology and Systematics* **7**, 1–32 (1976).

[10]R. Lande and S. J. Arnold, "The measurement of selection on correlated characters," *Evolution* **36**, 1210–1226 (1983).

[11]J. G. Kingsolver, H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, and A. Hoang, "The strength of phenotypic selection in natural populations," *American Naturalist* **157**, 245–261 (2001).

[12]K. Ritland, "Marker-inferred relatedness as a tool for detecting heritability in nature," *Molecular Ecology* **9**, 1195–1204 (2000).

[13]R. Doerge, "Mapping and analysis of quantitative trait loci in experimental populations," *Nature Reviews Genetics* **3**, 43–52 (2002).

[14]M. Lynch and B. Walsh, Genetics and Analysis of Quantitative Traits. Sinauer Associates, Sunderland, MA, 1998.

[15]D. L. Remington, M. C. Ungerer, and M. Purugganan, "Map-based cloning of quantitative trait loci: progress and prospects," *Genet.Res.* **78**, 213–218 (2001).

[16]J. D. Terwilliger and K. M. Weiss, "Linkage disequilibrium mapping of complex disease: fantasy or reality?," *Curr. Opin. Biotechnol.* **9**, 578–594 (1998).

[17]J. K. Pritchard and M. Przeworski, "Linkage disequilibrium in humans: Models and data," *American Journal of Human Genetics* **69**, 1–14 (2001).

[18]B. Stirling, G. Newcombe, J. Vrebalov, I. Bosdet, and H. D. Bradshaw, "Suppressed recombination around the MXC3 locus, a major gene for resistance to poplar leaf rust," *Theor. Appl. Genet.* **103**, 1129–1137 (2001).

[19]L. R. Cardon and J. I. Bell, "Association study designs for complex diseases," *Nat. Rev. Genet.* **2**, 91–99 (2001).

[20]B. Walsh, "Quantitative genetics in the age of genomics," *Theoretical Population Biology* **59**, 175–184 (2001).

[21]S. H. Strauss, R. Lande, and G. Namkoong, "Limitations of Molecular-Marker-Aided Selection in Forest Tree Breeding," *Canadian Journal of Forest Research* **22**(7), 1050–1061, 1050–1061 (1992).

[22]M. Pigliucci, "Ecological and evolutionary genetics of *Arabidopsis*," *Trends in Plant Science* **3**, 485–489 (1998).

[23]N. J. Risch, "Searching for genetic determinants in the new millennium," *Nature* **405**, 847–856 (2000).

[24]D. E. Reich and E. S. Lander, "On the allelic spectrum of human disease," *Trends in Genetics* **17**, 502–510 (2001).

[25]G. C. Johnson, L. Esposito, B. J. Barratt, A. N. Smith, J. Heward, G. Di Genova, H. Ueda, H. J. Cordell, I. A. Eaves, F. Dudbridge, R. C. Twells, F. Payne, W. Hughes, S. Nutland, H. Stevens, P. Carr, E. Tuomilehto-Wolf, J. Tuomilehto, S. C. Gough, D. G. Clayton, and J. A. Todd, "Haplotype tagging for the identification of common disease genes," *Nat. Genet.* **29**, 233–237 (2001).

[26]E. R. Martin, E. H. Lai, J. R. Gilbert, A. R. Rogala, A. J. Afshari, J. Riley, K. L. Finch, J. F. Stevens, K. J. Livak, B. D. Slotterbeck, S. H. Slifer, L. L. Warren, P. M. Conneally, D. E. Schmechel, I. Purvis, M. A. Pericak-Vance, A. D. Roses, and J. M. Vance, "SNPing away at complex diseases: analysis of single-nucleotide polymorphisms around APOE in Alzheimer disease," *Am. J. Hum. Genet.* **67**, 383–394 (2000).

[27]L. Kruglyak, "Prospects for whole-genome linkage disequilibrium mapping of common disease genes," *Nat. Genet.* **22**, 139–144 (1999).

[28]K. M. Weiss and J. D. Terwilliger, "How many diseases does it take to map a gene with SNPs?," *Nat. Genet.* **26**, 151–157 (2000).

[29]D. Altshuler, M. Daly, and L. Kruglyak, "Guilt by association," *Nature Genetics* **26**, 135–137 (2000).

[30]J. C. Weber and R. F. Stettler, "Isoenzyme variation among ten [riparian] populations of Populus trichocarpa Torr. et Gray in the Pacific Northwest," *Silvae Genet.* **30**, 82–87 (1981).

[31]C. L. M. Alstrom-Rapaport, Y. C. Wang, G. Roberts, and G. A. Tuskan, "Identification of a RAPD marker linked to sex determination in the basket willow (*Salix viminalis* L.)," *Journal of Heredity* **89**, 44–49 (1998).

[32]S. H. Strauss, W. H. Rottmann, A. M. Brunner, and L. A. Sheppard, "Genetic engineering of reproductive sterility in forest trees," *Molecular Breeding* **1**, 5–26 (1995).

[33]G. G. Simpson, A. R. Gendall, and C. Dean, "When to switch to flowering," *Annu. Rev. Cell Dev. Biol.* **15**, 519–550 (1999).

[34]S. H. Strauss, R. Meilan, S. P. DiFazio, A. M. Brunner, and J. Carson, Tree Genetic Engineering Research Cooperative (TGERC) Annual Report:2000–2001. Forest Research Laboratory, Oregon State University, Corvallis, Oregon, 2001.

[35]D. N. McLetchie and G. A. Tuskan, "Gender determination im Populus," *Norwegian Journal of Agricultural Sciences Supplement* **18**, 57–66 (1994).

[36]A. L. Lawton-Rauh, E. S. Buckler, and M. D. Purugganan, "Patterns of molecular evolution among paralogous floral homeotic genes 8," *Mol. Biol. Evol.* **16**, 1037–1045 (1999).

[37]W. H. Rottmann, R. Meilan, L. A. Sheppard, A. M. Brunner, J. S. Skinner, C. Ma, S. Cheng, L. Jouanin, G. Pilate, S. H. Strauss, C. P. Ma, and S. P. Cheng, "Diverse effects of overexpression of LEAFY and PTLF, a poplar (Populus) homolog of LEAFY/FLORICAULA, in transgenic poplar and Arabidopsis," *Plant J.* **22**, 235–245 (2000).

[38]S. P. DiFazio, Measuring and Modeling Gene Flow from Hybrid Poplar Plantations: Implications for Transgenic Risk Assessment. PhD Thesis, Oregon State University, Corvallis, OR, 2002. http://www.fsl.orst.edu/tgerc/dif_thesis/difaz_thesis.pdf.

[39]R. F. Stettler, "Variation in sex expression of Black Cottonwood

and related hybrids," *Silvae Genet.* 42–46 (1971).

[40]G. A. Tuskan, L. E. Gunter, Z. Yang, T. M. Yin, M. Sewell, and S. P. DiFazio, "Characterization of Microsatellites Revealed by Genomic Sequencing of *Populus trichocarpa*," *Can. J. Forest Res.* in press (2004).

[41]M. Slatkin, "A Measure of Population Subdivision Based on Microsatellite Allele Frequencies," *Genetics* **139**, 457–462 (1995).

[42]J. Felsenstein (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Produced and Distributed by Author.

[43]M. Slatkin and L. Excoffier, "Testing for linkage disequilibrium in genotypic data using the expectation-maximization algorithm," *Heredity* **76**, 377–383 (1996).

[44]L. A. Sheppard, A. M. Brunner, K. V. Krutovskii, W. H. Rottmann, J. S. Skinner, S. S. Vollmer, and S. H. Strauss, "A DEFICIENS homolog from the dioecious tree black cottonwood is expressed in female and male floral meristems of the two-whorled, unisexual flowers," *Plant Physiol.* **124**, 627–639 (2000).

[45]D. Gordon, C. Abajian, and P. Green, "Consed: a graphical tool for sequence finishing," *Genome Res.* **8**, 195–202 (1998).

[46]B. Ewing and P. Green, "Base-calling of automated sequencer traces using phred. II. Error probabilities," *Genome Research* **8**, 186–194 (1998).

[47]D. A. Nickerson, V. O. Tobe, and S. L. Taylor, "PolyPhred: Automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing," *Nucleic Acids Res.* **25**, 2745–2751 (1997).

[48]D. Hartl and A. G. Clark, Principles of Population Genetics. Sinauer Associates, Inc., Sunderland, MA, 1997.

[49]A. J. Flavell, S. R. Pearce, J. S. P. Heslop-Harrison, and A. Kumar, "The evolution of Ty1-copia group retrotransposons in eukaryote genomes," *Genetica* **100**, 185–195 (1997).

[50]F. J. Cho, M. Mindrinos, D. R. Richards, R. J. Sapolsky, M. Anderson, E. Drenkard, L. Dewdney, T. L. Reuber, M. Stammers, N. Federspiel, A. Theologis, W. H. Yang, E. Hubbell, M. Au, E. Y. Chung, D. Lashkari, B. Lemieux, C. Dean, R. J. Lipshutz, F. M. Ausubel, R. W. Davis, and P. J. Oefner, "Genome-wide mapping with biallelic markers in Arabidopsis thaliana," *Nature Genetics* **23**, 203–207 (1999).

[51]S. Aparicio, J. Chapman, E. Stupka, N. Putnam, J. Chia, P. Dehal, A. Christoffels, S. Rash, S. Hoon, A. Smit, M. D. S. Gelpke, J. Roach, T. Oh, I. Y. Ho, M. Wong, C. Detter, F. Verhoef, P. Predki, A. Tay, S. Lucas, P. Richardson, S. F. Smith, M. S. Clark, Y. J. K. Edwards, N. Doggett, A. Zharkikh, S. V. Tavtigian, D. Pruss, M. Barnstead, C. Evans, H. Baden, J. Powell, G. Glusman, L. Rowen, L. Hood, Y. H. Tan, G. Elgar, T. Hawkins, B. Venkatesh, D. Rokhsar, and S. Brenner, "Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes," *Science* **297**, 1301–1310 (2002).

[52]P. Dehal, Y. Satou, R. K. Campbell, J. Chapman, B. Degnan, A. De Tomaso, B. Davidson, A. Di Gregorio, M. Gelpke, D. M. Goodstein, N. Harafuji, K. E. M. Hastings, I. Ho, K. Hotta, W. Huang, T. Kawashima, P. Lemaire, D. Martinez, I. A. Meinertzhagen, S. Necula, M. Nonaka, N. Putnam, S. Rash, H. Saiga, M. Satake, A. Terry, L. Yamada, H. G. Wang, S. Awazu, K. Azumi, J. Boore, M. Branno, S. Chin-bow, R. DeSantis, S. Doyle, P. Francino, D. N. Keys, S. Haga, H. Hayashi, K. Hino, K. S. Imai, K. Inaba, S. Kano, K. Kobayashi, M. Kobayashi, B. I. Lee, K. W. Makabe, C. Manohar, G. Matassi, M. Medina, Y. Mochizuki, S. Mount, T. Morishita, S. Miura, A. Nakayama, S. Nishizaka, H. Nomoto, F. Ohta, K. Oishi, I. Rigoutsos, M. Sano, A. Sasaki, Y. Sasakura, E. Shoguchi, T. Shin-i, A. Spagnuolo, D. Stainier, M. M. Suzuki, O. Tassy, N. Takatori, M. Tokuoka, K. Yagi, F. Yoshizaki, S. Wada, C. Zhang, P. D. Hyatt, F. Larimer, C. Detter, N. Doggett, T. Glavina, T. Hawkins, P. Richardson, S. Lucas, Y. Kohara, M. Levine, N. Satoh, and D. S. Rokhsar, "The draft genome of Ciona intestinalis: Insights into chordate and vertebrate origins," *Science* **298**, 2157–2167 (2002).

[53]A. Ching, K. S. Caldwell, M. Jung, M. Dolan, O. S. Smith, S. Tingey, M. Morgante, and A. J. Rafalski, "SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines," *Bmc Genetics* **3** (2002).

[54]S. Strauss, R. Lande, and G. Namkoong, "Limitations of molecular-marker-aided selection in forest tree breeding," *Canadian Journal of Forest Research* **22**, 1050–1061 (1992).

[55]D. E. Reich, M. Cargill, S. Bolk, J. Ireland, P. C. Sabeti, D. J. Richter, T. Lavery, R. Kouyoumjian, S. F. Farhadian, R. Ward, and E. S. Lander, "Linkage disequilibrium in the human genome," *Nature* **411**, 199–204 (2001).

[56]T. M. Yin, S. P. DiFazio, L. E. Gunter, D. E. Riemenschneider, and G. A. Tuskan, "Large-scale Heterospecific Segregation Distortion in Populus Revealed by a Dense Genetic Map," *Theoretical & Applied Genetics* (accepted) (2004).

Genomic Characterization of Belowground Ecosystem Responses to Climate Change

S. P. DiFazio,[1] J. Zhou,[1] L. E. Gunter,[1] C. C. Brandt,[1] R. J. Norby,[1] J. C. Schryver,[2] and J. F. Weltzin[3]
[1]*Environmental Sciences Division*
[2]*Computer Science and Mathematics Division*
[3]*Department of Ecology and Evolutionary Biology, University of Tennessee*

Belowground processes are key to predicting the trajectory and effects of global change on terrestrial ecosystem responses to atmospheric and climatic perturbations. However, these processes remain poorly understood because dated and cumbersome methodologies are being applied to highly complex and heterogeneous biological and geochemical systems. We are developing quantitative, species-specific molecular assays to analyze the plant root composition of soil cores, thereby characterizing belowground competitive interactions with unprecedented precision. We are also assessing functionally significant changes in plant-associated microbial communities by assaying the relative abundance and expression of nitrogen-cycle genes using microarray technology. This research will provide a unique view into the inscrutable yet important world of belowground biota, thus contributing to a mechanistic understanding of responses of entire plant and microbial communities to climate change.

With the rapid development of high-throughput techniques for genotyping and expression analysis, exponentially expanding sequence databases, and rapidly developing computational and bioinformatics resources, the field of genomics has the potential to revolutionize ecological research by enabling the study of obscure and recalcitrant systems with unprecedented precision.

The project has three main tasks: (1) development of quantitative real-time PCR (QRT-PCR) assays that can be used to determine the abundance of different plant species in root samples, (2) development of a microarray for assaying abundance and diversity of organisms involved in nitrogen cycling, and (3) application of the developed techniques to determine the relationship between plant species abundance and microbial diversity in an experiment involving seven plant species and altered [CO_2], temperature, and moisture (the OCCAM facility).

Task 1.1. DNA extraction from roots. We have propagated all seven plant species involved in the OCCAM facility and determined that processing with a ball mill provides the most uniform and consistent tissue for DNA extraction. After testing a wide variety of extraction methods, we determined that a silica gel column provides the most consistent yields of high-quality DNA from the target species.

Task 1.2. Species-specific PCR assays. We designed PCR primers for each target species and two reference species using publicly available nuclear DNA sequences to amplify a species-specific, 200-bp PCR product. We optimized PCR conditions for each species and tested primers against each of the target species as well as nine weed species that occur in or near the study plots. Our PCR assays were highly specific for the target species, even differentiating *Trifolium pratense* from the two congeneric weeds *T. repens* and *T. campestre* (Fig. 1).

Task 1.3. QRT-PCR using reference samples. We have completed pure dye calibrations on the BioRad iCycler and are currently analyzing QRT-PCR conditions on a standard thermocycler using Sybr Green I label in the

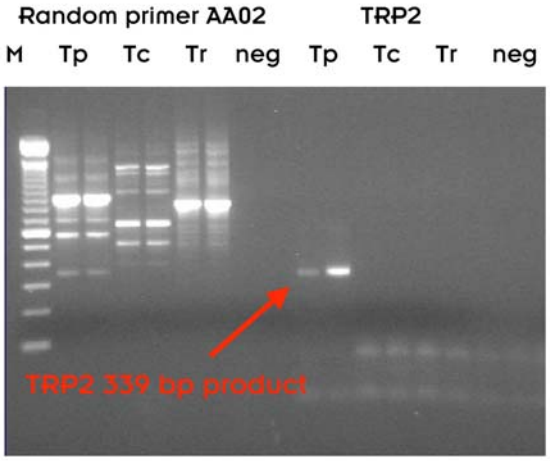


Fig. 1. Specificity of Trifolium pratense *(Tp) primers. DNA from* T. campestre *(Tc) and* T. repens *(Tr) was not amplified with the Tp-specific TRP2 primer (targeting a region of the polyphenol oxidase 2 gene). Amplification with the random decamer primer AA02 indicated that the DNA from all species was of sufficient quality to ensure amplification with matching primers.*

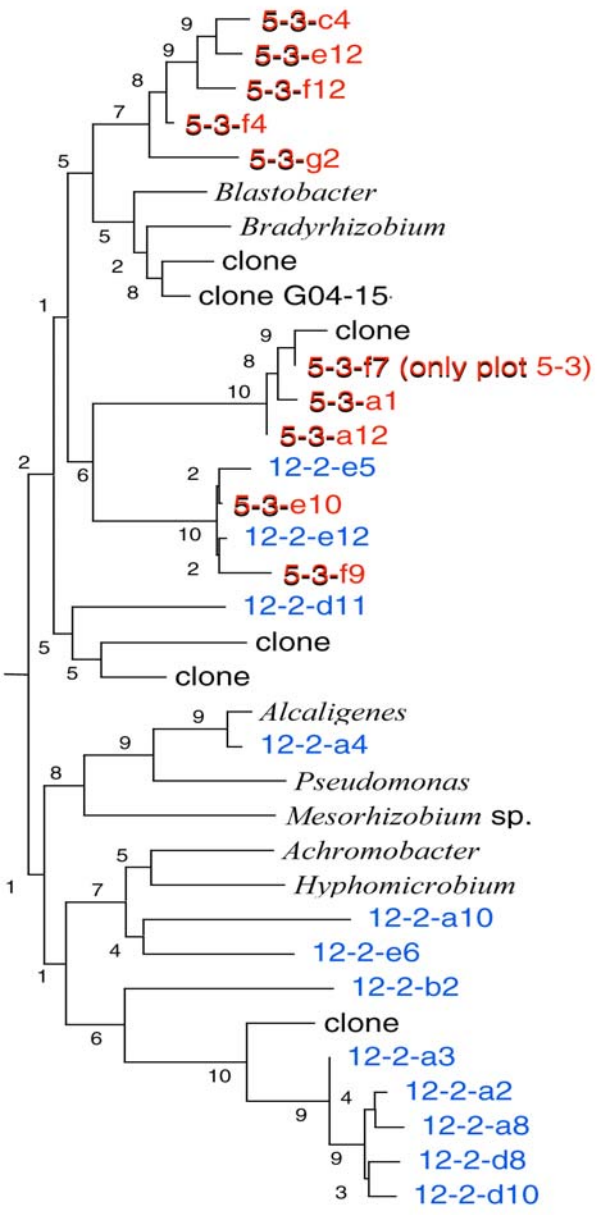


Fig. 2. Neighbor joining tree showing relationships among PCR-amplified fragments of nir*K (Nitrite Reductase K) genes from plots 5 (blue) and 12 (red) from the OCCAM study and previously characterized* nir*K genes. The two plots contained substantially different nitrogen-reducing bacterial communities, and many of these organisms were previously unrepresented in public sequence databases.*

reactions and are in the initial stages of performing QRT-PCR on our salmon reference sample. *Future Plans*: We will derive QRT-PCR calibration curves for each of our target species in isolation and in mixtures with DNA from other species. We will then test the assay on field-collected samples spiked with known quantities of DNA from reference species. Finally, we will test species quantification using mixtures of known quantities of roots.

Task 2.1. DNA and RNA extraction from root-associated soil. Our previous protocol for the extraction of nucleic acids was modified to account for samples derived from soil cores, and in particular to enrich for rhizosphere-associated microbes. With a soil sample of 50 g, a root wash yielded 3 to 5 µg of DNA, and the surrounding soil yielded approximately 100 µg. Because of required replicates and dye switching for microarray experiments, approximately 20 µg is needed.

Task 2.2.1. Database sequences. We have extracted all known *Amo*A, *Nir*S, *Nir*K, and *Nif*H genes from public sequence public databases.

Task 2.2.2. PCR amplification of N-cycling genes from root-associated soil. Clonal libraries have been constructed from PCR amplification of pooled OCCAM rhizosphere DNA using conserved primers for nitrogen cycle genes. We have sequenced portions of these libraries to characterize the microbial communities of two of the OCCAM plots prior to treatment. We found that there were substantial differences between plots for the *nir*K gene and that 84% of the sequences discovered were not represented in current databases (Fig 2).

Task 2.3.1. Microarray construction. More than 5000 unique probes have been designed and spotted onto oligonucleotide arrays. *Future Plans*: Available freeware is being modified for the design of unique probes to account for the alignment of segments common to all sequences, comparison to our growing sequence databases, and comparison back to public databases to test "overall" uniqueness of probes.

Task 2.3.2. Microarray hybridization and scanning. We have tested the arrays using DNA from bulk soil from OCCAM cores. Preliminary results indicate that *nir*K, *nir*S, *amo*A, and *nif*H are present. Two *amo*A sequences (n = 48) and four *nif*H sequences (n = 220) predominated, and these were principally derived from Tennessee groundwater and forested upland soils.

Task 3.2. Data Analysis. We will analyze treatment effects and associations between plant species and microbial diversity using a variety of linear and nonlinear multivariate techniques, including multivariate analysis of variance, principal components analysis, partial least squares, correspondence analysis, canonical correspondence analysis, and artificial neural networks.

Ecosystem genomics is an exciting new field of inquiry where the tools and information from the field of genomics are providing unprecedented insights into the structure and functioning of biological communities. This project is a good example of such work. The methods we are developing will allow ecologists to gain unprecedented insights into the composition of belowground communities, and the responses of those communities to perturbations such as elevated CO_2 and climatic change.

ENVIRONMENTAL SCIENCE AND TECHNOLOGY

Seed Money Fund

Metabolic Profiling: A Required Element in Functional Genomics

T. J. Tschaplinski,[1] G. J. Van Berkel,[2] and J. P. Miles[3]
[1]*Environmental Sciences Division*
[2]*Chemical Sciences Division*
[3]*Center College, Danville, Kentucky*

The emerging science of metabolomics or metabolite profiling offers tremendous potential to discover novel genes and assign function to those genes. Metabolic profiling determines the consequences of a targeted change in gene activity and has the potential to provide information on gene function and how it affects the complex biochemical network. The creation of a comprehensive gas chromatography–mass spectrometry (GC-MS) library of trimethylsilyl (TMS) derivatives of organic constituents is essential for the near-term development of metabolic profiling in poplar (*Populus*). A proof-of-principle study was initiated to demonstrate that broad-spectrum metabolite analyses can be achieved, with derivatization protocols effective at reproducibly silylating the 300–450 metabolites, with the number of metabolites analyzed and throughput rates increased by the concurrent development of direct-infusion electrospray mass spectrometry (DI-ES-MS). The research is organized around three tasks that include (1) optimization of GC-MS derivatization and analytical protocols for high throughput; (2) creation of a mass spectral library of known *Populus* metabolites, including the characterization of unknown compounds; and (3) extension of analyses to the liquid chromatography–mass spectrometry (LC-MS) platform to further increase the number of metabolites analyzed and improve characterization of unknowns. The project has resulted in substantial progress towards meeting the goals of all three tasks that places ORNL in a more competitive position to pursue outside funding for metabolomics research.

Introduction

Sequencing of microbial and plant genomes will require ascribing function to the vast amount of draft gene sequence data to establish precise relationships between gene structure and phenotypic variation. Changes in gross phenotypes (e.g., branching, flowering time, growth rate) are the integrated result of changes in the pattern of many interacting genes. In many cases, these complex networks can only be deconvoluted through detailed analysis of changes in metabolites, the precise endpoints of complex biochemical pathways. Assignment of gene function will require broad-spectrum biochemical analyses that be used in high-throughput phenotyping. Key elements in achieving high throughput that we are addressing include establishing a comprehensive metabolite database of mass spectra and minimizing time for sample preparation for GCMS, and maximizing the number to metabolites analyzed by developing DI-MS protocols.

Technical Approach

The research is organized around three tasks that include (1) optimization of gas chromatography–mass spectrometry (GC-MS) derivatization and analytical protocols for high throughput; (2) creation of a mass spectral library of known *Populus* metabolites, including the characterization of unknown compounds and establishment of deconvolution strategies to separate overlapping compounds; and (3) extension of analyses to the liquid chromatography–mass spectrometry (LC-MS) platform to further increase the number of metabolites analyzed and improve characterization of unknowns. Low volatility and thermal instability result in many metabolites not being amenable to analysis by GC/MS. In order to include such difficult-to-derivatize metabolites in a profiling strategy, including those metabolites with higher molecular weight or high polarity, LC-MS protocols were tested. These compounds include sugar phosphates and nucleotides involved in energy transduction. Development of such protocols will obviate the need for derivatization and expand the range of compounds that can be included in metabolite profiling. Phosphorylated metabolites are involved in primary C metabolism, the fixation of carbon dioxide to carbohydrates (e.g., Calvin-Benson-Basham cycle) or the metabolism of carbohydrates to provide energy (e.g., glycolysis and the pentose phosphate pathway). Furthermore, the relative concentrations of phosphorylated metabolites, and especially the nucleotides, are essential to evaluate the metabolic status of an organism. The importance of determining

phosphorylated metabolites by LC-ES-MS is confirmed by a recent publication from the metabolomics research group at the Max Planck Institute.[1]

Though chromatography of phosphorylated metabolites is problematic, the importance of adenylates (the nucleotides, ATP, AMP, etc.) in energy transduction has led several research groups to attempt the protocol development with varying degrees of success. The analysis of such adenylates with reverse-phase, high-performance liquid chromatography (HPLC) is problematic (without using ion-pairing reagents). The concentrations of buffer salts are often 0.1 molar and greater, which would be incompatible with electrospray mass spectrometry (ES-MS). An alternative approach to reverse-phase HPLC is hydrophilic interaction chromatography (HILIC). This approach relies upon partition of components between a somewhat polar mobile phase and a water phase immobilized onto a support. The nature of the stationary water layer is not well known. Elution order is similar to that found in normal-mode chromatography—nonpolar components elute first and more polar (hydrophilic) components are retained longer. Mobile-phase gradients begin with less polar solvent (commonly acetonitrile) and progress to increasing proportions of aqueous buffer. The approach has recently been used by Tolstikov and Fiehn (2002). They evaluated both TSK Gel Amide 80 column from TosoHass and PolyHYDROXYETHYL A column from PolyLC and found the former to be slightly better. Our approach was to test several reverse-phase approaches and focus the most attention on HILIC.

Results and Accomplishments

Task 1. Protocol refinement to minimize time of sample preparation and analysis was conducted by contrasting an aggressive derivatization protocol (stronger TMS donors, higher temperatures, and longer duration) with the lower-temperature methoximation followed by trimethylsilylation, used by the Max Planck researchers. The objective was to find the method that optimizes sample preparation time and spectral output (i.e., maximum metabolites observed). The stability of the low-molecular-weight organic acids was better with methoximation, but there was a tradeoff in less efficient derivatization of higher-molecular-weight phenolic glycosides. Tradeoffs between derivatization efficiency and number of metabolites that can be characterized will have to be made, depending on whether the analyses are targeted to a class of compounds or remain broad spectrum. Included in this task was a major proof of principle, which involved comparing the metabolite profiles of polar compounds of activation-tagged transgenic *Populus tremula* x *alba* with the non-transformed hybrid. The transgenic line (B18) displayed the branchy phenotype. The largest difference between the transgenic and the non-transformed control was a 4-fold increase in a metabolite that matched the predicted fragmentation pattern of ent-7α-hydroxykaurenoic acid, a key precursor of gibberellins (GAs), a class of growth regulators that regulate height growth in plants. The accumulation of a key GA precursor in a branchy mutant suggest GAs play a role in determining apical dominance, previously thought to be controlled by the antagonistic interaction of the plant growth regulators, auxin, and cytokinins. Our first attempt to determine a metabolic phenotype of a transgenic mutant poplar clone was successful.

Task 2. Effort was focused on the establishment of a comprehensive GC-MS library of TMS derivatives of organic constituents, particularly those of *Populus*. Many of the metabolites of *Populus* are shared with willow (*Salix* sp.). Leaf and stem samples of a number of poplar and willow species were first extracted, TMS derivatized, and analyzed by GC-MS. Many metabolites were identified and key fragments and retention time assigned to those compounds that remain unknown for future deconvolution analyses. A substantial database that includes the fragmentation patterns of many compounds was established that now includes ~ 450 metabolites with approximately half of TMS-derivatives of those metabolites not in commercially available databases (Wiley and NIST). The fragmentation pattern of known poplar compounds were predicted to aid in identifying unknowns. Recent metabolic profiles of *Arabidopsis* were used to confirm the presence or absence of positively identified metabolites in *Populus*. The fragmentation patterns of TMS-derivatized metabolites from our previous GC-MS analyses of *Populus* were matched to that of the Max Planck Institute databases.

Task 3. Reverse phase is certainly the most widely used mode in HPLC due to the wide range of analytes and sample matrices addressed and the chromatographic reproducibility offered by modern bonded-phase columns. Our work in this mode acknowledged that retention of the ionic phosphorylated metabolites would be difficult and non-ionic but polar metabolites might be retained irreversibly under the conditions suitable for the anionic components. The results indicate the column operated under two modes. Under conditions of relatively high organic modifier, none of the seven test compounds were retained very long and elution does appear to be in order of decreasing polarity. For instance, glucose phosphate eluted early, and ATP with the presumably more hydrophobic adenine portion of the molecule was more strongly retained. When operated with a mobile phase containing a high proportion of aqueous buffer, the retention order is changed but not simply reversed. Column efficiency was disappointing.

Hydrophilic interaction liquid chromatography (HILIC) has been suggested to offer potential advantages relative to reverse-phase chromatography for mixtures of very polar and ionic substances. The first column that we investigated for HILIC mode was the ThermoHypersil-Keystone BioBasic AX column. This is a bonded phase ion exchange column, with an imine functionality (R=C=N-H). This polar phase allows a water layer to form, which is necessary for the HILIC mode. Of course, for analyte compounds that are anionic, the column continues to show ion exchange behavior. Analyses were conducted using an ammonium acetate mobile phase in which pH was either not controlled or was adjusted to pH 8. Only after we adjusted the buffer to pH 10 did we see retention behaviors that could be acceptable. The separation was better than what we experienced in the reverse-phase mode. Unfortunately silica backbone column packings are well known to be unstable in high pH conditions; since the silica is eventually dissolved, the backbone collapses, and back-pressure increased catastrophically. Further work in this system was discontinued due to this shortcoming.

The use of a "neutral" HILIC column was evaluated to avoid the difficulty of confounding anion exchange and partition behavior, and hopefully exhibit only HILIC mode. The polyhydroxyethyl A stationary phase offered by PolyLC, Inc., has an aspartamide functionality which is claimed to present enough hydrophilic character to immobilize a significant water layer and to not present any anion exchange behavior. In fact, the column is also advertised as suitable for steric exclusion separations. The critical issues for the use of this column then are the proportion of water to acetonitrile and the salt content of the mobile phase that will be necessary to yield acceptable separations. Since a larger-pore-size column will exhibit lower capacity, we conducted some preliminary experiments with a 300A column. Retention of AMP was about a third that for the 100 A pore column, all other conditions identical. The retention of mono-anionic species might well be satisfactory, but di-anionic and presumably tri-anionic substances are retained excessively even at salt concentrations and water/acetonitrile conditions that would lead to poor ES-MS sensitivity.

Ion-pair reverse-phase chromatography was evaluated using two reverse-phase columns and triethylammonium acetate as the ion pair reagent. Peak widths were unacceptably large for serious consideration.

In light of difficulties in identifying a chromatographic system that would give adequate separation under conditions amenable to ES-MS, we chose to approach the larger metabolomics question from a different perspective. All previous work was in pursuit of an analytical system that would reliably identify and quantify a considerable proportion of the metabolites in a sample; a task that demands at least "good" separation of the mixture. An alternative approach to the larger question of metabolomics is to simply identify a technique that can reliably detect differences between samples. Direct-infusion ES-MS would seem to be an ideal approach. The capability to scan a wide mass range yielding semi-quantitative intensity vs mass to charge information would provide a unique sample "fingerprint." The potential criticism of this approach is the variation in sensitivity among species in, for instance, the negative ion mode electrospray mode. This difficulty can actually be turned into an advantage if the sample is examined in both the negative and positive electrospray modes and the negative and positive atmospheric pressure chemical ionization (APCI) modes. Leaf tissue extracts from two samples known to be genotypically distinct poplar clones were injected directly using the four ionization modes identified above. Differences were readily detected by the DI-ES-MS and APCI-MS experiments (Figs. 1 and 2), which lead us to conclude that fingerprinting by mass spectrometry is a potentially fruitful high-throughput avenue to pursue.

Summary and Conclusions

The testing of the different derivatization protocols confirms the need to tailor the protocols to the matrix being analyzed to maximize the amount of metabolite data available in the samples. The expansion of our TMS database gives ORNL a decided edge in being able to identify the metabolic perturbation induced by the introduction of a genetic mutation. Elucidation of many of the currently unidentified metabolites and expansion of that database will continue. Database expansion will serve to further improve our capability to conduct metabolic phenotyping. The completion of this study has brought us closer to being able to conduct such phenotyping at high-throughput rates.

As powerful and rapid as GC-MS is and may become as we migrate our analyses to high-throughput GC time-of-flight MS, further capability development was needed to expand the range of metabolites that can be detected in heterogeneous extracts and expand the spectral (mass) dynamic range of the detectable metabolites. DI-MS protocols were developed as rapid, high-throughput screening tools, where derivatization will not be required and the suite of metabolites profiled is further broadened. A wide array of metabolites is directly injected (infused) into a MS or tandem MS that is operated with multiple ionization modes, including positive and negative mode ES and positive and negative mode APCI. The analyses target metabolites that are otherwise typically difficult to derivatize or analyze by GC-MS because of large size (>1000 Da), too many reactive functional groups, and low volatility. Preliminary DI-MS analyses contrasting extracts

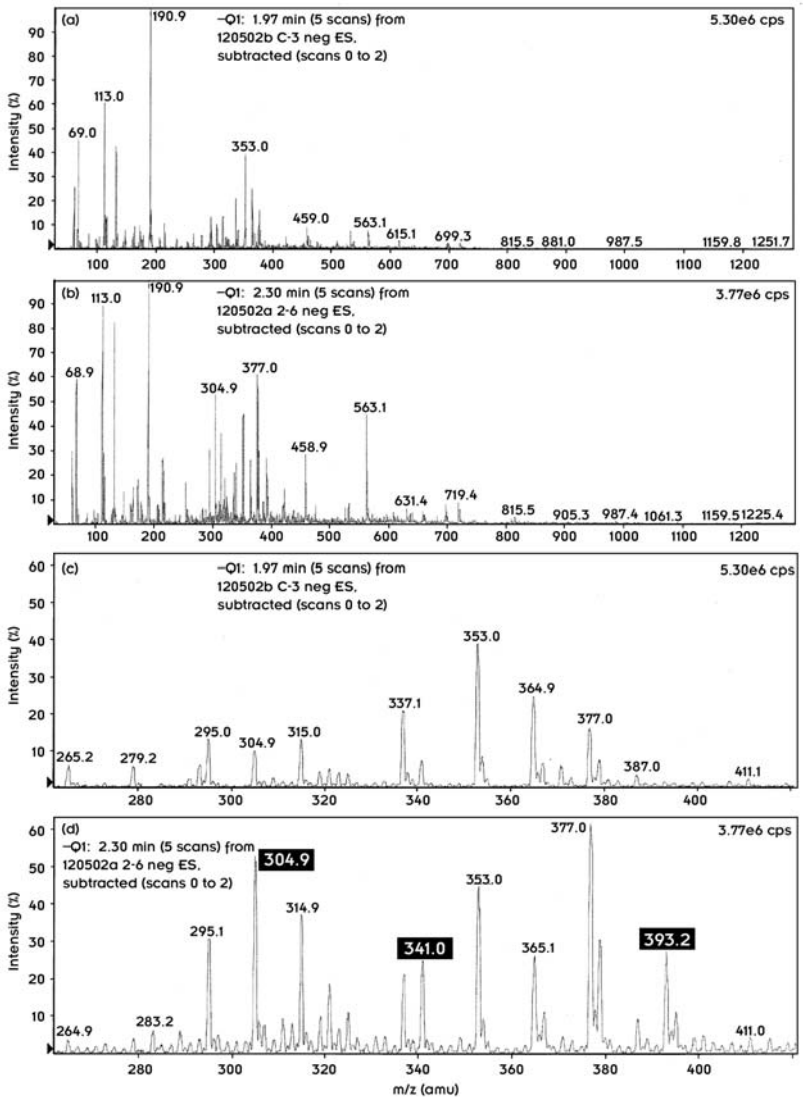


*Fig. 1. DI-MS-ES **negative** ion mode of extracts of (A) Populus tremula x alba nontransgenic (control) versus (B) activation-tagged mutant (GA_2ox). The magnified scale m/z scale (~250–450) is shown in (C) and (D), respectively. Note increased abundance of specific isotopomers in the mutant (D).*

from transgenic versus nontransgenic poplar clones have been very promising, with DI-MS readily extending the spectral dynamic range of the analyses from the 1000 m/z Da cutoff in GC-MS analyses up to 3000 Da. To extend this work to very high throughput, we can now apply what has been learned in this study to direct nanoelectrospray infusion analysis, making use of a NanoMate 100 that incorporates ES Chip technology (Advion BioSciences, Ithaca, NY). ORNL is currently one of only a very few labs in the world to have this most cutting-edge sample-handing, nanoelectrospray ion source system. Finally, the studies convinced us that separation and analysis of charged metabolites, such as adenylates, are not likely to be achieved by current reverse-phase HPLC if detection is by ES-MS. This includes the HILIC separations that are currently receiving considerable attention at several labs. As a result of the separations we observed, our approach is now to pursue chemical electrophoresis to resolve or separate charged metabolites, followed by MS or MS/MS analyses for detection and characterization.

The combined activities place ORNL in a more competitive stance as the funding agencies recognize the need to incorporate metabolomics data in functional genomics research programs in order to characterize how the biochemical networks are interrelated and to elucidate the function of unknown genes.

The most likely offices and programs to fund the follow-on research include DOE's Genomes to Life and OBER. The activities conducted over the course of this study allowed the continued development of a robust TMS-based library of mass spectra for GCMS analyses, and the initial development of deconvolution strategies that are essential for near-term metabolomics studies. The database and GCMS profiling approach were immediately (soon after the initiation of the LDRD project) included as a component of a large, multi-institution functional genomics proposal to a Request-for-Proposals solicited by OBER titled "Terrestrial Carbon Sequestration Using the Poplar Program." Our proposal titled "Genome-Enabled Discovery of Carbon Sequestration Genes in Poplar" (G.A. Tuskan, PI) was successfully funded at $1.7 M/yr for 3 yr, with approximately $1M/yr remaining at ORNL. Additionally, the aforementioned GCMS capabilities and the outcome of the LCMS protocol development were used to develop a large $3M (total funds requested) proposal to the Genomes-to-Life Program to further develop high- throughput techniques for metabolite profiling. Unfortunately, this latter proposal was not successfully funded. We will pursue follow-on funding with both NIH and NSF in this fiscal year.

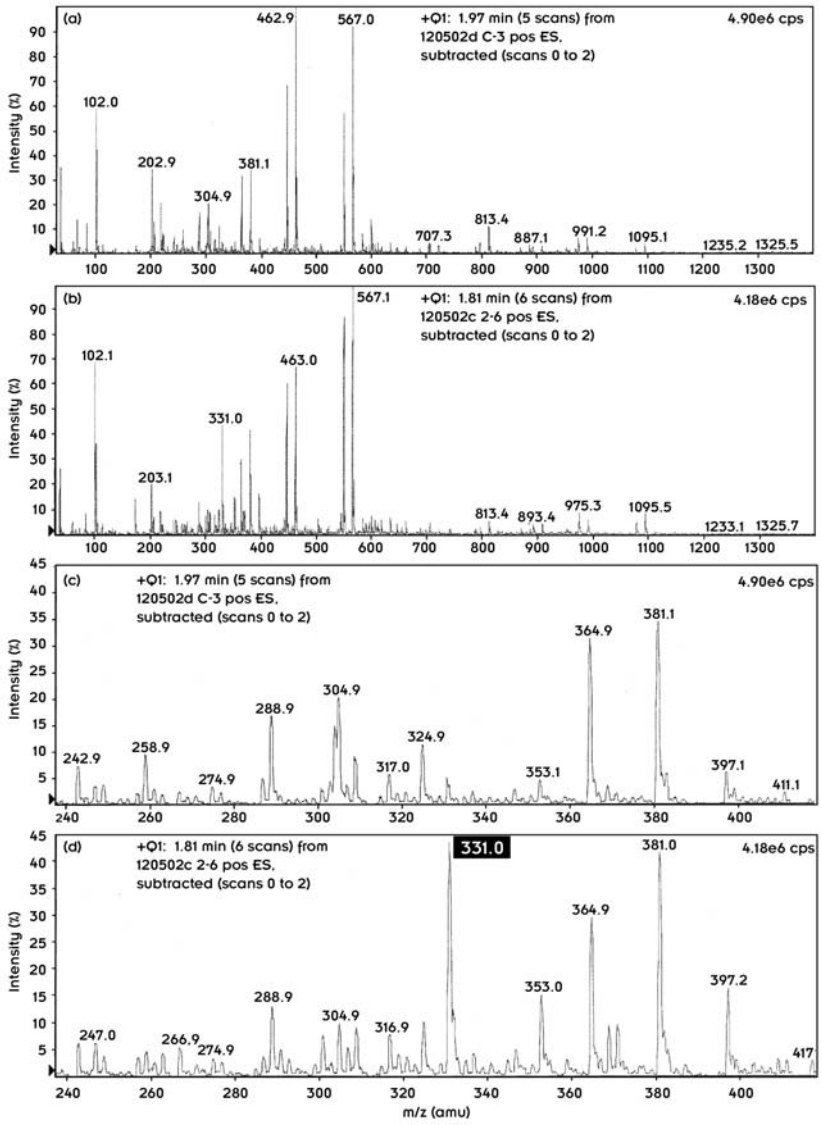


Fig. 2. DI-MS-ES ***positive*** *ion mode of extracts of (A)* Populus tremula *x* alba *nontransgenic (control) versus (B) activation-tagged mutant (GA_2ox). The magnified scale m/z scale (~240–450) is shown in (C) and (D), respectively. Note increased abundance of specific isotopomer 331 m/z in the mutant (D).*

Reference

[1]V. V. Tolstikov and O. Fiehn, "Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry," *Analytical Biochemistry* **301**, 298–307 (2002).

Proof-of-Principle Test for Endophyte-Mediated Effect on the Uptake of Metals by Tall Fescue

A. J. Stewart and M. Z. Martin
Environmental Sciences Division

A preliminary study was conducted to assess the opportunity for modifying a plant-endophyte system with the purpose of enhancing phytoremediation. Leaf material from endophyte-infected (E+) and endophyte-free (E–) tall fescue populations was analyzed by inductively coupled plasma mass spectrometry (ICP-MS) for 11 metals to determine if the presence of a fungal endophyte (*Neotyphodium coenophialum*) increased the ability of the grass to accumulate metals from soil. Of the seven metals with concentrations high enough to be reliably compared, most were higher in the E+ compared with the E– tall fescue. Additionally, the concentrations of Cu, Cr, Ni, and Pb were two to four times greater in the E+ compared to the E– grasses. However, very few of the differences were statistically significant, probably due to the low concentrations of metals in the soils of the test plots, the low number of samples analyzed, and the high analytical detection limits for several of the metals. Although the effect of the fungal endophyte on metal accumulation in tall fescue is uncertain, additional, more rigorous testing to assess its potential role in enhancing phytoremediation is warranted. An alternative analytical technique, laser-induced breakdown spectroscopy (LIBS), was used successfully for the first time to detect metals in grasses. Further research is needed to make this method more quantitative for routine use in the analysis of metals in biological material.

Introduction

The presence of endophytes in plants may enhance their ability to take up and accumulate metals from contaminated soils. Endophytic fungi that live within plant tissues can have large effects on their host, such as increased drought and heat tolerance,[1,2] lower rates of herbivory,[3] and greater tolerance to Al and Zn in soil.[4,5,6] By investigating the metal-accumulating mechanisms and attributes of a plant-endophyte system, it might be possible to eventually enhance the plant's phytoremediation potential. The well-studied model system of a grass (tall fescue, *Festuca arundinacea*) and its endophyte (*N. coenophialum*) was selected to determine if the presence of the fungal endophyte influenced the host's ability to accumulate metals from soil. The proposed study was a proof-of-principle test of this concept, and a necessary step that was required before a credible proposal could be submitted in response to an agency solicitation on phytoremediation.

Technical Approach

Four random samples of tall fescue for metals analysis were collected in December 2002 from each of seven grass seed production plots at the University of Tennessee Knoxville Experiment Station. Two of the plots contained E– forms of the grass and five plots contained E+ forms. The samples were dried, milled, and acid digested. The digestates were centrifuged, neutralized, and diluted (to reduce interference by nitrate) before they were analyzed by ICP-MS. A National Institute of Standards and Technology (NIST) reference material (apple leaf, SRM 1515) was digested and analyzed in the same manner for QA/QC purposes. The instrument was calibrated by analyzing blanks and external standards, both of which were also run during sample analysis to verify the instrument's calibration. Data from the blanks, standards, and NIST reference material were used to select the most appropriate mass for evaluation in cases where more than one mass per element was available. The elements of interest included Mg, Ca, Cr, Fe, Mn, Co, Ni, Cu, Zn, Cd, and Pb.

In addition to ICP-MS, an alternative analytical technique, LIBS,[7] was used to determine if metals of interest, such as Cd, Ni, and Zn, could be resolved in the milled plant material with minimal additional sample processing. To explore this laser application, 400- to 500-mg quantities of milled plant material were pressed into pellets (diameter: 1.3 cm; thickness: 4–7 mm) and analyzed individually with no further preparation. The laser used for LIBS was operated at a pulse of 23 mJ, an excitation wavelength of 532 nm, a delay of 4 μs, and a gate setting of 10 μs. Data were accumulated for 10 laser pulses with a 50-μm spectrometer slit width.

The data obtained by ICP-MS were evaluated using one-way analysis of variance (ANOVA). The plots

contained tall fescues that differed with respect to infection status (E+ or E–) and location (a proxy for soil type). The tests for effects due to plot position provided information on the effect of spatially linked edaphic factors on metal accumulation by the tall fescue-endophyte system.

Results and Accomplishments

Substantial differences were found in the concentrations of metals in E+ and E– tall fescue populations, but most of these were not statistically significant. For example, Cu, Cr, Ni, and Pb were two to four times higher in the tall fescue plots that were infected by the fungal endophyte compared to tall fescue that was endophyte free ($p > 0.05$). In another comparison involving E+ versus E– tall fescue in different seed production plots, E+ tall fescue had higher levels of Ca (18%), Fe (55%), Mn (54%), Ni (39%), Cu (21%), and Zn (13%) than E– fescue. However, in this comparison, only Fe and Mn were statistically significant ($p < 0.05$). Finally, two widely separated plots containing E+ tall fescue on different soil types were tested for spatial effect. In this case, the differences for Ca and Mg were large and significant ($p < 0.02$), but the differences for Mn, Cu, Ni, and Zn were not statistically significant ($p > 0.05$). No investigators have yet measured metals in tall fescue for the purpose of determining how the endophyte affects the host plant's ability to accumulate metals. Indeed, many investigators either do not know that the tall fescue endophyte exists or are unaware that it has a significant impact on the physiology of the host.[8] Although limited in scope, this preliminary study will contribute to the knowledge of endophyte-plant systems and to the importance to plant ecology of fungal-plant mutualisms.

The ICP-MS method was limited in its detection of low concentrations of metals (e.g., Cd) due to the need to digest samples, which diluted them. With LIBS, a sample can be analyzed directly with no dilution. Thus, the study provided an opportunity to assess LIBS as an analytical method especially suited for phytoremediation studies. With minimal sample preparation, LIBS unequivocally detected the presence of Cd, Mg, and Zn (Fig. 1). Other metals detected by LIBS have emission peaks that are outside of the 252- to 286-nm window shown in Fig. 1. Additional research is required to make this method of analysis quantitative enough for routine analysis of metals in plant tissues. However, such efforts should advance phytoremediation significantly, because metals could be analyzed at low cost and at near real time in the field. Such a technique would allow investigators to optimize phytoextraction by providing (1) better accounting of spatial variation in the amount of metal extracted, which is needed for closure on mass-balance calculations; and (2) better monitoring of the time-course of metal accumulation in aboveground plant tissues during the growing season.

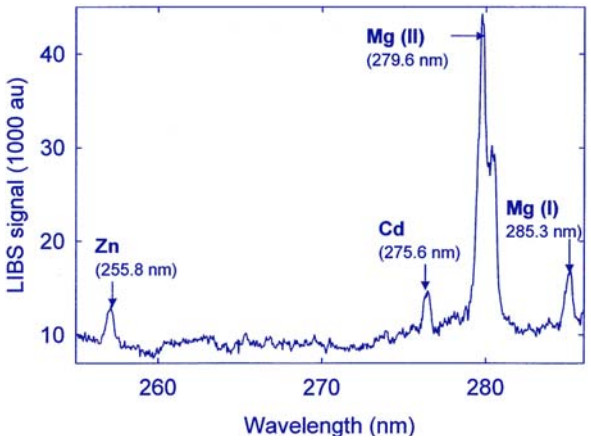


Fig. 1. Laser-induced breakdown spectroscopy (LIBS) emission spectrum for dried, milled tall fescue leaf samples. Peak-height data (Y-axis) are arbitrary units. Gate and delay times, excitation wavelength, and excitation energy can be varied to optimize the detection of metals at other emission wavelengths.

Summary and Conclusions

A preliminary study was conducted to assess the opportunity for modifying a plant-endophyte system with the purpose of enhancing phytoremediation. The study consisted of sampling and analysis to determine if an endophyte-infected grass (tall fescue) had different concentrations of metals than noninfected grass. Leaf material was collected from infected and noninfected tall fescue production plots at the University of Tennessee Knoxville Experiment Station and analyzed for 11 metals using ICP-MS and LIBS, a new experimental technique that had not previously been used to analyze metals in plant material. The tests for effects of infection status on the accumulation of metals in tall fescue leaf tissue were inconclusive. Sampling and analysis of selected metals in infected vs noninfected grasses did not provide strong evidence to support or refute the hypothesis that the presence of *N. coenophialum* affects the metal content of aboveground tall fescue tissues. Although the concentrations of some metals (Cu, Cr, Ni, and Pb) were two to four times higher in the E+ grass, the differences between the E+ and E– plots were not statistically significant. The LIBS technique unequivocally detected the presence of Cd, Zn, and Mg in tall fescue samples, but additional research is needed to make the method more quantitative for routine use in analyzing metals in plant tissues.

Although the results of this preliminary study provided only weak evidence of differences in metal accumulation between tall fescue populations with and

without a fungal endophyte, they emphasize the need for a more definitive and rigorous evaluation of endophyte-mediated effects on metal uptake in grasses. Many federal agencies, including DOE support research on phytoremediation. The use of grasses and other plants to sequester contaminants remains an attractive remediation alternative. It is a relatively low-cost, low-risk cleanup approach with an excellent record of acceptance by the public. By studying metal-accumulating mechanisms and attributes of a plant-endophyte system, it may be possible to develop a system that can specialize in the extraction of certain metals from soil. Even if that goal is not achieved, investigators who work with tall fescue must be alert to the possibility of endophyte-mediated effects, given the increasing use of tall fescue in phytoremediation studies and in the revegetation of disturbed soils.[8]

Efforts to obtain follow-on funding will be focused on federal solicitations for proposals related to phytoremediation. A draft manuscript describing the results of this preliminary study has been completed and will be submitted to the *Journal of Environmental Quality*. Also, this preliminary study has demonstrated the potential value of LIBS, and the data collected in this study will support future efforts to obtain funding to continue the development and application of this innovative technology.

References

[1]D. Malinowski and D. Belesky, "Adaptations of endophyte-infected cool-season grasses to environmental stresses: Mechanisms of drought and mineral stress tolerance," *Crop Sci.* 40, 923–940 (2000).

[2]R. Redman, K. Sheehan, R. Stout, R. Rodriquez, and J. Henson, "Thermotolerance generated by plant/fungal symbiosis," *Science* **298**, 1581 (2002).

[3]K. Clay, S. Marks, and G. Cheplick, "Effects of insect herbivory and fungal endophyte infection on competitive interactions among grasses," *Ecology* **74**, 1767–1777 (1993).

[4]D. Malinowski and D. Belesky, "Tall fescue aluminum tolerance is affected by *Neotyphodium coenophialum* endophyte," *J. Plant Nutrition* **22**, 1335–1349 (1999).

[5]F. Monnet, N. Vaillant, A. Hitmi, A. Coudret, and H. Sallanon, "Endophyte *Neotyphodium lolii* induced tolerance to Zn stress in *Lolium perenne*," *Physiol. Plantarum* **113**, 557–563 (2001).

[6]D. Zaurov, S. Bonos, J. Murphy, M. Richardson, and F. Belanger, "Endophyte infection can contribute to aluminum tolerance in fine fescues," *Crop Sci.* **41**, 1981–1984 (2001).

[7]M. Martin, M. Cheng, and R. Martin, "Aerosol measurement by laser-induced plasma technique: A review," *Aerosol Sci. Technol.* **31**, 409–421 (1999).

[8]S. Humphries, K. Gwinn, and A. Stewart, "Effects of endophyte status of tall fescue tissues on the earthworm (*Eisenia fetida*)," *Environ. Toxicol. Chem.* **20**, 1346–1350 (2001).

Permeable Environmental Leaching Capsules for Nondestructive In Situ Evaluation of Contaminant Immobilization Techniques in Soil

B. P. Spalding and S. C. Brooks
Environmental Sciences Division

A nondestructive technique using permeable environmental leaching capsules (PELCAPs), which encapsulate radioisotope-spiked soil within a water-permeable polyacrylamide matrix, is being tested as an in situ field method to measure immobilization of contaminants. PELCAPs have been retrieved and replaced in ground and surface waters many times during which they have been assayed nondestructively using gamma spectroscopy. As a proof of principle, soils contained in PELCAPs were labeled with two radioisotopes (^{85}Sr and ^{134}Cs) and have been leached in laboratory tests as well as deployed in ground and stream waters at Oak Ridge field sites. In situ retention of both isotopes within PELCAPs, containing thermally treated soil with high degrees of radioisotope immobilization, has been demonstrated up to three months in field tests.

The objective of this investigation is to develop and demonstrate an inexpensive, direct, and effective in situ technique to monitor contaminant immobilization nondestructively in the field using radioisotope-spiked soil contained within PELCAPs. The PELCAP technical approach has potential advantages for the assessment of induced and/or natural contaminant availability within soils, including the following:

- Nondestructive measurement of the amount of immobilized contaminant in a soil over time, thereby avoiding the necessity for repeated, costly, and destructive soil sampling;
- Direct comparison of several immobilization treatments, including a no-treatment control, under identical field conditions within the same well;
- Internal calibration of PELCAP leaching relative to specific reference tracers (^{85}Sr and ^{134}Cs) which have well-characterized environmental leaching behavior;
- Accelerated remedial alternatives evaluations by deploying PELCAPs in uncontaminated regions of a site to avoid possible masking by continued contaminant flux;
- Correlation of PELCAP behavior with widely accepted laboratory selective extraction and isotopic dilution protocols.

This PELCAP soil contaminant availability research should establish five research findings: (1) develop PELCAPs from which ^{85}Sr and ^{134}Cs freely diffuse with coefficients similar to those found in water; (2) develop PELCAPs which retain soil particles greater than 0.1 μm; (3) show that soil available forms of radionuclides in PELCAPs leach to a similar degree as unencapsulated soil in laboratory sequential extractions; (4) show that soil immobilized forms of radionuclides in PELCAPs behave similarly when leached in situ in field groundwater as when leached by laboratory methods; and (5) establish the methodology for PELCAP contaminant availability analysis, PELCAP durability for extended field deployment, and precision of PELCAP performance under field conditions.

A standard formulation for preparing PELCAPs in a polyacrylamide matrix has been established and tested using laboratory sequential extractions with ^{134}Cs and ^{85}Sr spiked soils. The sequential extraction behaviors of these radioisotopes from encapsulated soils and from the same but nonencapsulated soils were quite similar. Two suites of 54 PELCAPs each (spiked with isotopes, with and without soil, and with ambient or thermally stabilized soils) have been deployed in both stream water and groundwater. The natural in situ leaching of the isotopes from PELCAPs by natural waters in the field has taken several months to reach equilibrium but has behaved similarly to the laboratory-extracted PELCAPs. Stability of the immobilized radioisotopes in these field-deployed PELCAPs will continue to be monitored for up to six months, and this performance is a critical parameter to support their planned environmental application.

The PELCAP soil contaminant availability methodology is applicable to DOE's mission to predict the impacts of long-term stewardship of its waste disposal sites and practices. By providing a low-cost methodology to assess long-term performance of contaminated soil, these research findings should enable DOE managers to make judicious choices in selecting remedial actions.

Construction of a Gene-Prediction Algorithm in *Populus*: Adding a New Dimension to Complex Biology

L. E. Gunter[1] and F. W. Larimer[2]
[1]*Environmental Sciences Division*
[2]*Life Sciences Division*

The goal of this project is to create a unique annotation resource for the *Populus* genome sequencing effort. We will develop *Populus*-specific models for the functional analysis of genes using novel information contained in a library of full-length cDNA clones derived from various *Populus* tissues. A collection of accurately characterized, full-length cDNA sequences will allow us to develop a proof of principle for automated gene-prediction algorithms that are critical for recognition of *Populus* genes. The design and implementation of a precise gene-prediction tool for the annotation of the *Populus* genome will open new avenues of functional genomics at ORNL.

The Joint Genone Institute has recently announced its completion of the complete sequence for *Populus*, the first tree and only the third plant to have its genome sequenced. The goal of this proposal is to fully exploit the opportunity afforded by the whole-genome sequencing effort by designing and testing a novel annotation tool that will establish ORNL as a leader in *Populus* functional genomics. The development of this type of informatics tool will furnish us with a resource that will enhance our ability to address future DOE missions in functional genomics of *Populus* and increase the scientific and technical vitality and visibility of ORNL through securing a national and international role in *Populus* genomics. This knowledge could have profound implications for carbon management issues related to long-lived perennials, conversion of biomass to fuels and products, and the deployment of plant-based systems for phytoremediation. The objectives of this research are to (1) provide sequence input of unique *Populus* structural genes to gene-prediction models and (2) design a customized version of GrailEXP to provide unique and consistent exon finding and gene model algorithms for *Populus*.

A large collection (>1,000) of full-length cDNAs from libraries derived from a variety of tissues of *Populus trichocarpa* "Nisqually-1" will be sequenced using dye terminator chemistry, then assembled into full-length reads. Raw sequence data files will be analyzed to assign quality scores to sequenced fragments prior to assembly. Only nucleotides assigned a quality score of Q40 with a 2X or greater coverage will be used to assemble full-length cDNA sequences that will be used to construct the gene-prediction algorithm.

Populus cDNA and EST sequences will be aligned with genomic draft sequence. From these alignments, we will extract a description of exon size and number distribution, coding exon hexamer distribution, intron size distribution, intron hexamers, and splice site structure. These results will be used to train the Perceval (Protein-coding Exon, Repetitive, and CG-Island EVALuator) neural net for exon finding and the Galahad Gene Message Alignment routine for splice junction optimization. If strictly patterned intron structures are recognized, we will configure the Gawain (Gene Assemblies with Alignment Information) gene model assembler accordingly. After the construction of a set of high-quality gene models, we will extend the collection of *Populus* signal motifs using a recently developed motif finding tool, MIRA. This motif finding routine is a background rareness–based iterative multiple-sequence alignment algorithm for regulatory element detection. In addition to motif finding in raw sequence, this tool can utilize clustering information from expression or proteomics data to enhance its search specificity.

To date over 800 full-length poplar cDNAs have been identified from the libraries. Of these, 320 have been completely sequenced and assembled, and another 352 are awaiting internal sequencing for closure. Genome Canada will be contributing another ~1000 full length cDNA sequences to this effort. These cDNAs will be aligned with the poplar genome sequence, which has recently been completed, and for which a draft assembly will be available by late November or December 2003. In addition to the cDNAs, Stefan Jansson from Sweden has submitted 30,000 contributed ESTs to the assembly. This will enhance the gene prediction models and allow more accurate gene calling.

Preliminary hexamer frequency plots have been derived from the first 320 fully assembled cDNAs. Of

these, 122 have high-identity alignments with their Arabidopsis homologs. The sequences are sufficiently similar to use the Arabidopsis gene models to simulate the splicing patterns of poplar. From these analyses, we have generated preliminary profiles of coding exons, coding hexamers, and splice site structure. Exploratory studies have been carried out using an Arabidopsis-trained gene finder against ~1mbp of preliminary poplar assemblies. These results indicate that the poplar intron structure and distribution is sufficiently different from Arabidopsis to require retraining. Alignment of the cDNA library against the full poplar assembly will be necessary to develop the intron profiles for a full training set.

The development of this novel annotation model for poplar will act as a bridge to allow current research to move beyond the sequence and focus on a comprehensive understanding of biological and ecological systems. Our primary purpose in this approach is to generate a capability for engaging in the types of scientific research that DOE considers significant, that is genome-enabled carbon sequestration, phytoremediation, ecosystem science, etc., for which solid annotation of the poplar genome is a prerequisite.

Sounds of Rapids as an Attractant for Migratory Fish

C. C. Coutant,[1] M. S. Bevelhimer,[1] and G. R. Wetherington[2]
[1]*Environmental Sciences Division*
[2]*Engineering Science and Technology Division*

Downstream-migrating salmon and other migratory fish have difficulty finding their way past dams. This project is exploring whether the complex underwater sound of a natural river's rapids serves as an attraction for downstream orientation. If true, artificially generated sound could be used to guide migrants to fish bypasses or through reservoirs where migratory cues are lost. The objectives are to (1) obtain recordings and characterizations of underwater sounds using field measurements and sound-characterization software, and (2) identify the capability of such complex sounds (played back recordings) to orient fish in quiet water in the laboratory and eventually in the field.

The project is aimed at developing effective fish guidance systems that would divert migrating fish away from hydropower turbines (where many are killed) without spilling large amounts of water for fish passage. We are exploring whether fish could be guided by simulating a natural behavior cue. Fish would be guided to alternative passage routes around the dam that successfully pass more fish with less water. Spilling water for salmon passage on the Columbia River, for example, reduces water available for generation and costs the DOE/Bonneville Power Administration over $1 million per day in lost generation. The DOE Hydropower Program's goal of a 10% increase in hydropower generation in the United States could be met by simply converting spill in the Federal Columbia River Power System to generation.

We began this project by setting up three of the classical "U-tank" behavioral-choice experimental tanks in the ORNL Aquatic Ecology Laboratory. We were able to modify existing artificial stream channels, which resulted in considerable cost savings. The tanks consist of a long U, with approximately 10-m-long, 0.3-m-wide, and 0.3-m-deep arms. A speaker connected to a Compact Disk player is placed in one of the two arms (or neither, for a control). The entire U-tank area is enclosed by curtains to avoid visual interference. Experimental fish are introduced at the base of the U, where the arms join. Fish position is recorded with a video multicamera array that provides coverage of the whole U. Video frames are recorded at set intervals (e.g., every 2 seconds) for the duration of an experimental trial. We have explored the use of the experimental equipment (using local stream fish species) but have not yet conducted formal trials.

We have made tentative arrangements for several potential field test sites at hydropower dams, pending completion of successful laboratory tests to be conducted during FY 2004.

This project, if successful, will benefit DOE and other federal agencies [e.g., Corps of Engineers, Bureau of Reclamation, NOAA Fisheries, Fish and Wildlife Service)] concerned with maximizing use of renewable hydropower with minimal detrimental effects on fisheries resources.

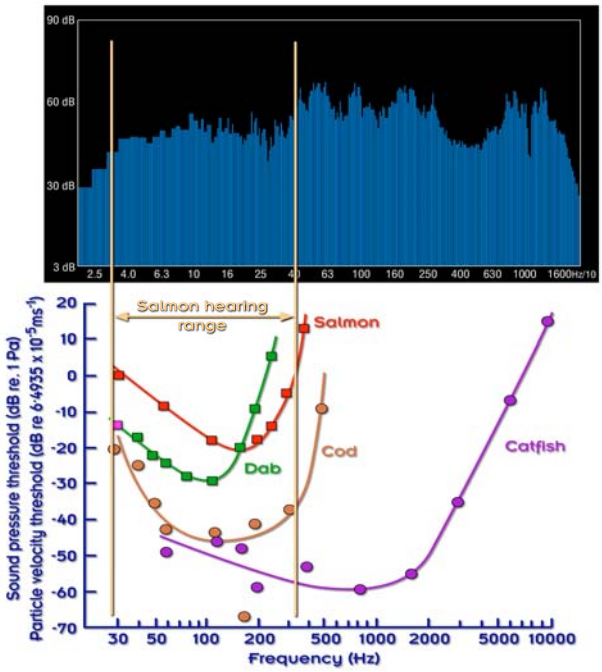


Fig. 1. Underwater sounds of rapids (audiogram in upper panel) occur at frequencies to which many fish species are sensitive (lower panel; the lower the threshold, the greater the sensitivity). The ORNL research will test whether migratory fish are attracted to such sounds, which might be used for guiding fish past obstacles such as dams.

NUCLEAR SCIENCE AND TECHNOLOGY

Director's R&D Fund

Innovative Safety Technologies for Generation IV Reactor Designs

D. T. Ingersoll,[1] J. C. Gehin,[1] M. D. Muhlheim,[1] R. E. Pevey,[2] B. R. Upadhyaya,[2]
R. T. Wood,[1] and A. B. Wollaber[2]
[1]*Nuclear Science and Technology Division*
[2]*University of Tennessee*

The objective of this project was to develop new analysis tools and capabilities that will enable ORNL to be a significant contributor in emerging programs to develop advanced fission reactors and, specifically, to extend our ability to design advanced reactor systems that have enhanced safety features while simultaneously improving system reliability and economics. The project focused on three related but distinct aspects of nuclear plant safety: probabilistic risk assessment (PRA), reactivity control, and component surveillance and diagnostics. In all three cases, the new capabilities were demonstrated for a specific advanced reactor concept, the International Reactor Innovative and Secure (IRIS), led by Westinghouse Electric Company. New capabilities generated from the project include the development of a graphical-based software tool for performing PRA studies on concept-level reactor designs, the assembling of a state-of-the-art reactor physics analysis system for studying reactivity control in light water reactors, and the development of an architectural framework for advanced diagnostics and monitoring systems. Demonstrations of these tools for IRIS include the 100-fold decrease in estimated core damage frequency by proper selection of safety system components, the design of an IRIS core that uses erbium as a burnable poison to greatly reduce the level of soluble boron for reactivity control, and an integrated demonstration of advanced software simulation models and prototypic wireless communication hardware to detect and isolate faults in a helical coil steam generator.

Introduction

A resurgence of the nuclear energy industry in the U.S. and its growth worldwide is being driven by three major factors: an increasing demand for electricity, an increasing concern regarding the global warming effects of fossil fuels, and a new urgency to replace our dependence on foreign supplies of energy with domestic resources. The ability of the nuclear power industry to meet these expectations will depend substantially on our ability to build new plants that have increased safety and reliability relative to existing plants, while also remaining economically competitive with alternative energy sources. Increased nuclear plant safety places a greater demand on creative design solutions and advanced technology. The objective of this project is to develop new analysis tools and capabilities that will enable ORNL to design advanced reactor systems that have enhanced safety features and improved system reliability and economics. This was accomplished by addressing three aspects of plant safety: (1) the novel application of probabilistic risk assessment (PRA) methodologies to concept-level trade-off studies, (2) the development of advanced computational methodologies for the study of innovative alternatives to reactor reactivity control, and (3) the development of an innovative approach for in situ surveillance and diagnostics to ensure plant integrity during extended periods of operation. In all cases, new capabilities were demonstrated using a specific advanced reactor concept: the International Reactor Innovative and Secure (IRIS) concept,[1] which is a leading advanced reactor concept that has a "safety by design" philosophy and special design characteristics that make it especially attractive for deployment in developing countries.

Technical Approach and Results

The following sections describe the technical approach for each of the three major tasks in this project. Also described are results from the application and demonstration of the new tools and methodologies to the IRIS design.

Task 1: Probabilistic Risk Assessment

Current PRA analyses are performed primarily to demonstrate the safety of the plant and are typically performed when the design is approximately 50% complete and also 90–100% complete. It is difficult and costly at these points in the design process to make changes to the design, and changes are done only if significant vulnerabilities in the design are demonstrated by the PRA. A much better approach is to perform PRAs throughout

the design process, especially in the formative stages when individual components and systems are being selected. This allows for safety and reliability to be incorporated into the design from the very outset. Task 1 of this project was to develop an easy-to-use software tool that can facilitate the selection of nuclear plant components based on their impact on the predicted probability of system failure, core damage, radiological release, etc.

The overall architecture of a new concept–level PRA tool was defined, and a working version of the *VisualBasic* driver/overlay program was constructed. Considerable planning had to be invested to determine how the event trees and fault trees are combined in order to allow the user to not only evaluate different system design options but also to select alternatives to the underlying probabilistic models and component data. Extensive use was made of data from the PRA of Westinghouse's AP-600 certified reactor design.[2] The one-button architecture was designed to evaluate combinations of design and event options through sets of fault tree modules in a dynamic PRA. Changing the design is easily performed by picking a design alternative from a drop-down menu. The new program, designated RBOT (Reliability-Based Optimization Tool), automatically inserts the correct fault tree modules(s) into the PRA model in *FaultTree+*,[3] relinks the correct support systems, and recalculates the parameters of interest.

Data structures were defined that permit the accumulation of design alternatives and resulting failure probabilities, which facilitates the ability of the code to perform reliability-based design optimization.

To demonstrate the power of the new tool, RBOT was used to assess the sensitivity of the predicted IRIS core damage frequency to design choices for the valves in the automatic depressurization system (ADS) and the emergency heat removal system (EHRS). Choices included using either air-operated valves (AOV) or motor-operated valves (MOV), whether to configure the system with the valves initially open or closed, the number of valve trains, and the capacity of the relief valves. Eleven component choices in the ADS and EHRS generate a total of 160 unique design choices. Using RBOT, we were able to easily evaluate the impact of the 11 component choices, which resulted in a selection of preferred components that was not intuitive. Specifically, although AOVs have higher reliability than MOVs, when used in an integrated system, they resulted in a higher core damage frequency (CDF). This is because the common cause failure for air-operated systems is higher than for electrically driven systems. Figure 1 compares the failure probability for AOVs and MOVs when treated individually or within an integrated system. It was determined that for the EHRS, using two valve trains with one AOV and one MOV provided a 42% higher reliability and a 23% reduction in CDF. Using this

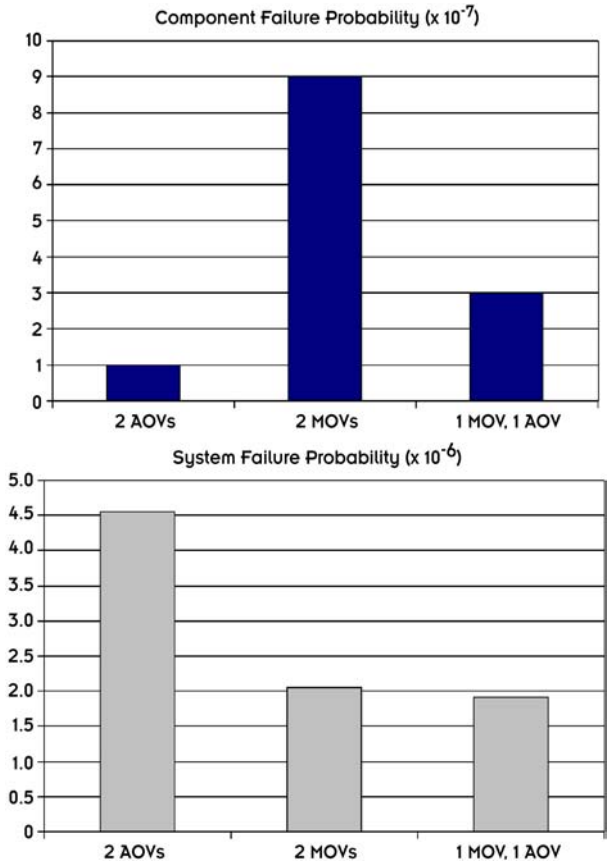


Fig. 1. Failure probabilities for combinations of air-operated valves (AOV) and motor-operated valves (MOV) taken as isolated components or integrated into the emergency heat removal system.

input and other results from our application of RBOT, the IRIS PRA specialists were able to reduce preliminary estimates of the CDF by nearly two orders of magnitude from 1×10^{-6} to 2×10^{-8} incidences per reactor-year.

Task 2: Reactivity Control Options

For the analysis of reactivity control options, a state-of-the-art core analysis methodology was assembled and validated against IRIS-specific benchmark problems. For pin and assembly level analysis, the HELIOS code[4] was selected, which is a highly configurable collision probability code. The NESTLE nodal diffusion code[5] was used for basic core analysis and the study of reactor transients. A code linking the HELIOS lattice code with the NESTLE code was developed for this project and tested with several single-assembly benchmarks. An evaluation was made of different versions of the FORMOSA core loading optimization code[6] to determine which version was the best suited for IRIS analysis. Standard versions of FORMOSA exist for pressurized water reactors (PWR) and for boiling water reactors (BWR), but the innovative IRIS design incorporates some

features of both. The FORMOSA-P code was selected and adapted for this project and used to perform general core design analysis, minimization of power peaking, maximization of core lifetime, and analysis of two-batch core designs. Four IRIS-specific benchmarks were analyzed, and results compared with results from other IRIS team members: two pin cell benchmarks, one assembly benchmark, and a whole core benchmark. Benchmark results showed good agreement with the consensus results.

Using the newly assembled analysis package, we studied the feasibility of using distributed burnable poisons (BP) in the fuel pins as an alternative to relying on soluble boron in the coolant as a means of compensating for the excess reactivity that must be included in the fuel. Eliminating the need for soluble boron not only improves the safety of the core by eliminating the potential for a positive void reactivity coefficient and eliminates the potential for a boron dilution accident but also significantly improves the economics of the plant by eliminating boron chemistry systems. Several BPs in various homogeneous and heterogeneous arrangements were considered, including gadolinium (Gd), erbium (Er), and zirconium-diboride (ZrB_2). Preliminary design studies were performed for a standard 17 × 17 fuel assembly by varying the Er and Gd content and pin distribution to obtain an "ideal" reactivity profile, that is, an infinite multiplication factor (k_{inf}) of 1.0 throughout the fuel burn-up cycle. Figure 2 shows the burnup-dependent profiles for several distributions of Er in an IRIS assembly. Although an acceptable profile could be achieved, it required that the BP be loaded into 32 pins at an unacceptably high weight percent (>20%). The result of this preliminary analysis indicates that Er is the preferred BP since it can be distributed more evenly throughout the core and results in lower power peaking factors. Currently, the IRIS design retains soluble boron due to other considerations; however, the use of Er as a burnable poison permits a reduced boron concentration in the coolant, which improves the safety performance of the core.

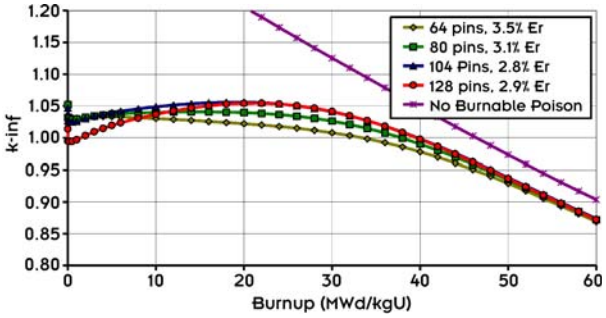


Fig. 2. Infinite multiplication factor for a 17 × 17 pin fuel assembly with different distributions of erbium burnable poison.

Task 3: In Situ Monitoring and Diagnostics

The third major task involved three related studies: (1) developing an architectural framework for an advanced diagnostics and monitoring system, (2) exploring in situ diagnostic techniques capable of automatically checking the integrity of critical sensors and reactor components, and (3) performing an integrated demonstration of a control system with fault detection. A conceptual architectural framework for control and communications was developed and adapted to IRIS. The supervisory control structure is hierarchical with a recursive nature among the vertical levels. Each node in the hierarchy (except for the terminal nodes at the base) is a separate supervisory module. Each module makes decisions appropriate for its level in the hierarchy and passes the decision and necessary supporting data to the modules above. The hierarchical structure is depicted in Fig. 3. Plant-wide use of secure, wireless data communication and the use of "open systems" (i.e., nonproprietary communication standards) were emphasized and recommended. The use of wireless systems where feasible will significantly reduce both capital and maintenance costs.

Steady-state and dynamic models of a helical coil steam generator (HCSG)[7] were developed to generate simulated "measurement" data for both normal and off-normal conditions. Using subspace identification techniques, a ninth-order linear state space model of the HCSG was developed for the dynamic simulation of full power operation. The model was able to predict the steam generator pressure, the cold leg temperature, the steam outlet temperature, the saturated boiling length, and the sub-cooled length with acceptable accuracy. A robust dynamic parity space approach was then developed to design residual generators for fault detection and isolation (FDI). Using the residuals from multiple FDI variables, the methodology was able to reliably detect and isolate

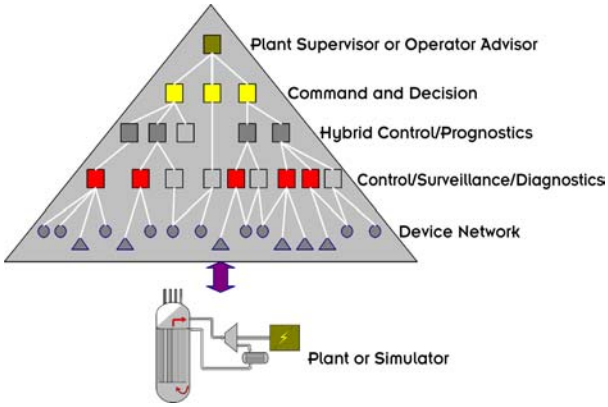


Fig. 3. Schematic of a hierarchical control system with progressive layers of decision-making functions.

the simulated fault of a steam temperature sensor during steady state or transient operation of the HCSG.

As a final demonstration of the new methodology, an experiment was performed, with the dynamic model of the HCSG running on one computer connected to a wireless transmitter and the FDI module running on a second computer connected to a wireless receiver. Temporary faults in the steam temperature sensor were injected into the simulation, and in all cases, the FDI module was able to identify the fault within 10 sec and also recognize the return to normal conditions.

Summary and Conclusions

The project was highly successful both in terms of technical accomplishment and in attracting interest in follow-on work for ORNL. With many universities, laboratories, and vendors now exploring a wide range of advanced reactor concepts, the availability of the concept-level PRA tool developed by this project will be highly valuable for producing safe, reliable, and cost-effective plant designs. In addition to positively impacting the design of the IRIS safety systems, we have received interest from the U.S. Nuclear Regulatory Commission and from Westinghouse for potential sponsorship of its further development. The assembling of a powerful reactor core analysis package was directly beneficial to the evolving design of the IRIS core and has positioned ORNL well to participate in the design of other advanced reactor systems. Finally, the development and demonstration of new in situ plant monitoring and diagnostics methods will help to facilitate plant designs with longer fuel cycle and maintenance cycle times without compromising plant safety and has positioned ORNL as the lead organization within the IRIS consortium for plant control. Additionally, our active participation in the IRIS team has led to DOE-funded collaborations with other consortium members.

References

[1]M. D. Carelli, "IRIS: A Global Approach To Nuclear Power Renaissance," *Nuclear News*, 32–42 (September 2003).

[2]*Simplified Passive Advanced Light Water Reactor Plant Program, AP600 Probabilistic Risk Assessment,* Rev. 7, Westinghouse Electric Corporation, June 1996.

[3]Isograph Reliability Software, *FaultTree+ V10*, 1986–2003.

[4]R. J. J. Stammler and A. A. Ferri, "HELIOS: Angularly Dependent Collision Probabilities," *Nuc. Sci. and Eng.* **122**, 16–31 (1992).

[5]P. J. Turinsky, et al., "Code Abstract—NESTLE: A Few-Group Neutron Diffusion Equation Solver Utilizing the Nodal Expansion Method For Eigenvalue, Adjoint, and Fixed-Source Steady State and Transient Problems," *Nucl. Sci. Eng.* **120** (1995).

[6]D. J. Kropaczek and P. J. Turnisky, "In-Core Nuclear Fuel Management Optimization for PWRs Using Simulated Annealing," *Nucl. Tech.* **95** (1991).

[7]L. Cinotti et al., "Team Generator of the International Reactor Innovative and Secure," *in Proc. 10th Int. Conf. on Nuclear Engineering (ICONE-10)*, Arlington, Virginia, April 14–18, 2002.

Advanced High-Temperature Test Loop for Materials Compatibility in Advanced High-Temperature Reactors

D. F. Williams,[1] D. F. Wilson,[2] G. D. Del Cul,[1] L. M. Toth,[3] J. Caja,[3] and J. P. Renier[1]
[1]*Nuclear Science and Technology Division*
[2]*Metals and Ceramics Division*
[3]*Consultant, Metals and Ceramics Division*

A primary hurdle for use of molten salts in advanced high-temperature reactors (~1000°C) is materials performance. To this end, materials/molten-salt compatibility was evaluated in a thermal-convection loop and redox potentials of buffers to further reduce corrosion were determined.

The neutronics results show that very attractive salt and buffer systems exist, which have acceptable responses to the most common nuclear stability perturbations. The electrochemical results indicated that there are suitable redox buffers (Yb, V) that can be used to minimize corrosion at very high temperatures. Further, a stable reference electrode has been identified. In the presence of very low impurities in this Group IA-halide salt (FLiNaK), corrosion of Hastelloy-N was minor. However, molybdenum fluoride, which has an extremely low solubility in this salt, was transported around the loop. A mechanism based on the change in coordination with temperature is proposed.

Introduction

The desire to increase the operating temperatures of Advanced high-temperature reactors (AHTRs) to achieve greater efficiencies gives rise to the need for materials systems that possess good compatibility with external air and the internal environment of the reactor and/or heat exchangers. Molten fluoride salts, because of their low vapor pressures at high temperatures, are candidates for use in these AHTRs. Materials compatibility in molten fluoride salts is dependent on the salt constituents being more stable than the fluorides of the structural containment material. In addition, mass transfer due to thermal gradients must be evaluated. This project was designed to assess system compatibility at high temperatures under dynamic flow conditions and to conduct preliminary experiments.

Technical Approach

Evaluation of materials for use in AHTRs and/or heat exchangers is a complex activity that involves many factors. In addition to materials mechanical properties and external oxidation resistance, neutronic viability and molten salt/materials compatibility must be considered. The technical approach screens candidate materials and salt systems for application to AHTRs, evaluates their chemical and neutronic viability, and establishes prototypic system compatibility at high temperatures under the flow conditions developed in a thermal-convection loop.

Thermodynamic Screening and Buffer Chemistry

Screening of likely candidates was performed based upon (a) their properties at high temperatures, (b) their likely resistance to corrosion (based upon thermodynamic projections), and (c) their neutronic properties. A requirement for material compatibility is that fluorides of the salt constituents be more stable than fluorides of the containment materials. Thus, alkali metals (e.g., Li) are the preferred cations for salt constituents, and the more noble metals (e.g., Ni) are the preferred container constituents. As a consequence, the experimental part of this project was focused on "FLiNaK," LiF-NaF-KF" (46.5-11.5-42 mol %) as a prototypical coolant salt. It is also known that the thermodynamic potential for corrosion can be minimized by maintaining the salt in a net reducing condition by use of an oxidation-reduction ("redox") buffer. Hence, buffer elements, which were studied by cyclic voltammetry (CV), were added as a fluoride (VF_3, EuF_3, UF_4, YbF_3) at concentrations between 10^{-2} M and 10^{-3} M. Before use, they were purified by flowing F_2 in a closed loop that included a NaF bed to retain HF. These buffers were not employed in the thermal convection loop test.

A three-electrode system was used for the CV determinations. The working and counter electrodes were made of iridium wire 1 mm in diameter welded to 3.175 mm stainless steel rods. The iridium wire was immersed approximately 25 mm into the molten salt. The reference electrode was adapted from a design by Kontoyannis.[1] The electrode consists of a graphite cylinder

(12.7-mm diameter) that was drilled (6.35-mm internal diameter), leaving a flat membrane at the bottom (1 mm thick) and 3.17-mm-thick walls. The top of the graphite electrode was threaded to accommodate a boron nitride cap. The cap had a threaded hole in the center to accommodate the nickel electrode, which consisted of a thin nickel rod (1.59 mm in diameter) connected to a stainless steel rod (3.175 mm in diameter). The cavity inside the graphite electrode was filled with NiF_2 (0.1 M) dissolved in FLiNaK. Further details are presented elsewhere.[2]

Neutronic Screening

A preliminary neutronic screening of the proposed redox buffers was performed. The neutronic models at ORNL for the modular high-temperature gas-cooled reactor were modified for use with a molten salt coolant (in place of the conventional helium coolant). For this screening analysis, adequate neutron cross-section libraries exist to evaluate the spectral range of interest for the redox buffers considered and for all other salt constituents. In addition to evaluation of new buffers, the neutronic performance of an alternate to the well-established FLiBe salt system (2LiF-BeF2) was evaluated. "FLiNaRb" (LiF-NaF-RbF: 45-10-45 mole%, 430°C m.p.) has some very attractive features. It promises very significant neutronic improvements over the most common competitor to FLiBe, "FLiNaK" (LiF-NaF-KF: 46.5-11.5-42 mole%, 454°C m.p.). Because it does not contain Be, it is a relatively nonhazardous chemical.

Materials Evaluation

As outlined earlier, alkali metals (e.g., Li) are some of the preferred cations for fluoride salt constituents, and the more noble metals (e.g., Ni) are the preferred container constituents. Thus, the materials evaluation work focused on FLiNaK as the molten salt and Hastelloy-N as the material of containment. In addition, this alkali halide mixture permits a most sensitive investigation of thermal-gradient, mass-transfer effect, because of the greater solubility of the structural material fluorides in this IA-halide (FLiNaK) than in salts containing IIA-halides (FLiBe). Hastelloy-N is a nickel-based material that was originally designed for use in the Molten Salt Reactor Experiment, and as a result, there is a large data base of materials properties.

Because of the significant role impurities in the salt can play in the corrosion behavior of the material/salt system, extreme care was taken in the manufacturing, purification, and handling of the salts. After preparing a mixture of reagent-grade components to produce the LiF-NaF-KF (46.5-11.5-42 mol %) composition, moisture, hydroxides, and sulfates were removed by sparging with a mix of HF/H_2 at 600°C.

A thermal convection loop was used to evaluate materials compatibility issues of the FLiNaK/Hastelloy-N system. The temperature gradient and flow field in a loop are necessary to determine the importance of a mass-transfer effect. A Hastelloy-N loop, shown in Fig. 1, was used. The construction of this loop, which contained 32 Hastelloy-N specimens in each of the hot and cold legs, is detailed elsewhere.[3] The Hastelloy-N thermal convection loop was connected to salt supplies and equipped with clam-shell heaters on the hot leg and calrods on the cold leg, thermocouples, pressure gauges, a helium gas supply and associated gas purification system, and an array of alarms. Freeze valves at the bottom of the loop maintained the salt in the loop during testing. Analytical ports and instrumentation (Fig. 2) consisted of a double ball valve(DBV) arrangement for removing salt samples and electrodes in the surge tank for CV measurements.

Prior to the start of testing, the Hastelloy N loop was cleaned of residual oxides in two steps. First, the loop was flushed and then filled with argon/4% hydrogen, held at 850°C for 2 h, and then allowed to cool under argon/4% hydrogen environment. Secondly, the loop and one of the salt supplies (a "flush" salt) was heated to 550°C and the salt pushed into the loop by pressurizing the salt supply vessel with helium. The loop was then heated to 650°C

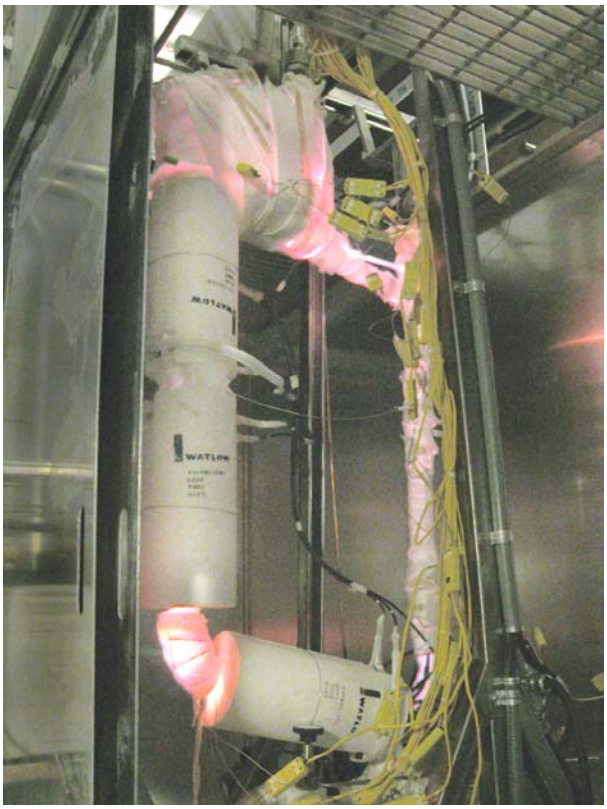


Fig. 1. Loop showing clamshell heaters (left), surge tank (top), and thermocouples. Salt supplies are at the bottom of the loop (not shown).

Fig. 2. Assembly at top of surge tank showing double ball valve for salt sampling and three electrodes for cyclic voltammetry.

and maintained at temperature overnight. The next day, the "flush" salt was allowed to flow back into its salt supply vessel.

Following the above procedure, the loop was then filled with the working salt from the second salt supply vessel. The working temperature of 800°C at the top of the hot leg was established, power to the cold leg calrods turned off, and flow established. The system was operated for 3048 h.

Periodically, electrochemical analyses were performed on the salt using three-electrode, cyclic voltammetry. When not in use, the nickel reference electrode was pulled into the void space above the loop surge tank. Following these measurements, salt samples were collected by submerging a small nickel crucible into the melt through the DBV arrangement (Fig. 2) and withdrawing the sampler into the DBV system under a helium atmosphere. The samples were analyzed for cation content.

Results and Accomplishments

Thermodynamic Screening and Buffer Chemistry

Table 1 shows the preliminary standard electrode potentials for the following redox couples: Yb(III)/Yb(II), U(IV)/U(III),V(III)/V(II),and Eu(III)/Eu(II) in FLiNaK at 610°C. The behaviors of Eu and Yb appeared to be simple and reversible. The behavior of vanadium was more complex and irreversible and needs further analysis. The irreversible behavior of the U(IV)/U(III) system in FLiNaK was briefly evaluated, and the results are similar to those of Clayton.[4] Ytterbium and vanadium are suitable redox buffers for minimizing containment corrosion at high temperatures.

Table 1. Standard electrode potential in FLiNaK at 610°C with Ni/NiF_2 reference electrode

Redox couple	E_o (V)
Yb(III)/Yb(II)	–1.45
U(IV)/U(II)	–1.2
V(III)/V(II)	–0.9
Eu(III)/Eu(II)	–0.64

Neutronic Screening

A computational model for an Advanced High Temperature Reactor was defined, and critical nuclear parameters (voiding coefficients, temperature feedback) for the candidate molten salt coolants were calculated. In Table 2, coolant salts with attractive physical properties that are also nontoxic and do not generate tritium are presented. The buffer systems with low void coefficients are presented in Table 3. Natural vanadium and enriched ytterbium show a zero void coefficient. Calculation of the time-dependent fuel temperatures (and Doppler feedback)

Table 2. Coolant void reactivity coefficient using 1 mol % natural vanadium
(ΔK/K for whole core 100% voiding)

Compounds	Composition	Void %
	Alkali-Fluorides (IA)	
LiF-NaF-RbF	(45-10-45) (0.01 at % Li6)	0.3
LiF-NaF-KF	(11.5-46.5-42) (0.01 at % Li6)	0.5
LiF-RbF	(43-57) (0.01 at % Li6)	0.2
	Alkali + Alkaline Earth Fluorides (IA + IIA)	
LiF-BeF2	(66-34) (0.01 at % Li6)	–0.1
NaF-BeF2	(57-43)	0.0
	Alkali + Zirconium fluorides (IA + Zr)	
NaF-ZrF4	(50-50)	0.0
LiF-NaF-ZrF4	(42-29-29) (0.01 at % Li6)	0.1
RbF-NaF-ZrF4	(40-23-37)	0.2
RbF-NaF-ZrF4	(50-8-42)	0.2

Table 3. Coolant void reactivity coefficient with 1 mol % buffer
(ΔK/K for whole core 100% voiding)

Buffer	Enrichment	Void %
Vanadium	Natural	0.0
Thulium	Natural	5.0
Ytterbium	Natural	0.7
Ytterbium	100 wt % Yb172	0.0
Ytterbium	100 wt % Yb176	0.0
Samarium	Natural	28.5
Samarium	100 wt % Sm144	0.0
Samarium	100 wt % Sm148	0.05
Samarium	100 wt % Sm154	0.1
Samarium	100 wt % Sm153	5.8
Europium	Natural	25.5

after an accidental voiding of the coolant needs to be completed to determine if the desired negative void coefficient can be obtained.

Materials Evaluation

Results of the 3048-h thermal convection loop test are presented in Fig. 3. As in previous work,[5] the Hastelloy-N specimens (surface area of 10.8 cm^2) at higher temperature lost weight and those at lower temperatures gained weight. However, the weight changes were significantly smaller than those previously observed. As shown in Fig. 4, the cross-section of a specimen at 815°C (highest temperature and weight loss) exhibits a very shallow depth of attack and no voids. In contrast, a specimen that was 566°C and gained weight (Fig. 4) exhibits a deposit on the surface. As shown in Fig. 5, there is a deposit on the surface of the maximum-weight-loss specimen, which is significantly thinner than that on the specimen that gained weight. Further examinations revealed that there were deposits on all specimens, with the amount deposited increasing with decreasing

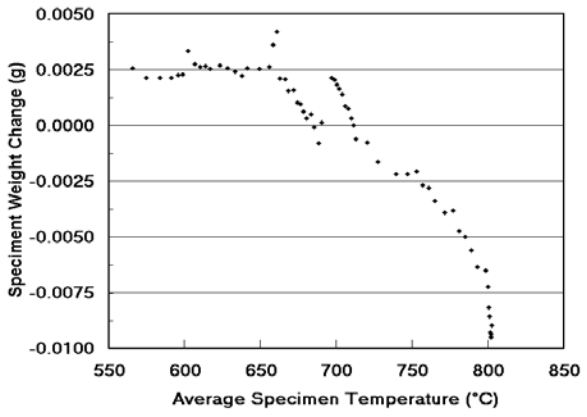


Fig. 3. Weight change as a function of temperature. Highest and lowest temperatures are in the hot and cold leg, respectively.

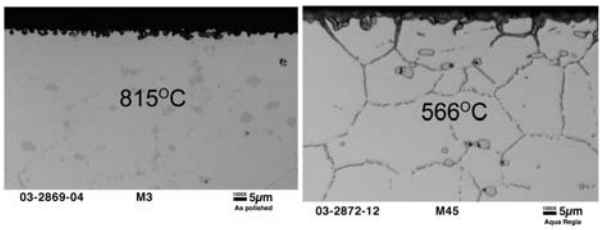


Fig. 4. Specimen at 815°C lost weight and showed a shallow depth of attack whereas the specimen at 566°C gained weight and showed a layer on its surface.

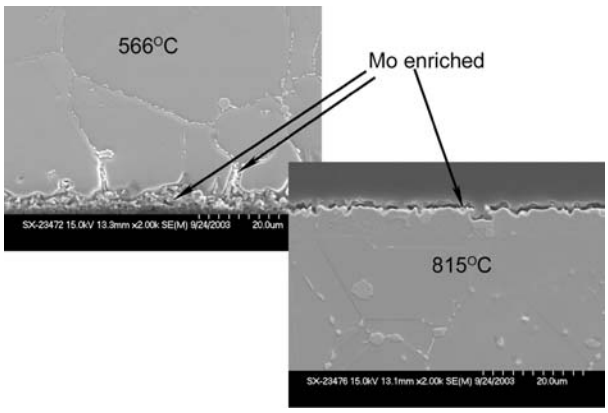


Fig. 5. SEM photo shows a very thin layer on the surface of the specimen that lost weight at 815°C. This layer is much thinner than that on the specimen at 566°C. These layers are enriched in molybdenum.

temperature. These layers are significantly higher in molybdenum content than the base material. There is also transport of the molybdenum along the grain boundaries (Fig. 5). In keeping with the absence of voids and no enhancement of chromium in the deposited layer, the chromium content in the salt was low and over the lifetime of the test averaged 167 ppm as compared to less than 40 ppm in the as-prepared salt. The molybdenum content of the salt averaged less than 60 ppm, as compared to less than 40 ppm in the as-prepared salt. These results clearly demonstrate that corrosion can be significantly reduced by reducing the impurity levels in the salt. Also, these results raise the prospect of a mechanism of molybdenum transport based on the unusual properties of some lower valence molybdenum compounds at the different temperatures, which would allow the molybdenum to disproportionate and deposit at all temperatures in the test loop. This proposed mechanism, somewhat supported by the delay before the CV scans showed a minor change in behavior (Fig. 6), needs further investigation.

Summary and Conclusions

The results from this project removed two of the fundamental barriers to pursuit of very-high-temperature

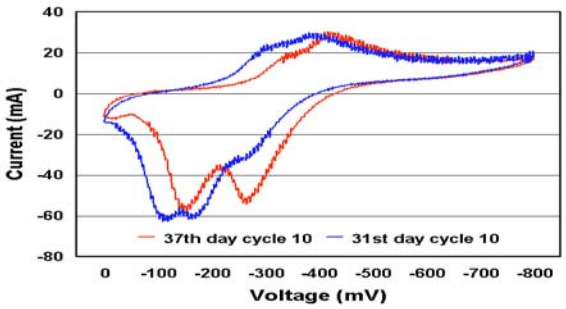


Fig. 6. Cyclic voltammetry data showing splitting of the peaks with increasing time. This is indicative of a slow increase of a species in solution.

molten salt coolants. The neutronics results showed that attractive salt and buffer systems that have acceptable responses to the most common nuclear stability perturbations exist. The electrochemical results indicate that there are very suitable redox buffers (Yb, V) that can be used to minimize corrosion at very high temperatures. Further, a stable nickel reference electrode has been identified. In the presence of very-low-salt impurities in this Group IA-halide salt (FLiNaK), which has a greater solubility for the fluorides of structural materials than Group IIA-halides, containment corrosion was very minor. However, molybdenum fluoride, which has an extremely low solubility in this salt, was transported around the loop. A mechanism based on the change in coordination with temperature is proposed. Additional studies are necessary in order to confirm the mechanism and to understand the effect on the long-term behavior of containment material.

The instrumented loop capability established on this project and the associated research findings provide several opportunities for follow-on funding. First and foremost, this work impacts the direction of the Department of Energy, Office of Nuclear Energy (DOE-NE), Generation IV Program especially, the Next Generation Nuclear Plant (NGNP) concept in which it is proposed that high-temperature (1000°C) heat be used to produce hydrogen. Fluoride salts are attractive as the heat transfer working fluids because of their low vapor pressures. The neutronics, buffers, and corrosion results of this work all support the desired outcomes of the NGNP concept. Hence, funding for the use of molten salts as heat transfer media and a molten-salt-cooled reactor concept will be pursued. A more detailed engineering evaluation of the entire concept, as well as the natural follow-on of extensive convection loop tests at even higher temperatures, will be proposed. Additional sources of support will be sought from programs that have stated an interest in fluid-fueled concepts. For fission concepts, this is the Advanced Accelerator Applications Program (formerly Accelerator Transmutation of Waste Program) and DOE-NE. Fluid-fueled concepts are especially attractive for transmutation of minor actinides because of the difficulty in conventional fabrication of fuels from these intensely radioactive materials. Work in both thermal (fluoride) and fast (chloride) systems would be pursued. The interest in fast-spectrum devices is especially intense, so a focus on chloride systems that are inherently more corrosive would be a natural extension of this work. Operations at more reducing conditions could be proposed for minor actinide systems in chlorides and fluorides and may be the only hope for reliable containment of chloride salts. The work on corrosion in chloride systems would also have significant application to pyrochemical processing schemes. The Molten Salt Reactor Experiment coolant salt constituted with Li-7, FLiBe, is widely regarded as the premier tritium breeding salt for fusion applications. The fusion community has many of the same corrosion concerns as the proponents of molten salt cooled reactors and fluid-fueled concepts. The instrumented convection loop capability can establish the corrosion resistance in their model systems, and funding will be pursued in this area.

References

[1]C. G. Kontoyannis, *Electrochim. Acta* **40**, 2547 (1995).

[2]G. D. Del Cul, D. F. Williams, L. M. Toth, and J. Caja, "Redox Potential of Novel Electrochemical Buffers Useful for Corrosion Prevention in Molten Fluorides," *Proceedings of the Thirteenth International Symposium on Molten Salts*, P. C. Trulove, H. C. DeLong, R. A. Mantz, G. R. Stafford, M. Matsunaga, eds., PV202-19, p. 431, *The Electrochemical Society Proceeding Series*, Pennington, N.J., 2002.

[3]D. F. Wilson, et al., *Hastelloy N Thermal Convection Loop: Status Report*, ORNL/TM-2003/112, Oak Ridge National Laboratory, May 2002.

[4]F. R. Clayton, G. Mamantov, and D. L. Manning, *J. Electrochem. Soc.* **121**, 86 (1974).

[5]J. R. Keiser et al., *The Corrosion Resistance of Type 316 Stainless Steel to Li_2BeF_4*, Oak Ridge National Laboratory Report, ORNL/TM-5782, 1977.

Breakthrough Multi-Megawatt Space Reactor Power System Design

S. R. Greene,[1] A. L. Qualls,[1] E. D. Blakeman,[1] S. H. Kim,[1] J. O. Johnson,[1] J. S. Neal,[1] K. W. Childs,[1]
J. C. Conklin,[2] and P. J. Otaduy[2]
[1]*Nuclear Science and Technology Division*
[2]*Engineering Science and Technology Division*

A Space Reactor Power System (SRPS) optimization tool is under development that uses evolutionary computational algorithms called genetic algorithms (GA) to optimize the design of SRPS subsystems for a nuclear electric propulsion (NEP) space vehicle. The tool is designed to work in conjunction with the Nuclear Electrical Vehicle Optimization Tool (NEVOT), which is being jointly developed by ORNL, Marshall Space Flight Center, and Arnold Engineering Development Center. The tool is being employed by ORNL to produce a pre-conceptual design for a multi-megawatt space reactor power system. Although the tool was developed for a space power application, the methodology can be applied to any multi-discipline, multi-objective optimization problem.

The objectives of this work are to (1) develop a breakthrough methodology for designing multi-discipline, multi-objective engineering systems, (2) to re-establish ORNL as a leading center for the design and development of multi-megawatt Rankine Space Reactor Power Systems (SRPS), and (3) to pilot the Tri-Lateral Alliance with Arnold Engineering Development Center and Marshall Space Flight Center.

The SRPS design optimization tool employs detailed subsystem design codes in conjunction with a genetic algorithm (GA) optimization tool to produce self-consistent, optimized, detailed SRPS designs. A series of design codes have been written to perform system evaluations after the important features of the system have been selected by the GA. The best solutions are found in the GA methodology by removing solutions with poor evaluations from a generation and repopulating the next generation with combinations or mutations of the better designs over many successive generations until only the best solutions remain.

The power of the methodology is that (1) the systems are created without sequential design information, which allows novel approaches to be investigated; (2) the algorithm does not require modification to search for the optimal solutions to different missions, constraints, or different definitions of fitness; (3) more sophisticated analyses, models, and approaches can be easily included into the methodology as they become available; and (4) the results are detailed, self-consistent designs of optimized integrated SRPS subsystems. These features make the SRPS optimization tool a flexible and powerful tool for performing subsystem trades studies and optimization for a broad range of design problems.

An integrated low-fidelity first-generation SRPS evaluation tool has been developed and is being used to facilitate the development of the NEVOT optimization tool at the Marshall Space Flight Center. A higher-fidelity second-generation SRPS evaluation module is being developed at ORNL, and initial results are expected during FY 2004. Second-generation design modules have either a sophisticated engineering design algorithm or a realistic physics evaluation of the subsystem. The liquid metal reactor module was derived from a modified version of the ALKASYS Design code.[1] The Rankine liquid metal power conversion module and the heat rejection module are derived from a set of programs developed to predict Nuclear Electric Propulsion subsystem performance and mass developed under NASA program RTOP 593-72. The potassium power conversion code[2] and the heat rejection code[3] were modified slightly to operate at lower power; however, they are largely unchanged. The shield and Brayton system power conversion design modules were written from first principles at ORNL. Integration software has been written to allow modules with different levels of sophistication to be selected for use in the evaluation, which permits the algorithm to scope down to better solutions with simple models before using the more computational-intensive second- and third-generation algorithms for fine tuning the designs.

The designs created are not derived through an iterative sequential design approach. They are evaluated after being created. By developing the tool without engrained sequential design logic, the GA algorithm has more flexibility to design and combine subsystems, which allows it to find new and unique design solutions. The algorithm does not have to be changed to search for the

optimal solutions to different problems or to emphasize a different set of optimization parameters. The result is a flexible and powerful tool for performing subsystem trade and optimization studies that can be applied to any multi-objective, multi-discipline science application.

In addition to developing a power design tool, this project will result in a state-of-the-art multi-megawatt SRPS design and new and improving relations with our Tri-lateral Alliance partners. It has already generated interest and attention from potential sponsors for follow-on funding including JPL and DOE.

References

[1]J. C. Moyers and J. P. Nichols, *ALKASYS, A Computer Program for Studies of Rankine-Cycle Space Nuclear Power Systems,* ORNL/TM-10427.

[2]G. A. Johnson, *Potassium-Rankine Power Conversion Subsystem Modeling for Nuclear Electric Propulsion (Task Order 10), NASA Contractor Report CR-191134*, 1993.

[3]M. P. Moriarity, *Heat Pipe Cooled Heat Rejection Subsystem Modeling for Nuclear Electric Propulsion (Task Order 18), NASA Contractor (Rockwell International) Report CR-191132*, 1993.

NUCLEAR SCIENCE AND TECHNOLOGY

Seed Money Fund

Variance Reduction Tools for Making Monte Carlo Radiation Treatment Planning Clinically Useful

D. E. Peplow and J. C. Wagner
Nuclear Science and Technology Division

Radiation treatment planning (RTP) for cancer therapy requires the calculation of radiation dose distributions in a computed tomography–scan representation of the body for a given set of beam locations and parameters. This is a difficult problem, which is normally done using approximate pencil-beam convolution methods. These methods lack accuracy near interfaces of dissimilar materials, such as lung/tissue or tissue/bone interfaces. Monte Carlo (MC) methods can be used to very accurately calculate dose distributions for RTP; however, the computer time required to achieve reasonable stochastic uncertainties is currently too great to enable routine clinical usage. Variance reduction methods can be applied to many MC problems to get lower uncertainties for a given calculation time. This proposal was to explore and develop variance reduction techniques from more typical nuclear engineering applications for improving the efficiency (speed) of MC simulations for RTP.

Introduction

The physician's goal in RTP is to maximize dose to the primary target volume (e.g., tumor region) while sparing, or at least reducing as much as possible, dose to the surrounding healthy tissues. Dose to nearby "critical structures" (such as the spinal cord, bladder, optic nerve, etc.) must be kept low or severe injuries to the patient may result.

The current computational tools are based on a simple model: a thin "pencil" beam of radiation striking a large tank of water. The penetration, lateral spread, and ultimate dose to each portion of the tank of water has been parameterized from experimental measurements or time-consuming computer calculations (e.g., analog MC). This one response is then applied many times to the patient geometry, and the effects are totaled to find the patient dose distribution. This method is very fast but is accurate only for patients who look like a large block of water. For real patients, with inhomogeneities such as lungs and sinus cavities of low density and bones and teeth of high density, the pencil beam approach breaks down quickly. Radiation transport depends heavily on the density and the atomic composition of the material. Using an ordinary computed tomography (CT) scan, physicians can develop a detailed model of density and material type for each patient with millimeter resolution. Pencil beam algorithms do not take advantage of this information.

Monte Carlo algorithms[1,2] have been used for many years to calculate dose distributions from radiation treatments.[3] Monte Carlo techniques are considered the most accurate since they simulate the detailed physics of radiation interactions and transport through the regions of the patient—each voxel of the CT scan. The drawback of these methods is the long running times required to calculate the dose in each voxel to a reasonable statistical uncertainty. We consulted with Julian Rosenman, MD, and his staff at the University of North Carolina (UNC) Radiology Department on the needs of clinicians. They run their MC dose calculator, MCRTP, on 100 PC's for 10 to 20 min. The expressed goal for routine clinical use is to have the dose calculator run on a single computer and take 1 to 2 min.

Technical Approach

Variance reduction (VR) methods can be applied to MC radiation transport calculations to reduce the computational time required to achieve the same statistical uncertainty of the final answers as the original, unmodified code (called the 'analog' code). These methods generally reduce the time spent following particles that will not contribute much to the final answer and instead focus on the particles that will. Hence, these techniques require a priori knowledge of particle importance. Each particle is assigned a weight, a value that represents the probability of the particle being followed by the analog code. Some of these methods are fairly simple—such as stopping particles that do not have enough energy to escape the voxel they are currently in, or forcing photons to interact with the patient at least once (and not pass through). More powerful techniques, such as stratified sampling (starting the simulation from a more uniform set of source particles) and the point detector algorithm (finding the contribution of the current particle to every dose response at each interaction), can also be used. These types of methods

have been used to speed up mammographic image simulation codes by orders of magnitude.[4] Of course, the dose distribution problem is more complex than the imaging problem, and it was recognized at the onset of this work that a similar speed up may not be possible.

Other VR approaches are more complex—such as using a faster, non-MC calculation to get an estimate of the dose distribution and then using that estimate in the MC to help guide particle selection, ultimately giving a lower uncertainty in the final result. These methods address the main difficulty associated with using VR techniques, namely, the determination of the problem-dependent VR parameters. A number of recent efforts[5] have focused on the development of automated VR methods based on statistically or deterministically generated importance functions. These efforts have generally been based on the recognition that the adjoint function (i.e., the solution to the adjoint form of the Boltzmann transport equation) has physical significance as a measure of the importance of a particle to some objective function (e.g., dose in some region).[6] It is this physical interpretation that makes the adjoint function well suited for use as an importance function for VR of MC calculations. Hence, many of the recent automated VR efforts that utilize approximate adjoint solutions substantially improve the calculational efficiency and reliability of traditional nuclear applications, while at the same time significantly reduce the time and experience requirements of the user. Based on the success with traditional nuclear application, automated VR was applied to the RTP simulation problem.

Results and Accomplishments

For each voxel, the dose D_i and the stochastic uncertainty σ_i are calculated. A standard figure of merit (FOM) is defined to be the reciprocal of the product of the calculation time T and the square of the uncertainty (variance). For the radiation treatment problem, the dose is calculated in millions of voxels, so the idea of FOM needs to be recast as an average. Similar to Kawrakow,[7] this is done by defining an average relative error s over a group of voxels as

$$s = \frac{1}{N}\sum \frac{\sigma_i}{D_{highest}}$$

where the sum is only over the N voxels with a dose between some minimum dose D_{min} and some maximum dose, D_{max}. $D_{highest}$ is the dose of the highest dose voxel in the entire problem. This allows for the problem to be portioned into low-dose, medium-dose, and high-dose areas. The FOM for groups of voxels is then $1/Ts^2$. Doses and uncertainties from codes with VR were then compared to the analog calculation, an example of which is shown in Fig. 1.

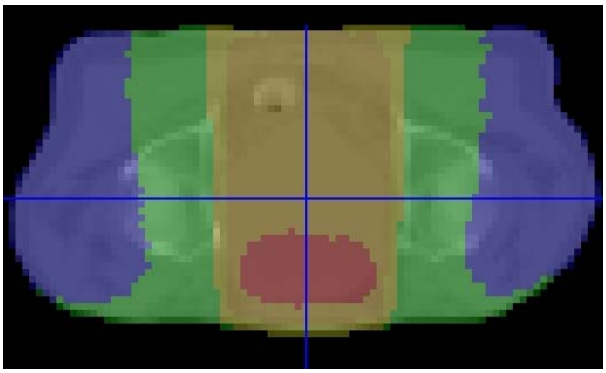


Fig. 1. One slice of a Monte Carlo dose calculation for a single beam prostate therapy. Dose colors are overlayed on the black and white CT values. Red denotes above 90% maximum dose, orange above 50%, yellow above 10%, green above 1%, and blue above 0.1%.

Standard methods for VR fall into categories of modifying the first interaction of the source photons, importance sampling of photons and electrons, particle survival biasing, and changing the photon scatter distributions. For each category, several methods were tried, and for each method listed in Table 1, several sets of parameters were tried. The increase or decrease in FOM listed for each method describes generally how the method performed for all the different cases of parameters. For example, the performance of path-length shortening scheme with respect to one of its parameters is shown in Fig. 2.

Most of these methods changed the distribution of photons throughout the patient. In photon-only problems, some of the above methods gave speed-ups of nearly a factor of 4. But with the small voxel sizes, electrons must be modeled. The electrons deposit dose in many neighboring voxels, so controlling which voxels the photons interact in does not necessarily help reduce the variance on the final energy tally.

In parallel with the above work, automated VR methods based on both deterministic- and stochastic-based importance functions were developed, implemented, and tested with two RTP problems, including one used in a previous VR study.[7] Applicable baseline models and results were generated for both photon-only and coupled photon-electron transport simulations, and metrics for computational efficiency were developed. The methods were developed with the goal of achieving more efficient global dose convergence (i.e., uniform density of simulated particles with the aim of achieving a uniform distribution of statistical uncertainty) without manual intervention (i.e., no manual "tweaking" of VR parameters). The goal of global optimization was based on consultations with the staff at the UNC Radiology Department.

The automated VR methods are based on the concept of importance sampling and represent different strategies for assigning parameters for the standard *weight-window*

Method	Increase/decrease in FOM (%)		
	Low dose	Med. dose	High dose
Modifying the first interaction of the source photons			
Preventing long first interaction distances	6	6	6
Forcing an interaction before leaving the geometry	0	0	15
First interaction distribution in the patient—uniform	–20	40	–20
First interaction distribution in the patient—front	–40	0	35
First interaction distribution in the patient—rear	–50	40	–50
Stratified sampling	0	20	–20
Stratified sampling with variance check	–40	–50	–60
Path-length shortening (Fig. 2)	–25	–25	40
Advanced path-length shortening	–30	30	60
Importance sampling			
Oversampling the edge of the beam	–10	–10	15
Rouletting the core of the source beam	15	15	–35
Source sampling: more spot photons	–40	0	5
Source sampling: more non-spot photons	20	–5	–5
Importance sampling using the source energy	–20	10	20
Electron importance as function of energy	–15	–15	–15
Particle survival biasing			
Interaction splitting	–10	–10	–10
Biasing secondary production	–30	–25	–25
Splitting low energy electrons	0	0	0
Photon scatter distribution sampling			
Compton scatter uniform distribution	–35	–35	–35
Compton scatter right angle biasing	–50	–50	–50

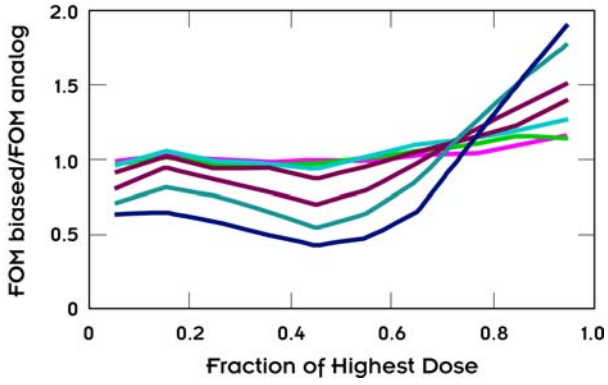


Fig. 2. Results for different levels of biasing (each color) using the path-length shortening scheme. The scheme improves FOM's for higher dose regions at the expense of medium- and lower-dose regions.

and *source-biasing* VR techniques. The weight-window technique is a space- and energy-dependent facility by which splitting and roulette functions are performed, while the source biasing technique enables preferential selection of the source space. The primary methods considered are referred to as (1) individual voxel adjoint–based VR (IVABVR), (2) global adjoint–based VR (GABVR), and (3) global forward–based VR (GFBVR). IVABVR utilizes the consistent adjoint–driven importance sampling (CADIS) methodology,[5] which is based on "traditional" deterministic adjoint functions. This method was applied to evaluate optimization potential of calculated dose in specific voxels and/or localized groups of voxels, while the GABVR and GFBVR methods attempted to reduce overall computational effort by achieving global convergence. For the IVABVR method, the maximum efficiency gain on a voxel-by-voxel basis was less than a factor of 5 and typical gains were less than a factor of 2. For a number of voxels, reductions in efficiency, as compared to the analog (unbiased) case, were observed; this was particularly true for cases that included electron transport.

The GABVR method involved weighting the voxel-wise source for the deterministic adjoint calculation based on the ratio of the maximum to individual voxel response (dose). This method attempted to weight the low-dose regions more heavily (in the adjoint calculation) and thus generate an importance map that would result in global convergence. The GFBVR method, which is similar to work in ref. 8, assigns the importance map directly based on the ratio of maximum to individual voxel response (dose). This method attempts to create a uniform density of simulated particles throughout the problem (patient) in a more direct manner than the GABVR method. Although the GABVR and GFBVR methods are well founded and have been shown to be effective for other problems, neither method demonstrated sufficient efficiency improvement to warrant further development for RTP simulations. This finding is attributed to the computational overhead associated with generating and using the detailed space- and energy-dependent importance map, the inability of the underlying method (the weight window technique) to achieve the desired uniform density of simulated particles, and the inherent characteristics of the RTP problem.

Summary and Conclusions

A VR method that worked for all voxels of the radiation-dose calculation problem without sacrificing any MC accuracy was not found. Methods that improve traditional photon and neutron MC problems did not perform well for this problem, confounded by the electron transport across the small regions.

The results of this work are generally consistent with the results of others who have developed and/or applied a wide variety of different techniques with the goal of improving the efficiency of RTP simulations and hence further demonstrate the obstinate nature of this problem with respect to improvements in computational efficiency.

Despite the lack of significant improvements in computational efficiency, this work is valuable in that traditional MC VR methods were evaluated and adjoint-based VR methods were developed for RTP simulations

for achieving global VR. It is expected that some of these methods will prove effective for other nuclear applications, including deep-penetration problems in which detailed information is sought.

References

[1]J. F. Briesmeister, editor, MCNP™—A general Monte Carlo N—particle transport code, version 4C. LA-13709-M, Los Alamos National Laboratory, Los Alamos, New Mexico, 2001.

[2]L. L. Carter and E. D. Cashwell, Particle-transport simulation with the Monte Carlo method. Washington: Technical Information Center, Office of Public Affairs, U.S. Energy Research and Development Administration, 1975.

[3]A. F. Bielajew, H. Hirayama, W. R. Nelson, and D. W. O. Rogers, History, overview and recent improvements of EGS4, SLAC-PUB-6499 (also NRC-PIRS-0436, KEK Internal 94-4), Stanford (CA): Stanford Linear Accelerator Center, 1994.

[4]D. E. Peplow and K. Verghese, "Digital Mammography Image Simulation Using Monte Carlo," *Medical Physics* **27**:568–579 (2000).

[5]J. C. Wagner and A. Haghighat, "Automated Variance Reduction of Monte Carlo Shielding Calculations Using the Discrete Ordinates Adjoint Function," *Nuclear Science and Engineering* **128**, 186 (1998).

[6]G. I. Bell and S. Glasstone, *Nuclear Reactor Theory*, New York, Van Nostrand and Reinhold, 1970.

[7]I. Kawrakow and M. Fippel, "Investigation of variance reduction techniques for Monte Carlo photon dose calculation using XVMC," *Physics in Medicine and Biology* 45:2163–2183 (2000)..

[8]M. A. Cooper and E. W. Larsen, " Automated Weight Windows for Global Monte Carlo Particle Transport Calculations," *Nuclear Science and Engineering* **137**, 1–13 (2001).

Identifying a Suitable Methodology to Extract Meaningful Results from Dose Volume Histograms using Statistical Analysis

B. L. Kirk,[1] R. W. Counts,[2] D. E. Peplow,[1] D. A. Wolf,[2] M. P. Langer,[3] and M. J. Rivard[4]
[1]*Nuclear Science and Technology Division*
[2]*Computer Science and Mathematics Division*
[3]*Department of Radiation Oncology, Indiana University*
[4]*Department of Radiation Oncology, Tufts University*

External-beam radiation therapy, teletherapy, is used as one of the main treatments for many kinds of cancer. The challenge of teletherapy is to deliver the radiation to effectively eradicate the tumor while minimizing damage to normal tissues and subsequent side effects. The radiation treatment plan includes a set of dose volume histograms (DVHs). Each DVH is a rank percentile curve that displays the fraction volume of a structure (target or some normal organ) that will receive a radiation dose of some value or greater. In this study, we analyzed DVHs and post-treatment bladder complication ratings from 61 patients treated for localized prostate cancer. The study focused on determining if linear discriminant analysis (LDA) could be used to identify features of the bladder DVHs that correlate with post-treatment complications as determined using the Common Toxicity Criteria (CTC) scale. Because of the limited range of CTC ratings and the large variation within CTC groups, LDA was not successful in identifying features of the DVHs that are highly correlated with the CTC toxicity ratings. LDA may perform much better using a larger volume of patient data collected in systematic and controlled manner. Critical to the success of future studies are precise definitions of the CTC groups to be distinguished so that the groups exhibit differences that make discrimination viable.

Introduction

External-beam radiation therapy, teletherapy, has been used since the late 1950s as one of the main treatments for many kinds of cancer and is usually administrated over a period of 5–8 weeks. Because the radiation used to destroy the cancerous cells travels through the body to reach the target, normal tissues which lie along the path to the target receive some of the radiation dose and, as a result, complications due to the radiation may develop in the normal tissues. The challenge of radiation treatment has been to improve the planning and delivery of radiation beams to more effectively target the tumor while keeping damage to normal tissues and subsequent side effects at an acceptable level.

During the past decade, significant advances in the planning and delivery of radiation therapy have been obtained using more energetic radiation generators, advances in computational speed, improved computer graphics, and the wide availability of whole-body computed tomography (CT) scans. One such advance was the development of three-dimensional conformal radiation treatments (3D-CRT). This type of treatment shapes the radiation field to conform to the individual patients' tumor and is delivered with a higher degree of precision than earlier types of teletherapy. Conventional 3D-CRT utilizes a cross-fire, four-field box technique that centers the radiation beams on the target but also includes substantial portions of the intervening normal tissues within the radiation fields.

A newer, more advanced form of radiotherapy delivery, intensity-modulated radiation therapy (IMRT), represents a significant improvement in treatment conformality. IMRT uses many thin beams (beamlets), having small fields to precisely irradiate the target while reducing exposure to other nearby organs. In addition, IMRT allows the intensity of each beamlet to be modulated so that the intensity of the radiation can be varied during the treatment delivery. IMRT produces dose distributions that are far more conformal than can be obtained with standard 3D-CRT. While IMRT reduces radiation exposure to healthy tissues compared to 3D-CRT, these toxicities are not entirely eliminated. The move from standard 3D-CRT to IMRT has resulted in a larger volume of normal tissue being exposed to lower doses.

IMRT allows the radiation oncologist to specify desired doses or dose limits to be delivered to specific regions of normal tissues. Using these goals, specialized software calculates the beam intensity maps that are to be delivered. Once the treatment plan is calculated, a set of dose volume histograms (DVHs) is obtained which depicts the volumes of target and normal tissues that will receive a particular radiation dose. A DVH is a rank percentile

curve that displays the fraction volume of a structure (target or some normal organ) that will receive a radiation dose of some value or greater. Many such curves can be developed by the physician and radiation physicist, and the curve that is ultimately selected as the treatment plan for an individual patient is the decision of the radiation oncologist and medical physicist. A representative set of DVHs used for IMRT planning is presented in Figure 1.

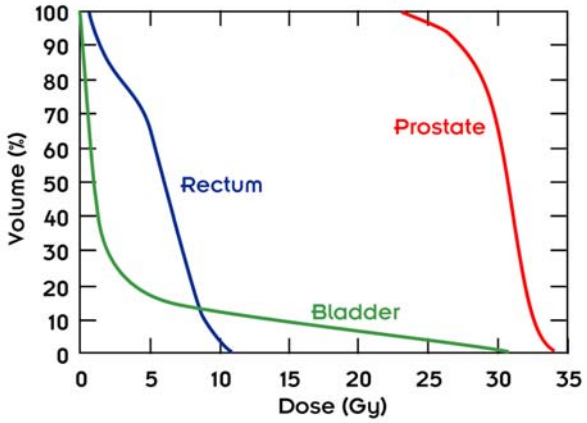


Fig. 1. An example of cumulative dose-volume histograms for IMRT treatment of prostate cancer.

The availability of new methods for treatment planning allows novel distributions of dose across structures of interest to be created. A basic problem of the prescribing physician is to review the created dose distributions to determine whether they are safe to prescribe. The DVH allows the vast amount of information contained in a three-dimensional dose distribution to be reduced to a two-dimensional framework to facilitate review and inter-comparisons among cases. However, building reliable predictors of injury from a set of two-dimensional plots remains a difficult theoretical problem. The purpose of this study was to determine if liner discriminant analysis (LDA) can be successfully applied to identify features of one tissue, the bladder, that correlate with complications after IMRT for prostate cancer. If features of the DVH can be identified that may be used to predict complications, this information would be valuable to aid future IMRT treatment planning.

Technical Approach

Clinical data for this study were provided by the Department of Radiation Oncology at Tufts-New England Medical Center (Tufts-NEMC). The data were obtained retrospectively from patient charts and completely de-identified before sending to the ORNL for analysis. Informed consent was not contained as the approach was approved by the Institutional Review Boards at both Tufts-NEMC and ORNL. Data consisted of treatment planning information and post-treatment bladder complication ratings from 61 patients treated for localized prostate cancer. The irradiation technique consisted of a combination of four co-planar fields using 3D-CRT techniques followed by IMRT. Patient target doses for the four-field treatment varied from 45 to 55 Gy (mean of 45.3 Gy), while target doses for the IMRT varied from 10–28 Gy (mean of 26.4 Gy). The total target dose ranged from 65 to 73 Gy (mean of 71.5 Gy).

Patient treatment information for the four-field treatment was limited to the target dose as DVHs were not consistently available. However, detailed information was provided for the IMRT plans and included the following: (i) the target dose [Gy and % isodine line], (ii) DVH plots, (iii) the prostate total volume [cm^3], and (iv) the bladder total volume [cm^3]. Also provided was the bladder toxicity rating according to the Common Toxicity Criteria (CTC) scale version 2.[1] Table 1 presents a general description of the CTC toxicity scale along with the number of patients assigned to each grade.

A key question in radiation therapy is the following—given that a certain total dose of radiation must be given to kill the cancer cells, are there certain dose distributions that are less likely to lead to post-treatment complications? It has been concluded by the National Institutes of Health and the American Association of Physicists in Medicine that the ability to evaluate and compare the complex dose distributions with respect to expected clinical outcomes has lagged far behind the advances in hardware and software for delivering the treatment.[2] Previous studies have addressed this question by analyzing DVHs to determine if they contain features that correlate to post-treatment complications.[3,4,5] These studies applied techniques such as mechanistic model building, direct comparison of DVHs, permutation tests, and logistic regression. In this study we applied linear discriminant analysis (LDA) to determine if there is a predictive relationship between the DVDs and post-treatment complications. Our approach to discrimination makes use of the entire DVH for each patient, which is in contrast with that published in studies where researchers select only a few points on the curve for discrimination.

Table 1. General patient complication assignment according to the CTC toxicity scale

CTC grade	Complication description	No. of patients
0	No complication	30
1	Minor symptom requiring no treatment	20
2	Minor symptom requiring medication	10
3	Symptom requiring minor surgical intervention	1
4	Symptom requiring major surgical intervention	0

Results and Accomplishments

Thirty-two predictor variables were defined to describe bladder DVHs. The variables representing IMRT boost doses beginning at 0.1 Gy and ending at 31.1 Gy in increments of 1.0 Gy were established. The variable values were defined as the corresponding bladder volumes (either % volume or absolute volume) from the DVHs. Because only one patient was rated toxicity grade 3, three groups were defined: CTC grade 0, 1, and 2/3.

Linear discriminant analysis (LDA) was used to determine if the predictor variables defined from the DVHs could be used to identify combinations of dose and volume that separate DVHs into risk groups for bladder toxicities. Analyses were performed using both percent-of-bladder volume and absolute-bladder volume. Because three CTC groups were defined, two linear discriminant functions (LDFs), denoted as LD1 and LD2, were defined for the three groups of CTC toxicity grades.

The dose-volume values used in the LDA came from the discretization of the DVH curve resulting in a large number of correlated predictor values. It is well known that LDA using high-dimensional, correlated predictors results in an over-fitting of the data. That is, the LDA tends to give overly optimistic results. A variable selection procedure was used to reduce the dimensionality of the problem. We approached the variable selection through all-subsets binary 0,1 regression. (Note that for the two-group case, a standard regression analysis on a binary variable indicating group membership is equivalent to LDA.) Each pair of CTC groups, CTC 0 vs. CTC 1, CTC 0 vs. CTC 2, and CTC 1 vs. CTC 2, was considered. In each of these cases, one group was coded 0 and the other group coded 1 and an all-subsets regression analysis was performed using the possible 32 predictors. The best predictor sets of size 1, 2, and larger were evaluated. This analysis suggested that more than three variables was unwarranted and that a set containing doses 1.1 Gy, 8.1 Gy, and 17.1 Gy along the DVH was among the best (see Fig. 2). Using this three-variable set in an LDA of all three groups resulted in an overall 29 of 61 (47.5%) correct classification rate. A cross-validation procedure was used to estimate how well the LDA might perform for future sample. The cross validation consisted of two steps. First, one patient was omitted and linear classification functions (LCFs) were determined on the remaining 60 patients. Then these LCFs were used to classify the omitted unit into one of the three CTC groups. This process was carried out 61 times, and the number of omitted units correctly classified was used to estimate the classification rate. This procedure resulted in 26 of 61 (42.6%) being correctly classified. The procedure as described above was repeated, replacing percent-of-bladder volume with absolute-bladder volume. Again, a three-variable set containing doses of 1.0, 11.0, and 14.1 Gy was selected for the discriminant analysis. The LDA resulted in an overall 28 of 61 (45.9%) correct classification rate (cross validated rate of 41.0%). Table 2 summarizes the results of these analyses.

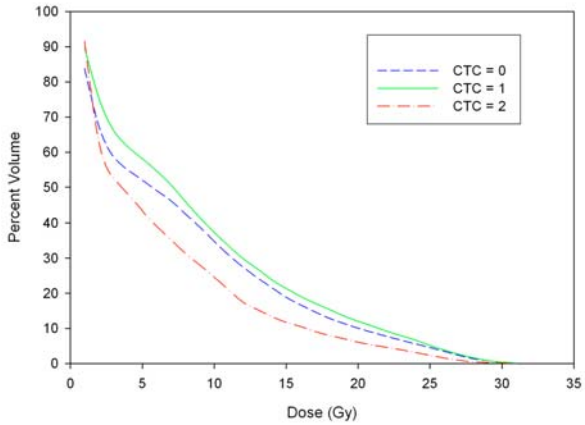


Fig. 2. CTC group means versus dose for percent bladder volume.

As indicated from the results displayed in Table 2, this application of LDA was not successful in identifying features of the DVHs from the IMRT boost doses that are highly correlated with the CTC toxicity ratings. We believe that the lack of success was largely due to two reasons. First, the data used in this study covered a limited range of CTC ratings. The data included only one patient with a CTC = 3 and none with CTC = 4. Second, there was a great deal of variability among DVH curves within each CTC group. Given the limited range of CTC ratings and the large variation within CTC groups, it is doubtful that any procedure would identify combinations of dose and volume that separate DVHs into risk groups for bladder toxicities.

Summary and Conclusions

An important issue facing the biomedical medical research community and recognized by both the National Cancer Institute and regulatory authorities is how to ensure the safety of patient treatment. New devices in radiotherapy offer the potential to create novel dose distributions that may enhance the likelihood of tumor sterilization, but determining which of the new dose distributions are safe is a pressing problem. This study has demonstrated that linear discriminant analysis is an appropriate methodology

Table 2. Summary of linear discriminant analyses

Variables used in the Discriminant Analysis	Dose values selected (Gy)	Classification rate	Cross-validated classification rate
DVH % volume	1.1, 8.1, 17.1	47.5%	42.6%
DVH absolute volume	1.0 , 11.0, 14.1	45.9%	41.0%

for identifying features of radiotherapy DVHs and may be effective in predicting post-treatment complications. The study has also revealed the importance of the quality of the data used in evaluating the methodology.

Large databases of dose volume histograms for structures of interest that include hundreds or thousands of patients are being assembled, often as part of National Cancer Institute (NCI)–sponsored projects in single institutions or cooperative groups. Mathematical and statistical techniques to analyze the vast information data bases are wanting. A pressing need is to derive from the data set reliable predictors that will provide a greater measure of patient safety, allow new treatment delivery techniques to be implemented with more assurance to patients and prescribing physicians, and provide cleaner estimates of complication rates to support the design of large-scale clinical trials. There is now no recognized way to derive such predictors from the acquired dose distributions and to form from them constraints on the dose volume histogram to more effectively guide the treatment planning process. Linear discriminant analysis offers an opportunity to do so, using sophisticated tools developed in the mathematics and statistics communities that remain to be fully explored in clinical radiation oncology.

The analysis done in the report manifests the applicability of Department of Energy (DOE) research into real applications—in this case health care. The potential impact of such a study will also benefit the National Institutes of Health (NIH) programs. As a result, the investigators are applying for follow-on funding to the NIH. A recent report[2] emphasizes the importance of the analysis of the uncertainties in DVH data. A proposal to NIH is being prepared.

References:

[1]A. Trotti, R. Byhardt, J. Stetz, C. Gwede, B. Corn, K. Fu, L. Gunderson, B. McCormick, M. Morris, T. Rich, W. Shirley, and W. Curran, "Common Toxicity Criteria: Version 2.0. An improved reference for grading the acute effects of cancer treatment: impact on radiotherapy," *Int. Journal Radiation Oncology Biol. Phys*. **47**, 13–47 (2000).

[2]J. Deasy, A Niemierko, D. Herbert, D. Yan, A. Jackson, R Haken, M. Langer, and S. Sapareto, "Methodological Issues in Radiation Dose-Volume Outcome Analyses: Summary of a Joint AAPM/NIH Workshop," *Medical Physics* **29**(9), 2109–2127 (2002).

[3]A. Hartford, A. Niemierko, J. Adams, M. Urie, and W. Shipley, "Conformal Irradiation of the Prostate: Estimating Long-Term Rectal Bleeding Risks using Dose-Volume Histograms," *Int. J. Radiation Oncology Biol. Phys*. **36**(3), 721–730 (1996).

[4]M. Skwarchuk, A. Jackson, M. Zelefsky, E. Venkatraman, D. Cowen, S. Levegrun, C. Burman, Z. Fucs, S. Leibel, and C. Ling, "Late Rectal Toxicity after Conformal Radiotherapy of Prostate Cancer (I): Multivariate Analysis and Dose-Response," *Int. J. Radiation Oncology Biol. Phys*. **47**(1), 103–113 (2000).

[5]A. Jackson, M. Skwarchuk, M. Zelefsky, D. Cowen, E. Venkatraman, S. Levergrun, C. Burman, G. Kutcher, Z. Fucs, S. Liebel, and C. Ling, "Late Rectal Toxicity after Conformal Radiotherapy of Prostate Cancer (II): Volume Effects and Dose-Volume Histograms," *Int. J. Radiation Oncology Biol. Phys*. **49**(3), 685–698 (2001).

Development of a Three-Dimensional Radioisotope Depletion Method Using Monte Carlo Transport

M. D. DeHart and L. M. Petrie
Nuclear Science and Technology Division

Accurate calculation of the depletion of nuclear materials requires careful determination of the neutron flux density and spectrum in the region(s) of interest. Increasing complexity in reactor designs, evolutionary concepts, and nonreactor applications such as safeguards, security, and nonproliferation are beginning to require robust geometrical modeling capabilities to capture neutron transport for complex configurations. Monte Carlo transport methods offer the type of flexibility needed for such applications but will present other difficulties not encountered in deterministic transport methods. The objective of this project is the development of a prototypic code sequence to perform Monte Carlo-based depletion. As part of this effort, approaches will be developed to relate the stochastic uncertainty of the Monte Carlo method to the uncertainty in isotopic predictions, minimize uncertainty via variance reduction, and reduce the computational effort associated with burnup-dependent iterations.

Deterministic solutions have advantages over Monte Carlo simulations for the transport solutions used in coupled depletion analyses because of their ability to generate an accurate spatial distribution of fluxes over a complete problem domain. On the other hand, Monte Carlo methods provide powerful geometric modeling capabilities in three-dimensional domains. Monte Carlo calculations require a large number of neutron-particle simulations to converge on an accurate system response. In order to obtain reasonable local neutron fluxes and power density distributions, significantly more computational effort must be invested. This effort is compounded in a depletion calculation, where transport solutions must be repeated in an iterative sequence alternating with depletion calculations to update isotopic cross sections and inventories. And while deterministic solutions are based on fluxes that are converged to a specified degree over the full problem domain, the nature of Monte Carlo simulations makes it extremely difficult to obtain accurate fluxes in locations that are far removed from the most reactive region of an analysis domain. Since the accuracy of the neutron flux is therefore a function of position in a Monte Carlo simulation, the accuracy of the depletion solution is likewise spatially distributed. If effective depletion based on Monte Carlo transport is to be successfully implemented, these deficiencies must be recognized and addressed.

Coupled Monte Carlo–point depletion methods have been developed elsewhere with varying degrees of success. MOCUP and its more widely used successor MONTEBURNS are based on the MCNP Monte Carlo code coupled with ORIGEN2 for depletion calculations. The German KENOREST system utilizes KENO V.a combined with HAMMER and ORIGEN2. Japan's Toshiba Corporation recently published a description of an in-house system based on MCNP and ORIGEN2. Japan's MVP-BURN and SWAT2 depletion sequence couple the MVP Monte Carlo code to ORIGEN2. Approaches based on MCNP are limited by the slow solution time of that code relative to other Monte Carlo codes such as ORNL's KENO V.a. ORIGEN2 is an older depletion code with known deficiencies and is no longer supported for future development. Additionally, none of these packages consider the effects of error propagation nor implement variance reduction methods to accelerate the solution. Finally, few if any of these systems are maintained under a controlled QA environment.

The current work at ORNL seeks to implement the Monte Carlo transport codes KENO V.a and KENO VI within the TRITON depletion driver module of SCALE, and to study and implement methods to improve the efficacy of iterative depletion based on Monte Carlo transport solutions. Error propagation from the results of Monte Carlo simulations will become a key attribute of such a system. ORNL possesses both adjoint and auto-differentiating versions of ORIGEN-S as developmental tools that have been used previously in propagating data uncertainties into the calculation results. Transitioning this capability to track the effects of flux uncertainties is the next logical step.

This project is still relatively young. Although initial work was started in FY 2003, the majority of the project will be completed in FY 2004. To date, ORNL has modified the TRITON sequence to accept KENO V.a input

and to perform transport solutions driven by TRITON. The SCALE utility code KMART, originally developed to post-process KENO V.a calculations, has been adapted to provide collapsed cross sections and fluxes required by TRITON for setting up ORIGEN-S depletion calculations. TRITON is able to perform full KENO V.a-based depletion calculations for simple fuel pins and is undergoing debugging in application to full assembly depletion analysis.

Potential users of our new depletion codes include the National Nuclear Security Administration, the Office of Nuclear Energy, Science, and Technology, the Office of Science, and the Office of Civilian Radioactive Waste Management within the Department of Energy (DOE). Outside the DOE, other organizations and agencies that would potentially benefit from the results of this work include The Department of Homeland Security, the Department of Defense, and the National Aeronautics and Space Administration. The Nuclear Regulatory Commission has a strong interest in this work and has been following ORNL's progress. The advantages to potential users of a SCALE-based depletion sequence are multifold: SCALE is an easy-to-use system developed around an understandable and simplified input form, ORNL provides formal quality assurance procedures, source-code control, validation and verification of code components that meet many sponsors' needs and ongoing development, testing, and updates. ORNL's SCALE project team provides training and support for users, and ORNL has the capacity to support and protect both proprietary and classified work. Most importantly, ORNL has long been recognized as a world leader in highly accurate and effective Monte Carlo methods (KENO family, MORSE), depletion (ORIGEN family) and automated code sequences (SCALE and its TRITON sequence). ORNL staff have long maintained a productive and cooperative relationship with many of the potential sponsors cited above.

PHYSICS

Director's R&D Fund

Detector Development for Fundamental Neutron Physics at the Spallation Neutron Source

G. Greene,[1] T. V. Cianciolo,[1] K. Rykaczewski,[1] R. Grzywacz,[1] W. M. Snow,[2]
W. D. Bowman,[3] S. Penttila,[3] and W. S. Wilburn[3]
[1]*Physics Division*
[2]*Indiana University*
[3]*Los Alamos National Laboratory*

Low-energy, cold neutrons have been employed in a wide variety of investigations that shed light on important questions in particle and nuclear physics. Among the most important experiments in this field are those related to the precise measurement of the parameters that describe the beta decay of the free neutron. This includes the determination of the correlations between the decay particle momenta and between the decay particle momenta and the neutron spin. The purpose of this project is the development of a detector that can be used to provide a highly accurate determination of these decay correlation coefficients. Such measurements would provide a detailed test of the standard model and new information that could address a current discrepancy among data from nuclear beta decay and high-energy physics experiments. The work supported here concerns the development of the silicon detectors, front-end electronics, and electromagnetic fields that comprise the neutron decay detector.

The challenge in the measurement of the neutron beta decay correlations lies in the development of a detector that can measure the complete kinematics of the decay products that result from the decay of a polarized neutron. This includes the determination of the direction of emission of the decay proton and electron and their energies. We are pursuing a novel approach that will allow the measurement of the electron and proton in the same detector as well as provide accurate timing of delay between the detection of each particle. The detector will consist of segmented silicon surface barrier detectors, high-speed digital signal processing front-end electronics, and a combination of electric and magnetic fields that carefully define the particle trajectories. We have made substantial progress in the development of each of these components.

Silicon Detectors. The silicon detectors must satisfy the following three criteria: (1) a sensitivity of <30keV, (2) an energy resolution of ≤10keV up to ~1MeV, and (3) a capability of few-nanosecond timing. We tested a series of thick (~2-mm) silicon surface barrier detectors and demonstrated charged-particle detection with sensitivities and timing resolution that satisfy conditions (2) and (3). In a separate experiment, we demonstrated that the thin (~200-μm) silicon detectors with very thin entrance windows provide energy sensitivity as required for condition (1). This work demonstrates that a thick detector with a thin window should be able to meet all the criteria. We have identified manufacturers who are able to provide such detectors.

Digital Signal Processing Electronics. In a series of tests with radioactive sources, we demonstrated an ability to reach noise-limited timing resolution with a 40-MHz digital signal processor based data acquisition system. While the performance of such a 40-MHz system is somewhat marginal for our needs, this work has given us confidence that a 100-MHz system will be well suited. As a result we have begun the development of such a system in collaboration with a commercial vendor. A prototype board has been designed, and fabrication is under way. This system has the feature that it is fully scaleable to the approximately 200 channels required for a final experiment with a segmented detector.

Electric and Magnetic Field Design. Using a number of simulation techniques including the particle trajectory calculation programs Simian© and Giant as well as custom analytical software, we have developed a system of electric and magnetic fields that define the decay particle trajectories and provide suitable particle timing. The system developed is unique in its ability to utilize the proton time of flight to provide complete kinematic information of the decay particle. It was further demonstrated that the field geometry is sufficiently homogeneous to provide immunity from all identified systematic effects.

The measurements and simulations have provided an initial design for a neutron decay correlation coefficient detector that satisfies the technical requirements of the next generation of experiments anticipated at the Spallation Neutron Source. Because the design is based on a straightforward extension of existing commercial technologies, there should be no fundamental problems in its realization.

Three-Dimensional Neutron Structural Microscopy: Design and Demonstration

S. Spooner,[1] B. C. Larson,[1] G. E. Ice,[2] J. D. Budai,[1] J. Tischler,[1] and C. R. Hubbard[2]
[1]*Condensed Matter Sciences Division*
[2]*Metals & Ceramics Division*

We have designed and tested achromatic neutron-beam focusing mirrors and spatially resolved neutron Laue-diffraction techniques as steps toward enabling nondestructive, spatially resolved, three-dimensional (3D), structural microscopy investigations of coarse-grained polycrystalline materials. Kirkpatrick-Baez neutron super-mirrors tested at the pulsed-neutron LANSCE facility showed that ~1 × 3 mm^2 neutron beams could be focused to ~0.1 × 0.2 mm^2 diameters, a flux gain of ~75 assuming a reflectivity of 0.7 for each mirror. Depth-resolved, unfocused-white-beam Laue diffraction measurements performed at the IPNS facility using assembled polycrystals demonstrated that our traveling-aperture technique provides both the crystal structure and the lattice orientation with <1 mm spatial resolution. Similar measurements on cylindrically indented copper showed that plastic deformation induced lattice rotations could be spatially resolved in 3D. Arrangements have been made to demonstrate focused, 3D, depth-resolved neutron microscopy at the Chalk River steady-state reactor.

Although advanced materials processing relies almost entirely on controlling microstructural features such as grain boundaries, nondestructive 3D probes of the local lattice structure, strain and orientation are not currently available. This project aims at developing 3D neutron structural microscopy with submillimeter resolution. Our approach employs high-resolution neutron beam focusing and depth-resolved Laue diffraction measurements. The scientific motivation is the investigation of the local 3D inter- and intra-grain interactions using both spallation and reactor white-beam sources. Three-dimensional studies are needed to link materials theory and computer modeling with experimental microstructures.

Our analysis of non-dispersive (achromatic) Kirkpatrick-Baez (KB) neutron optics showed that mirror focusing will provide increased neutron intensity with smaller beam diameter than simple apertures for beams focused to 0.5 mm or less. Guided by the modeling, components of a neutron KB mirror pair were designed and assembled into a working independent unit, as shown in Fig. 1. Each mirror consists of a SiO_2 substrate with M3 multilayer coatings, and mechanical bending generates an elliptical reflecting surface. Laser testing of the focus was completed, and the first neutron test of the KB mirror was carried out at the Los Alamos Neutron Science Center (LANSCE) pulsed neutron source. The mirror assembly successfully focused a monochromatic 1- × 3-mm^2 beam to half-intensity widths of ~0.1 × 0.2 mm^2, demonstrating a flux gain of ~75 assuming a reflectivity of 0.7 for each mirror, consistent with the original objectives.

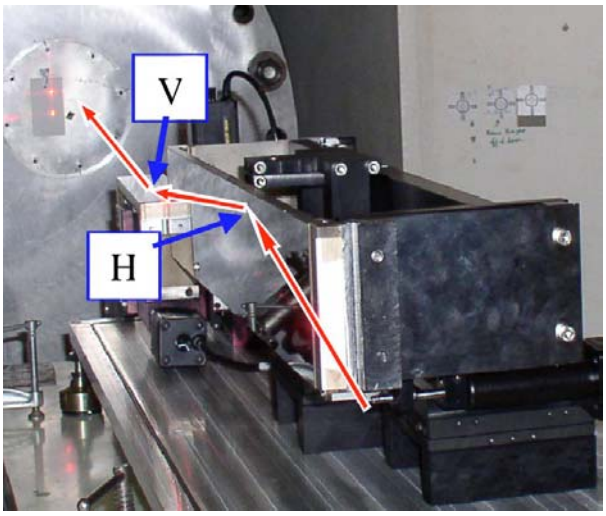


Fig. 1 Neutron-optics apparatus used to experimentally evaluate KB mirror focusing performance at LANSCE. H and V indicate the horizontally and vertically focusing mirrors. The red arrows show the neutron path as the beam reflects first from the H mirror and then the V mirror.

Depth-profiling of neutron Laue diffraction and data analysis techniques were adapted from our recently developed synchrotron X-ray profiling procedures[1] and were tested in two experiments carried out using the Single Crystal Diffraction (SCD) beamline at the relatively low-intensity Intense Pulsed Neutron Source (IPNS) at Argonne. The first experiment demonstrated the ability of 3D Laue diffraction microscopy to spatially resolve

lattice symmetry, chemical structure, and crystal orientation as a function of depth along an incident unfocused 1-mm-diameter neutron beam. As shown schematically at the bottom of Fig. 2, a series of copper or niobium single-crystal plates of 0.5-, 1.0-, or 2-mm thickness were stacked perpendicular to the neutron beam direction. Analysis of Laue images from this "synthetic polycrystal" revealed the depth-resolved crystal orientations and the niobium-versus-copper structure of individual plates with 1-mm spatial resolution. A second neutron microscopy experiment was performed on the same beam line to demonstrate the ability of the technique to determine lattice rotations in a large copper crystal that was cylindrically indented to a depth of 0.5 mm using a 5-mm-diameter steel rod. Measurements made as a function of depth along a direction perpendicular to the indent axis and as a function of distance below the indent yielded a spatially resolved map of the local lattice rotations (range ~0–5°) inside the deformed sample, successfully demonstrating our ability to nondestructively map 3D microstructures.

Combining the developments described above, the final test of the neutron microscopy concept is planned for early 2004 at the Neutron Reactor Universal (NRU)—a steady-state neutron source at Chalk River Laboratories. A white beam will be deflected from the main beam by a mirror contained within the monochromator drum of the L2 diffractometer. This beam will be focused by our KB mirror to a submillimeter diameter, and the scattering gauge-volume will be scanned through large-grained polycrystalline specimens. A high-resolution image-plate area detector will record the Laue patterns. Important objectives in these experiments will be (1) evaluation of the resolution achievable in lattice strain measurements and (2) a relative comparison of steady-state versus pulsed-neutron (time-of-flight) microscopy techniques.

In summary, ray-trace modeling has shown show the capabilities of KB mirrors for focusing to submillimeter dimensions. A neutron-optics apparatus was constructed and tested, achieving focusing to a beam width of 0.1–0.2 mm. Depth-resolved Laue diffraction experiments at the IPNS demonstrated the ability to nondestructively map microstructures inside bulk specimens. We believe that combining neutron-beam focusing, depth-profiling, and high-resolution area detectors will enable important new 3D micro- and macro-neutron materials investigation capabilities. The advent of more powerful neutron sources such as the SNS will be ideally suited to exploit these capabilities. In engineering technology, the determination of the stress state within large-grained materials will become possible for the first time. Individual grain morphologies and microstructures could be monitored during thermal or deformation processing. In biology, the application of focused beams to macromolecular crystallography has the potential to significantly reduce the size of the required specimens. The construction of high-pressure cells for smaller samples in a focused beam could also be greatly simplified. These examples represent but a few areas where focusing optics and 3D structural microscopy can have a major impact on neutron science and materials investigations.

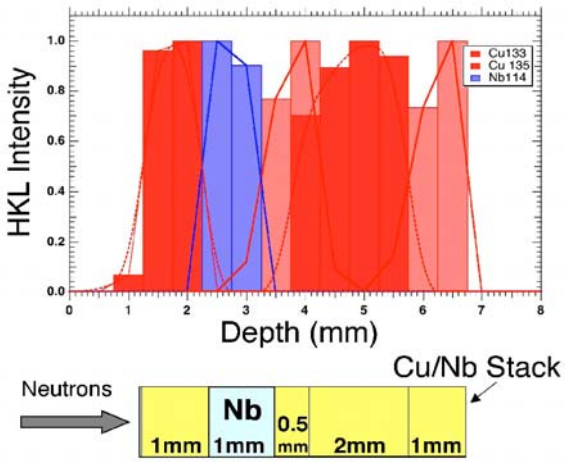


Fig. 2 Experimental demonstration of depth-resolved and BCC/FCC lattice-structure-resolved neutron microscopy from a stack of copper (FCC) and niobium (BCC) single-crystals simulating a two-phase polycrystalline material.

Reference

[1]B.C. Larson, W. Yang, G. E. Ice, J. D. Budai, and J. Z. Tischler, "Three-Dimensional X-Ray Structural Microscopy with Submicrometre Resolution," *Nature* **415**, 887 (2002).

Neutron-Rich Radioactive Ion Beam Production with High-Power Electron Beams

J. R. Beene,[1] W. T. Diamond,[2] J. D. Fox,[1] D. W. Stracener,[1] and M. Saltmarsh[1]
[1]*Physics Division*
[2]*Chalk River Nuclear Laboratories*

We are investigating the production of neutron-rich radioactive ion beams (RIBs) using high-power electron beams. This concept has never been employed at an operational facility, but numerical investigations and preliminary experiments indicate that there is potential for very high intensities of particularly interesting RIB species at a cost much lower than that of a comparable-performance conventional facility. We propose to measure RIB yields of key species, in a well-understood environment such that relative production rates and approximate absolute rates can be deduced out to very neutron-rich species. These measurements will provide a first quantitative experimental evaluation of electron-beam RIB production.

Introduction

Much of the scientific focus in nuclear physics for the next decade will be on the use of neutron-rich radioactive ion beams (RIB) to explore novel nuclear structure effects associated with weak nucleon binding, to address important questions in nuclear astrophysics, and to provide the opportunity for certain critical cross section measurements of importance to the nuclear safeguards and stockpile stewardship community. ORNL is the site of the Holifield Radioactive Ion Beam Facility (HRIBF), which will provide the nucleus on which a RIB physics research community will be developed and nurtured in anticipation of a next-generation U.S. facility. The science that has been done over the last several years at HRIBF is remarkable. However the HRIBF in its current form cannot meet the level of performance that will be required of it over the next decade. We have developed a phased approach to upgrading our facility that fits well with our understanding of possible funding profiles that are acceptable to the DOE Office of Nuclear Physics (DOE-NP). As part of this phased approach, we have identified an entirely new direction[1] for RIB production using a high-power electron beam which could improve the capabilities of the HRIBF by two orders of magnitude in critical areas, at modest cost.

To assess the potential of this concept, we propose to use a high-power beam of electrons from the Oak Ridge Electron Linear Accelerator (ORELA) to induce photo-fission in a uranium foil target array. Fission products emitted from the uranium foils will be stopped in high-pressure helium gas and captured by NaCl aerosol particles in the gas. The flow of aerosol-laden helium will transport the fission products over a distance ~12–15 m to a skimmer system, where the aerosols will be separated from the helium and injected into an appropriate ion source for subsequent acceleration and mass analysis. Transit times below 1 second are expected, with essentially no chemical-species dependence. Consequently we will be able to make systematic yield measurements of a variety of elements over a wide isotope range extending to very neutron-rich nuclei. This program can be accomplished at very modest cost because of the availability of ORELA, and a diligent canvassing of the North American nuclear physics community for surplus hardware that can be used to assemble our experimental system.

A target suitable for high-power (at least 10 kW) irradiations at ORELA has been designed, fabricated, and pressure tested to 15 atmospheres. The multifoil target contains 20 uranium foils inside of a water-cooled housing made from high-purity copper with input and outlet connections for the transport of salt-loaded helium gas. The transport of fission products from this target has been checked with a ^{252}Cf source and flowing helium at 10 atmospheres. Preliminary analysis indicates that the transport efficiency is around 40% along a 30-m, 0.75-mm-diameter stainless tube.

Substantial effort was devoted to adapting the HRIBF electron beam plasma ion source to function as an ion source/skimmer. We have demonstrated that we cannot achieve the necessary He throughput with this source and decided instead to adapt an existing Bernas-Nier source which was developed at Chalk River Nuclear Laboratories (CRNL) as a He-jet source. A completed source is available for testing as of September 30, 2003.

Design of the system to contain the production target, and insert it into the ORELA target room, has been completed. The device will be inserted in one of the neutron beam flight tubes (Flight Tube # 9) and will

intersect the electron beam between the vacuum window and the neutron target. The design includes vertical and horizontal position adjustment over a large range. The target can be retracted out of the electron beam line during normal ORELA operation. A significant effort has been placed into considerations of remote handling of the target after the experiments are finished. The design is complete and is undergoing final checking in preparation for fabrication. Fabrication of the insertion device will be completed in the second quarter of FY 2004.

All equipment necessary for the completed experimental setup has now been procured or identified. A major savings was achieved by locating a complete helium recirculation system at Los Alamos Nuclear Laboratories (LANL). Arrangements to transfer it to ORNL took far longer than expected and resulted in a substantial delay in the project. Testing and commissioning of the helium recirculation system was well under way by the end of FY 2003. We will begin assembly of the complete experiment system in the first quarter of FY 2004. The setup will be fully operation in the third quarter of 2004. We plan to begin measurements no later than the end of the third quarter and complete them by the end of fiscal year. A postdoctoral fellow has been identified to participate in this work. We have been able to carry out this project in a very cost-effective manner because we have been able to identify and obtain specialized equipment ideally suited to our needs from other laboratories, at essentially zero cost in dollars, but a finite cost in time delay.

The first phase of the HRIBF upgrade plan has just been funded by DOE-NP at a level of $4.7M for FY 2004–2005. This initial project does not involve a production accelerator upgrade and hence is not directly coupled to this LDRD project; however, the existence of this R&D program on an innovative direction for RIB production which might eventually be implemented at a later stage was viewed favorably by both program office staff and project reviewers. If this project is successful, we will have a strong case for adding a modest electron accelerator to the upgraded HRIBF complex. The total additional cost of the electron-accelerator addition to the upgrade program would be about $10M. The total upgrade package for the HRIBF, including the electron accelerator, should approach $20M. We have already been awarded the first-phase funding. A second-phase funded at about $5M has a high probability. The direction we go with the third phase of the upgrade, and the probability for funding, depends on the outcome of this work. The project may also provide results of value to nuclear safeguards and stockpile stewardship programs of DOE and Homeland Security.

Reference

[1]W. D. Diamond, "A Radioactive ion beam facility using photofission," *Nucl. Instrum. and Methods Phys. Res.* **A432**, 471 (1999).

Probing Explosive Nucleosynthesis Through Measurements at the Holifield Radioactive Ion Beam Facility

D. W. Bardayan, J. C. Blackmon, and M. S. Smith
Physics Division

Unstable nuclei play an influential role in many astrophysical processes occurring in exploding stars such as novae, supernovae, and X-ray bursts. To understand these events, we must understand the reactions involving unstable, radioactive nuclei. The best way to probe such reactions is by using beams like those available at the Holifield Radioactive Ion Beam Facility (HRIBF). We are performing studies of proton-induced reactions on radioactive fluorine and phosphorus nuclei and neutron-induced reactions on radioactive neutron-rich nuclei. The large detector arrays and the ion beams at the HRIBF are being used to carry out these measurements.

The understanding of the reactions on and structure of unstable nuclei is critical to understanding cataclysmic stellar explosions such as novae, supernovae, and X-ray bursts. The goal of this project is to measure the properties of and reaction rates on several important radioactive nuclei. The information gained will help address issues such as the gamma-ray emission from novae, which is important to understanding results from new space-based telescopes, and of the composition of meteorites, which sample the universe outside the solar system.

In particular, we will study the rate of the $^{18}F + p \rightarrow {}^{15}O + \alpha$ reaction by measuring the rate directly at a variety of energies and by studying the elastic scattering of $^{18}F + p$. We will also study the proton-capture rates of ^{29}P and ^{30}P. These reactions form compound nuclei ^{30}S and ^{31}S, respectively, which we will study via the $^{32}S + p \rightarrow t + {}^{30}S$ and $^{32}S + p \rightarrow d + {}^{31}S$ reactions at the HRIBF. Finally, the transfer of neutrons to heavy neutron-rich radioactive beams will also be studied. This is important because in supernovae, neutron-rich nuclei are processed to even heavier nuclei by a series of neutron captures and beta decays called the r-process. To understand this processing, we must understand the neutron-captures rates on these nuclei, and we will study these rates via measurements of neutron-transfer reactions on radioactive beams at the HRIBF.

Several requirements of the measurements make them difficult, if not impossible, to be studied at other facilities besides the HRIBF. Only at the HRIBF is a ^{18}F beam available with sufficient quality to perform the high-precision measurements necessary for astrophysics. The $^{32}S + p$ studies require high-energy proton beams coupled with large-area silicon-detector arrays. Such a combination is only available in the United States at the HRIBF. The neutron-transfer studies on r-process nuclei are only possible at the HRIBF. These studies require beams of neutron-rich nuclei accelerated to energies high enough to overcome the electrostatic repulsion of the positively charged nuclei. Such beams are only available at the HRIBF.

Much progress has been made on the preparation and implementation of these studies. The elastic-scattering rate of $^{18}F + p$ has been measured over a large energy range. Properties of important states in the compound nucleus ^{19}Ne have been deduced, and the results presented at scientific meetings. To study the ^{32}S + p reactions, special target materials are required. Elemental sulfur is unstable under vacuum conditions, and thus targets made from sulfide compounds are required. Other requirements of the targets resulted in the only suitable compounds being ZnS and FeS, of which commercially made targets are not available. Special production techniques were developed locally and suitable targets have now been produced. Finally, the neutron-transfer reaction $^{82}Ge + d \rightarrow p + {}^{83}Ge$ has been measured at the HRIBF. This is the first neutron-transfer measurement ever made on an r-process nucleus. The techniques developed highlight a roadmap to further studies at future facilities such as the Rare Isotope Accelerator (RIA), a proposed facility endorsed by the Nuclear Science Advisory Committee at the highest priority for future construction.

The DOE Office of Nuclear Physics supports research in low-energy nuclear physics with a new component involving reactions of astrophysical interest. This component is growing due to the recent availability of radioactive ion beams. This project is central to DOE's desire to continue growth in that area and to develop techniques applicable for use at future facilities such as RIA. This project also advances goals of the National Science Foundation (NSF), which supports nuclear measurements of astrophysical interest.

PHYSICS

Seed Money Fund

A Novel Device for Quantitative Single-Atom Detection of Carbon-14

F. W. Meyer
Physics Division

In this seed money project, the feasibility of developing a novel, compact apparatus for the quantitative detection of ^{14}C at the part-in 10^{12} level has been investigated. Carbon-14–labeled compounds are widely used in the pharmaceutical industry (e.g., as tracers to determine the fate of these compounds in vivo). Tracer studies of metabolic products in breath exhalations are used in the diagnosis of certain metabolic disorders. The sensitivities of most present methods are inadequate to permit utilization of sufficiently small quantities of ^{14}C to avoid the issues of radioactive waste and contamination, both of which are unacceptable for environmental, health and safety, and financial reasons. For these reasons, there is significant current interest in exploring other high-sensitivity ^{14}C detection methodologies that permit reduction of ^{14}C labeling to levels slightly above background. Conventional accelerator mass spectrometry (AMS) is at present the only approach that offers sufficiently high sensitivity to avoid the above-radiological issues, but it requires large-scale facilities that are usually not dedicated to a single task, with correspondingly high cost. The AMS technique further entails time-consuming sample preparation prior to the actual measurements and so is not suited to quasi–real time monitoring of ^{14}C levels.

Introduction

The natural abundance of ^{14}C in "modern" samples is about 1.18×10^{-12} per ^{12}C atom, and this determines the background level from which the levels of tracers used (e.g., in the pharmaceutical industry) must be distinguishable. The main difficulty in single-atom detection of ^{14}C arises from the isobaric interferences due to atomic ions (e.g., ^{14}N) and molecular ions (e.g., $^{12}CH_2$ and ^{13}CH). In conventional AMS,[1] the approach consists of using a negative ion source to eliminate the ^{14}N contamination, since it does not support a stable negative ion, accelerating the negative ion beam in a tandem accelerator to high energy (few MeV), and then dissociating molecular ions isobaric with $^{14}C^-$, also present in the ion beam, either in a foil or gas target. Subsequent stages of electrostatic and magnetic analysis are then used to isolate the ^{14}C ions before their detection. Conventional AMS requires large facilities, usually not dedicated to a single task, with correspondingly high cost, and entails time-consuming sample preparation prior to the actual measurements.

Technical Approach

For the present proof-of-principle project, the ORNL Multicharged Ion Research Facility (MIRF)[2] electron cyclotron resonance (ECR) ion source was used for the production of a multicharged carbon beam with charge state of +3 or higher to eliminate molecular isobar interference at mass 14. After magnetic selection of the desired charge state, the multicharged ion beam, which will still be dominated by ^{14}N multicharged ions of the same charge state, is directed at grazing incidence on a LiF single-crystal surface, where efficient negative ion formation takes place without appreciable energy loss of the scattered beams. The different scattered charge states are dispersed in the first stage of a large-acceptance-angle, two-stage electrostatic analysis system. The second, high-resolution, stage further spatially separated the desired $^{14}C^-$ ions from other scattered charge states prior to their detection on a two-dimensional, position-sensitive detector (2-D PSD). The analysis system transmitted more than 50% of the negatively charged projectiles produced during the grazing scattering interactions with the LiF surface. Since ^{14}N does not form a stable negative ion, interference due to ^{14}N could be effectively eliminated. Unique characteristics of the apparatus are its small size, low cost, high efficiency (i.e., throughput), and ease of sample preparation, in comparison to conventional AMS hardware. As a result, this apparatus[3] should find great utility in such applications as quasi–real time monitoring of ^{14}C-based chemical tracer uptake in biological systems, atmospheric pollution studies, cancer research, medical diagnostics, and other biomedical studies.

Results and Accomplishments

During FY 2001, a careful analysis of the mass rejection capability of the ORNL MIRF magnetic analysis system was performed. It was verified that adjacent mass

rejection is more than sufficient to meet the stringent requirements of this project. In addition, the 2D-PSD used in the final stage of ^{14}C detection (i.e., after the negative ion formation by surface scattering) was extensively refurbished and prepared for installation into the surface scattering chamber.

During FY 2002 the detailed design and fabrication of the ^{14}C detection apparatus, shown schematically in Fig. 1, were carried out. The work performed included design and fabrication of a large-area LiF sample mount that was heatable to 400°C. The sample holder, shown in Fig. 2, is supported by a 3-axis goniometer with rotation to permit accurate orientation relative to the incident multicharged ion beam. In addition, two large-acceptance, spherical-sector electrostatic deflectors were designed and fabricated. The first, a 45° analyzer, served to separate the negatively charged ^{14}C ions produced by the grazing interaction with the LiF target from the dominant scattered neutral ^{14}N flux. The second, a higher resolution 90° analyzer, served to energy-analyze the desired negatively charged ^{14}C ions prior to detection for further signal-to-noise enhancement. Both analyzers are double-focusing and thus permit maximum collection rate of the desired ions. To house these analyzers and detector, suitable vacuum chambers were designed and fabricated as well.

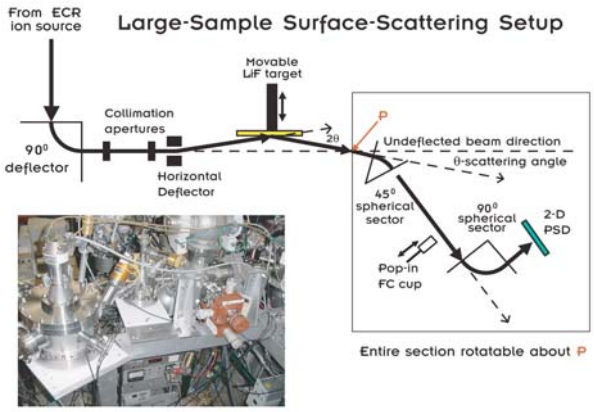


Fig. 1. Schematic of the ^{14}C detection apparatus developed for this project. The scheme entails grazingly scattering up to 3-mm-diam, μA-intensity ion beams from a large single-crystal sample, and a two-stage, large-acceptance-angle, double-focusing analysis for separation of $^{14}C^-$ ions produced during the grazing surface interactions from the more intense scattering products.

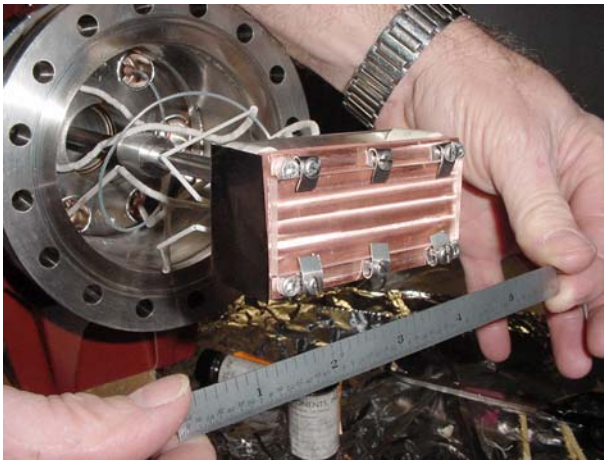


Fig. 2. Large sample mount showing 69-mm × 29-mm × 3-mm cleaved LiF target mounted on a heated sample stage capable of being heated to 400°C.

During FY 2003 the apparatus was assembled and tested. Beam currents of 0.2–0.5 pμA of $^{12}C^{q+}$ (q = 2–4) were achieved on the large LiF sample at incidence angles as low as a few degrees. The grazing scattering interactions produced up to 20 pnA of $^{12}C^-$. The observed C^- yield was found to be significantly lower than expected on the basis of our earlier O^- production measurements[3] but still corresponded to a particle flux of about 1.2×10^{12}/s. From the natural abundance of ^{14}C quoted earlier, this translates to a $^{14}C^-$ flux of about 0.2 Hz. However, background events were found to dominate the $^{14}C^-$ count rate by about two orders of magnitude. The smaller-than-expected C^- yield was found to be due to a subtle atomic physics effect having to do with differences in the statistical weights of the neutral parent states from which C^- and O^- can be formed.[4] Choice of an alkali halide target with a higher-lying valence band is expected to increase the yield at least a factor of two. The source of the above noted background was found to be F^- sputtered particles. Due to the close mass proximity of mass-14 projectiles and the mass-19 F target constituents, the maximum energy of the sputtered F^- particles E_{sp} is very close to the primary beam energy E_0 ($E_{sp} = 0.977\ E_0$). As a result the F^- particles could not be effectively discriminated against in the second electrostatic analyzer and thus resulted in the observed background,[5] as shown in Fig. 3. By using a heavier alkali halide target, such as RbI, both the low-negative-ion yield and sputtered-particle background problems can be eliminated. The top of the valence band of RbI lies almost 6 eV higher in energy than LiF. Also, with its much higher halogen mass (iodine is mass 127), the maximum energy of sputtered I^- particles will be only about one-third of the primary beam energy, which can thus be effectively discriminated against in the second analyzer. As illustrated in Fig. 4, already for mass-40 incident projectiles, for which the maximum sputtered F energy is reduced to about 88% of the primary beam energy, the mass-14 backgrounds were found to be reduced by more than two orders of magnitude to levels less than 1 Hz.

Summary and Conclusions

In this project, proof-of-principle of a low-energy, compact scheme for quantitative ^{14}C detection has been successfully demonstrated. The next stage of development

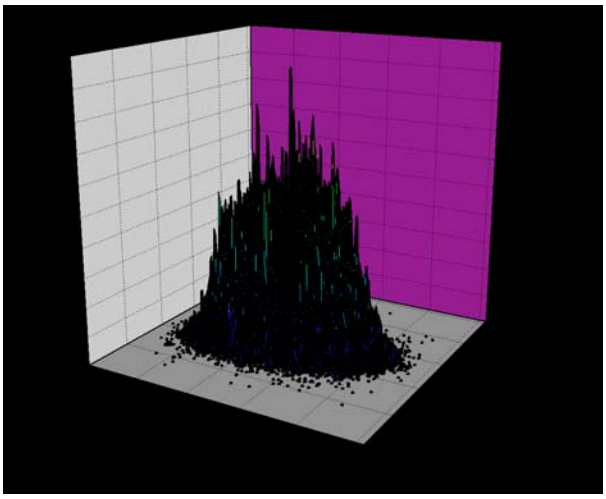


Fig. 3. Intensity distribution on the 2-D PSD when 27-keV triply charged mass-14 ions of μA intensity are incident on the LiF target; the high background is due to full-energy F^- sputtered particles indistinguishable from $^{14}C^-$ ions of the same energy produced during the grazing LiF surface scattering.

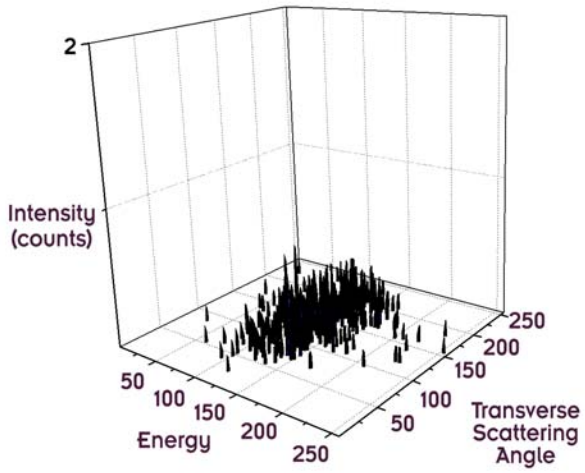


Fig. 4. Intensity distribution on the 2-D PSD when 27-keV mass-40 ions are incident on the LiF target; note greatly reduced background due to absence of full energy F^- sputtered particles.

can therefore now be envisioned. Together with implementation of a heavier alkali halide target such as RbI to increase the negative-ion yield and to greatly reduce background levels, the use of a compact all-permanent magnet ECR source could result in a truly table-top-size system. The addition of front-end delivery systems to handle real-time introduction of liquids and combustion products into the ECR source would make this device of great utility in the pharmaceutical industry or as a biomedical research tool. Accelerator mass spectrometry (AMS) is the only presently available method sufficiently sensitive to permit utilization of sufficiently small quantities of ^{14}C to avoid the issues of radioactive waste and contamination, both of which are unacceptable for environmental, health, and safety and financial reasons. AMS facilities, however, are expensive, not portable, and do not have real-time processing capability. For these reasons, the pharmaceutical industry is currently exploring other high-sensitivity ^{14}C detection methodologies that permit reduction of ^{14}C labeling to levels slightly above background.

A proposal for further development of the approach demonstrated in this project will be prepared for submission to the DOE Office of Biological and Environmental Sciences or NIH. Subsequently or alternatively, an industrial partner could be sought for commercialization. The apparatus developed in this manner should find broad application in biomedical research using ^{14}C tracers, and, with some sensitivity enhancement, in tracer studies of atmospheric chemistry and transport, ocean mixing, erosional processes and glacial recession, diffusion through soils, diesel exhaust pollution, lubricant consumption and degradation, wear analyses of graphite composite materials, and various petroleum industry problems.[6] One local application of the device could be in the ecosystem-scale ^{14}C pulse labeling experiment on carbon sequestration in soil,[7] made possible by the 1999 ^{14}C release from one of the Oak Ridge Reservation incinerators. If sufficient improvement in sensitivity over the initial design goal (10^{-12}) is achieved, then additional possibilities might include radiocarbon dating applications in the areas of paleoclimatology and archaeology.

References

[1]D. Elmore and F. M. Philips, *Science* **236**, 543 (1987).

[2]F. W. Meyer, "ECR-based Atomic Collision Physics Research at ORNL MIRF," in *Trapping Highly Charged Ions: Fundamentals and Applications,* J. Gillaspy, ed., Nova Science Pub, New York, 2000, pp. 117–164.

[3]U.S. patent 6,455,844, issued September 24, 2002.

[4]F. W. Meyer, Q. Yan, P. Zeijlmans van Emmichoven, I. G. Hughes, and G. Spierings, *NIM* **B125**, 138 (1997).

[5]F. W. Meyer, "A Novel Device for Quantitative Single-Atom Detection of Carbon-14," Book of Abstracts, XXIII ICPEAC, Stockholm, Sweden, July 23–29, 2003.

[6]J. C. Davis, *NIM* **B*92***, 1 (1994).

[7]"Carbon-14 levels discovered on ORR provide a unique scientific opportunity," *ORNL Reporter*, Issue 44, p. 1, December 2002.

A Miniature Electrostatic Molecular-Ion Storage Ring

M. E. Bannister and D. R. Schultz
Physics Division

Cold molecular ions are needed to experimentally determine key collision rates required for fundamental physical and chemical studies, fusion energy research, aeronomy, astrophysics, lighting, and plasma processing, for example. The objective of this project has been to produce proof of concept for a novel, compact source of molecular ions which can very cost-effectively produce rotationally/vibrationally cold molecular ions necessary for research underpinning these diverse applications. In brief, molecular ions were formed in a surface-wave plasma source where their internal temperature is lowered through adiabatic cooling in a supersonic expansion. The feasibility of injecting these ions into a small, cryogenically cooled electrostatic storage ring was studied. Storage for milliseconds in the cooled ring would allow radiative relaxation of the ions' excited states, further cooling them to the temperatures required for fundamental and applied investigations. Alternative technologies rely on large magnetic or electrostatic storage ring facilities that cost hundreds of thousands to millions of dollars and are not presently available in the United States.

Introduction

European nations and Japan invested significant funds in heavy-ion storage rings with electron cooling in the last two decades while the United States did not, forcing U.S. researchers seeking to perform atomic and molecular physics experiments enabled by these devices to request and compete for time abroad. In addition to experiments exploring electron-atomic-ion collisions, the rings have made possible study of collisions of electrons with cool molecular ions. Cooling of the internal vibrational energy of the molecular ion proceeds by injecting ions from an ion source and allowing them to circulate in the ring long enough to radiate away their internal energy. Studying such collisions is central to understanding processes in a wide range of environments including fusion reactors, astrophysical and aeronomical plasmas, lighting, and technical plasma processing devices.

In particular, we focused on demonstration of proof of concept for a cool molecular ion source coupled to a miniature electrostatic storage ring that is extremely compact and inexpensive compared to existing rings. This device could bring a significant capability to ORNL and to the U.S. research community, opening opportunities for key research now only possible using large, foreign storage-ring facilities. The development steps included the design and testing of a novel cool ion source, simulation of the storage ring operation, assembly of the small electrostatic ring, and testing of the output performance (e.g., ion temperature, beam intensity, collimation) of the device via spectroscopic measurements.

Technical Approach

A key component of this project is the source that produces the cool molecular ions to be stored in the miniature electrostatic storage ring. The molecular ions are formed in a surface-wave-sustained discharge that is expanded adiabatically into vacuum through a supersonic nozzle. Microwaves propagated from a waveguide launching structure called a surfatron supply the energy necessary to maintain the plasma column. These surface waves travel down the discharge tube, which is effectively a plasma waveguide, and, if sufficient microwave power is applied, reach the supersonic nozzle at the end of the tube. For given flow conditions in the expansion, determined by the nozzle diameter and the gas pressure in the tube, the plasma parameters such as electron density and temperature can be controlled in the nozzle and expansion region by adjusting the microwave power. The microwave circuit of the surfatron launcher has three separate adjustments to optimize the coupling of the microwaves to the plasma column so that the reflected power from the surfatron is typically less than 5% under operating conditions.

The supersonic nozzle was formed at the end of the 4.0-mm-I.D. quartz tube by heating the end of an open tube until it drew closed on itself. This closed end was then ground away until a small opening, less than 100 μm in diameter, was made. During source operation, the nozzle end of the discharge tube was cooled by distilled water sent through a coaxial cooling jacket by a refrigerated circulator. Additional cooling of the surfatron and discharge tube were provided by a combination of chilled water and forced air.

Extraction of the molecular ions from the supersonic expansion region was accomplished by an small metal annulus located around the nozzle end of the discharge tube. Application of a voltage of up to a few kilovolts on this annulus forced the ions away from the nozzle where they were focused into a beam by a cylindrical lens. Using additional cylindrical lenses and deflectors, the beam was transported more than 1 m to either a Faraday cup used for total current measurements or to a channel electron multiplier used for time-of-flight (TOF) beam diagnostics. Short bunches of ions used in the TOF measurements were produced by rapidly pulsing (pulse width ~300 ns) the voltage on one of the ion deflectors to allow ions to pass only during a narrow window of time.

Results and Accomplishments

We summarize here some typical results from the characterization of the surfajet ion source. The dependence of the plasma parameters in the nozzle expansion region on the absorbed microwave power was evidenced by measurements of the total ion current as a function of power. Figure 1 shows this dependence for a nitrogen discharge with a pressure of 1.2 Torr expanding through an 84-μm-diam nozzle. Until the absorbed power reaches approximately 810 W, essentially no ions survive into the expansion. As the power increases above that threshold level, the ion current extracted at 1 kV increases linearly. Emission of visible light from the discharge extends all the way to the nozzle for power levels above this critical value. These two observations strongly suggest that in order for ions to exist in the expansion region, sufficient microwave energy must propagate past the nozzle end of the discharge to continue ionizing the gas, because ions formed before the nozzle recombine due to collisions in the nozzle.

Time-of-flight spectra were recorded for various levels of power above the threshold for the discharge conditions described above. Two such spectra are shown in Fig. 2 with power levels corresponding to points labeled A and B on Fig. 1. For an absorbed power just above the threshold value for ion production, the TOF spectrum (dashed curve in Fig. 2) shows only molecular ions, namely N_2^+ and impurity H_2O^+. At the higher power level, the TOF spectrum (solid curve in Fig. 2) also shows the presence of atomic ions, N^+, comprising approximately 12% of the extracted beam. Despite the obvious change in plasma parameters in the expansion region, no difference was seen in the width of the molecular ion peak, which is related to the translational temperature of the beam.

The miniature electrostatic molecular-ion storage ring was simulated in three dimensions using the computer program SIMION.[1] The base design was a scaled-down version of an electrostatic ion storage ring operating in Japan.[2] Results from these simulations indicated that building such a miniature ring with a circumference of 1.4 m was feasible provided that electrodes were machined and assembled with sufficient accuracy. It was determined that the addition of cryogenic cooling to the elements of the ring, however, would greatly increase the complexity of the construction and operation.

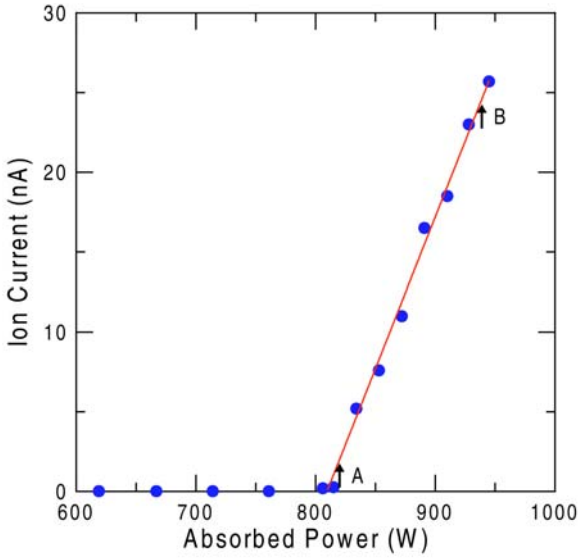


Fig. 1. Total current extracted at 1 kV from the surfajet molecular ion source as a function of the microwave power absorbed in a surface-wave-sustained nitrogen plasma. The gas pressure was 1.2 Torr and the supersonic nozzle had a diameter of 84 μm.

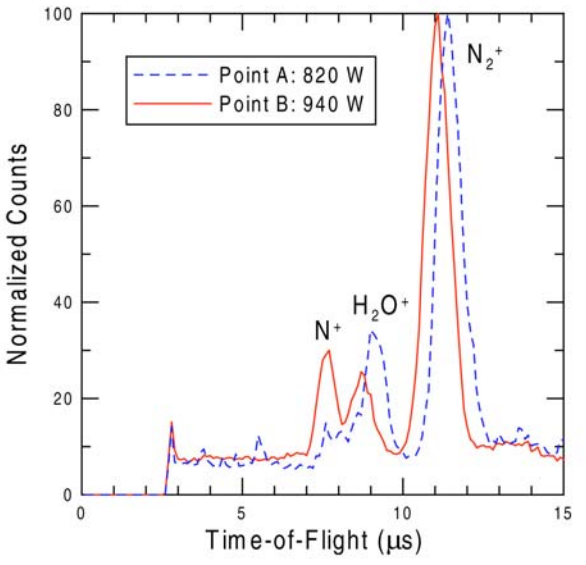


Fig. 2. Time-of-flight spectra for ion beams extracted at 1 kV from the surfajet molecular ion source. The dashed curve was measured for 820 W absorbed power (labeled point A on Fig. 1); the solid curve was for 940 W absorbed power (point B on Fig. 1).

Summary and Conclusions

The development of the surfajet source for cool molecular ions yielded beams sufficient for injection into a miniature electrostatic ion storage ring or for separate collision experiments (greater than 100 nA at 1-keV energy). Time-of-flight diagnostics on the extracted beams were inconclusive about the degree of cooling of the ions by the supersonic expansion; follow-on spectroscopic measurements will be required to quantify the internal cooling of the extracted ions. Simulations of the miniature electrostatic ion storage ring showed the feasibility of such a device, but the time and funding required turned out to be beyond the scope afforded by this project.

Once the internal cooling of the molecular ions is fully characterized, the surfajet source is quite suitable for collision experiments with electrons, atoms, and surfaces. It can be adapted for use on the high-voltage ion source platform under construction at the Multicharged Ion Research Facility in the ORNL Physics Division, making a variety of cool, well-characterized molecular ions available at energies up to 250 keV for experiments such as measurements of dissociative recombination, excitation, and ionization by electron impact.

This project on development of a cold molecular-ion source has already led to additional funding from the DOE Office of Basic Energy Sciences (OBES) at a level of $200K/year for the study of electron-impact dissociation of molecular ions. This OBES-funded experimental program includes support for continued research on characterization and optimization of the surfajet as a source of cold molecular ions. A well-characterized source will also be advantageous for pursuit of future funding from NASA for experiments on cold molecular ion species of astrophysical significance.

References

[1]D. A. Dahl, *SIMION 3D Version 7.0*, INEEL-95/0403, Idaho National Engineering and Environmental Laboratory, Idaho Falls, Idaho, 2000.

[2]T. Tanabe, K. Chida, K. Noda, and I. Watanabe, "An Electrostatic Storage Ring for Molecular Science," pp.797–799 in *Proceedings of the Seventh European Particle Accelerator Conference*, ed. M. Regler, 2000.

Ultrafast-Laser-Produced Radioactive Ions

C. R. Vane,[1] Y. Liu,[1] H. F. Krause,[1] J. R. Beene,[1] D. R. Schultz,[1] K. Flippo,[2]
A. Maksimchuk,[2] and D. Umstadter[2]
[1]*Physics Division, ORNL*
[2]*Center for Ultrafast Optical Science, University of Michigan, Ann Arbor, Michigan*

Significant production of radioactive isotopes that can be used in fundamental research or for medical purposes has been demonstrated through novel experiments in which the isotopes were created by nuclear reactions driven by light particles accelerated in ultrafast/ultra-intense laser irradiation of solid targets. Measurements were carried out to explore optimization of yields of several radioactive species by choice of laser parameters, as well as driver target materials and geometries. These experiments were accomplished and yield results were made absolute using detailed diagnostics and analyses based on nuclear spectroscopy of the produced radionuclides.

Introduction

Short-lived radioisotopes and ions (that is, ions with unstable nuclei) are of intrinsic interest in nuclear physics, of great importance in astrophysics, and of great practical use, particularly in medical applications. Consequently, research centers such as ORNL's Holifield Radioactive Ion Beam Facility (HRIBF) are dedicated to producing these exotic species and exploring in detail their properties. In addition, hospitals and medical research facilities generate short-lived radiotherapeutic and diagnostic isotopes by using commercial cyclotrons or by collecting radioactive samples generated at nuclear reactors. The focus of the work initiated here was aimed at exploring and potentially creating a new source for radioisotopes and ion beams (RIBs) based on laser acceleration of light particles. Such sources may have many significant advantages over conventional methods for producing these isotopes.

The recent advent of ultrafast (femtosecond pulses), ultra-intense (multiterawatt or higher) lasers has unexpectedly lead to the observation of highly accelerated, well-collimated electron and proton beams emanating from solid, gas, and cluster targets subjected to high-power laser pulses. The ultra-intense laser pulses rapidly accelerate electrons within the target, creating charge separation and enormous transient electric and magnetic fields. The electric fields accelerate protons or deuterons occupying hydrogen-bearing surface layers on the target. These discoveries have opened the possibility of creating compact, efficient sources of radioisotopes using the laser-accelerated lighter particles as drivers of nuclear reactions in isotope production targets.[1] The very rapid advances being made in capabilities of ultrafast/ultra-intense lasers also greatly leverage development and future performance of any ultrafast-laser-produced radioisotope source.

Technical Approach

Our approach was to develop the technology necessary for proof-of-concept production of several short-lived radionuclides through the use of intense (10-TW), ultrafast (400-fs) pulses of laser light focused on solid targets. Protons and deuterons accelerated to high energies by interaction of the laser pulses focused on these primary driver targets were used to bombard secondary targets and produce radioactive isotopes.

Simulation studies were performed at ORNL for optimized production of several potential radioisotopes from laser-driven high-energy proton and deuteron impact on elemental targets, as well as for extraction of the fraction of radioactive ions transmitted during recoil out of thin targets. Calculations were also performed to give expected yields of the radioactive ions that could be slowed in a He buffer gas and then extracted using electric fields. Expertise at the HRIBF was used to establish reliable calculations of ion beam extraction from high-energy impact on isotope targets, using methods employed for cyclotron-accelerated protons. Following these simulations, an apparatus was designed and constructed and three experiments were carried out at the Center for Ultrafast Optical Science (CUOS), University of Michigan at Ann Arbor. These experiments were performed in collaboration with Professor Donald Umstadter and his colleagues at the University of Michigan, especially Kirk Flippo, whose Ph.D. thesis (2003) was based partly on work in this project. The measurements were all performed using 1.053-nm, 10-TW, 400-fs laser pulses from a hybrid

Ti:sapphire/Nd:phosphate glass chirped pulse amplification laser to accelerate protons and deuterons out of thin-foil driver targets and onto isotope production targets. The laser was focused before each shot (~7 min between shots) to approximately 6-μm spot diameter. The corresponding maximum focused intensity was ~4×10^{18} W/cm^2.

All exposures were performed in vacuum. The isotope production targets were either counted in place or extracted from the vacuum chamber and monitored for nuclear-decay gamma-ray emission using scintillation detectors. Measured yields of specific gamma rays were then used to calculate production yields of the identified isotopes. Employing appropriate nuclear cross sections, estimates were also extracted for the initial fluences of the respective accelerated particles (*p* or *d*) driving the nuclear reactions. With these in hand, one can estimate production yields for a much wider range of interesting radionuclides.

Protons and deuterons were accelerated to MeV energies by laser pulses focused on the front surfaces of thin mylar, aluminum, copper, and titanium foils. The basic setup is shown in Figs. 1 and 2. The deuteron surface coatings were produced either by laying down a layer of deuterated polystyrene in solution on mylar or aluminum foils, or by covering aluminum foils with a thin layer of sublimated titanium and then backfilling with deuterium gas to produce a few monolayers of TiD. Foil thicknesses typically ranged from 1 to 50 μm. The laser-ejected protons and deuterons from these thin foils were directed onto thick isotope production targets of enriched ^{11}B, or naturally occurring ^{12}C or ^{63}Cu to produce radioisotopes ^{11}C, ^{13}N, or

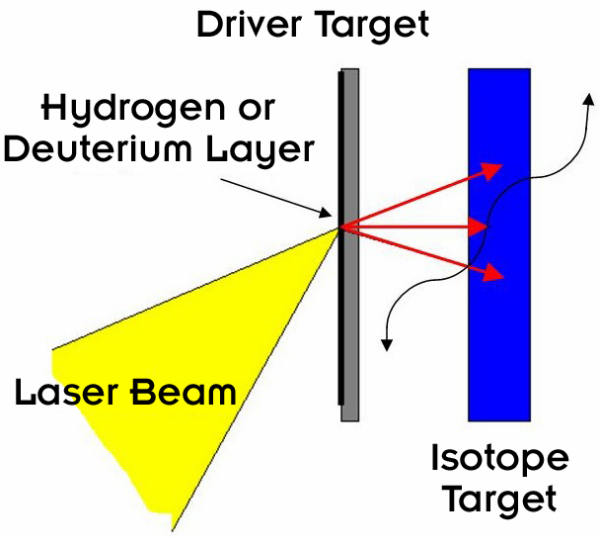


Fig. 1. Schematic of the two-step process for radionuclide production using ultrafast, ultra-intense lasers. The laser is focused on a thin driver foil where protons or deuterons are accelerated to high energy and emitted from the opposite side normal to the primary target. The fast particles are directed to a second isotope production target chosen to yield the selected radioisotope.

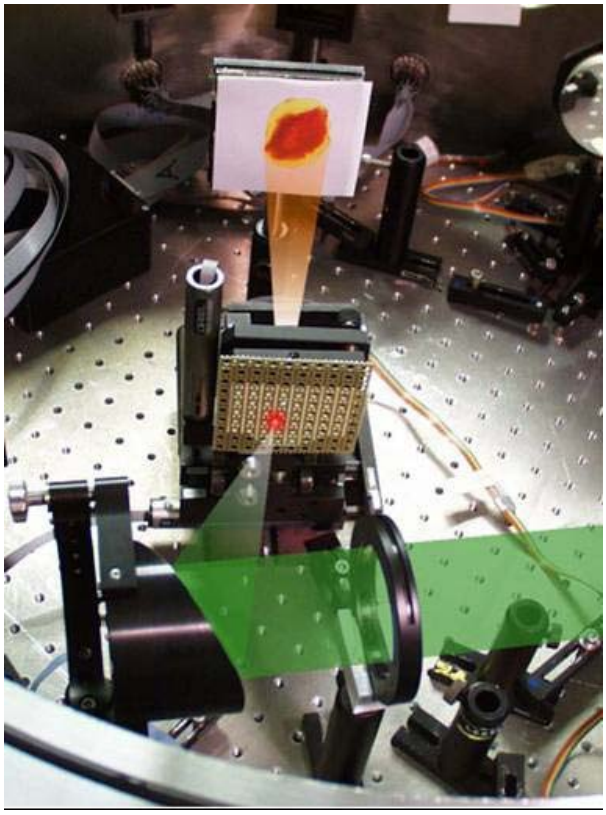


Fig. 2. Collimated beams of fast protons, observed from light emitted by passage through the vacuum chamber rest gas. The protons, which originate either from hydrogen-bearing impurities or from implanted deuterium on the front side of the target, are accelerated and exit the back side in a direction normal to the target surface. Electric field gradients ~10 GV/cm are inferred from measured maximum proton energies (up to ~5 MeV), which can be explained by charge separation due to "vacuum heating."

^{63}Zn, respectively. Gammas rays emitted from the decaying radioisotopes after a laser shot were initially counted in-place inside the vacuum chamber by NaI scintillations detectors located there, but the exceptionally high electromagnetic noise echoing after the shots and activation of the NaI materials made this method untenable. A target transport apparatus was constructed to permit fast extraction and retrieval of the isotope targets, and gamma-ray counting was carried out using Ge(Li) or NaI detectors mounted in a relatively noise-free environment outside the chamber after each laser pulse.

To optimize target material thickness for production of fast protons, tests were performed using a Thomson combined electric and magnetic field spectrometer[2] constructed by Kirk Flippo at the University of Michigan to roughly measure the accelerated heavy-particle energy distributions. For the conditions of this laser system, the proton yield-optimized thickness was found to be approximately 12.5 μm for aluminum foils. These thicknesses were subsequently used in measurements of

radioisotope production from a variety of driver target materials (mylar, Al, Ti, and Cu), again to examine effects on radioisotope production through (*p*,*n*) and (*d*,*n*) reactions.

Several previous researchers examining pulsed laser heavy-ion acceleration have shown that protons on the driver-target surface inhibit production of high-energy particles of any other species.[3] To reduce the hydrogen-bearing material contamination on thin metal targets, some of the foils were heated prior to laser exposure. In an attempt to observe these effects, we performed $^{12}C(d,n)^{13}N$ reaction yield measurements using 12.5-µm-thick aluminum foils coated with a deuterium-loaded titanium layer as a function of driver-target temperature. The driver foils were tested at temperatures from 20 to 210°C using a resistive heater attached to a copper mounting frame. The thick carbon isotope target used was maintained at 20°C.

Results and Accomplishments

Results of these studies to optimize high-energy particle acceleration and subsequent radioisotope production will be reported in a number of forthcoming publications,[4–6] For all the isotopes examined here, decays proceed primarily by positron emission. Subsequent annihilation of the beta leads to emission of two 511-keV gammas from the targets. Coincident counting of the 511-keV gammas was performed for each target over several lifetimes of the respective radionuclide. Time-dependent yields of singles and coincident 511-keV gammas were corrected for counting efficiencies and converted to initial numbers of the decaying parent nuclei. For the (*p*,*n*) reaction-generated nuclei and targets at room temperature, the best measurements gave approximately 10^5 radioactive ^{11}C and ^{63}Zn atoms produced per laser shot via $^{11}B(p,n)^{11}C$ and $^{63}Cu(p,n)^{63}Zn$ reactions, respectively. Identifications of specific isotopes were confirmed by decay rates of the gamma-ray singles rates. Figures 3 and 4 show examples of measured time dependences of the gamma-ray count rates for these isotopes.

For the (*d*,*n*) reaction–generated ^{13}N nuclei, where the target material (aluminum) and thickness (12.7 µm) were optimized, approximately 4×10^5 radioactive ^{13}N were produced per laser shot via $^{12}C(d,n)^{13}N$. We concentrated somewhat on improving production of these isotopes because ^{13}N is of special interest for nuclear astrophysics studies of hot CNO fusion cycle stars. We note that in contradiction to prior expectations, no enhancement in ^{13}N production was observed for aluminum driver targets heated to 210°C, although this temperature may not have been high enough to effectively remove all hydrogen contamination in the poor vacuum conditions experienced in the target chamber(~10^{-5} Torr).

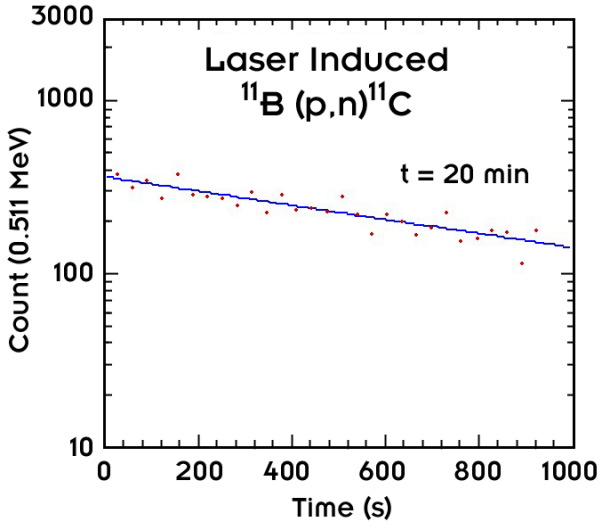


Fig. 3. Measured 511-keV gamma counts vs time after laser shot for decay of ^{11}C from $^{11}B(p,n)^{11}C$.

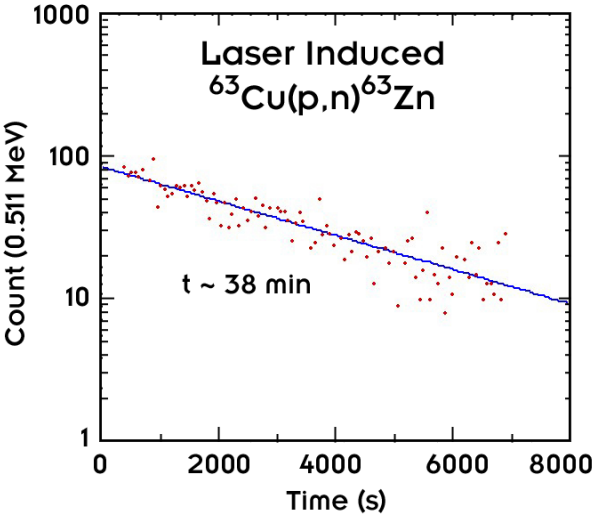


Fig. 4. Measured 511-keV gamma counts vs time after laser shot for decay of ^{63}Zn from $^{63}Cu(p,n)^{63}Zn$.

The ~10^5 ^{11}C, ^{13}N, and ^{63}Zn produced per laser shot represent a significant yield of radioactive nuclei, especially for multi-hertz laser systems. Assuming 10 pulses per second, as planned for a new laser system presently installed and undergoing initial testing at CUOS, the yield will conservatively be ~10^6/s, and result in equilibrium production of ~10^9 ^{13}N in a few tens of minutes. Similar yields of other important isotopes will also be possible.

From the measured isotope yields and known (*p*,*n*) and (*d*,*n*) cross sections, one can extract approximate incident intensities of the accelerated protons and deuterons. For example, $^{12}C(d,n)^{13}N$ is especially simple, since the cross section is essentially constant with deuteron

energy above about 0.8 MeV at 150 mb, as shown in Fig. 5. Assuming a deuteron energy distribution inferred from measurements for protons (actually, essentially half the proton end-point maximum energy), and correcting for energy loss in the isotope production ^{12}C target, we estimate that approximately 10^9 deuterons above 0.8 MeV were generated per laser pulse. That corresponds to a particle current of ~$10^{20–21}$/sec emitted from a ~10-μm spot during the few psec duration of particle emission after the laser pulse!

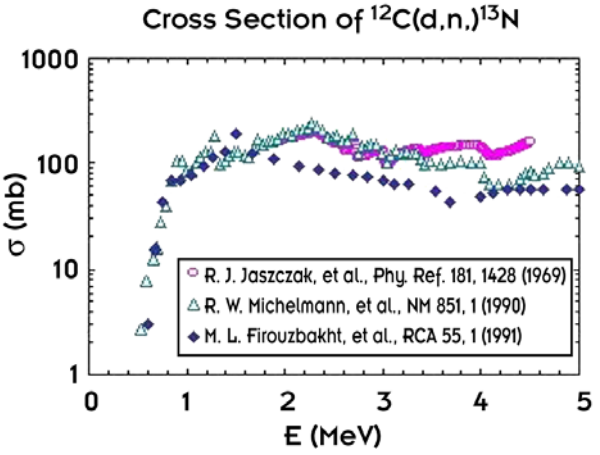


Fig. 5. Measured cross sections for deuteron impact $^{12}C(d,n)^{13}N$ reaction.

Summary and Conclusions

As demonstrated in this project, tabletop-sized, high-intensity, ultrafast pulse lasers have the potential for successful applications as sources of radioisotopes and, potentially, as high-quality ion beams of these isotopes. Higher laser power results in higher yields of accelerated particles and higher maximum energies, which will also make thicker isotope production targets and radionuclide yields possible. Similar experiments and simulations at larger, higher-power facilities have indicated that useful GBq quantities of short-lived medical isotopes (e.g., ^{11}C and ^{19}F, important for positron-emission tomography, for example) can be produced using 1-kHz lasers, perhaps available as tabletop systems only a generation or so away.[7]

The experience and results gained here have had significant influence on plans for future applications of laser-based radioactive ion beams. A new Field Work Proposal (ERKBL01 "Research and Development in Support of RIA") has been submitted to DOE Office of Nuclear Physics, based on research toward development of laser ion sources for radioactive ion beams (RIBs). Research outlined in this proposal concentrates on using laser resonance ionization properties for selectivity in extraction of RIBs produced by proton impact from high-energy accelerators. However, future work may also address laser pulse influence in direct production and preferential magnetic selection and acceleration of selected species.

More work is also planned to explore the feasibility of using high-quality, multi-MeV electrons generated by intense laser pulses on solid and gas targets for production of bremsstrahlung photons. Photonuclear reactions initiated by these high-energy photons have been studied rigorously for bremsstrahlung produced by electrons from accelerators. Photofission reactions of heavy nuclei are the only known mechanisms for producing neutron-rich isotopes, which are of primary interest to HRIBF researchers.

References

[1]K. Nemoto, A. Maksimchuk, S. Banerjee, K. Flippo, G. Mourou, D. Umstadter, and V. Y. Bychenkov, "Laser-triggered ion acceleration and table top isotope production," *Applied Physics Letters* **78**, 595 (2001).

[2]M. J. Rhee, "Compact Thomson Spectrometer," *Rev. Sci. Instrum.* **55**(8), 1229–1234 (1984).

[3]M. Hegelich et al., "MeV Ion Jets from Short-Pulse-Laser Interaction with Thin Foils," *Phys. Rev. Lett.* **89**, 085002 (2002).

[4]A. Maksimchuk, K. Flippo, H. Krause, G. Mourou, K. Nemoto, D. Shultz, D. Umstadter, R. Vane, V. Y. Bychenkov, G. I. Dudnikova, V. F. Kovalev, K. Mima, V. N. Novikov, Y. Sentoku, and S. V. Tolokonnikov, "High-Energy Ion Generation by Short Laser Pulses," *Plasma Phys. Rep*. (submitted).

[5]K. Flippo, T. Lin, V. Wong, M. Rever, K. Nash, A. Maksimchuk, and D. Umstadter, "Dependence of High-Intensity Laser Produced Proton Beams on Target Material Properties," *Phys. Rev. Lett.* (submitted).

[6]K. Flippo, A. Maksimchuk, T. Lin, M. Rever, and D. Umstadter, "Detailed Study of Deuteron and Heavier Ion Species Acceleration from Front and Rear Surfaces of High-Intensity Thin-Film Target Interactions*," Phys. Rev. Lett* (to be submitted).

[7]S. Fritzler, et al., "Proton Beams Generated with High-Intensity Lasers: Applications to Medical Isotope Production," *App. Phys. Lett.* **83**, 3039 (2003).

Density Matrix Renormalization Group for Interacting Fermions

T. Papenbrock, T. Barnes, D. J. Dean, and M. R. Strayer
Physics Division

Within this project we developed a numerical program (Fortran 90 code) based on the density matrix renormalization group (DMRG) to compute the low-lying states of spin chains and quantum dots with high accuracy. As an application, we investigated the critical behavior of the alternating spin-1/2 antiferromagnetic Heisenberg chain and computed the critical exponents of the ground-state energy defect and the excitation gap as the uniform chain is approached. We also applied the DMRG for electronic structure calculations in quantum dots and gave a proof-of-principle of this method.

Introduction

Fermions are the relevant degrees of freedom in physical systems ranging from macroscopic materials (metals and semiconductor devices) through mesoscopic and nanoscale electronic structures down to subatomic systems such as nuclei and hadrons. Understanding these dynamical multifermion systems is greatly complicated by the presence of interparticle interactions, which induce strong correlations in the many-body wave function of the system. As a result, simple approaches, such as mean field theory, fail to give an accurate description of many physical properties. In quantum dots, for example, it is now understood that fluctuations in conductance peak spacings cannot be understood in a mean field description. Another interesting example is given by quantum spin chains that exhibit critical behavior. At a critical point, the correlation length diverges and the theoretical description requires the understanding of many degrees of freedom.

Ab initio calculations of the properties of such systems are usually restricted to a rather small number of active fermions, since the exact diagonalization have computer memory and CPU time requirements that grow exponentially with system size. This exponential growth implies that the size of the system that can be treated using these traditional methods increases rather slowly with improvements in computer hardware. Although the use of Monte Carlo techniques allows the approximate solution of much larger Bose problems than can be treated using exact methods, when applied to dynamical fermion systems the Hamiltonian Monte Carlo approaches encounter a "minus-sign problem" that quickly degrades the accuracy of the results as the system size is increased. Several different approaches to overcome the minus-sign problem can be implemented, but none is robust across differing fermionic systems. There is clearly a great need for an improved numerical method for the study of large dynamical fermion problems in many fields of science.

Technical Approach

The DMRG is a promising alternative to conventional numerical methods in the study of large dynamical fermion problems.[1] This method scales favorably with system size when compared to exact diagonalization and has already been applied quite successfully to one-dimensional systems in condensed matter.[2] More recently the DMRG was also applied to problems in quantum chemistry and nuclear physics. The DMRG is based on a powerful truncation scheme that never lets the size of the problem get out of hand. Within DMRG, the model space is grown to very large sizes by the sequentially adding of more and more single-particle orbitals/lattice sites to the problem. Only the most important many-body states are kept, while less important states/correlations are truncated. The implementation of this algorithm is rather straightforward for spin chains and similarly homogeneous systems. For interacting many-body systems like quantum dots, questions concerning the optimal order of orbitals and partitioning of the model space arise. These issues influence the convergence of the method.

In our work we mainly followed the literature for DMRG application on spin chains.[1,2] However, we found that the convergence of the method improves notably when dimers (i.e., two spin-1/2 lattice sites) instead of individual lattice sites are used as elementary blocks. This procedure reduces the need of sweeps with the finite algorithm in the case of the spin-0 ground state of the alternating spin-1/2 antiferromagnetic Heisenberg chain. For the case of quantum dots, we restricted ourselves to two-dimensional spherical dots with harmonic confinement and transferred the ideas of Refs. 3 and 4 to our situation. In particular, we keep angular momentum as a good quantum number and use the Wilson renormalization group approach for the initial warm-up sweep. Orbitals are included corresponding to their distance from the Fermi surface. Particular care has to be taken to correctly implement the

Pauli principle. We developed both serial and MPI-based parallel codes for the DMRG algorithm. Standard library routines (BLAS, LAPACK, ARPACK) were used whenever possible to achieve both performance and portability.

Results and Accomplishments

We have developed a Fortran 90 program that solves a variety of interacting fermion problems using the DMRG algorithm. The main code was designed for single-processor machines, and a parallelized version has been tested and used on the high-performance computers of the Center for Computational Sciences (CCS) at ORNL. The program is portable and has been tested and used on a variety of computers (Compaq, IBM, Intel).

Spin Chains

We used our DMRG programs to study the critical behavior of the alternating spin-1/2 antiferromagnetic Heisenberg chain. This problem is of scientific interest since theoretical predictions for the critical exponents involve uncontrolled but widely accepted approximations that have to be checked numerically; it is also a numerical challenge due to the divergence of the correlation length and the necessary control of finite size effects. Considering increasingly long chains and decreasingly small alternation, we computed the critical exponents for the ground-state energy defect and the gap (see Fig. 1). Our numerical results are consistent with a pure power-law behavior. They agree with the theoretical predictions of power-law plus logarithmic corrections once we allow for a rather large (of order 100) and non-universal scaling parameter inside the logarithms.[5] We believe that our study is the most accurate numerical work on this subject. Our results may motivate further theoretical study to confirm the numerical values of our scaling parameters.

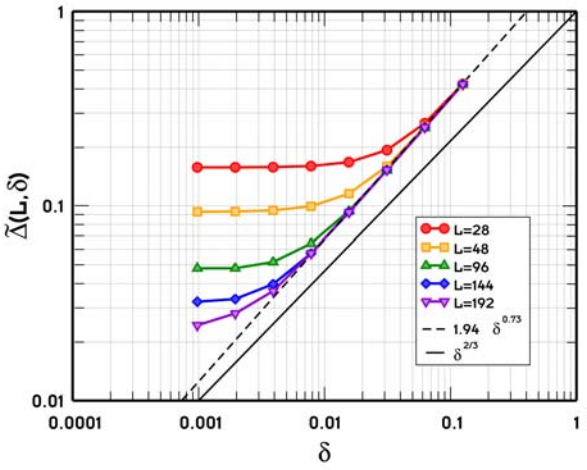


Fig. 1. Critical behavior of the singlet-triplet gap as a function of alternation d for spin chains of length L. *The dashed line shows a power law fit, which gives an exponent* $\delta = 0.73$. *The full line shows the renormalization-group exponent* $\delta = 2/3$.

Quantum Dots

We also performed numerical computations for two-dimensional spherical quantum dots with harmonic confinement and Coulomb interactions between the electrons. The DMRG algorithm converges toward the results of exact diagonalizations as the number of retained states is increased. This proof-of-principle shows that the DMRG is applicable also to interacting fermion problems in more than one dimension. However, the numerical effort is considerably larger for these systems than for the spin chains since the two-body interaction also couples distant single-particle orbitals and considerably more states have to be kept to obtain converged results. Further research is necessary to optimize the order of the single-particle orbitals and the algorithm for the initial sweep.

Nuclear Shell Model

We also applied the DMRG to nuclear shell model calculations. In these applications, the method encountered convergence problems due to difficulties during the warm-up phase of the algorithm. We hope to overcome these difficulties eventually. An analysis of this problem led two of us to propose and investigate a related method, the wave function factorization, for large-scale nuclear structure problems.[6]

Summary and Conclusions

Within this Seed Money Fund project, we developed a DMRG-based Fortran 90 program for numerical calculations in interacting fermion systems and applied it to computations in spin chains and quantum dots. The program is available for serial and parallel computers and is easily portable. We used the DMRG to study in detail the critical behavior of the alternating spin-1/2 antiferromagnetic Heisenberg chain. Our numerical results for the critical exponents of the gap and the ground-state energy defect agree with theoretical predictions provided rather large scaling parameters are introduced in the logarithms. They are the most accurate numerical results available for the study of this problem. We gave a proof-of-principle application of the DMRG to interacting electrons in harmonically confined spherical quantum dots. The results of this project open the avenue for a computational solution of different interacting fermion systems ranging from spin chains to quantum dots to atoms and molecules.

This project made the DMRG algorithm available within ORNL and thereby added to the Laboratory's competency in computational modeling of interacting

quantum systems. We presented the results of this project in seminars to the computational materials science group at ORNL's CCS and to the neutron scattering group at ORNL's High Flux Isotope Reactor (HFIR). Our DMRG program has already become useful inside ORNL: very recently, it was employed to interpret the data of a neutron scattering experiment at HFIR.[7]

References

[1]S. R. White, *Phys. Rev. Lett.* **69**, 2863 (1992).

[2]I. Peschel, X. Wang, M. Kaulke, and K. Hallberg (Eds.), *Density-Matrix Renormalization Group*, Springer, Berlin, 1999.

[3]G. Fano, F. Ortolani, and L. Ziosi, *J. Chem. Phys.* **108**, 9246 (1998).

[4]T. Xiang. *Phys. Rev. B* **53**, 10445 (1996).

[5]T. Papenbrock, T. Barnes, D. J. Dean, M. V. Stoitsov, and M. R. Strayer, *Phys. Rev. B* **68**, 024416 (2003).

[6]T. Papenbrock and D. J. Dean, *Phys. Rev. C* **67**, 051303(R) (2003).

[7]A. Zheludev, T. Masuda, B. Sales, D. Mandrus, T. Papenbrock, T. Barnes, and S. Park, eprint cond-mat/0310741.

Engineered Entanglement in Two-Photon States

V. A. Protopopescu, W. P. Grice, and Y.-H. Kim
Computer Science and Mathematics Division

The process of spontaneous parametric down-conversion (SPDC) is a convenient source of entangled photon pairs for use in various quantum information applications. For many of these applications, a pulsed source is required, but the polarization entanglement tends to suffer when the process is pumped by an ultrafast laser. Indeed, the experiments described below show that ultrafast-pumped SPDC produces pairs in which the photons have different spectral characteristics. We developed two methods for correcting the deleterious effects of these spectral differences. In the first, we specify the experimental parameters needed to produce a spectrally symmetric state, and in the second, we change the relationship between polarization and the distinguishing spectral information. Using the latter method, we have demonstrated the highest degree of polarization entanglement ever achieved in ultrafast-pumped type-II SPDC.

Introduction

The notion of entanglement was first introduced by Einstein, Podolsky, and Rosen (EPR) in a paper that analyzes the predicted measurement results in a *gedanken* experiment for a pair of specially correlated particles.[1] The quantum state of the pair is such that neither particle could be considered to be in a state independent of the other, regardless of the spatial separation of the two entities. Entangled states of two particles (photons) are most commonly generated in the process of type-II spontaneous parametric down-conversion (SPDC), in which a high-energy UV pump photon is split inside a nonlinear optical crystal into two orthogonally polarized lower-energy daughter photons. To date, most of the SPDC experiments have employed monochromatic (cw) lasers to pump the nonlinear crystals. With this type of pumping scheme, the emission time of a photon is highly correlated with that of its sibling since the photons are always emitted in pairs. The time at which a pair may be emitted, however, is completely random within the relatively long coherence time of the cw pump laser.

However, for applications such as the generation of multiphoton entangled states or quantum teleportation, well-timed photon pairs (within 100 fs) are a fundamental requirement. For other technologies, such as quantum cryptography, better timing control would represent an improvement over the state of the art. A seemingly natural solution is to use an ultrafast laser (femtosecond pulsed laser) to pump the SPDC process. Unfortunately, this method is not without its problems. The most notable of these is that the polarization entanglement in pulsed type-II SPDC is of poor quality; that is, the two-photon polarization state from SPDC pumped by an ultrafast laser behaves more like a mixed state, as opposed to a pure entangled state.[2]

From the study of the temporal and spectral properties of the two-photon wavefunctions, we found that the poorer quality of the polarization entanglement in pulse pumped type-II SPDC is likely due to differences between the group velocities of the photons involved in the SPDC process.[3] As a result, the two photons are emitted with different temporal and spectral properties. The primary objective of the project described herein was to develop a method for correcting the spectral dissimilarities and subsequently to provide the first experimental generation of high-intensity pulsed two-photon states maximally entangled in polarization. This was to be accomplished via a novel approach to the problem. Rather than passing the photons through narrow-band spectral filters in order to eliminate spectral differences (which would have reduced the count rates to very low levels), we sought to address the problem at the source by "engineering" a two-photon source to emit photons with the desired properties. A secondary objective was to verify the predictions of the theoretical treatments of ultrafast-pumped SPDC through the first-ever measurements of the joint spectrum of the two-photon state.

Technical Approach

In simplest terms, a polarization-entangled state consists of two photons whose polarizations are correlated in every measurement basis, even though their individual polarizations are undefined. The most direct method for the generation of such a state involves the geometric superposition of two definite polarization states emitted from a type-II SPDC crystal, yielding the state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left[|H\rangle_1|V\rangle_2 - |V\rangle_1|H\rangle_2\right],$$

where the subscripts denote two distinct directions of propagation and where H and V refer to horizontal and vertical polarization, respectively. In each path, a given photon may have either polarization, with the orthogonal polarization to be found in the conjugate path. A two-photon state constructed in this way may be considered maximally entangled only if the horizontally and vertically polarized photons are identical in all respects except polarization. If the polarization information is masked, it must be impossible to distinguish the two constituent terms.

As stated above, it is quite straightforward to generate maximally entangled photon pairs using cw-pumped SPDC. However, all efforts involving this type of technique have come up short when an ultrafast pump has been employed. Experimental studies have shown that such a method yields partially entangled states, with the degree of entanglement dependent on pump bandwidth and crystal length.[2] It had been suspected, but not directly confirmed, that the diminished entanglement was due to spectral and temporal differences between the two photons.[3,4] More precisely, the vertically polarized photons were suspected to have different properties from the horizontally polarized photons. An important task in this project was to experimentally verify these differences through the measurement of the joint spectrum. This measurement is relatively straightforward (albeit tedious), entailing simultaneous measurements of the spectra of the individual photons. The two photons were directed to two independent monochrometers monitored by single-photon detectors. The joint spectral intensity was then recorded as the normalized coincidence rate measured as a function of the wavelength settings of the two monochrometers. The data shown in Fig. 1 represent the first-ever measurements of the joint spectra of two-photon states. Measurements were made both for cw-pumped down-conversion [Fig. 1(a)] and for ultrafast pumped down-conversion [Fig. 1(b)]. The asymmetry in the second plot is a clear indication that the two photons have different spectral characteristics, even though their center wavelengths are identical.

The spectral asymmetry shown in Fig. 1(a) is the source of the poor polarization entanglement in ultrafast-pumped type-II SPDC. The theoretical treatments that predict this spectral structure also hold the key for overcoming the problem.[5] The asymmetry occurs because of the birefringence of the down-conversion crystal. The properties of the crystal determine the phase-matching conditions for the down-conversion process, that is, all of the ways that the energy of a pump photon may be distributed to the two daughter fields. Because the crystal is birefringent, the fields for the orthogonally polarized

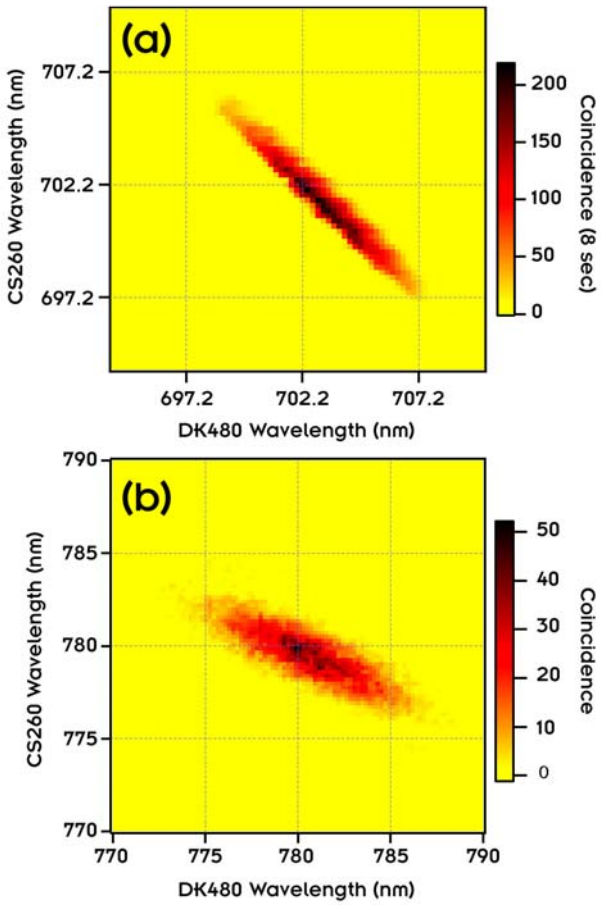


Fig. 1. Measured joint spectra for the two-photon states from (a) cw-pumped SPDC and (b) ultrafast-pumped SPDC.

photons are shaped by two different sets of crystal properties, resulting in an asymmetric phase-matching condition. This asymmetry is not seen in cw-pumped down-conversion because the very narrow pump field tends to "enforce" symmetry on the interaction. It is only when the pump bandwidth becomes large—as in ultrafast-pumped SPDC—that the asymmetry becomes apparent. Our approach to the problem was to examine the parameters affecting the down-conversion process in order to identify a configuration that would yield a symmetric phase-matching condition. It was found that the desired symmetry could be achieved if the group velocities of the three interacting fields satisfied the relation $v_{pump}^{-1} = \frac{1}{2}\left(v_1^{-1} + v_2^{-1}\right)$. After modeling several different crystal materials, a suitable candidate was found in beta barium borate (BBO) pumped at 757 nm. The calculated joint spectrum for this configuration, which is shown in Fig. 2, exhibits the desired symmetry.

An alternative approach for eliminating the effect of distinguishing spectral information was also developed during the course of this project. Here the idea is to permit

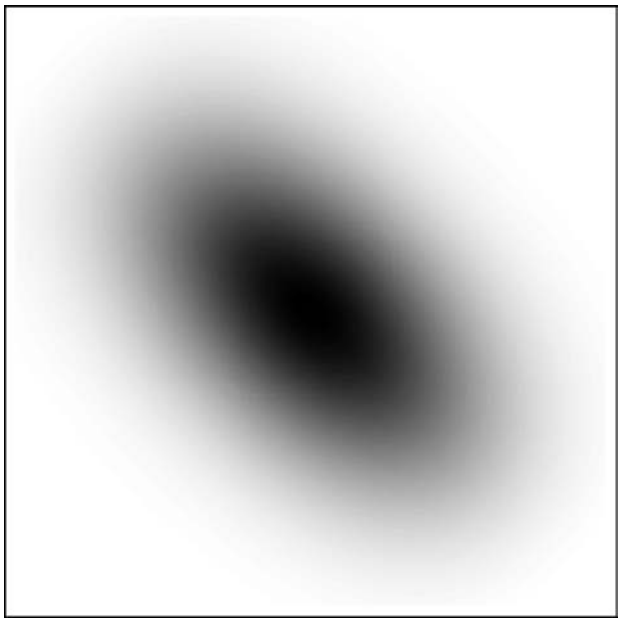


Fig. 2. Calculated joint spectrum with engineered symmetry.

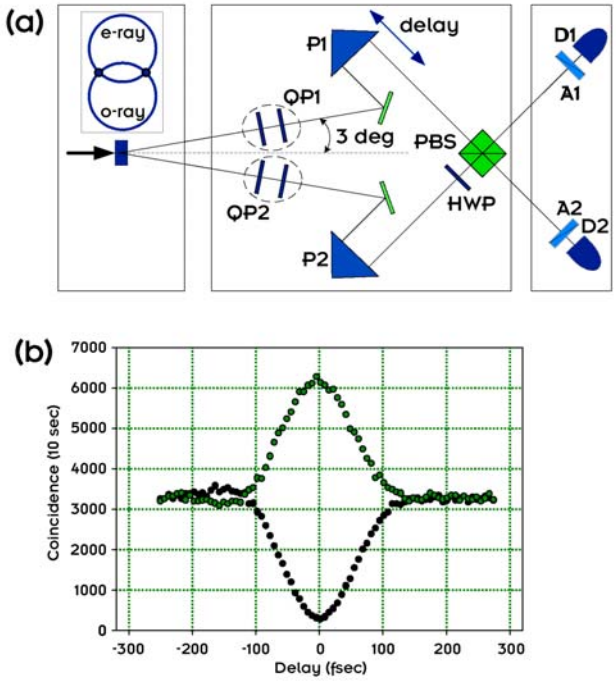


Fig. 3. (a) Two photons produced via ultrafast-pumped type-II SPDC meet at a polarization beamsplitter (PBS) after a half-wave plate (HWP) rotates the polarization in one arm by 90°. (b) Coincidence rates measured between detectors D1 and D2 as a function of delay. Polarization analyzers A1 and A2 were both set to 45° for the upper curve, while A2 was rotated to –45° for the lower curve.

spectral and temporal differences between the photons and focus, instead, on their correlation with polarization. By shifting the spectral correlation to path, rather than polarization, the spectral differences would yield no information regarding the polarization of a particular photon.[6] This "reassignment" of the correlation was accomplished in the set-up shown in Fig. 3(a). The photons exit the crystal as described above with a horizontally polarized photon in path 1 and a vertically polarized photon in path 2 ($|H\rangle|V\rangle$), or vice versa ($|V\rangle|H\rangle$). Before the photons are directed to a polarization beamsplitter (PBS), one of the photons has its polarization rotated by 90° so that the photons are identically polarized. Thus, they are either both reflected or both transmitted at the PBS, yielding the polarization-entangled state

$$|\psi\rangle_{\text{OUT}} = \frac{1}{\sqrt{2}}\left[|H\rangle_1|H\rangle_2 - |V\rangle_1|V\rangle_2\right].$$

The important thing to note here is that the photon that exited the crystal with horizontal (vertical) polarization always ends up in path 1 (2). Only one set of spectral properties is present in each path, then, even though either polarization may be found. Thus, the spectral properties are correlated with path, rather than with polarization. This results in a high degree of polarization entanglement, as shown in Fig. 3(b). The two curves represent the coincidence rates as the photons are passed through crossed (lower curve) and parallel (upper curve) polarizers oriented at 45° with respect to the horizontal. When the path lengths are balanced (central region), the visibility is greater than 90%, indicating a high degree of polarization entanglement.

Results and Accomplishments

The work described above represents a number of important contributions to the fields of quantum optics and quantum information science. These are described in detail in six published papers[6,7] and two additional manuscripts either in preparation or under review.

The measurements of the joint spectra are notable for a number of reasons, not the least of which is that they are the first of their kind. As researchers progress from proof-of-principle experiments to real quantum applications (e.g., quantum cryptography, quantum computing, etc.), they will need to know more and more about the quantum states they use. For this reason, the joint spectrum measurement technique is expected to become an important tool. Extending the measurement to the temporal domain would enhance this capability even further, as it would enable the complete chrono-spectral characterization of the two-photon state. Another significant contribution to be attributed to the measurements of the joint spectra is the validation of the theoretical treatments of ultrafast-pumped SPDC. Prior to these measurements, the theory could be tested only indirectly. The data obtained and analyzed within this project provide a means to measure directly the spectral properties of the photon pairs and are in very good agreement with the theory.

Specification of the parameters needed for a symmetric joint spectrum is expected to be important for a number of quantum applications. Any application requiring a pulsed source of polarization-entangled photons will benefit, since a symmetric joint spectrum is needed for a high degree of polarization entanglement. In addition, it has recently been shown that photon pairs with a positively correlated joint spectrum can be used in quantum-enhanced clock synchronization. The positive correlation is possible only if the joint spectrum is symmetric.

Finally, the most significant contribution is likely to be the technique for redistributing distinguishing spectral information to transform mixed states into polarization-entangled states. Although this approach was not conceived before the project began, it is more subtle than the originally proposed method and still meets the stated objective of the project, namely, the first experimental generation of high-intensity pulsed two-photon states that are maximally entangled in polarization. Moreover, it meets this objective in a way that addresses the more general problem of managing information in entangled systems. Whatever form the first quantum computer takes, it is certain that the constituent systems will contain many degrees of freedom. In order to isolate the degree of freedom of interest, it will be necessary to control correlations with other variables in the system. In this case, the detrimental correlation between spectral properties and polarization (the degree of freedom of interest) was transformed to an innocuous correlation between spectral properties and path. We expect that this concept could be extended to other degrees of freedom and other systems. It is the generalization of this concept that is likely to have the most far-reaching effects.

A number of federal agencies are working toward the development of applications involving entangled photons. It is clear that the work described here will aid in those efforts, and several agencies have expressed interest. Quantum computing is being championed by the Defense Advanced Research Projects Agency. The National Reconnaissance Office has expressed interest in quantum-enhanced clock synchronization and in quantum cryptography. Also interested in quantum cryptography are the National Security Administration and the Advanced Research and Development Activity (ARDA). Based in large part upon the results obtained within this project, we have already obtained ARDA funding at the level of $150 K for the development of a pulsed two-photon source for quantum cryptography. Funding is expected to continue at this level for at least one more year.

Summary and Conclusions

Using a novel measurement technique developed for this project, it was shown that the spectral properties of the two photons produced in the process of ultrafast-pumped spontaneous parametric down-conversion are not identical, even when the center wavelengths are the same. These results provide the first direct validation of the multi-mode theory describing the process. Left unchecked, the spectral differences tend to reduce the degree of polarization entanglement achievable with the usual entangling techniques. Two methods have been developed to counteract these effects and improve polarization entanglement. In the first method, which is known as spectral engineering, the joint spectral properties of the photon pair are shaped through judicious choices of relevant experimental parameters. In the second method, the photons are allowed to retain their different spectral properties, but the spectral correlation with polarization is shifted to a correlation with direction of propagation. Using this technique, near-perfect polarization entanglement was achieved for the first time in ultrafast-pumped type-II down-conversion.

References

[1]A. Einstein, B. Podolsky, and N. Rosen, *Physical Review* **47**, 777 (1935).

[2]G. Di Giuseppe et al., *Phys. Rev. A* **56**, R21 (1997); W. P. Grice et al., *Phys. Rev. A* **57**, R2289 (1998).

[3]T. E. Keller and M. H. Rubin, *Phys. Rev. A* **56**, 1534 (1997).

[4]W. Grice and I. Walmsley, *Phys. Rev. A* **56**, 1627 (1997).

[5]Y.-H. Kim and W. P. Grice, *J. Mod. Opt.* **49**, 2309 (2002).

[6]Y.-H. Kim et al., *Phys. Rev. A* **67**, 010301(R) (2003).

[7]Y.-H. Kim and W. P. Grice, to appear in *Phys. Rev. A*; Y.-H. Kim, *Phys. Lett. A* **315**, 352 (2003); Y. H. Kim, *JOSA B* **20**, 1959 (2003); Y. H. Kim, *Phys. Rev. A* **68**, 013804 (2003); Y. H. Kim and W. P. Grice, *Phys. Rev. A* **67**, 065802 (2003).

Development of Readout Electronics Architecture for a Silicon-Strip Vertex Detector Upgrade to the PHENIX Experiment

V. Cianciolo, T. C. Awes, M. Bobrek, C. L. Britton, W. L. Bryan, K. F. Read, P. W. Stankus, A. L. Wintenberg, and G. R. Young
[1]*Physics Division*
[2]*Engineering Science and Technology Division*

The Relativistic Heavy Ion Collider (RHIC) at Brookhaven National Lab (BNL) smashes particle beams into each other in an attempt to create conditions of energy and density that have not existed since shortly after the Big Bang, thereby melting protons into their constituent quarks and gluons. The PHENIX experiment, one of the RHIC detectors which collects data on these collisions, has provided an intriguing series of results that has electrified the nuclear physics community. A multi-layer Silicon Vertex Detector would greatly extend PHENIX's scientific impact, and the goal of this project is to develop a viable concept for the outer layers of such a detector that meets stringent requirements on the detector thickness, radiation tolerance, and interfaces with a novel BNL-developed sensor and the PHENIX data acquisition system (DAQ).

A Silicon Vertex Detector would greatly extend the PHENIX scientific impact in the areas (key to characterization of the extreme conditions created in RHIC collisions) of suppression of particle production at large transverse momentum, suppression of quarkonia production, characterization of open heavy-flavor production, and gamma-tagged jet production.

Radiation damage concerns argue against using a double-sided silicon detector technology. Material thickness limits argue against multiple layers of silicon microstrip detectors in different orientations, and channel count argues against silicon pixel detectors. A novel alternative, developed by colleagues at BNL and RIKEN, is a strip-pixel detector which is essentially a pixel detector which has individual pixels ganged together into multiple strips on different metallization layers. This allows for two-dimensional readout with a single-sided process. Our goal was to develop a credible concept for the readout electronics that connected these sensors to the existing PHENIX DAQ.

Two options were available for the signal processing chip. The first was to a scheme based on a modification of the TGV family of application-specific integrated circuits (ASICs) developed previously by ESTD members of this project in conformance with PHENIX DAQ specifications and currently in use throughout PHENIX. An alternative was to use an ASIC (SVX4), developed by Fermi National Lab (FNAL), which might be suitable in its current form, thus eliminating the cost and risk associated with any chip development.

Through literature studies and extensive discussions with BNL and FNAL engineers we believe that the SVX4 will be compatible with both the sensor and with the PHENIX DAQ. We have also developed a small testboard that we will use to verify compatibility with the PHENIX DAQ. As a follow-on to this project, funds may be available from our Japanese colleagues to incorporate this testboard in a full system test (with a source and in a beam) of the sensor. We have evaluated the bandwidth requirements for the system and found design options compatible with the existing data-collection modules (DCMs) and with proposed DCM upgrades. We have evaluated the minimum thickness that a detector based on these sensors is likely to require. The result is that more material is required than desired—this information is being incorporated into simulations of the entire PHENIX detector and preliminary indications are that the required amount of material is acceptably small. Similarly we have determined that the SVX4 alternative does not have event selection (triggering) capability before digitization. Again, this information is being incorporated into simulations of the entire PHENIX detector and preliminary indications are that such triggering is not necessary. Finally, we have developed a novel concept for the bus required to bring signals and power to and from the signal processing chips. This concept greatly simplifies assembly and reduces the amount of material in the detector of the detector and readout electronics.

Toward Neutron Star Merger Simulations: Gravitational Waves, Heavy Element Nucleosynthesis, and Gamma-Ray Bursts

C. Y. Cardall,[1,2] M. W. Guidry,[1,2] A. Mezzacappa,[1] and R. J. Toedte[3]
[1]*Physics Division*
[2]*The University of Tennessee*
[3]*Computer Science and Mathematics Division*

Neutron star mergers are associated with three phenomena of high interest in astrophysics: gravitational waves, heavy element nucleosynthesis, and gamma-ray bursts. Advanced simulation and analysis tools are needed to shed light on neutron star mergers and their phenomenology. Our objective in this project is to develop a set of computational tools, *GenASiS* (*Gen*eral *A*strophysical *Si*mulation *S*ystem), that will position us to study neutron star mergers with massively parallel simulations on distributed memory computer architectures. As initial stages to *GenASiS*, we are developing an adaptive-mesh radiation hydrodynamics code with self-gravitation and associated visualization tools.

The purpose of this project is to seed the development of a program to study neutron star mergers with massively parallel simulations on distributed memory computer architectures.

Neutron star mergers feature two self-gravitating bodies composed of dense nuclear matter, moving at high speed; this calls for adaptive-mesh radiation hydrodynamics, a frontier in scientific computing. The idea of an adaptive mesh is to employ high resolution only where needed in order to conserve memory and computational effort. In this problem, an adaptive mesh would allow much of the "empty space" between the two neutron stars to be handled with low resolution. Radiation transport is so expensive—both in terms of memory and computational demands—that an adaptive mesh is the only way it could be implemented in three spatial dimensions. Adaptive-mesh refinement for hydrodynamics had been around for some time, but adaptive-mesh refinement for radiation hydrodynamics remains a frontier to which significant contributions can be made.

Before this project began in June 2003 we had begun developing *GenASiS* (*Gen*eral *A*strophysical *Si*mulation *S*ystem), an adaptive-mesh neutrino radiation hydrodynamics code; our work in this project will bring *GenASiS* to a level of development that will permit credible proposals to simulate neutron star mergers. The project is to be completed in FY2004. We have four tasks:

1. Parallelize our adaptive-mesh hydrodynamics code and add a realistic dense nuclear matter equation of state. By the end of FY2003 we had added the realistic equation of state, written code that partitions the mesh in such a way that each processor has its share of spatial zones at each level of refinement, and written code that exchanged mesh refinement and coarsening information between all processors.
2. To include self-gravity, we must implement a Poisson solver on adaptive mesh. At the end of FY 2003 we had worked out the finite differencing of the Poisson equation on an adaptive mesh in coordinates sufficiently general to allow Cartesian, spherical, and cylindrical coordinates to be used. We have also become familiar with Portable, Extensible Toolkit for Scientific computation (PETSc), whose routines we will use to solve the distributed sparse matrix system resulting from the finite differenced Poisson equation.
3. Add realistic neutrino interactions to our parallel, fixed-mesh neutrino radiation transport code; incorporate number and energy transfer to the fluid; and make this code adaptive. Our work in FY 2003 focused on adding realistic interactions and number and energy transfer. The realistic interactions are being implemented using the GlobalArrays library to distribute segments of the high-dimensional interaction kernels among processors; code has been written which allows each processor to grab the pieces of the table it needs. Code infrastructure to handle number and energy transfer is partially completed.
4. Implement visualization tools for our adaptive mesh. This work will take place in FY2004.

In addition to enabling a program of neutron star merger studies, this work has many benefits to DOE and other agencies. It addresses DOE's need for research in advanced scientific computing. Moreover, neutron star mergers involve interesting nuclear physics issues, including the dense matter equation of state, neutrino

interactions, and heavy element (r-process) nucleosynthesis (the subject to be studied by the Rare Isotope Accelerator proposed to DOE). Developments in radiation transport are useful in studies of combustion, climate, and stockpile stewardship—all of interest to DOE. This project also will be relevant to projects undertaken by the National Science Foundation (NSF) and the National Aeronautics and Space Administration (NASA). The NSF is funding the Laser Interferometer Gravitational Wave Observatory that is comprised of two detectors of gravitational radiation from astrophysical sources. Neutron star mergers are expected to be a primary source of gravitational waves, and "templates" computed from simulations that are needed to distinguish gravitational wave signals from noise. The NSF is also funding a Physics Frontier Center for gravitational wave phenomenology at Penn State, to which our work will also contribute. NASA has several orbiting observatories, both active and planned, sensitive to various wavelengths of electromagnetic radiation; these study the enigmatic gamma-ray bursts, of which neutron star mergers are one of two proposed "central engines."

AUTHOR INDEX

INDEX OF PROJECT NUMBERS

DIRECTOR'S R&D FUND RESEARCH THRUST AREAS INDEX OF PROJECTS

NATIONAL SECURITY

NEUTRON SCIENCES INITIATIVE

TERASCALE COMPUTING AND SIMULATION SCIENCE INITIATIVE

GENERAL