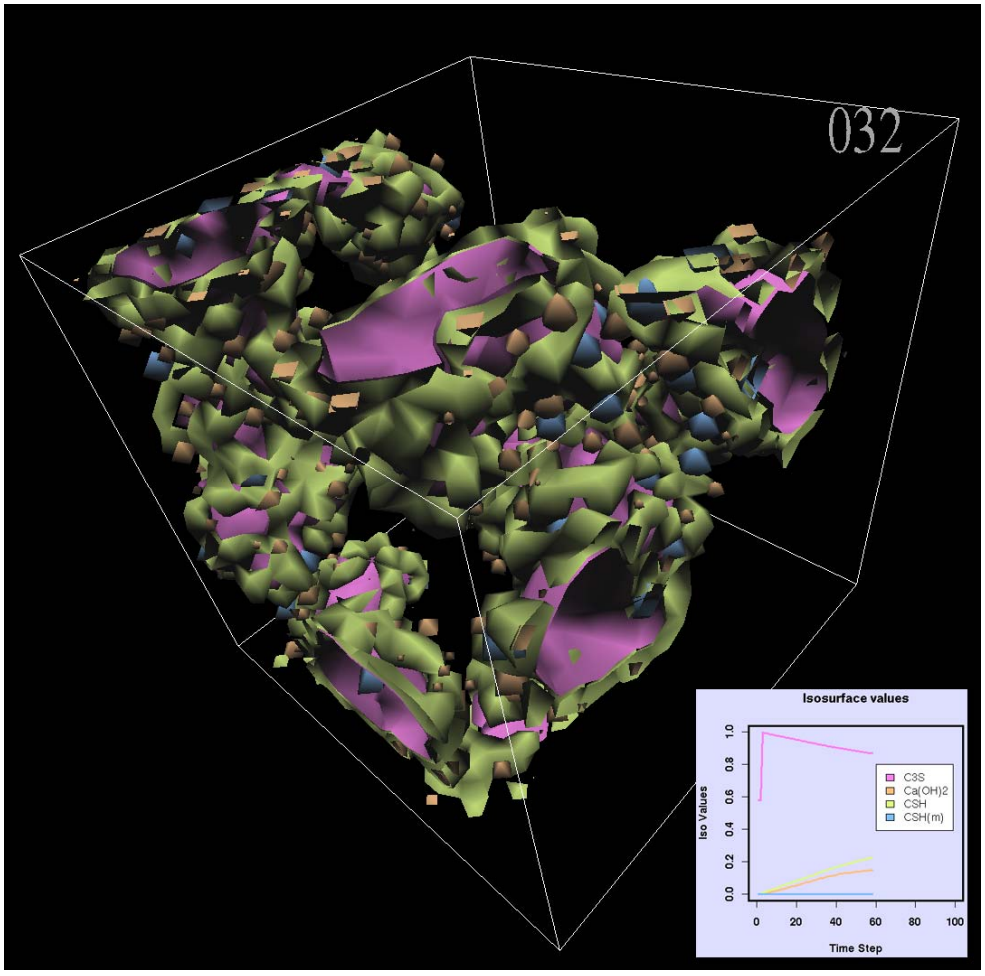
Mathematical and Computational Sciences Division

Summary of Activities for Fiscal Year 2007



Information Technology Laboratory
National Institute of Standards and Technology
U.S. Department of Commerce

January 2008

INFORMATION TECHNOLOGY LABORATORY

NIST National Institute of Standards and Technology U.S. Department of Commerce

Abstract

This report summarizes the technical work of the Mathematical and Computational Sciences Division (MCSD) of NIST's Information Technology Laboratory. Part I (Overview) provides a high-level overview of the Division's activities, including highlights of technical accomplishments during the previous year. Part II (Features) provides further details on eight particular projects of particular note this year. This is followed in Part III (Project Summaries) by brief summaries of all technical projects active during the past year. Part IV (Activity Data) provides listings of publications, technical talks, and other professional activities in which Division staff members have participated. The reporting period covered by this document is October 2006 through December 2007.

For further information, contact Ronald F. Boisvert, Mail Stop 8910, NIST, Gaithersburg, MD 20899-8910, phone 301-975-3812, email boisvert@nist.gov, or see the Division's web site at http://math.nist.gov/mcsd/.

Cover photo. Visualization and analysis of the microstructure a computational model of cement hydration showing four distinct phases. This is the result of research performed by William George, Steve Satterfield, and Edith Enjolras of MCSD in collaboration with Jeffrey Bullard of the NIST Building and Fire Research Laboratory.

Acknowledgement. We are grateful to Robin Bickel for collecting the information and organizing the first draft of this report.

Disclaimer. All references to commercial products in this document are provided only to document how results have been obtained. Their identification does not imply recommendation or endorsement by NIST.

Table of Contents

Part III

Project Summaries

Mathematics of Metrology

Computable Error Bounds for Delay Differential Equations

David Gilsinn
Sita Ramamurthi

See feature article, page 25.

The Lipschitz Exponent as an Image Metrology Tool

Alfred Carasso, MCSD
Andras Vladar (NIST MEL)

Recent work by A. Carasso and A. Vladar, to appear in *Optical Engineering*, further develops the notion of Lipschitz exponents by applying it to several areas of image analysis. This important MCSD image metrology tool can be helpful in both hardware and software applications. As a hardware tool, it can be used to quantify and monitor the performance of such imaging systems as electron microscopes, where the shape of the electron beam deteriorates over time. In recent years, digital image restoration has become pervasive in science and engineering. Lipschitz exponents can be used to evaluate and rank order the performance of competing image restoration algorithms.

The Lipschitz exponent α of an image, where $0 < \alpha < 1$, measures the degree of fine structure and other non-differentiable features in an image. Images that are of bounded variation, or smoother, have $\alpha = 1$. The value of α decreases systematically with increasing image roughness. Most natural images are not of bounded variation, and often have α values less than 0.5. The value of α changes when an image is degraded. The behavior of α when that image is restored by various procedures is of prime interest, as it indicates the extent to which significant fine structure has been recovered.

The Lipschitz exponent can be found by convolving the given image, $f(x,y)$, with successively narrower Gaussians, to form the sequence of blurred images $G_t f$, as t tends to zero. Here t is a width parameter, and $G_t f$ tends to f as t tends to zero. The *slowness* with which this happens is indicative of the non-smoothness of $f(x,y)$. Let $\mu(t)$ denote the L_1 relative error in approximating f by $G_t f$. The value of α can be found by plotting $\log(\mu(t))$ vs $\log t$, as t decreases from $t=1$, and evaluating the slope of that curve in the "slowly varying" region immediately to the left of $\log t=0$. A NIST patent covering this methodology is currently being processed.

The experiments in Figs. 1 and 2 deal with synthetically degraded images, and the $\mu(t)$ vs t traces *A, B, C, D* correspond to images (a), (b), (c), and (d) respectively. In both these figures, Lipschitz exponents were obtained by estimating trace slopes on the interval $-7 \le \log t < 0$.

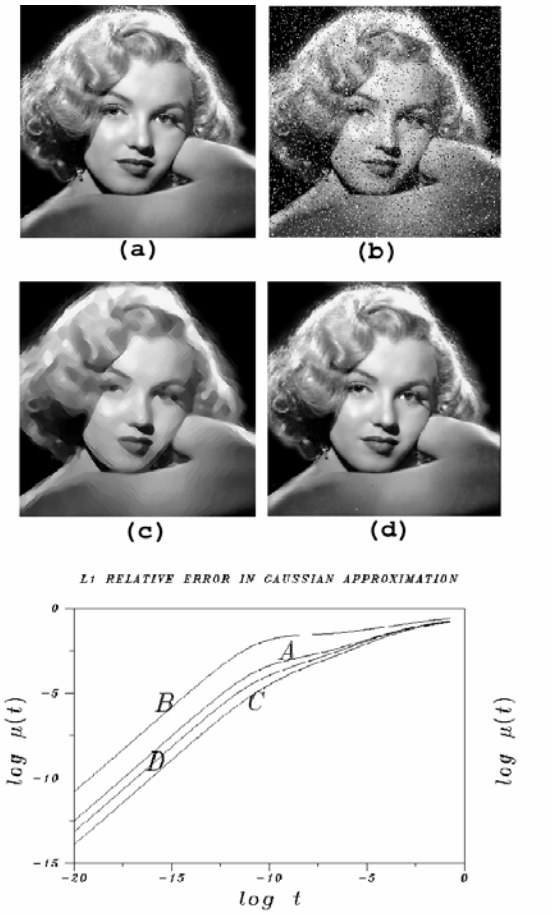


Figure 1. *Image Denoising. Noise addition can artificially lower Lipschitz exponent α, while some noise removal algorithms can eliminate texture and increase α. Traces A, B, C, and D correspond to images (a), (b), (c), and (d) respectively. Exponents α obtained by estimating trace slopes on $-7 \le \log t < 0$.*

Fig. 1 involves noising and denoising the *512 x 512* Marilyn Monroe image. The original image (a) has trace *A* with $\alpha = 0.591$. Salt and pepper noise was added to produce image (b) with trace *B* and $\alpha = 0.302$. Two distinct denoising methods were examined. The 'total variation' method produced image (c), where a good deal of texture has been eliminated along with the noise. This led to a substantially larger $\alpha = 0.714$ in trace *C*. The second denoising method used

was 2D median filtering in image (d). This is a good approximation to the original, and trace *D*, with α = 0.645, is closer to trace *A* than is trace *C*.

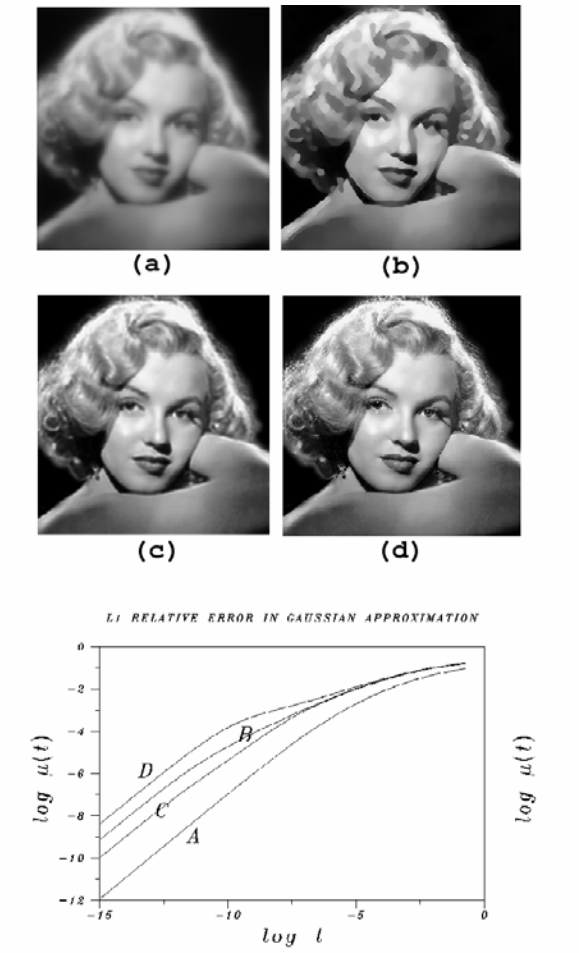


Figure 2. *Image Deblurring. Not all deblurring algorithms are equally effective at recovering image texture*

Our second example, summarized in Fig. 2, involves blurring and deblurring the Marilyn Monroe image. Fig. 2(a) with trace *A* and *α = 0.887* results from synthetic blurring of Fig. 1(a) by convolution with a Lorentzian density. Blurring without adding noise increases α, as blurring is generally a smoothing operation. Three mathematically distinct methods of deblurring image (a) are examined. The Marquina-Osher total variation method produces image (b) with trace *B* and α = 0.695, higher than the original value of *0.591*. Image (c) with trace *C* and α = 0.714, results from *200* Lucy-Richardson iterations. Both of these procedures smooth out the image and fail to recover fine structure. Image (d) with trace *D* and α = 0.616 is quite close to the original. It was produced using the "slow evolution from the continuation boundary" (SECB) procedure.

Use of Lipschitz exponents will play a significant role in ongoing NIST research. State of the art software tools, such as Caltech curvelet denoising, Mumford-Shah regularization, and Bremen minimum-norm blind deconvolution can be evaluated and applied to NIST problems.

Systems Identification and Parameter Estimation

Bert W. Rust

Identifying a system of ordinary differential equations (ODEs) describing the dynamical relationships between (1) man-made carbon dioxide emissions $P(t)$, (2) atmospheric carbon dioxide concentrations $c(t)$, and (3) global temperatures $T(t)$ would provide a useful tool for evaluating mitigation strategies for global warming. Last year I demonstrated that changes in $T(t)$ vary linearly with changes in $c(t)$ and that the warming is accelerating. That acceleration is driven by the ever increasing emissions $P(t)$. Attempts to model the $c(t)$ measurements by assuming that a constant fraction of the emissions remain in the atmosphere failed because that fraction has apparently changed during the 151 year period for which measurements are available. The relationship between $c(t)$ and $P(t)$ can be modeled by an underdetermined linear system in which the annual $P(t)$ measurements are used to form the matrix, the $c(t)$ measurements are the elements of the right hand side vector, and the elements of the solution vector are the unknown annual values of the fraction remaining in the atmosphere. Unfortunately, the $c(t)$ record has many gaps, so the matrix has dimensions 83×153. This problem would be hopelessly underdetermined were it not for two additional considerations: (1) the elements of the solution vector are constrained to lie between the values 0 and 1, and (2) it is not necessary to determine the fraction for each individual year, but rather is sufficient to determine average values over suitably chosen subintervals of time. O'Leary's BRAKET-LS algorithm was used to compute these constrained averages and their confidence interval bounds. The results for three equally spaced subintervals of [1850, 2000] are shown in Fig. 1.

It is clear that a single value of about 0.377 will suffice for the whole twentieth century and that a statistically significantly higher value of about 0.545 is required for the earlier period. It is not clear that the jump from the higher to the lower value occurred exactly at epoch 1900.0, and future work will be directed toward a more accurate estimate of this threshold time. Future work on this project will also concentrate on refining the ODE for $P(t)$ and on simultaneously fitting the solutions of all three equations to their respective time series.

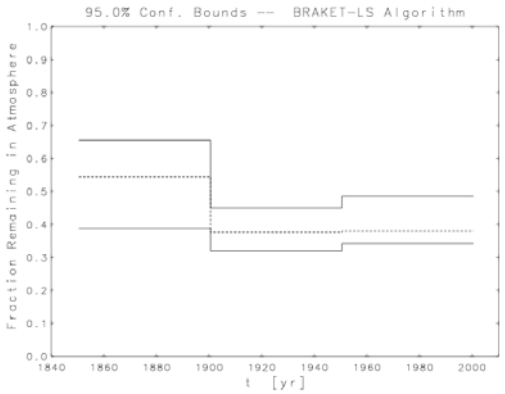


Figure 1. *The fraction of man-made carbon dioxide emissions that remains in the atmosphere. The dashed lines are the BRAKET-LS estimates of the average values on the subintervals [1850,1900], [1900,1950], and [1950,2000]. The corresponding solid lines define 95 % confidence intervals for the estimates. The interval bounds are not symmetrically spaced about the estimates because the elements of the solution vector were constrained by the calculation to lie in the interval [0,1].*

Sparse Representations in High Dimensional Geometry

Bradley Alpert
Yu Chen (New York University)

Medical diagnostic systems often produce images that lack quantitative information or comparability across imaging systems, which can lead to difficulties in diagnosis and long-term management of disease that requires patient monitoring. Elaborate measurements are taken, digitally processed, then reported in essentially qualitative form. For MRI, functional MRI, and some other processes, formation of the images relies on filtering and Fourier transformation.

Although the computational recovery of functions from their Fourier representations is considered routine, aside from possible efficiency concerns arising from unequispaced data, standard approaches assume that the functions are smooth and can be recovered with the discrete Fourier transform (computed via the FFT). When this assumption fails, as is evident from Fourier data that are not small at the highest frequencies measured, some sort of attenuation, or belling, scheme is typically used. This procedure tends to blur sharp features and, being more art than science, can lead to different functions (images) in different systems. Although this problem, and the Gibbs phenomenon, have received attention from mathematicians in recent years, recently developed methods tend to work well only in an asymptotic sense; they do not exploit available data efficiently.

Despite inherent limits on resolution that can be obtained from truncated Fourier data, a change of assumptions from *smooth* to *piecewise smooth* can lead to significantly improved recoveries. Procedures implementing this idea are not completely established even for functions (signals) in one dimension; they are yet more challenging in two and three dimensions, where discontinuities may be expected to occur along mostly smooth curves or surfaces. This constraint, which magnifies the advantage of the piecewise smoothness assumption, must be appropriately reflected in the methods used. Alpert and Yu Chen are conducting extensive numerical experiments to understand this environment and to develop reliable procedures for these problems.

A related, yet more general, challenge is parsimonious, or sparse, representation and recovery of functions under assumptions appropriate to an application. Although there is considerable current interest in the mathematical community in these problems L^1-norm minimization, compressed sensing, sparse representation in high-dimensional geometry), and much recent progress (initiated in pivotal work by Candes, Romberg, and Tao) most of the methods being explored are limited to linear spaces. While linearity is a natural starting point, strong evidence suggests that image recovery cannot be done this way yet nevertheless is within reach.

The initial approach involves applying the sparse recovery methods, and experimenting with others, to Fourier data generated from numerical models of scatterers. The deficiencies of the resulting recoveries are characterized and point to ways that the representation space should be altered and the recovery procedures extended.

Sequential Importance Sampling and the Markov Chain Monte Carlo Method

Isabel Beichl
Francis Sullivan

The Markov Chain Monte Carlo Method (MCMC) has been studied by many researchers and an extensive theory has been developed. But getting quantitative answers to problems is still difficult because MCMC can be slow, and determining its rate of convergence is far from obvious. On the other hand, a less well-studied Monte Carlo method, sequential importance sampling (SIS), can sometimes be very fast but rigorous bounds on variance are lacking.

As one example, for the monomer/dimer problem Kenyon, Randall and Sinclair developed an MCMC method which has not been successfully used

for practical computation, while we have developed an SIS method and computed quantities of interest. So, the situation is: MCMC can be trusted but may be very, very difficult to use, whereas SIS gives a numerical answer, but it's hard to verify that the answer is correct. The research question is how to combine the best of both methods. We have recently developed the following.

1. A way to use our SIS method to reliably estimate optimal parameters such as fugacities (i.e., probability of accepting proposed moves) for MCMC, and so to avoid the trial and error aspect of deciding what fugacities to use.
2. A way to use the SIS method to calculate the mixing time for the MCMC method efficiently. The mixing time tells us the convergence rate and thus determines the number of steps that must be taken between samples.

Quantum Information

Making Optical "Schrödinger Cat" States

Scott Glancy
Emanuel Knill
T. Gerrits (NIST EEEL)
T. Clemment (NIST EEEL)
Sae Wo Nam (NIST EEEL)
Richard Mirin (NIST EEEL)
H. M. Vasaconcelos (Universidade Federal do Ceará, Brazil)

See feature article, page 28.

Funding: *NIST Innovations in Measurement Science program.*

Quantum Computing with Ion Traps

E. Knill
D. Leibfried (NIST PL)
D. Wineland (NIST PL)
Ion Storage Group (NIST PL)

Quantum computers have the potential for significantly speeding up many useful algorithms. However, building quantum computers is challenging. Currently available quantum devices can realize computations with only a few quantum bits (qubits) and steps. Atomic qubits in ion traps are currently considered one of the leading candidates for realizing large quantum computers. We are collaborating with the Ion Storage Group on experiments to determine the suitability of ion traps for large scale information processing and to verify the basic principles of the needed quantum control. We are also participating in research on new approaches to realizing quantum information in ion traps.

To determine the quality of quantum control in the context of quantum information processing, we developed a randomized strategy for determining an effective error probability per quantum gate and implemented it with one-qubit gates in an ion trap. The basic idea is that it suffices to determine the increase in error probability of long sequences of gates as a function of the length of the sequence. Furthermore, because of what we know of fault-tolerant quantum architectures, it suffices to consider a special subset of gates for which the final answer can be efficiently predicted using classical computation. Using this strategy, we determined that the basic one-qubit gates in our test configuration had an effective error probability of only 0.5%. Since the sources of the errors are well understood, we expect that this error probability can be greatly reduced by better stabilization of the needed lasers and other technological improvements. Once sufficiently high-quality two-qubit gates are available, we hope to repeat this experiment with two or more ion qubits.

One of the problems of the traditional scheme for manipulating the states of ion qubits is that it requires many independently and precisely controllable laser beams. We participated in an analysis of an alternative approach developed by D. Leibfried based on static laser beams where the manipulations are realized by moving ions. Since precise ion movement is required in any case, this has the potential for greatly reducing the complexity of ion trap quantum computers. The analysis shows that the necessary quantum gates can be implemented in parallel with sufficiently low error using static laser beams. Because the beams can be derived by multiple passes of a single beam, the power requirement is also reduced significantly.

One of the goals of our collaboration is to demonstrate so-called logical quantum information encoded in a quantum error-correcting code capable of correcting all one-qubit errors. Such a demonstration should include preparation of logical states, application of logical gates and measurement of logical information. The logical gates need to be performed with sufficiently low error. The technology needed for such a demonstration is currently being developed and tested.

[1] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland, Randomized Benchmarking of Quantum Gates, arXiv:0707.1032, *Physical Review A*, to appear, 2007.

[2] D. Leibfried, E. Knill, C. Ospelkaus, D. J. Wineland, Transport Quantum Logic Gates for Trapped Ions, *Physical Review A* **76** (2007), pp.032324/1—12.

Funding: *Quantum Information ACI Initiative (2007).*

Quantum Computing Theory

Scott Glancy
Emanuel Knill
Adam Meier
Yanbao Zhang
R. Somma (LANL)
G. Ortiz (Univ. of Indiana)

The implementation of quantum computers is based on a substantial body of theoretical work showing the utility of quantum algorithms and providing techniques for protecting quantum devices from inevitable noise. Our contributions to quantum computing theory include work on better understanding the limitations as well as the power of quantum computers and better methods for measuring observables in quantum physics simulations on quantum computers.

One of the most promising applications of quantum computers is to the simulation of quantum physics. In most such simulations, the desired answer is the expectation of an observable. The simplest way to determine such an expectation with a given accuracy $1/n$ is to run the simulation $1/n^2$ times, make a von Neumann measurement of the observable each time and take the average of the answers. For sufficiently well understood observables, it was known that the resources needed could be reduced from $1/n^2$ to $1/n$ by taking advantage of quantum mechanical effects. We showed that essentially the same improvement is possible for any observable under mild assumptions. An important part of our algorithm for determining the expectation of an observable is to provide rigorous confidence intervals, where the efficiency of the algorithm is logarithmic in the inverse p-value associated with the confidence interval. This greatly improves the performance of the conventional measurement schemes based on phase estimation or the quantum Fourier transform, which require resources that grow linearly with the inverse p-value.

The applications of quantum information to communication include quantum cryptography and quantum secret sharing. Such applications make possible protocols that cannot be realized using only classical information processing. One reason for the power of quantum information in this setting is the presence of quantum non-locality, which are correlations with no classical analogue. Experiments that verify the presence of quantum non-locality have been performed, but are still subject to "loop holes", that is, a reasonable but unverifiable assumption is required to ensure that there is no classical explanation for the data. Removal of the loop holes requires better detectors and/or better choices of experimental configurations. We are developing a computational framework based on information theory to compare and optimize experimental configurations. An important part of the framework is a toolkit for statistical analysis of experimental data that can provide explicit p-values for rejecting all possible classical explanations for the data.

In previous years we have developed and simulated fault-tolerant quantum computing architectures that can tolerate surprisingly high error rates per quantum gate. This year, we have started to develop new simulation algorithms for fault-tolerant architectures. The goal of our effort is to greatly improve the efficiency of the simulations and to enable simulation of architectures based on quantum systems with more than two dimensions. So far we have implemented the C++ libraries for the basic operations of such quantum systems and for data structures needed to keep track of the necessary quantum gates and their errors. Once we have tested and refined the C++ libraries, we will use them to compare different architectures. Future research is aimed at simulations of universal gate sets and non-stochastic error models.

[1] E. Knill, G. Ortiz, and R. D. Somma, Optimal Quantum Measurements of Expectation Values of Observables, *Physical Review A* **75** (2007), pp. 012328/1-13.

Funding: *NIST Innovations in Measurement Science. Quantum Information ACI Initiative (2007).*

Preparing Ancillary States for Quantum Computation

Bryan Eastin

The promise of quantum computing is unlikely to be realized without effective quantum error correction, but quantum error correction itself requires the application of quantum operations. To perform these in a robust manner it is necessary to utilize ancillary systems prepared in known quantum states. Below a certain threshold, the effective, that is, encoded, error rate decreases as the number of quantum systems in the encoding increases, but the difficulty of preparing the necessary ancillary states typically increases with the size of the encoding as well. Employing standard methods and codes, the direct implementation of even relatively small quantum encodings is impractical. In the literature on error thresholds for quantum computation, this problem is circumvented by employing a layered (or concatenated) approach to error correction and state construction. Codes and procedures that permit more direct state construction are intriguing, however, because they ought to require fewer quantum systems for their implementation.

In the coming year I plan to investigate new procedures and codes relevant to the construction of large ancillary states. With regards to new procedures,

I hope that the completion of my present work on inferring the location of propagated errors based on the measurement of complementary errors and knowledge of the preparation circuit will facilitate the construction of a broad class of quantum states. My coding research will focus primarily on low-density parity-check (LDPC) quantum codes, a class for which state preparation is dramatically simplified, but at the cost of less favorable code parameters. Both the development of new LDPC codes and more detailed error analyses of existing codes are included among my goals.

Funding: *NIST/NRC Postdoctoral Research Program.*

Adaptive Finite Element Modeling of Two Confined and Interacting Atoms

William F. Mitchell
Marjorie A. McClain
Eite Tiesinga (NIST PL)

We have applied the parallel adaptive refinement and multigrid finite element code PHAML (see feature article, page 23) to a two-dimensional Schrödinger equation in order to study the feasibility of a quantum computer based on extremely cold neutral alkali-metal atoms. Qubits are implemented as motional states of an atom trapped in a single well of an optical lattice. Quantum gates are constructed by bringing two atoms together in a single well leaving the interaction between them to cause entanglement. Quantifying the entanglement reduces to solving for selected eigenfunctions of a Schrödinger's equation that contains a Laplacian, a trapping potential and a short-ranged interaction potential. The behaviors of the eigenfunction due to the trapping potential and interaction potential are on very different scales, which requires the use of adaptive refinement and high order finite elements.

A critical component of adaptive refinement algorithms is the a posteriori local error estimator, which is used to determine where the error is the largest, and hence which triangles should be refined to optimally reduce the error. Error estimators for elliptic boundary value problems were studied extensively in the 1980's and 1990's. The same error estimators are valid for elliptic eigenvalue problems like Schrödinger's equation. Moreover, a simple additional calculation gives an estimate of the error in the eigenvalue, too. We have added this capability to our code.

Use of high degree piecewise polynomials has proved to be crucial to get the high accuracy solutions required for correctly modeling the atom interactions. We are studying the use of a new technique, hp-adaptive refinement, to improve on this even further. With hp-adaptive refinement one adapts both the element size and polynomial degree locally. Properly done, this technique has a faster convergence rate than fixed-degree adaptive methods. Preliminary results with one strategy for hp-adaptivity look promising. We are continuing to study other strategies.

We have demonstrated the utility of our code for single-channel, time independent models through experiments in which we examined the effect of varying the scattering length, and of varying the trap aspect ratio. We have now extended the code to handle multi-channel problems. We solved a two-channel Feshbach model of potassium and performed experiments in which we varied the magnetic field surrounding the trapped atoms, and obtained results very close to a similar experiment of Paul Julienne (NIST PL). In the coming year we will further extend and examine the multi-channel capabilities, and begin studying time dependent models.

Funding: *Quantum Information ACI Initiative (2007).*

Fundamental Mathematical Software Development and Testing

Parallel Adaptive Multilevel Finite Elements

William Mitchell
Marjorie McClain
Eite Tiesinga (NIST PL)

See feature article, page 23.

OOF: Finite Element Analysis of Material Microstructures

Stephen A. Langer
Andrew C.E. Reid
Rhonald C. Lua (Penn State University)
Valerie R. Coffman (Cornell University)
R. Edwin Garcia (Purdue University)
Olga Kuznetsova (University of Maryland)

http://www.ctcms.nist.gov/oof/

The OOF Project, a collaboration between MCSD and NIST Materials Science and Engineering Laboratory, is developing software tools for analyzing real material microstructure. The microstructure of a material is the (usually) complex ensemble of polycrystalline grains, second phases, cracks, pores, and other features occurring on length scales large compared to atomic sizes. The goal of OOF is to use data from a micrograph to compute the macroscopic behavior of the material via finite element analysis. OOF is intended to be a general tool, applicable to a wide variety of microstructures in a wide variety of physical situations.

This year the OOF team released six new versions of the OOF2 program: 2.0.1, 2.0.2, and 2.0.3, and three pre-release (alpha) versions of 2.0.4. These programs were downloaded 1,453 times from the OOF website. The final version of OOF1 accounted for an additional 675 downloads.

The OOF customer base continues to be broad. We received help requests from about 60 users in FY07, representing 17 domestic and 17 foreign academic institutions, 3 domestic and 3 foreign industrial laboratories, and 2 domestic and 1 foreign government laboratory. (Some users did not identify themselves and were not included in the statistics.)

New features introduced to OOF2 this year include the following.

Subproblems. A typical OOF2 problem is to solve for the value of one or more fields (temperature, polarization, etc) at all points within a microstructure. OOF2 now can divide the job into subproblems, limiting the solution to a portion of the microstructure (just one material type, for example) or to a subset of the defined fields.

Periodic Boundary Conditions. Most real micrographs are not periodic, but simulated microstructures are often periodic. OOF2 can now generate a periodic mesh to perform calculations on periodic microstructures. Previous versions of OOF2 could use a periodic micrograph but could not guarantee that the finite element mesh or the computed fields would respect the periodicity.

Automatic Mesh Generation. The OOF2 user creates a finite element mesh and adapts it to fit the geometry of a microstructure by interactively applying a series of software tools. Version 2.0.3 introduced a method of doing this with very little user interaction. In the new method, the user specifies the length scales of the largest and smallest microstructural features of interest and a tolerance, and the program automatically applies a series of modification operations to generate a mesh that approximates the features. The method works well on a wide variety of microstructural geometries, as long as the features are distinctly delineated.

Orientation Mapping. Beginning with version 2.0.4, OOF2 has the ability to read orientation mapping data files. These files are created by electron backscatter diffraction (EBSD) experiments, and directly give the orientation of the crystal axes at points in a micrograph. Previously, OOF users had to invent orientations for the grains of a microstructure, or laboriously enter them by hand. Users can now start with the EBSD data instead of a conventional micrograph, and use actual orientations. OOF2 has built-in support for the orientation map file formats used by the two major suppliers of EBSD equipment, and can easily be extended to use other formats.

OOF2 is still under development. Current efforts are centered on solving for time-dependent and non-linear material properties. We have also begun work on OOF3D, which will compute three dimensional physical properties from three dimensional micrographs.

Sparse BLAS Standardization

Roldan Pozo
Iain Duff (Rutherford Appleton Labs)
Michael Heroux (Sandia National Laboratory)

http://math.nist.gov/spblas
http://www.netlib.org/blas/blast-forum

MCSD has played a leading role in the standardization effort for the Basic Linear Algebra Subprograms (BLAS) and continues to be a major contributor for the design and development of reference software and documentation. The BLAS are kernels for computational linear algebra subroutines comprising of fundamental matrix/vector operations common to most scientific computing applications. By standardizing such interfaces, computer manufacturers and software vendors can provide high-performance implementations especially suited to a specific hardware platform. By developing their applications in terms of BLAS, computational scientists can achieve high levels of performance and portability. The original BLAS, which were developed for dense vector and matrix operations from the late 1970s through the early 1990s, achieved this goal very well. To this, the BLAS Technical Forum (an international consortium of industry, academia, and government institutions, including Intel, IBM, Sun, HP/Compaq/Digital, SGI/Cray, Lucent, Visual Numerics, and NAG) formed an updated BLAS standard which addresses these extensions.

Among the most significant components of the updated BLAS standard is support for sparse matrix computations. R. Pozo of MCSD served as chair of the Sparse BLAS subcommittee during the standardization process, and NIST was first to develop and release a public domain reference implementation in ANSI C for early versions of the standard, which helped shape the final specification. After the standard was formally approved and accepted, the complete technical specification was published, and a special issue of the *ACM Transactions of the Mathematical Software* (TOMS) was devoted to the new BLAS standard, including a paper co-authored by R. Pozo and other subcommittee members providing an overview of the sparse matrix interface.

This year saw the development of a new C++ interface using abstract base classes, which define a specific interface and greatly simplifies the Sparse BLAS specification. The code interface is less than 600 lines of C++, yet captures the complexity vector and matrix operations with basic types (real and complex, single or double precision) and supports general types (e.g. intervals, multiprecision variables) beyond the original BLAST specification.

TNT: Object Oriented Numerical Programming

Roldan Pozo

http://math.nist.gov/tnt/

NIST has a history of developing some of the most visible object-oriented linear algebra libraries, including Lapack++, Iterative Methods Library (IML++), Sparse Matrix Library (SparseLib++), Matrix/Vector Library (MV++), and most recently the Template Numerical Toolkit (TNT). This package has downloaded to several thousand developers and is currently in use in several industrial and commercial applications.

TNT incorporates many of the ideas we have explored with previous designs, and includes new techniques that were difficult to support before the availability of ANSI C++ compilers. The package includes support for both C and Fortran multidimensional arrays, vector, matrices, and application modules, such as linear algebra.

The design of TNT separates the interface specification from the actual implementation. This allows library developers to create specialized modules that take advantage of particular hardware platforms, utilize vendor-specific libraries, or implement different C++ strategies, such as expression templates, or instrumented versions for debugging sessions.

Developments in the latest design of TNT (v. 3.0.8) provide support for both multidimensional arrays and integrate a linear linear algebra module of fundamental algorithms (LU, Cholesky, SVD, QR, and eigenvalues), as well as sparse matrix support (with computational operations via the Sparse BLAS). Also developed this year were example codes to interface with LAPACK high performance linear algebra library.

The TNT web site provides a basic implementation for testing and development, as well as links to other library packages that utilize the TNT interface. Full documentation and source code for all TNT components are available on-line.

SciMark, a Web-based Benchmark for Numerical Computing in Java

Roldan Pozo
Bruce Miller

http://math.nist.gov/scimark

The NIST SciMark benchmark continues to be one of the most widely used Java scientific benchmarks and is in the process of being included in the new SPECjvm2007 benchmark. SciMark consists of computational kernels for FFTs, SOR, Monte Carlo

integration, sparse matrix multiply, and dense LU factorization, comprising a representative set of computational styles commonly found in numerical applications. SciMark can be run interactively from Web browsers, or can be downloaded and compiled for stand-alone Java platforms. Full source code is provided, in Java and C programming languages for comparison under different compilers and execution environments. The SciMark result is recorded as megaflop rates for the numerical kernels, as well as an aggregate score for the complete benchmark. The current results database lists submissions form more than 3,000 submissions from computational platforms ranging from palm devices to high-end servers, and contains reports from nearly every operating system and virtual machine environment currently in use, including Solaris, FreeBSD, MacOS, Sun OS, IRIX, OSF1, Linux, OS/2, and Windows 95, 98, 2K, ME, NT, and XP.

SciMark and its kernel components have become a pseudo-standard in industry and academia. They were adopted by the Java Grande Benchmark Forum, Sun Microsystems used SciMark 2.0 to demonstrate the floating-point improvements to their JVM, and most recently are being included in in SPECjvm2007 benchmark.

As of January 2007, the record for SciMark is over 1 Glops, with some of the kernel benchmarks (notably the LU factorization) running over 2 Gflops on a single-processor PC.

Mathematical Knowledge Management

Math Search

Abdou Youssef
Bruce Miller
Daniel Lozier
Ronald Boisvert

See feature article, page 40.

Associated ITL Program: *Information Discovery, Use, and Sharing.*
Funding: *NIST Systems Integration for Manufacturing Applications (SIMA) program.*

Digital Library of Mathematical Functions

Daniel Lozier
Ronald Boisvert
Joyce Conlon
Peter Ketcham
Marjorie McClain
Bruce Miller
Frank W. J. Olver
Bonita Saunders
Abdou Youssef
Charles Clark (NIST PL)
Gloria Wiersma (NIST PL)
Qiming Wang (NIST ITL)
Richard Askey (U. of Wisconsin, Madison)
Michael Berry (U. Bristol, UK)
Leonard Maximon (George Washington U.)
Morris Newman (U. California, Santa Barbara)
Ingram Olkin (Stanford)
Peter Paule (J. Kepler U., Linz, Austria)
William Reinhardt (U. Washington, Seattle)
Nico Temme (CWI, Amsterdam)

30 authors under contract
25 validators under contract

http://dlmf.nist.gov/
http://dlmf-i.nist.gov/

Mathematical functions, from the elementary ones like the trigonometric functions to the multitude of special functions, are an integral part of all modern developments in theoretical and applied science. They are used to model natural phenomena in fields from quantum theory to astrophysics, formulate problems and solutions in engineering applications, and support numerical computations. To make effective use of mathematical functions, practitioners must have ready access to a reliable catalog of their properties.

Traditionally, in all fields of science, catalogs of relevant properties have existed in the form of massive published handbooks. These are still being produced and can be found on the desks of working scientists. An important example is the 1000-page Handbook of Mathematical Functions, edited by Abramowitz and Stegun, and published in 1964 by the National Bureau of Standards. Recently, however, the Web is showing great promise as a more advantageous method. A big potential advantage is that scientists can begin to integrate handbook data into documents and computer programs directly, bypassing the need for time-consuming and error-prone reentry of data, and by use of *metadata*, providing for much richness in Web interconnections, Web annotation, Web search, and so on. Another advantage is high-resolution graphics that users can rotate and view from any angle, giving them an unprecedented way of visualizing the complex behavior of mathematical special functions.

The Digital Library of Mathematical Functions has two main goals. First, we are reviewing the published literature on special functions, selecting the properties most relevant to current applications, and publishing an up-to-date handbook of the traditional sort. The handbook will consist of 33 chapters devoted to individual classes of special functions plus 4 chapters on algebraic and analytical methods, asymptotic approximations, numerical methods, and computer algebra. The most recent comprehensive handbook of special functions is the one by Abramowitz and Stegun, mentioned in the preceding paragraph. Still in print and in widespread use, it is badly out-of-date with respect to recent mathematical research, current scientific applications of special functions, and computational methods. Our second goal is to disseminate the same information, with significant augmentations, from a Web site at NIST. The augmentations include live links to available online software and references, a math-aware search capability, a facility for downloading formulas into word processors and computer software systems, and interactive visualizations.

The project is large, and the contributors fall into several categories. The *editorial board* consists of 4 principal and 8 associate editors. They are responsible for the selection and presentation of the technical information in book and Web formats. Since the beginning of the project, the principal editors have met frequently to review progress and to make midcourse corrections when necessary. *Authors* consist of expert individuals selected for their published research achievements and their ability to write for the intended

audience of scientists, engineers and mathematicians. Their contributions are being carefully edited and, in many cases, extensively revised by the principal editors to achieve uniformity of content and presentation across all chapters. *Validators*, like the authors, consist of expert individuals selected for their research accomplishments. Their responsibility is to check the work of the authors and editors, and to recommend corrections and improvements. This is a vital step to uphold the worldwide reputation of NIST as a reliable source of accurate, useful and timely scientific reference information. The *project staff* consists of highly qualified mathematicians and computer scientists whose responsibilities, broadly, are (i) construction of a mathematical database that encodes the entire technical content of the DLMF, (ii) application of advanced visualization methods and tools that enable users to display and manipulate complex functional surfaces, (iii) development of software tools to facilitate the production of the book and Web site, (iv) research into advanced techniques for the unambiguous translation of mathematical formulas and facts among different computer systems and application of these techniques into the DLMF when appropriate, (v) research into the frontiers of technical search methodology to enable effective queries involving fragments of technical mathematics and application of these techniques into the DLMF when appropriate. The *support staff* consists of individuals capable in the use of advanced mathematics document processors, symbolic and numerical computation packages, and bibliographic tools such as those provided by the American Mathematical Society.

By the end of calendar year 2007, all chapters but one (Computer Algebra) had been validated, and resolution of the validator's reports with the author's chapters had begun. Inclusion of Computer Algebra has been deferred until the other 36 chapters have been published. Validation confirms the mathematical correctness of the chapters. Resolution involves (i) integrating the validator's recommendations into the chapters, done at NIST, and (ii) obtaining approvals from the validators and authors that each chapter is ready for publication.

In July, 2007, the beta version of the DLMF Web site was places behind the NIST firewall for internal usage and feedback. The announcement ran in the NIST Technicalendar for 6 weeks. A number of useful comments were received about the mathematical contents and, especially, the variety of rich Web interconnections we have been developing: (i) math-aware search, visualization, and cross-referencing within the DLMF, and (ii) links to external Web sites for bibliographic data, full texts of referenced articles, public-domain software, and available documentation for proprietary software.

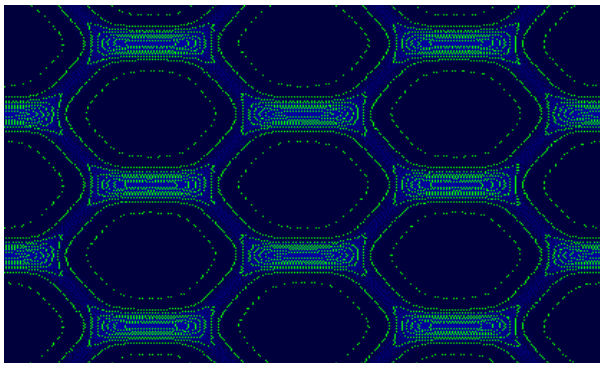


Figure 1. *The nonlinear Kadomtsev-Petviashvili differential equation (KP equation) is used to model the evolution of waves in shallow water. One family of such waves generates surface patterns that are two-dimensional, periodic, and hexagonal. These patterns are clearly visible, for example, when looking down into a swimming pool. They have been modeled with the KP equation. Shown above are level curves of an analytical solution derived in terms of Riemann theta functions. For properties of these functions see Chapter 21 of the DLMF. For a comparison of this analytical solution to experiment see Fig. 2 and the reference quoted there.*

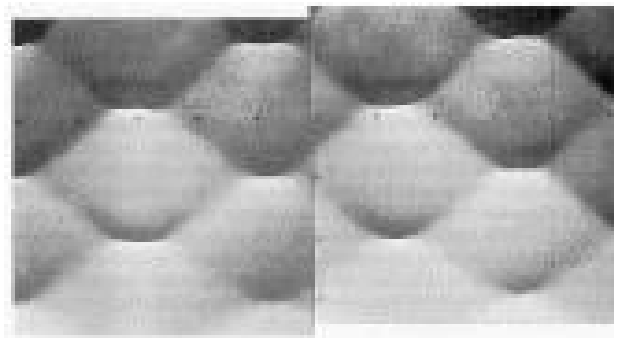


Figure 2. *Two-dimensional periodic waves in a shallow water wave tank. Taken from Joe Hammack, Daryl McCallister, Norman Scheffner and Harvey Segur, "Two-dimensional periodic waves in shallow water. Part 2. Asymmetric waves," J. Fluid Mech. (1995), Vol. 285, pp. 95-122. The caption reads "Mosaic of two overhead photographs, showing surface patterns of waves in shallow water".*

A number of innovative features will be present in the online DLMF. For example, in the DLMF, every symbol encountered in display equations is linked to its definition. To illustrate this, consider the equation

$$sn(x,k) = \sin(am(x,k)).$$

With just two clicks a Web user is able to "call up" the definitions of sn (a Jacobian elliptic function), x and k (real variables), sin (the trigonometric function), and am (the Jacobian amplitude function). In contrast, a book user would have to scan the text for the definitions or consult the index. In the DLMF this equation can be located using the search tool with a simple query like "sn am". It appears high up on the hit list.

The remaining work for the public release of the DLMF is expected to be complete in 2008. The release schedule (subject to change) is

March 31	Partial public release of Web site
June 30	Selection of publisher
Sept. 30	Delivery of files to publisher
Dec. 31	Full public release of Web site
	Book in print

Funding: *National Science Foundation. NIST Systems Integration for Manufacturing Applications (SIMA) program.*

Cultivating (Legacy) Mathematical Data

Bruce Miller
Michael Kohlhase (Jacobs University, Bremen)
Heinrich Stamerjohanns (Jacobs University, Bremen)

When the Digital Library of Mathematical Functions project was begun, a key issue was selecting the document format to use for preparing the material. The format had to support extensive and complex mathematical content, high-quality typesetting, and conversion to standard web formats. Above all, it had to be familiar to and convenient for the 40-50 authors, validators and editors who would be working on the project. One format so handily met all of these constraints but one, that LaTeX became the uncontroversial choice.

But, when no then-available system was sufficient to losslessly convert the DLMF source materials to XML and MathML, a parallel project was started to create one in-house, soon dubbed LaTeXML. A fruitful period of quiet but intense development ensued where LaTeXML and DLMF were able to leverage each other's strengths. Eventually, it became clear that this was a rather capable tool with broader application; the hammer thus set out in search of nails --- it didn't have to search for long.

While LaTeX may not be the universal document format, it certainly is the most broadly used within most of the scientific world, particularly where mathematics is involved. In fact it is actively used in those communities and a vast amount of so-called legacy material exists in that format. For example, the arXiv (http://arxiv.org/) contains almost a half million e-prints virtually all coming from LaTeX sources. Thus, beyond its obvious application to authors who wish to convert and present articles on the web, LaTeXML might serve to convert the arXiv to a more broadly usable XML form. Yet another DLMF spin-off, then, was the mass conversion of the arXiv to the generation of a large test suite to support the development of math-aware search engines and similar research in Mathematical Knowledge Management. The long-term goal of such work is that eventually scientists and engineers will be able to enter mathematics in search engines and find relevant articles as easily as they search for text terms today. Eventually, mathematical software, computer algebra systems, and similar may execute such searches automatically, behind the scenes, to accomplish their work.

During this year, the group at NIST, in collaboration with a group at Jacobs University in Bremen (formerly International University Bremen), began a project to apply LaTeXML to the arXiv corpus. A flexible mass-conversion and reporting system was developed at Jacobs and was applied to the task. A particular strength of the system is that it collects and summarizes conversion problems. This led to several iterations of refinement of the LaTeXML program --- it should be noted that LaTeX is both a powerful system and a quirky one that seemingly invites peculiar authoring idioms. At this point, the conversion of almost 300 thousand documents has been attempted, with over half being converted without error to intermediate XML form, and with almost 200 thousand converted to XHTML and MathML. The resulting improvements have, of course, come back to benefit the DLMF project.

Future work will focus on refining and improving the conversion rate, and getting broader coverage of the document styles used by various journals and publishers. Of course, we will also begin exploring the opportunities provided by the large data set.

The emphasis to this point has been on the presentational aspects of the mathematics. Future efforts will also be directed at developing methods and heuristics to synthesis a more semantic view of the mathematical material. This will greatly open the horizons for research into search methods and knowledge management.

Related to this work, Bruce Miller gave presentations at the Mathematical Knowledge Management Conference at Linz in June and at the International Congress on Industrial and Applied Mathematics at Zurich in July. Finally, Bruce Miller and Michael Kohlhase have been active participants in the W3C Math Working Group developing the third version of the MathML standard.

Associated ITL Program: *Information Discovery, Use and Sharing.*
Funding: *NIST Systems Integration for Manufacturing Applications (SIMA) program.*

Visualization of Complex Function Data

Bonita Saunders
Qiming Wang (NIST ITL)
Sandy Ressler (NIST ITL)
Daniel Lozier
Frank Olver

Effective visualizations can aid a researcher's understanding of high level functions that arise as solutions to problems in the mathematical and physical sciences. Consequently, a considerable effort has been devoted to developing state-of-the-art interactive 3D visualizations for the NIST Digital Library of Mathematical Functions. Using techniques from numerical grid generation and web-based visualization technologies such as the Virtual Reality Modeling Language (VRML) and X3D, we have completed over 200 3D visualizations for the digital library. A sample chapter on the gamma function is available externally at http://dlmf.nist.gov/Contents; internally, see http://dlmf-i.nist.gov/.

With preliminary visualizations completed for all chapters, we are now focusing on improving presentation quality while maintaining plot and data accuracy. Plot accuracy concerns the visual display of the data. We have used numerical grid generation techniques to create computational grids that successfully capture key surface features such as zeros, poles, and branch cuts, but we continue to work on adaptive enhancements to improve the rendering of surfaces with steep gradients. While primarily using VRML and X3D, we have also developed prototypes using alternative technologies such as embedded 3D and video capture to make the visualizations more accessible to users on platforms where no suitable VRML/X3D plugin is currently available.

To validate data accuracy we are computing the function values using codes from at least two independent sources, including standard computer algebra packages such as Mathematica and Maple, routines from repositories such as the NIST Guide to Available Mathematical Software (GAMS, http://gams.nist.gov/), and personal FORTRAN and C codes from the chapter authors themselves. Validation of static 2D image data is almost complete, and validation of data files for 3D visualizations has begun.

As the visualization work for the current phase of the NIST DLMF winds down, we will concentrate more on extending our work to on-demand generation of mathematical graphs and tables. This may be a stand-alone capability or ultimately be incorporated as part of the DLMF. We are searching the web for sites that may offer some of these capabilities and looking at the possibility of extending work on the live generation of tables completed by former NIST post-doctoral associate and faculty appointee Bruce Fabijonas.

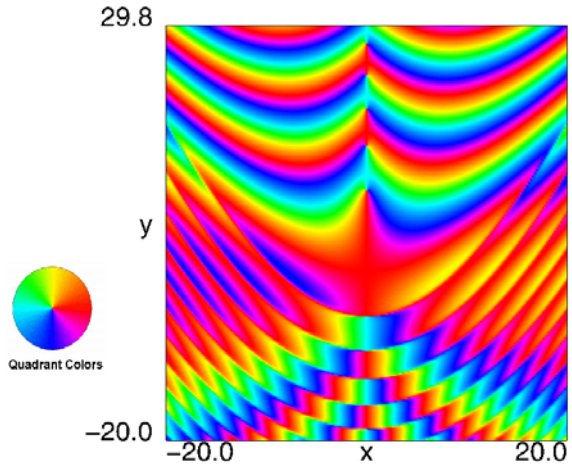


Figure 1. *Density plot of phase of swallowtail canonical integral function* $\Psi_3(x,y,3)$.

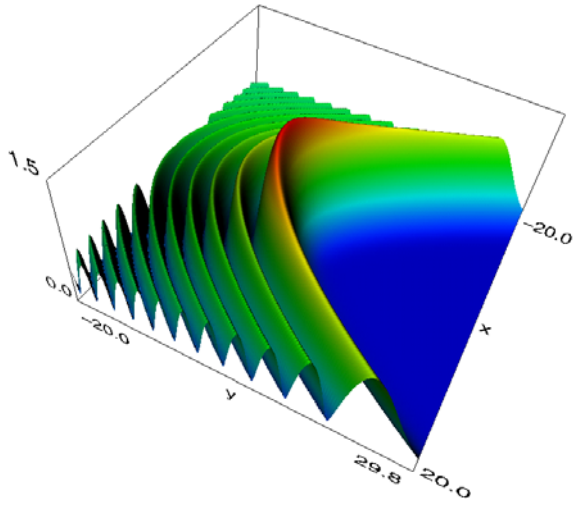


Figure 2. *Modulus of swallowtail canonical integral function* $|\Psi_3(x,y,3)|$.

[1] Q. Wang, B. Saunders, and S. Ressler, "Dissemination of 3D Visualizations of Complex Function Data for the NIST Digital Library of Mathematical Functions," *CODATA Data Science Journal* **6** (2007) (Supplement on the *Proceedings of the 20th International CODATA Conference*, Beijing, China, October 2006), pp. S146, S154.

[2] B. Saunders and Q. Wang, "From B-Spline Mesh Generation to Effective Visualizations for the NIST Digital Library of Mathematical Functions," *Curve and Surface Design: Avignon 2006, Proceedings of the Sixth International Conference on Curves and Surfaces,* Avignon, France, pp. 235-243.

High Performance Computing

Modeling the Rheological Properties of Suspensions

William George
Julien Lancien
Steve Satterfield
Marc Olano
Nicos Matrys (NIST BFRL)

See feature article, page 30.

Funding: *Virtual Cement and Concrete Testing Laboratory (VCCTL) Consortium.*

Computation, Visualization of Nano-structure and Nano-optics

James Sims
William George
Terrence Griffin
John Hagedorn
John Kelso
Marc Olano
Adele Peskin
Steve Satterfield
Judith Terrill
Garnett Bryant (NIST PL)

See feature article, page 33

High Precision Hy-CI Variational Calculations on Small Atomic Systems

James S Sims
Stanley A Hagstrom (Indiana University)

The ultimate goal of this work is to produce wave functions and energy levels of few electron atoms and ions with a precision that goes well beyond what can be achieved experimentally. In the past two decades, there have been breathtaking improvements in computer hardware and innovations in mathematical formulations and algorithms, opening up the real possibility that theoretical successes for the two electron atomic problem (He) and the two electron molecular problem (H2) can be extended to atoms (and molecules) with more than two electrons.

There are formidable bottlenecks to accomplishing this goal, as evidenced by the fact that in going from He (two electrons) to Li (three electrons) to Be (four electrons), the situation vis a vis high precision calculations degrades to the point that there are no calculations of the ground or excited states with an error of less than present day experimental uncertainty. This project attempts to remove the bottlenecks with a combination of mathematical formulations and algorithms revolving around a mathematical technique pioneered by Sims and Hagstrom many years ago (the Hylleraas-Configuration Interaction (Hy-CI) technique), and the use of extended precision, parallel computing, and efficient techniques for integral calculation and storage.

A notable accomplishment this year has been the completion of the second of three studies undertaken on math and computational science issues in Hy-CI calculations. “Math and computational science issues in high-precision Hylleraas-configuration interaction variational calculations II. Kinetic energy and Nuclear attraction integrals” has been published in J. Phys. B. At. Mol. Opt.Phys. A third (and final) study in the series, dealing with the requisite four-electron integrals, is nearly completed. This study has entailed the development of a new technique for the rapid evaluation of four-electron integrals utilizing recursion relations. Techniques for storing and retrieving the integrals as part of the matrix element assembly, the hardest part of the problem, are being devleoped. A hashing scheme has been worked out for this which, in combination with the recursion relations, resulted in 238,052 four electron integrals being calculated in 4.99 seconds in a 300 term Hy-CI calculation. These numbers are a validation that the recursion techniques are the only viable way to handle these integrals in a high-precision calculation.

The combination of these achievements this year and previous efforts culminating in an extended-precision parallel dense matrix generalized eigenvalue solver, should lead to a benchmark calculation on the Beryllium atom (the prototype for high precision atomic calculations for atoms with four or more electrons), resulting in theoretical predictions as accurate (or better) than modern experiments.

[1] James S Sims and Stanley A. Hagstrom, “Combined Configuration-Interaction-Hylleraas-Type Wave-Function Study of the Ground State of the Beryllium Atom,” *Physical Review A* **4**, 1971, pp. 908-916.

[2] James S. Sims and Stanley A. Hagstrom, “High Precision Hy-CI Variational Calculations for the Ground State of Neutral Helium and Heliumlike Ions,” *International Journal of Quantum Chemistry* **90** (6), 2002, pp. 1600-1609.

[3] James S. Sims and Stanley A. Hagstrom, “High Precision Variational Born-Oppenheimer Energies of the

Ground State of the Hydrogen Molecule," *Journal of Chemical Physics* **124**, 2006, 094101.

[4] James S. Sims and Stanley A. Hagstrom, "Math and Computational Science Issues in High-precision Hy-CI calculations II. Kinetic Energy and Electron-nucleus Interaction Integrals," *Journal of Physics B: Atomic, Molecular, and Optical Physics* **40**, 2007, pp. 1575-1587.

Screen Saver Science

William L. George
Julien Lancien

http://math.nist.gov/mcsd/savg/parallel/screen

The Screen Saver Science (SSS) project aims to develop a computing resource composed of a heterogeneous set of PCs, scientific workstations, and other available computers, that can be easily used by scientists to execute large highly distributed, compute intensive applications. Each individual computer makes itself available for participating in a computation only when it would otherwise be idle, such as when its screen saver is running. SSS is based on Jini, an open software architecture built on Java and intended for the development of robust network services. There are several goals to this project. First, we hope to utilize the idle processing power of the many PCs, workstations, and cluster nodes we have available here at NIST to execute production scientific codes. The compute power of personal PCs and workstations continues to increase and they have become increasingly capable of executing large compute intensive applications due to faster processors and larger main memories. Second, the SSS computing environment will allow us to develop and experiment with new highly parallel and distributed algorithms more suitable for emerging grid environments. Finally, the use of Java for scientific applications is of interest in general, and so the development of applications for SSS will give us the opportunity to explore this topic on actual production quality applications. Up until recently, this type of project would have required a large investment in software development just to become minimally functional and so was not practical, especially for a small team of programmers. However, with the introduction of Jini, and more specifically the Jini based network service called Javaspaces, the most difficult parts of this project have now become straightforward. Javaspaces is a portable, machine independent, shared memory system that expands upon the tuple-space concepts developed in the 1980s by David Gelernter of Yale University.

Over the past year we have developed a system of scripts to manage SSS workers on the compute nodes of the ITL/PL Linux cluster Raritan, thus enabling us to utilize every free cycle. We have also enhanced SSS in several basic ways. The user interface, specifically for submission of SSS computations to the Raritan cluster, and the retrieval of results has been simplified. The SSS infrastructure has been improved to simplify its building, packaging, and distribution, and to better support application development by multiple researchers

A user interface to SSS, to be used by those submitting computations to the SSS system as well as those simply allowing SSS tasks to run on their machine, has been completed. The interface is web-based, built on the 'Ruby-on-Rails' framework and uses jruby, a Java implementation of the Ruby scripting language. Using jruby, instead of the standard C-based Ruby implementation, allows us to more easily interact with both the Rails framework and the SSS system.

The OS specific parts of SSS have been reduced to a minimum. For example, we now use the Java Installer Builder, IzPack (http://izpack.org), to create a simple cross-platform (Linux, Mac, MS Windows) installer for SSS.

W. George and Guest Researcher J. Lancien have completed the initial draft of the SSS Users Manual and Administrators Guide.

Computational Modeling of Cement Paste Hydration and Microstructure Development

William George
Edith Enjolras
Jeffrey Bullard (NIST BFRL)

http://math.nist.gov/mcsd/savg/parallel/hydration/

Understanding the complex chemical changes that occur when cement powder is mixed with water is a long-standing but extremely challenging technological goal. Fundamental computational modeling of the hydration of cement is difficult because it involves a large number of coupled nonlinear rate equations that must be solved on a random three-dimensional spatial domain.

To address these challenges, we are applying a new computational model called HydratiCA which has several advantages over previous attempts to model cement paste hydration. HydratiCA is based on concepts of transition state theory and uses stochastic cellular automaton algorithms to simultaneously model 3-D reaction and transport phenomena. This allows us to track the detailed kinetics and equilibria that occur in a diverse range of cementitious systems. At the length scales required to finely resolve the reaction mechanisms and microstructure changes in cement paste,

HydratiCA must take small time steps (approximately 10^{-5} seconds) to remain numerically stable. In other words, tens of millions of time steps are required to simulate the behavior of cement paste for just one hour. Therefore, parallelization of the model is important for modeling systems that are large enough to be realistic, avoiding finite size effects, and still be able to complete the simulations in a reasonable amount of time.

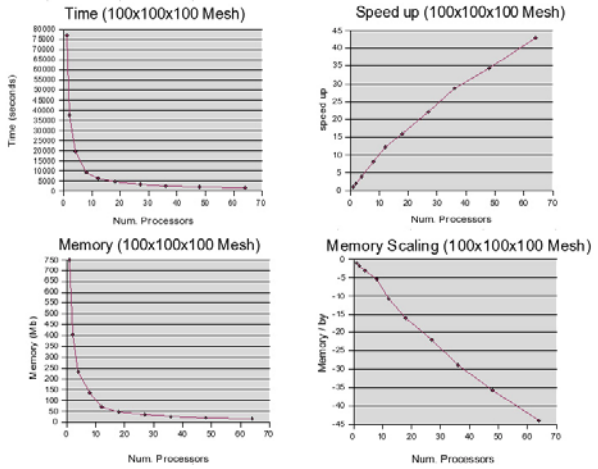


Figure 1. *Performance of the parallel HydratiCA code for a 100x100x100 mesh*

This year we parallelized the HydratiCA code using a spatial decomposition. This code has been tested on the Raritan linux cluster and shown to scale well. Performance of the parallel code can be seen in the figure below. NASA has granted us 200,000 hours of CPU time on its Columbia supercomputer at NASA Ames research Laboratory this year to study the scalability of the code, and to perform validation tests.

[1] E. Enjolras, W. George, J. Bullard, and J. Terrill, "Parallelization of HydratiCA," presented at the VCCTL semi-annual consortium meeting, Nov. 2007, NIST Gaithersburg.

[2] J. Terrill, W. George, T. Griffin, J. Hagedorn, J. Kelso, M. Olano, A. Peskin, S. Satterfield, J. Sims, J. Bullard, J. Dunkers, N. Martys, A. O'Gallagher, and G. Haemer, "Extending Measurement Science to Interactive Visualization Environments", Chapter in *Trends in Interactive Visualization: A-State-of-the-Art Survey*, edited by Elena Zudilova-Seinstra, Tony Adriaansen and Robert van Liere , to be published by Springer , UK.

Funding: *Virtual Cement and Concrete Testing Laboratory (VCCTL) Consortium.*

Physics Models for Transport in Compound Semiconductors

Miguel Rios
Adele Peskin
Terence Griffin
Judith Terrill
Herbert S. Bennett (NIST EEEL)

Physics models for carrier transport in semiconductors are essential inputs of computer programs that simulate the behavior of nanoelectronic and optoelectronic devices. Such simulations increase understanding, reduce times-to market, and assist in making selections from among competing or alternative technologies. As devices shrink in size to nanometers, performing experimental measurements becomes more costly and time-consuming. This means that computer simulations will become more essential for advances in future nanotechnologies.

Unlike many physics models that are based on using variations in parameters to fit experimental data, the NIST physics models developed in this project are based on quantum mechanical calculations with no fitting parameters to account for dopant ion effects and many-body physics effects. The calculations include many-body quantum effects and bandgap narrowing due to dopant ion carrier interactions. The many-body quantum effects treat both electron-electron and electron-hole interactions. The results are unique because other reported treatments for the electric susceptibility

(1) Do not treat these effects self-consistently,
(2) Are Taylor series expansions in either the ratio $R = Q^2/A$ or the inverse of R, where Q is the magnitude of the normalized wave vector and A is the normalized frequency used in the measurements, and
(3) Do not give the structure shown in the Fig. 1.

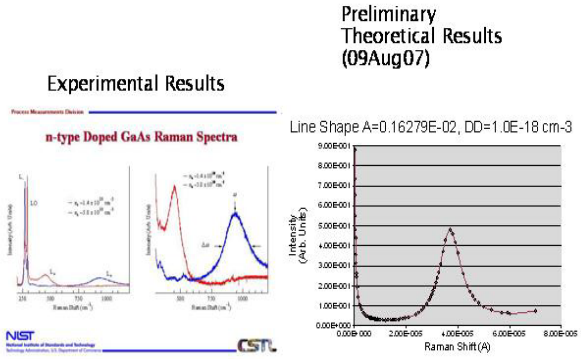


Figure 1. *Comparison of theoretical and experimental results.*

These results will change the way researchers and process engineers interpret non-destructive measurements to extract the carrier concentrations of GaAs

wafers. The wafer carrier concentration is a key figure of merit associated with a go/no-go decision for determining whether a wafer meets specifications and should undergo further processing.

This year the code was made completely open source. Also, a “first time ever” visualization of the theoretical complex electric susceptibility and Raman's line shape was computed and visualized. The visualization helps scientists to reduce the range of their data, reducing the time and cost of the experiments. This could lead to major improvements in the electronics and semiconductors fields; especially, for compound semiconductor nano-layers on silicon wafers. The comparison with experiment is shown in Fig. 1.

High Performance Visualization

Error Correction for Electromagnetic Motion Tracking Devices

John Hagedorn
Steve Satterfield
John Kelso
Whitney Austin
Judith Terrill
Adele Peskin

See feature article, page 36.

Computation, Visualization of Nanostructure and Nano-optics

James Sims
William George
Terrence Griffin
John Hagedorn
John Kelso
Marc Olano
Adele Peskin
Steve Satterfield
Judith Terrill
Garnett Bryant (NIST PL)

See feature article, page 33

Virtual Cement and Concrete Testing Laboratory

William George
Terence Griffin
John Hagedorn
John Kelso
Julien Lancien
Adele Peskin
Steve Satterfield
James Sims
Judith Terrill
Clarissa Ferraris (NIST BFRL)
Edward Garboczi (NIST BFRL)
Nicos Martys (NIST BFRL)

The NIST Building and Fire Research Laboratory (BFRL) does experimental and computational research in cement and concrete. MCSD has an ongoing collaboration with them to develop highly efficient parallel implementations of their modeling codes and in creating visualizations of their data. This work is done in the context of the Virtual Cement and Concrete Testing Laboratory (VCCTL) Consortium which we helped form in 2001. The NIST-led consortium consists of eight industrial members: BASF Admixtures (MBT), Ready Mixed Concrete (RMC) Foundation, Association Technique l'Industrie des Liant Hydrauliques (ATILH), National Stone Sand and Gravel Association (NSSGA), W.R. Grace, Sika Technology AG, Verein Deutscher Zementwerke eV (VDZ), and the Portland Cement Association. The overall goals of the consortium are to develop a virtual testing system to reduce the amount of physical concrete testing and to expedite the overall research and development process. It is expected that this will result in substantial time and cost savings to the concrete construction industry as a whole. MCSD continues as an active participant in the VCCTL.

Funding: *Virtual Cement and Concrete Testing Laboratory (VCCTL) Consortium.*

Visualization of Cement Paste Hydration and Microstructure Development

Steve Satterfield
William George
Edith Enjolras
Jeffrey Bullard (NIST BFRL)

http://math.nist.gov/mcsd/savg/vis/hydration/

When cement powder is mixed with water, complex chemical and microstructure changes occur. This is termed hydration. The NIST BFRL is developing a new a new computational model called HydratiCA. Visualization of the output of this model is important first for validation as model development proceeds. It is also important to facilitate understanding the distribution of phases in 3D. The output of the hydration numerical simulation is a 3D volume of data with percentage values for each of multiple material phases at each voxel location. Over the course of the simulation time, a series of data volumes is produced at the time intervals of interest. For each data set, an over all volume fraction is computed for each phase and plotted as a 2D graph. From the volume fraction values, a series of isosurface values is computed and also displayed as a 2D graph. See Fig. 1)

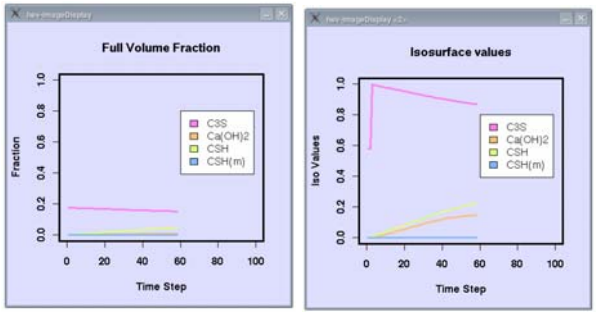


Figure 1. *Full Volume Fraction Plot (left) and Isosurface Values Plot (right)*

For each material phase a series of isosurfaces is generated. These time series of isosurfaces are combined (Fig. 2) into a 3D animation utilizing DIVERSE (diverse.sourceforge.net) and in-house developed software. These components (3D animation, 2D plots, interactions) are combined into a complete application for interactive exploration and analysis in the immersive visualization environment by the domain scientists.

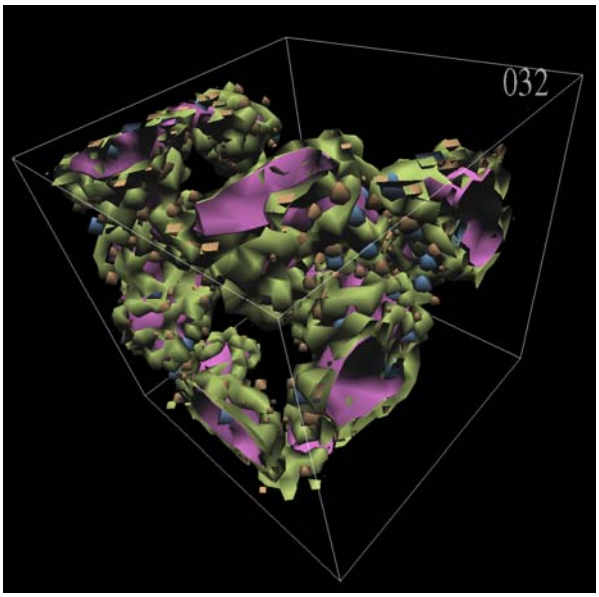


Figure 2. *Four combined phases (components of cement paste) at timestep 32.*

Future work in the area will evolve this application into a *Virtual Cement Analysis Probe (VCAP)*. The immersive visualization environment will be used to interactively probe the data and create application specific analysis and measurements. Additional software enhancements will allow alternate data representations such as volume rendering to augment the current isosurface representation.

Funding: *Virtual Cement and Concrete Testing Laboratory (VCCTL) Consortium.*

Three-D Desktop

John Kelso
Terence Griffin
John Hagedorn
Marc Olano
Adele Peskin
Steven Satterfield
Judith Terrill

http://math.nist.gov/mcsd/savg/vis/desktop

Immersive visualization systems have many advantages, but not all users have ready access to these systems. Additionally sometimes it's easier to look at new data on a desktop in the office rather than go down the hall to use the immersive system, and sometimes the immersive system is not available at all due to distance, time or other constraints.

To this end, SAVG is developing the "3D-desktop", a set of tools to help users more effectively use our immersive software in a desktop environment. These tools can not completely bridge the gap between desktop and immersive systems, but can help ameliorate the situation. Some of the tools will be general and can be used in any application, and some will be geared towards a specific application or type of application.

Our immersive software development API, DIVERSE, automatically allows applications to be used in both immersive and desktop systems. But, the API does not address user interface questions that are raised due to the differences between the two types of systems. Therefore our tools do not need to modify the underlying immersive application, but merely replace the immersive interaction techniques with ones more suitable for desktop use, and augment these techniques with additional tools to make up for the lack of intuitive presence necessitated by the desktop system.

Our progress so far on this project consists of the following general purpose tools:

1. ***Display Control Panel***. The display control panel allows users to:
   - Choose the type of frustum projection; parallel or perspective. A perspective view helps the user get a sense of depth, and a parallel view lets the user more easily determine the geometry of their data.
   - Modify the field of view of the perspective view, or width of the parallel view. The field of view can reduce or exaggerate the distortion of the geometry due to the perspective transformation.
   - Specify the distance to the near and far clipping planes. The near clipping plane discards all geometry in front of the plane, and the far clipping plane discards all geometry behind the plane.

Specifying their distance can help the desktop user get a sense of the relative depths of objects in the application.

2. ***3-View Displays.*** The 3-view display tool creates three additional windows, each sighting along each of the three major axes. Each of the X, Y and Z displays has its own control panel, a superset of the general display control panel, additionally allowing the user to specify the distance from the origin and position in the plane orthogonal to the line of sight. A center button will automatically center the geometry in the window.

3. ***Grids.*** The grid tool enhances the sense of depth by creating grids of markers in the virtual environment indicating the position of the wand, or any other positional object in the virtual environment. In addition, a roller can be used to specify the scale of the grid, making it finer or courser as needed for the application. Four types of grids can be independently selected.
   - *exact,* which displays the exact position of the wand with a red cross hair
   - *nearest,* which displays a bright cross-hair when the wand is close to a grid point
   - *proximity,* which displays a 5x5x5 grid of cross-hairs around the grid point nearest to the wand
   - *background*, which displays a static field of small cross-hairs, which can be used to determine the position of objects in the virtual environment by use of geometric occlusion.
   - unit cube, which displays a set of nested cubes centered around the nearest grid point
   - axes, which displays a set of five axis objects-the objects are centered and equally spaced along the +Y axis (going into the screen)

4. **hev-pickerRollers**. This tool allows the user to manipulate a generic picker object such as used by the grid too. The tool is a GUI control panel containing:
   - rollers for each of the picker's six degrees of freedom, and a numerical display for each roller. The displays give the screen or data coordinates of the picker, as specified below
   - radio buttons to select whether the picker moves in the screen's coordinate space, or the data's coordinate space.
   - a button to indicate if the picker should move with navigations in the virtual environment (i.e., move with the model) or remain stationary.

Fig. 1 shows a tape measure in a proximity grid with the exact position of the wand shown as a red crosshair.

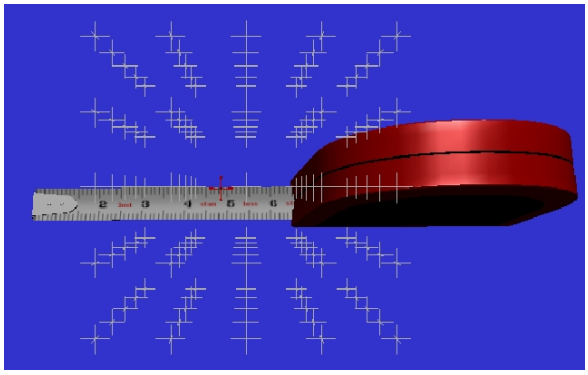


Figure 1. *A tape measure in a proximity grid with the exact position of the wand shown as a red cross hair.*

We have also developed application specific slider tools for the lung cancer project. The lung tumor sliders position a virtual lung tumor in an isosurface representing a lung. In the immersive system the user merely grabs and moves the tumor to the desired location by use of the tracked wand. Our desktop tool uses a set of sliders which move the tumor in any axis. A radio button allows the user to choose whether the tumor moves in the screen's coordinate space or in the data's coordinate space.

The 3D desktop project will continue to explore new user interface and tools to help enhance the immersive capabilities of desktop users.

Visualization of Network Dynamics

John Hagedorn
Cedric Houard
Dong Yeon Cho
Judith Terrill
Fern Hunt
Kevin Mills (NIST ITL)
E. Schwartz (NIST ITL)

http://math.nist.gov/mcsd/savg/vis/network/index.html

As part of the ITL Complex Systems program, we are creating visualizations and developing a software system that enables researchers to data visualizations from their model output. The purpose of these visualizations is to provide better understanding of the grid computing and network models. We have created a series of static and animated visualizations of network simulation data (Fig. 1). We have developed a data visualization tool that handles data from the grid computing models (Fig. 2). This tool is flexible and adaptable to a wide variety of types of data. We have also been using machine learning tools to explore the network simulation data. Fig. 3 is an example of using a hierarchical clustering tool on this data. Future work

will include the enhancement of the grid model visualization tool to provide additional modes of visualization, interactions, and data derivations. We will also be developing a similar flexible tool for visualizing time-dependent network model data. Machine learning tools will continue to be used to explore the data.

Associated ITL Program: *Complex Systems.*

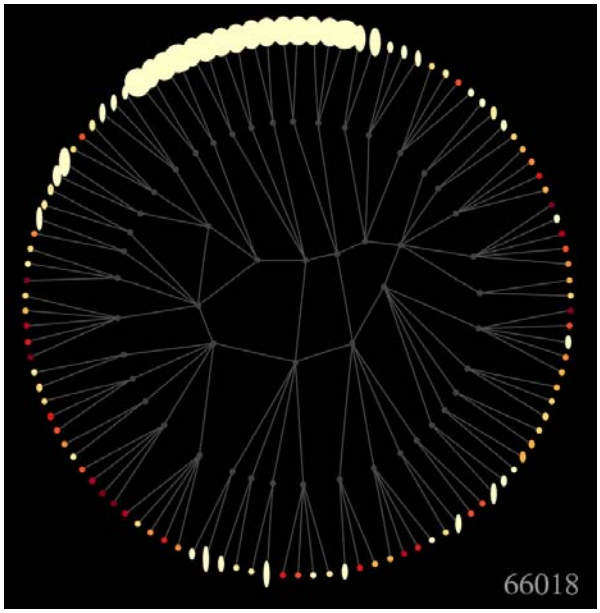


Figure 1. *Visualizations of network simulation data*

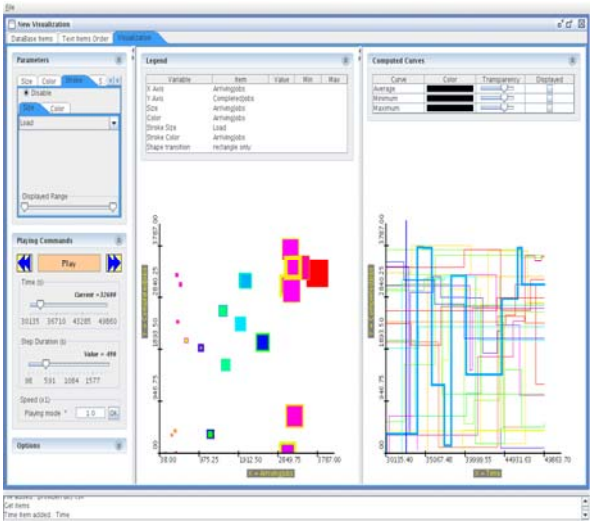


Figure 2. *Data visualization tool that handles data from the grid computing models*

Figure 3. *An example of using a hierarchical clustering tool on network simulation data.*

Monitoring Change in Lung Tumors

Adele Peskin
John Kelso
Terence Griffin
Judith Terrill
Alden Dima (NIST ITL)
Karen Kafadar (CU Denver)

http://math.nist.gov/mcsd/savg/vis/imaging/index.html

Detection of lung tumors by computerized tomographic (CT) imaging presents a number of challenges. There are currently a large number of different methods in use for measuring tumor growth by comparing CT scans taken at different points in time. A scarcity of useful data in the medical imaging community has led to variable results for tumor growth measurements. The work in our group for this project is two-fold: we are developing a common database of medical images containing tumors of known size on which to test measurement techniques, and we are working on a new method of measuring tumor growth using this data.

Our new tumor data is being created by embedding phantom tumors of known size into existing lung data. The starting data is a set of CT slices from the Cornell University lung database. A sample slice of this data is pictured in Fig. 1. We have developed a set of tools to visualize this lung data in our visualization system, both in the immersive environment and on the desktop. A 3D version of 40 slices of this data is shown in Fig. 2, a visualization of an isosurface created from the pixel values of the images.

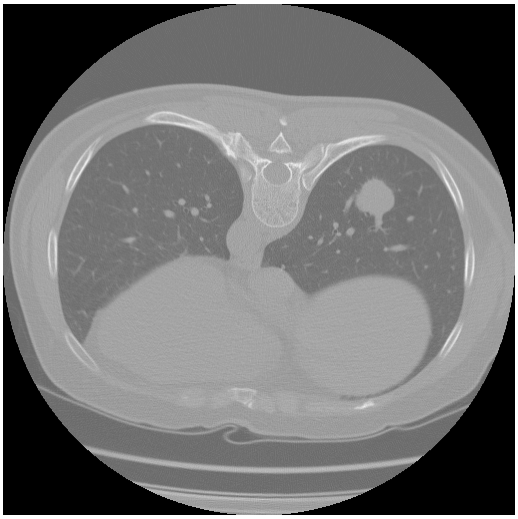


Figure 1. *Original Cornell lung data*

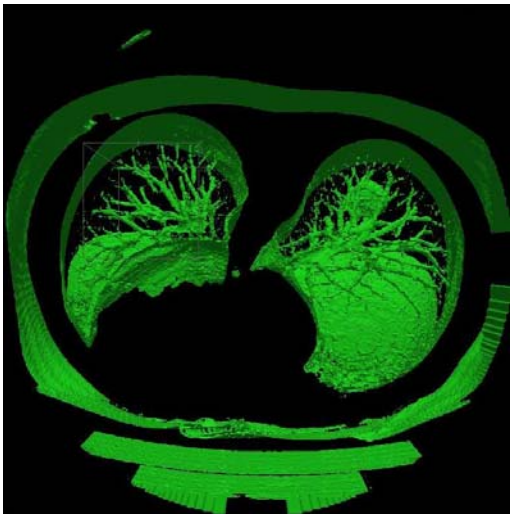


Figure 2. *Isosurface of 40 slices of this data*

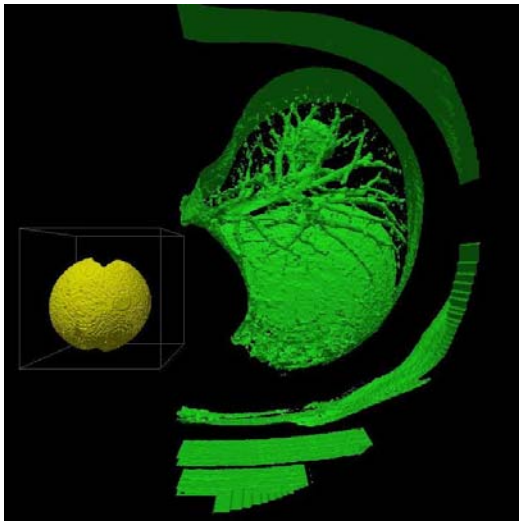


Figure 3. *Cornell lung data and the FDA phantom tumor.*

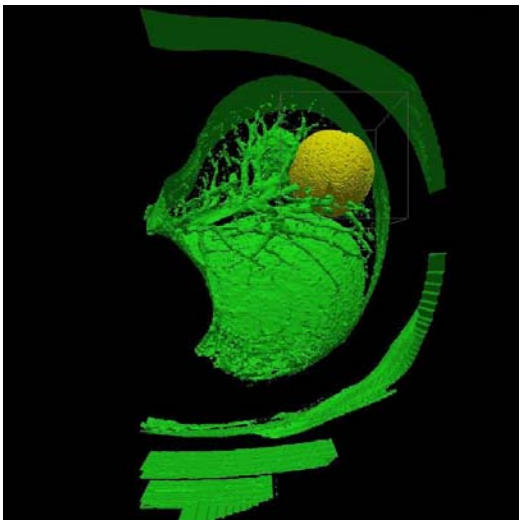


Figure 4. *Insertion*

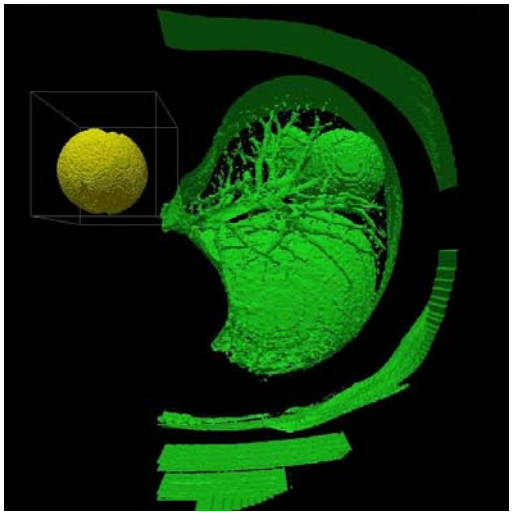


Figure 5. *Isosurface of the new lung data*

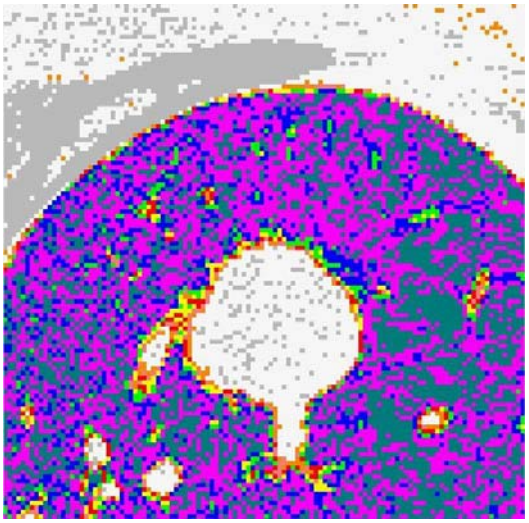


Figure 6. *Pixel distribution color-coded in one slice of the Cornell data set*

We have developed the tools to embed phantom tumors into this data, and output the modified slices in the same format (DICOM) in which we received them. An example of a modified image is shown in Figs. 3, 4, and 5, in which an isosurface from a phantom tumor from the FDA is shown, the phantom tumor is inserted into the data, and the data is then updated.

In order to make accurate measurements of tumors in this lung data, we are studying the pixel distributions in and around the tumors in the CT data, to understand which pixels can be said to be statistically inside the tumor and which pixels outside. Again, we use our visualization tools to help locate areas of specified pixels values, designated by a series of different colors in the image shown in Fig. 6 of the tumor located in Fig. 1. By understanding where areas of noise are located and the range of pixels values at tumor surfaces, we can characterize the shapes of these objects. The goal is to develop a method to measure tumor properties without requiring input from people reading the CT data, which often varies significantly from person to person.

Associated ITL Program: *Information Discovery, Use and Sharing*.

Mathematical Applications: Mechanical Systems and Processes

Application of Optimization Techniques to Design for Multi-Hazard Conditions

Florian Potra
Emil Simiu (NIST BFRL)

Traditionally, structures subjected to more than one potential hazard (e.g., wind and earthquake) have been designed independently for each individual hazard. This assures their safety under any of the potential hazards. However, techniques are currently being developed allowing designs to exhibit synergies, in the sense that design features that assure safety under any one of the hazards are utilized to improve performance under all hazards for which the structure must be designed. These techniques do not typically result in optimal designs with respect to cost or reliability.

Figure: *This solar energy collection array illustrates the type of structure being considered in this study.*

This project, which has just gotten underway, is aimed at developing an approach to design for multi-hazard conditions that assures optimality of the design for each individual hazard, as well as assuring optimality across the multiple hazards. To our knowledge this type of integrative optimization approach has never before been used in a structural engineering context. To test and illustrate the potential of our approach the example of a structural assembly developed as part of a novel renewable energy system is being considered. The assembly includes columns supporting water-filled pipes, heated by sun rays reflected onto the pipes by large arrays of computer-guided mirrors. Under constraints guaranteeing that the column design assures safety against buckling under the vertical loads transmitted by the pipes, modern mathematical optimization techniques are being applied to obtain optimal designs under (1) wind forces dependent upon the angle between the wind direction and the axis of the pipes, (2) direction-independent seismic forces, and (3) wind forces or seismic forces characterized by specified probabilities of exceedance.

Funding: *NIST Innovations in Measurement Science.*

Instability in Pipe Flow

G. B. McFadden
D. L. Cotrell (Lawrence Livermore National Lab)
B. J. Alder (Lawrence Livermore National Lab)

This study explores a classical problem in the stability of the flow of flow of fluid through a cylindrical pipe. This type of flow is relevant to a number of important applications, ranging from material transport in pipelines to blood flow in veins and arteries. The flow is driven by pressure gradients along the direction of the pipe, and is retarded by the effects of friction on the walls of the pipe. In general the nature of the flow depends on the ratio of inertial forces to viscous forces in the fluid, as measured by the dimensionless Reynolds number. For small Reynolds numbers (high relative viscosity) the flow is laminar. For large enough Reynolds numbers, however, the laminar flow is observed to become unstable, leading to more complicated ("secondary") flows. The resulting turbulence in the flow field gives rise to additional phenomena such as enhanced mixing and an increase in the pressure gradient required to drive the flow. The prediction of the onset of instability is often possible by employing linear stability theory to determine critical modes that grow temporally or spatially; such studies have been successful for a number of flow geometries, but not for the case of pipe flow.

For an idealized pipe of infinite length and circular cross-section a simple one-dimensional flow can be found from the underlying equations of motion whose stability can be examined through numerical means. There is a long known contradiction between the resulting linear stability results and the experimental observation that such flows become unstable at a Reynolds number of about 2000 for ordinary pipes. A hint that wall roughness may be important can be gathered from experiments which show that for smoothed pipes the onset of the instability can greatly exceed 2000. In this work we investigate the possibility that

the use of stick (or no-slip) boundary conditions in modeling the flow is responsible for the discrepancy between theory and experiment, because such models ignore the amplitude variations associated with the roughness of the wall. Once that length scale is introduced (here, crudely, through a corrugated pipe), linear stability analyses leads to stable vortex formation at low Reynolds number above a finite amplitude of the corrugation, and unsteady flow at a higher Reynolds number, where indications are that the vortex dislodges. Remarkably, extrapolation to infinite Reynolds number of both of these transitions leads to a finite and nearly identical value of the amplitude, implying that below this amplitude, the vortex cannot form because the wall is too smooth and, hence, stick boundary results prevail. The current work should serve as a general warning that stick boundary conditions, for example, in narrow biological channels where the surface roughness of the wall can be a significant fraction of the channel width, or, as another example, in the drag reduction problem, may be inappropriate. For the smooth walled case there exists a rigorous proof of stability for axisymmetric disturbances, with strong evidence that all linear perturbations decay for all values of the Reynolds number. Thus, there remains much interest in the cause of this transition and how one may affect the Reynolds number at which transition occurs.

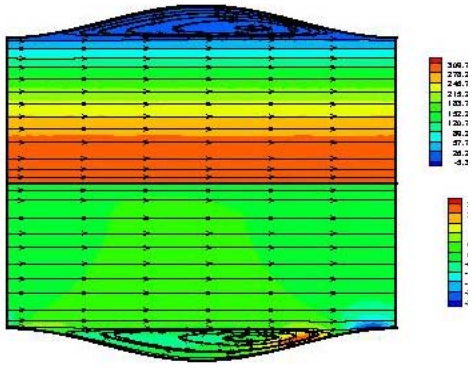


Figure 1. Base flow velocity contours. The upper portion of the figure shows axial velocity shading and streamlines, while the lower portion of the figure shows radial velocity shading and streamlines.

To investigate the linear stability of the base flow, a standard normal mode analysis is used which results in a system of homogeneous linear differential equations in the radial and axial coordinate for the velocity and pressure eigenfunctions. The disturbance velocity boundary conditions are zero on the pipe. Since the coefficients of this system are periodic in the axial direction, Floquet theory is used to consider disturbances whose structure is possibly incommensurate with corrugation wavelength. Discretization of the disturbance equations using finite-element methods leads to a sparse generalized algebraic eigenvalue problem, a minimum Reynolds number is sought for which at least one temporal eigenvalue is neutrally stable. Results show that for fixed corrugation amplitude and small Reynolds numbers the base flow is primarily unidirectional with significant streamline curvature seen only near the bulge and no vortex present in the domain. For sufficiently large Reynolds number the onset of vortex formation is observed in the bulge region (see Fig. 1); however, the flow is stable according to linear stability analysis both above and below this Reynolds number. If the Reynolds number is increased beyond this first transition, a value is reached at which the steady and axisymmetric bulge vortex flow transitions to an unsteady flow.

This work has been accepted for publication in the *Proceeding of the National Academy of Sciences*.

Materials Data and Metrology for Applications to Machining Processes, Frangible Ammunition, and Body Armor

Timothy Burns
Steven Mates (NIST MSEL)
Richard Rhorer (NIST MEL)
Eric Whitenton (NIST MEL)
Debasis Basak (Orbital Sciences Corporation)

The split-Hopkinson pressure bar (SHPB) is an instrument that is used in laboratories throughout the world to obtain material response data for the purposes of modeling and simulation of phenomena that involve rapid, permanent material deformation. Applications of SHPB data include weapons effects modeling such as armor penetration, safety studies such as simulation of crash-tests, measurement of the strength of ceramic materials, and the simulation of material forming operations in manufacturing.

Development of the NIST SHPB, also called the NIST Kolsky Bar after the man who made significant advancements in Hopkinson's original design, was initiated several years ago for the purpose of obtaining improved material response data for the modeling and simulation of high-speed machining operations. An SHPB test involves placing a thin, disk-shaped sample of the test material between two long, hardened cylindrical steel rods. Using precision alignment, the centerline of the sample disk is made to coincide with the centerlines of the two long bars. By means of propulsion by a compressed air gun, a shorter striker rod of the same material is launched so that it collides normally with the outside end of the incident bar, which conducts a compressive stress wave into the sample. The two long steel rods remain elastic in their response to the impact loading.

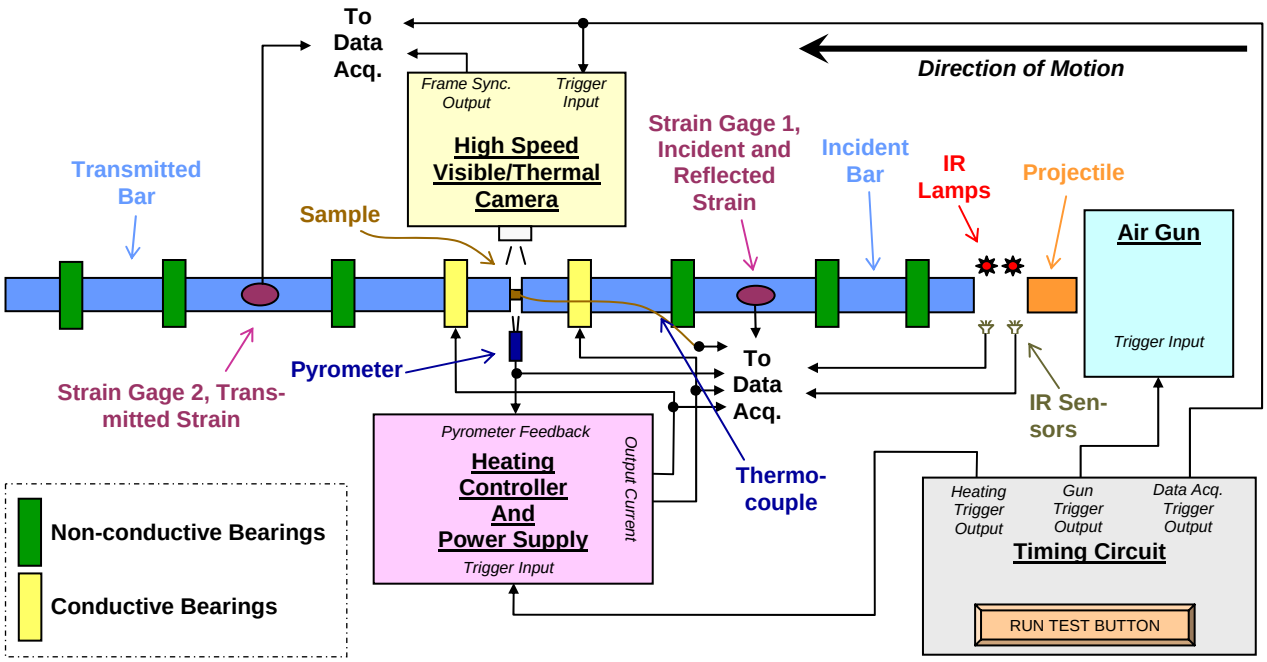


Figure: *The NIST pulse-heated split-Hopkinson (Kolsky) pressure bar.*

By ignoring radial effects, the experiment can be modeled using one-dimensional elastic wave theory. Due to an impedance mismatch at the bar-sample boundary, the input stress pulse splits into a tensile wave that is reflected back into the input bar, and a lower-amplitude compressive wave that travels through the sample and continues to propagate into the transmitted bar. When the compression wave enters the sample, it causes a permanent deformation of the material, known as plastic strain, at a rapid rate of deformation, or strain rate.

Unique features of the NIST Kolsky Bar facility are a capability for pulse-heating a test specimen from room temperature to a significant percentage of its melting temperature in tens of milliseconds prior to impact loading, and sophisticated instrumentation for measuring and controlling the sample temperature. These thermal capabilities, which are still being improved upon, have provided a means of simulating experimentally the extremely rapid heating that occurs in thin cutting regions during high-speed machining operations.

The development of the NIST SHPB was initially funded in large part by the Intramural ATP Program. Development of improved instrumentation for this work continues to be supported by MEL, MSEL, and ITL. The National Institute of Justice (DOJ), through OLES, the NIST Office of Law Enforcement Research, has also supported development of the NIST Kolsky Bar in its room-temperature configuration for the study of the dynamic response of frangible bullets.

During the present fiscal year, it was demonstrated that, for a carbon steel of interest in manufacturing, the pulse-heated material response is considerably stiffer that had previously been measured in tests in other laboratories, in which samples had been pre-heated slowly prior to impact. Post-test metallographic studies on the pulse-heated samples support the hypothesis that certain thermally-activated transformations in the material's microstructure have insufficient time to run to completion under the rapid heating conditions that are present in high-speed machining operations, with the result that higher cutting forces are required for these processes. This work has significant implications for the modeling and simulation of manufacturing operations that involve material removal by rapid machining.

Funding: *NIST Advanced Technology Program. National Institute of Justice (US DOJ via NIST Office of Law Enforcement Standards).*

Mathematical Applications: Electromagnetic Systems

Micromagnetic Modeling

Michael Donahue
Donald Porter
Robert McMichael (NIST MSEL)

http://math.nist.gov/oommf/

Advances in magnetic devices such as recording heads, field sensors, magnetic nonvolatile memory (MRAM), and magnetic logic devices are dependent on an understanding of magnetization processes in magnetic materials at the nanometer level. Micromagnetics, a mathematical model used to simulate magnetic behavior, is needed to interpret measurements at this scale. MCSD is working with industrial and academic partners, as well as with colleagues in the NIST MSEL, PL, and EEEL, to improve the state-of-the-art in micromagnetic modeling.

Michael Donahue and Donald Porter in MCSD have developed a widely used public domain computer code for doing computational micromagnetics, the Object-Oriented Micromagnetic Modeling Framework (OOMMF). OOMMF serves as an open, well-documented environment in which algorithms can be evaluated on benchmark problems. OOMMF has a modular structure that allows independent developers to contribute extensions that add to the basic functionality of OOMMF. The past year has seen two such contributions. The first, produced primarily at the Polish Academy of Sciences but in collaboration with researchers in ITL, adds support for periodic boundary conditions to OOMMF. The second is a spin momentum transfer module developed at the IBM Zurich Research Laboratory. OOMMF also provides a fully functional micromagnetic modeling system, handling both two and three-dimensional problems, with sophisticated extensible input and output mechanisms. OOMMF has become an invaluable tool in the magnetics research community. During fiscal year 2007 alone, the software was downloaded more than 3400 times, and use of OOMMF was acknowledged in 82 peer-reviewed journal articles.

OOMMF is part of a larger activity, the Micromagnetic Modeling Activity Group muMAG), formed to address fundamental issues in micromagnetic modeling through two activities: the development of public domain reference software, and the definition and dissemination of standard problems for testing modeling software. MCSD staff members are involved in development of the standard problem suite as well. There are currently four standard problems in the suite, testing both static and dynamic magnetization properties. The fourth standard problem, which tests magnetization dynamics, received a new submission this year, and new results and comparisons on this problem also figured prominently in a book chapter co-authored by one of ITL staff. Efforts have recently begun to draft a new standard problem to compare models and techniques for simulating the effects of spin momentum transfer.

In addition to the continuing development of OOMMF, the project also involves collaborative research using OOMMF. The project has provided modeling support for a four-laboratory NIST Innovations in Measurement Science (IMS) project (EEEL, MSEL, PL and ITL) on high sensitivity magnetic sensors, and for an ATP project on "Spin Momentum Transfer Oscillators for High-Frequency Nanoelectronic Applications." The project has also been instrumental in the 2007 IMS proposal (with MSEL and EEEL), "Magnetic nanostructures for post-CMOS electronics," which has been approved by the NIST Leadership Board for funding, pending budget availability. The MCSD micromagnetic project produced three journal papers, three conference presentations and two invited talks this past year.

Progress on computational issues has been a recent focus. A commonly used technique for computing the self-magnetostatic interaction tensor from analytic formulae has been determined to produce significant errors. Corrections of the worst of the errors, which occur in the far-field, have been developed employing asymptotic approximations. Techniques to minimize or control for errors introduced when the spatial resolution of the computation grid does not exactly represent the boundaries of the simulated material are in development. Also being pursued are code revisions to exploit the concurrency opportunities made possible by the increasing availability of multi-core hardware platforms.

Another focus area is the modeling of spin momentum transfer to magnetic domain walls arising from electric current flow. This spintronic effect is expected to play an important role in the near-term development of magnetic memory and logic devices. A project starting within ITL's Virtual Measurement Systems program is expected to produce a standard problem suitable to establish confidence in our ability to simulate this phenomenon.

Funding: *NIST Innovations in Measurement Science.*

Time-Domain Algorithms for Computational Electromagnetics

Bradley Alpert
Andrew Dienstfrey

http://math.nist.gov/AlgoCEM

Acoustic and electromagnetic waves, including radiation and scattering phenomena, are increasingly modeled using time-domain computational methods, due to their flexibility in handling wide-band signals, material inhomogeneities, and nonlinearities. For many applications, particularly those arising at NIST, the accuracy of the computed models is essential. Existing methods, however, typically permit only limited control over accuracy; high accuracy generally cannot be achieved for reasonable computational cost.

Applications that require modeling of electromagnetic (and acoustic) wave propagation are extremely broad, ranging over device design, for antennas and waveguides, microcircuits and transducers, and low-observable aircraft; nondestructive testing, for turbines, jet engines, and railroad wheels; and imaging, in geophysics, medicine, and target identification. At NIST, applications include the modeling of antennas (including those on integrated circuits), waveguides (microwave, photonic, and at intermediate terahertz frequencies), transducers, and in nondestructive testing.

The objective of this project is to advance the state of the art in electromagnetic computations by eliminating three existing weaknesses with time domain algorithms for computational electromagnetics to yield: (1) accurate nonreflecting boundary conditions (that reduce an infinite physical domain to a finite computational domain), (2) suitable geometric representation of scattering objects, and (3) high-order convergent, stable spatial and temporal discretizations for realistic scatterer geometries. The project is developing software to verify the accuracy of new algorithms and reporting these developments in publications and at professional conferences.

Recently Alpert and Dienstfrey have collaborated to attack the problem of fast eigenfunction transforms that arises in some time-domain electromagnetics computations (as well as elsewhere). In particular, representations of functions as expansions of eigenfunctions from Sturm-Liouville differential equations, which enable efficient application of certain naturally-occurring operators, must themselves be obtained typically by transforming from pointwise function values. Recent progress in fast algorithms for these transformations, from many researchers and resulting in methods analogous to the fast Fourier transform (FFT) to compute the discrete Fourier transform, has not yet achieved the efficiency to enable widespread acceptance of these new methods. Many of the methods rely on a divide-and-conquer approach that requires repeated interpolation of functions having a prescribed form and it is this operation that consumes the majority of the computation time in a transformation. Alpert and Dienstfrey have discovered, from an identity satisfied by the Green's function of a Sturm-Liouville equation, a new fast algorithm for these interpolations. This algorithm is expected to outperform significantly the fastest existing method for interpolation, which is based on the fast multipole method.

Although these eigenfunction expansions are more specialized than Fourier expansions, they are essential for efficient computation in certain special settings, which include spherical or elliptical geometry, plane-polar representations, bandlimited signals, and a number of more unusual problems. The recent work is expected to result in faster transforms, as well as fast applications of certain related operators, that arise in these problems.

Earlier work of this project has been recognized by researchers developing methods for computational electromagnetics and has influenced work on these problems at Boeing and HRL (formerly Hughes Research Laboratories). It has also been cited widely, including by researchers at Brown University, Caltech, University of Colorado, University of Illinois, Michigan State University, University of Texas, Yale University, University of Basel, and Tsing Hua University.

Laser Pulse Shape Measurement for Laser Guidance and Range Finding

Andrew Dienstfrey
Jack Wang (NIST ITL)
Paul Hale (NIST EEEL)
Rodney Leonhardt (NIST EEEL)

Laser target designators are used in a variety of military applications including range-finding and guidance technologies. As these technologies rely on measuring the reflectance of a far-away object in a battlefield environment with significant optical clutter, it is essential that the armed services have the capability to calibrate extremely sensitive *radiometers* (devices used for measurement of low-power optical signals). This capability is maintained through a coordinated effort involving private defense contractors, and measurement services provided by NIST. The current solution is predicated on the ability to model the relevant laser waveforms as having a *sufficiently broad, Gaussian shape*. In saying this, two fundamental constraints are in effect. One is that the laser waveform is symmetric about its peak. The other is that the ratio of the characteristic time of the laser waveform (e.g., the full-width-

half-max time) to the response time of the radiometer is much greater than one, i.e., the radiometer is fast relative to the optical signal it measures. These two assumptions allow one to model the waveform measurement process as a relatively simple rescaling operation.

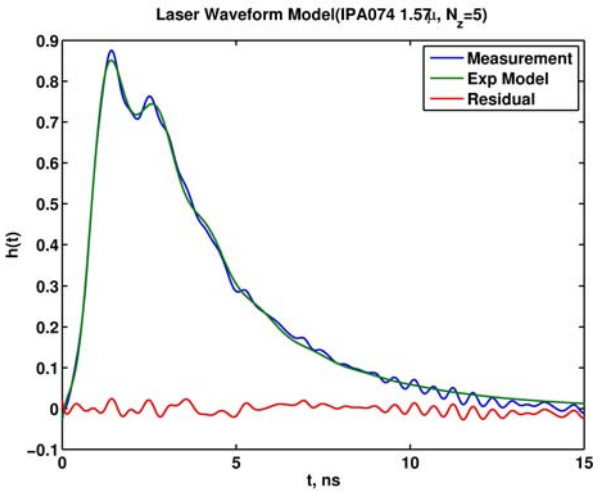


Figure 1. *Example of laser measurement, exponential model, and the residual error. The combination of the prominent secondary lobe and fast laser pulse (≈ 5ns) results in a non-trivial convolution between waveform and the radiometer response function.*

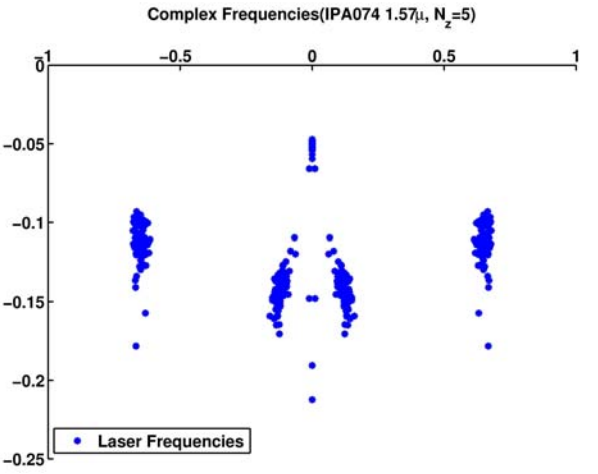


Figure 2. *Complex frequencies resulting from non-linear fit of measured waveform data for 1.57 μm laser.*

Recent changes in Department of Defense (DoD) policy require that next-generation laser target designators operate in the "eye-safe" wavelength region, $\lambda \approx 1.5$ μm. To meet this requirement, based on proprietary design considerations, the contractors that manufacture the test sets for verifying operation of these target designators now generate laser waveforms that are deviate substantially from a Gaussian shape. Furthermore, the characteristic times of the lasers have been measured to be on the order of 6 ns. By contrast, the next generation radiometers designed and built by NIST to meet the extreme sensitivity requirements of this application have impulse response times between 3 to 5 ns. Thus, both of the previous assumptions, symmetric laser pulse shape and relatively fast instrument response, are violated.

In the spring of 2007, the DoD Coordinated Calibration Group (CCG) contracted the NIST Waveform Metrology Team to perform a detailed study of options available to calibrate this next generation of laser target designators. As a first step, hundreds of representative laser waveforms were measured using a variety of techniques. One of these measurements involved running the laser in a high-power mode in which case the optical output could be measured with a "fast" photo-detector (as opposed to the radiometer which is extremely sensitive but slow.) After deconvolution of the detector response function from the measured waveform, we are confident that we have an accurate measurement of the laser waveform shape. Based on this shape, and subsequent analyzes, NIST is currently reviewing a hierarchy of calibration strategies for the CCG. These strategies entail different levels of complexity with corresponding cost-accuracy tradeoffs. NIST is considering performing extensive simulations to better quantify these trade-offs. Of fundamental importance for this effort is sufficiently accurate description of the laser waveform shape.

In the fall of 2007 Andrew Dienstfrey explored the possibility of modeling the waveforms as a sum of damped exponential modes,

$$h(t) = \sum_{1}^{N} \hat{h}_n \exp(i2\pi f_n t),$$

where the order N, expansion coefficients $\hat{h}_n$, and complex frequencies f_n are unknowns. Models of this form are suggested by physical motivations. Estimation of the parameters entails a nonlinear step which solves for the unknown frequencies, followed by a linear least-squares solve for the expansion coefficients. One such routine was implemented which accomplishes the nonlinear solve by eigenvalue analysis of a suitable matrix. A typical measured laser waveform, a five term exponential fit, and the residual are shown in Fig. 1. Anticipating that this degree of approximation error is acceptable for the present application, several hundred waveforms were analyzed. The resulting distribution of complex frequencies for a particular class of lasers is shown in Fig. 2. From this plot it appears the frequencies exhibit natural groupings in regions of the complex plane. This would suggest the possibility to model a "typical" laser waveform as a sum of damped exponentials with frequencies drawn from of an empirically determined multivariate, complex distribution.

In 2008 we will pursue developing and validating an exponential model for laser waveforms as described above. Pending success, the model will be used to validate and quantify the accuracy of calibration strategies that NIST will propose to the CCG. This research is partially sponsored by the 2004 competence proposal, "New Paradigms in High-Speed Waveform Metrology" and the 2007 Coordinated Calibration Group proposal, "Proposal for laser pulse shape measurement system to support 1.5 micron low-level pulsed-laser metrology".

Funding: *NIST Innovations in Measurement Science. US Department of Defense Coordinated Calibration Group.*

Modeling of Optical Spectra

Peter Ketcham
Eric Shirley (NIST PL)

Among the materials exploited by optical technology are crystalline materials with useful properties in the ultraviolet spectral range. These materials are incorporated into ultraviolet optical systems and optoelectronic devices that have relevance to photolithography and the semiconductor manufacturing industry. However, these materials are not fully understood and significant experimental uncertainties exist in their ultraviolet optical properties. In conjunction with the NIST Physics Lab, MCSD is performing first-principle calculations of the optical properties of these materials. Advanced theoretical models of the optical properties are based upon the absorption of a photon and the production of an electron-hole pair. Simpler models, which do not take electronic excitations into account, may fail to give quantitatively accurate results. The inclusion of electron-hole interactions in the theoretical model involves extensive computations which in turn demand high-performance computing systems.

P. Ketcham is currently collaborating with E. Shirley (NIST PL) to realize these theoretical models on multicore, multiprocessor, large-memory, high-performance computing systems. A message-passing approach has been chosen and the LAM implementation of the message-passing interface (MPI) provides the parallel programming software platform. Progress has been made in developing prototype parallel codes that call the MPI libraries from the C programming language under the Linux operating system.

Mathematical Applications: Chemistry and Biology

Modeling of Photochemical Reactions in a Focused Laser Beam

Fern Y. Hunt
A.K. Gaigalas (NIST CSTL)

Fluorescent techniques are widely used to measure and detect phenomena in the chemical, biomedical and material sciences. However the accuracy and sensitivity of these materials is severely limited by photodegradation — the photochemical reactions that transform excited fluorophores into a non-fluorescent species. Photodegradation is the result of a combination of processes that occur on multiple time scales. Therefore, experimental investigation and determination of the photodegradation rate has been quite difficult. For the past two years Fern Hunt has collaborated with A. Gaigalas and co-workers of CSTL to develop mathematical analyses that facilitate an experimental frequency domain based photodegradation measurement technique developed by Gaigalas *et al.*

Starting with a coupled pair of partial differential equations that model the dynamics of fluorescent decay in the course of the frequency domain experiment, we were able to obtain a dramatic simplification of the model using singular perturbation techniques and averaging. Our model incorporates the continuous entry of fluorescent fluid into the measurement apparatus at its base as well as the excitation of the fluorescent particles by a periodically fluctuating laser light located at the apparatus center. Previous work by others neglected the boundary conditions and assumed uniform illumination of the region.

Our simplifications can (in large part) be justified rigorously and as a result we derived a functional form that can be used to fit phase shift measurements as a function of laser frequency. A description of this analysis was finalized and appeared in [1]. This past year, we showed that the fit parameters are related to characteristics of the apparatus and physical constants of the experiment. This enabled us to obtain good quantitative as well as qualitative agreement with experimental results including known photodegradation rates [2].

[1] A.K. Gaigalas, F.Y. Hunt, and L. Wang, Modeling of Photochemical Reactions in a Focused Laser Beam, *Journal of Research of NIST* **4** (2007), pp. 191-208.

[2] A.K. Gaigalas, F.Y. Hunt, and L. Wang, Interpretation of Measurements of Photochemical Reations in Focused Laser Beams–Part 2, in preparation.

Optical Coherence Tomography for Biomedical Imaging

Andrew Dienstfrey
Tasshi Dennis (NIST EEEL)
Shelle Dyer (NIST EEEL)
Paul Williams (NIST EEEL)

Biomedical imaging techniques are increasingly being developed for diagnosis and monitoring of a wide range disease models. Well-known modalities include: MRI, PET scans, and quantitative x-ray tomography (qCT). Recently, optical coherence tomography (OCT) has been added to this list as it has been demonstrated to give quantitative structural information pertaining to biological scatterers. One proposed application has been to use OCT as a diagnostic tool for cancer.

Over a range of optical frequencies and electromagnetic constitutive parameters, it has been shown that cell nuclei are the dominant contributors to the back-scattered radiation. Thus, measurements of this scattering can be inverted for nuclei features, for example the characteristic size. As many epithelial cell cancers (e.g., cervical, esophageal, colon, skin, oral) exhibit enlarged cell nuclei as a precancerous condition, it is hoped that a fiber-based OCT technique could serve as an early diagnostic tool for these disease models. Additional biomedical uses of OCT currently being investigated include: imaging of cellular substructures, assessment of the efficacy of chemopreventative agents, and measurement of characteristics of large-scale intracellular organization.

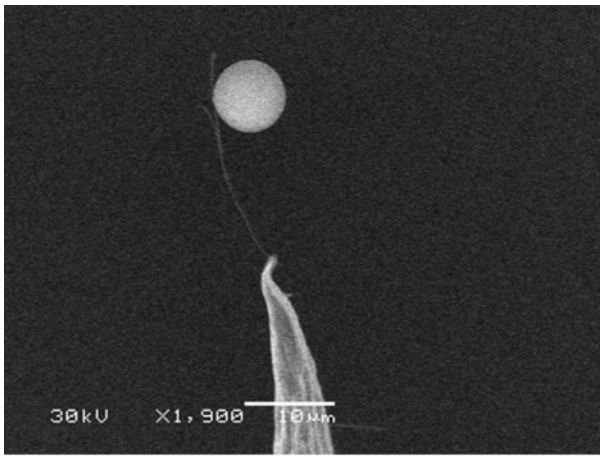


Figure 1. *A micrograph of a sphere-nanotube phantom created by NIST. The polystyrene sphere has a diameter of approximately 8 µm and is attached to a tungsten probe tip by a 100 nm double-wall carbon nanotube. This structure was used to compare measured scattering data with analytic results from Mie theory.*

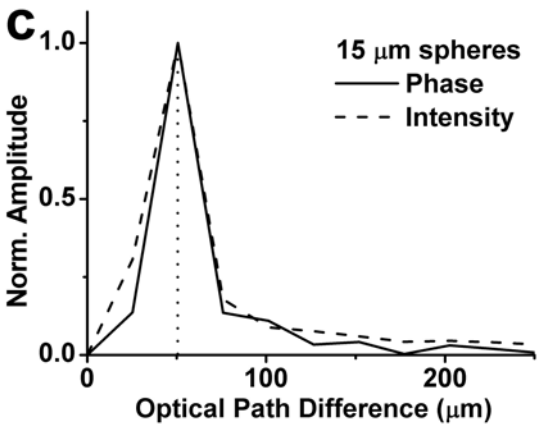


Figure 2. *Sizing distribution for 15 μm diameter (as stated by manufacturer) polystyrene spheres measured by NIST swept-source OCT. The optical path difference is the product of the sphere diameter and the refractive index. The peak corresponds to a sphere diameter of 15.5 pm 1.4 μm.*

This past year A. Dienstfrey continued collaboration with S. Dyer, T. Dennis, and P. Williams in NIST's Optoelectronics Division to investigate the quantitative potential of this technology. A phantom consisting of a single sphere scatterer was created and measured using the NIST-developed swept-source OCT system. The polystyrene sphere was approximately 8.5 μm in diameter and was attached to a tungsten probe tip by a multi-wall carbon nano-tube with a 100 nm diameter. (See Fig. 1). To our knowledge this is the first nanoscale spherical phantom to be developed and measured. This research could serve as a prototype for sensitive optical scattering measurements based on single sphere geometries at scales and optical parameters of biological significance. Returning to the spherical imaging measurements of last year, the data was analyzed more completely taking into account several sources of uncertainty. This analysis revealed that the current NIST OCT system is capable of measuring diameters of polystyrene spheres as small as 15.5 μm with an uncertainty of ± 1.4 μm or 9%; see Fig. 2). The results of this research were accepted for publication [1].

[1] T. Dennis, S. D. Dyer, A. Dienstfrey, S. Gurpreet, and P. Rice, Analyzing Quantitative Light Scattering Spectra of Phantoms Measured with Optical Coherence Tomography, *Journal of Biomedical Optics*, to appear.

Accuracy and Standards for X-ray Measurements of Bone Mineral Density

Andrew Dienstfrey
Tammy Oreskovic (NIST MSEL)
Lawrence Hudson (NIST PL)
Herbert Bennett (NIST EEEL)

Bone mineral density (BMD), as measured by dual-energy x-ray absorptiometry (DXA), is the principle biomarker used for assessing bone health and fracture risk, and for diagnosis and treatment monitoring of bone-related diseases such as osteoporosis. Currently no national, cross-manufacturer standards exist for assessment of the accuracy of DXA measurement of BMD. Fundamental problems include the questionable absolute accuracy of, and lack of cross-calibration among, DXA systems; cross-vendor studies of the same subject show that BMD results can vary by up to 18%. Bone-mineral densities reported for patients in g/cm^2 have no traceability to the international system of weights and measures. This lack of comparability stifles commercial competition—clinics are hesitant to switch to new hardware because it introduces an unknown connection to previous results. It also inhibits the free flow of patients and their histories. This situation threatens the credibility of a growing industry, and has implications for patient care, regulation, Medicare and insurance reimbursement, and drug discovery.

Improvements in DXA BMD measurement accuracy would have significant economic and societal impact. As stated by the National Osteoporosis Foundation, osteoporosis threatens an estimated 44 million Americans, 10 million of whom may already have the disease. Almost 34 million more are estimated to have low bone mass, placing them at increased risk for osteoporosis. The annual direct care costs for osteoporotic fractures range from $12 to $18 B per year. Indirect costs (e.g., lost productivity for patients and caregivers) likely add billions of dollars more. These costs may double or triple in the coming decades as the average age of the US population increases.

The Surgeon General's *Report on Osteoporosis and Bone Health* (2004) cites DXA variability as a fundamental limitation in BMD measurements. The NIST Bone Health Team was formed in 2005 as a United States Measurement Services (USMS) team to assess DXA technology. The problem is difficult technically due to: (1) an incomplete understanding of the X-ray physics of the soft and hard tissues responsible for bone-health and its structural and mechanical properties, and (2) proprietary algorithms used by manufacturers to analyze the radiograph scans, isolate regions of interest, and determine the areas of the defined regions.

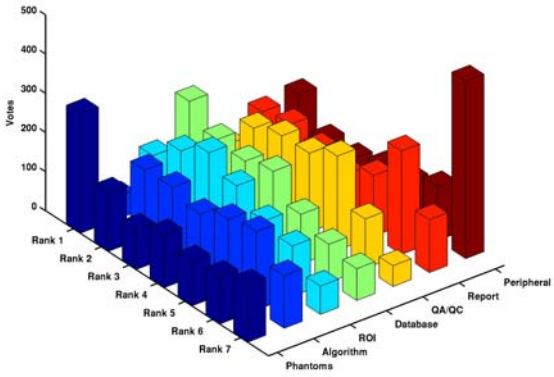


Figure 1. *More than 1000 persons from North America and Europe responded to the NIST-ISCD survey. Respondents were asked to rank seven items in order of importance: phantoms, image-analysis algorithms, region of interest selection, satabase, quality assurance and control, DXA report quantities, and peripheral DXA technology.*

It is important for NIST to develop a thorough understanding of the technology and its shortcomings prior to embarking upon and reaching any technical, measurement-based conclusions. In 2007, as a follow-on to the report documenting the 2006 USMS Workshop on Bone Health [1], NIST and the International Society of Clinical Densitometry (ISCD) designed and issued a web-based survey to further define the problems and potential NIST role in this technology. Over 1000 respondents ranked sources of error in order of importance. Preliminary results of the survey are shown in Fig. 1. In parallel, correspondences were established with world experts in DXA and bone health including researchers from: University of California at San Francisco, Johns Hopkins, Case Western University, and the Institute of Medical Physics at the University Erlangen-Nuernberg.

In 2008 we will complete the analysis of the survey, and work to achieve a consensus between those results and expert opinions. A proposed budget initiative is currently being developed by NIST's Physics Lab to provide support for this work in the future.

[1] H. S. Bennett, A. Dienstfrey, L. T. Hudson, T. Fuerst, and J. Shepherd, Standards and Measurements for Assessing Bone Health – Workshop Report Co-sponsored by the International Society for Clinical Densitometry (ISCD) and the National Institute of Standards and Technology (NIST), *Journal of Clinical Densitometry* **9**(4) (Oct. 2006), pp. 399-405.

.

Monitoring and Modeling Change in Lung Tumors

Adele Peskin
Javier Bernal
David Gilsinn
Terence Griffin
John Kelso
Judith Terrill
Alden Dima (NIST ITL)
Charles Fenimore (NIST ITL)
Karen Kafadar (CU Denver)

Lung cancer is a disease of uncontrolled cell growth in tissues of the lung. It is the most common cause of cancer related deaths in men and second most in women and is responsible for 1.3 million deaths worldwide annually. Tumors may be seen in chest X-ray and computed tomography (CT); see Fig. 1. The latter provides a sequence of 2D images by section (Fig 2). A diagnosis is usually confirmed by biopsy. Treatments include surgery, chemotherapy, and radiotherapy. With treatment, the five-year survival rate is 14%.

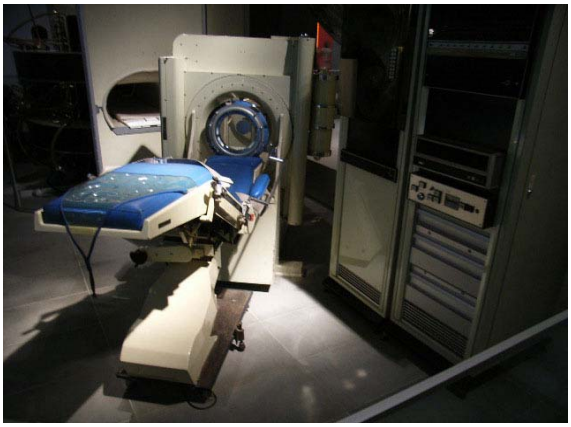


Figure 1. *A typical CT scanner.*

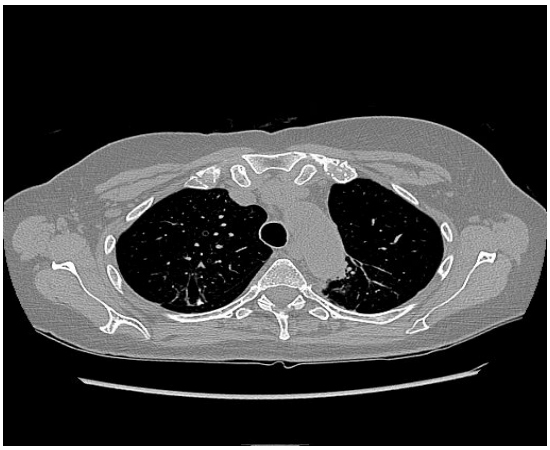


Figure 2. *CT scan of a lung section*

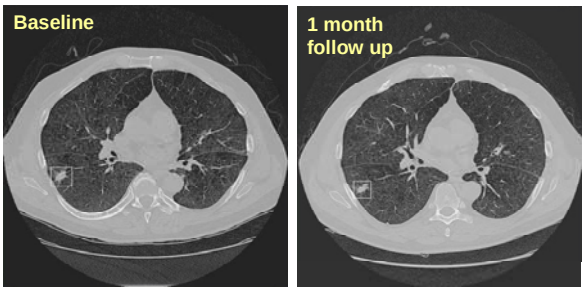


Figure 3. *Preliminary CT scan (left) with tumor identified by a box outline at lower left, and scan of the same patient one month later (right).*

Detection of lung tumors by computerized tomographic (CT) imaging presents a number of different challenges. There are currently a large number of different methods in use for measuring tumor growth by comparing CT scans taken at different points in time. A scarcity of useful data in the medical imaging community has led to variable results for tumor growth measurements. The work in our group for this project is two-fold: we are developing a common database of medical images containing tumors of known size on which to test measurement techniques, and we are working on a new method of measuring tumor growth using this data.

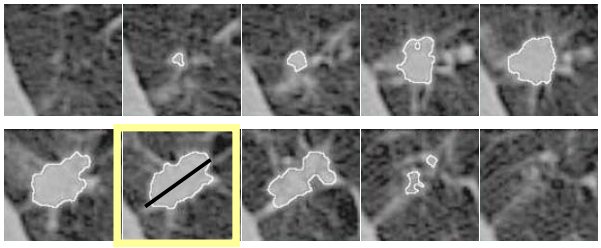


Figure 4. *A diameter measurement of the tumor from the baseline scans, one of which shown in Fig. 3 (left). Estimated diameter was 17.7 mm.*

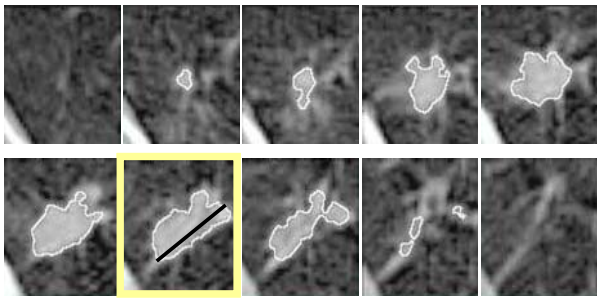


Figure 5. *A linear measurement of the same tumor one month later showing a slight decrease in the linear size. In this case the measurement was 17.1 mm, a decrease in size of about 3%.*

Trained medical radiologists examine the individual CT scans from a patient over several scan sessions and determine abnormal tissue. They assist drug companies in clinical test of new drugs. They determine the effect of drug treatments on patients. Radiologists currently estimate change by measuring linear distance across scans. Volume change is now considered a better measure of drug effect. The need for this change in tactics is illustrated in the following figures where linear measurements and volume measurements of an actual tumor are compared.

Figs. 4 and 5 show a small cancer tumor outlined in a box in the lower left of the images. These two images were taken a month apart. Radiologist performed linear measurements of the two tumors. These are shown as lines across small portions of ten scans covering the volume of the tumor identified above. These linear measurements are made across the CT scan section with what the radiologist considered the largest diameter portion of the tumor.

Solid models of the tumors from the preliminary scan and the one month follow up (Fig 6) show a much larger estimate of a decrease in the size of the tumor. In the case of these two scans the estimate of tumor reduction is nearly 40% by using volumes as opposed to 3% using linear measurements.

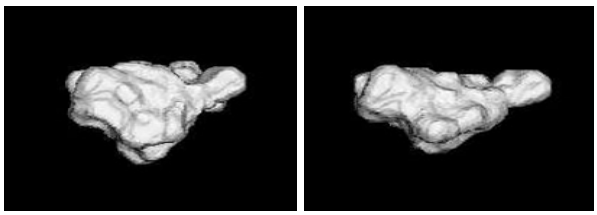


Figure 6. *Preliminary volume estimated as 886 mm^3 (left), and volume estimate one month later is 525 mm^3 (right)*

The Food and Drug Administration (FDA) is currently involved with the study of lung cancer and has developed simulated human torsos and simulated lung cancer tumors of known dimensions and volumes, called phantom nodules. Reference phantom nodules, with or without noise, such as spheres, ellipsoids, etc will be placed in these simulated torsos and CT scans made to determine whether accurate volumes can be estimated from CT scan slices.

This year we have begun an effort to study the measurement of lung cancer tumors. We are approaching this on two fronts. To be able to compare methods for estimating volumes a reliable source of realistic CT data with turmors of known volume is needed. To obtain this, we are embedding known, simulated or phantom tumors into existing CT scan data obtained from Cornell University. This work is currently going on in the MCSD Scientific Applications and Visualization Group.; see page 49. We are also studying methods to estimate tumor volumes from existing data.

Modeling Lung Tumors. Estimating volumes from 2-D scans is a challenging problem. For many of the CT scanners in use the spacing between scans can run 3 mm thus providing voids in between scans. The approach used in many CT software packages is to count

up the small volumes, or voxels, associated with pixels within the regions of interest. Depending on the size of these small volumes they may overlap non-tumor volumes at the boundaries and give a possibly erroneous volume estimate. As a result, we are investigating other methods of volume estimation. Two approaches based on extracting (*x, y, z*) data representing points on the boundary of the object are being considered.

Scans using fine resolution CT's, called micro CT's, will be used to scan reference objects and a segmentation algorithm will be applied to these scans to identify the objects. Looking at CT scans as density objects, a technique will be developed to identify boundary of an embedded object. From this, a series of (*x,y,z*) data points on the edge will be obtained. From this, estimates of the volumes of the embedded objects will then be computed in several ways in order to determine an optimal approach to volume estimation.

Two of the FDA phantom nodules have been given to NIST to have researchers in MEL use coordinate measuring machines (CMM's) to measure points on the surface and estimate volumes from the probed points. Since the phantom nodules where near spherical, MEL attempted to fit spheres to the data, but experienced relatively large residual errors. MEL sent their measured data to us to determine whether other methods could be used to model the data more accurately. We are currently considering using multi-dimensional B-splines to model the near spherical phantom nodules. Work is continuing along this line.

An early, although crude method, to estimate volumes directly from scan images relied on the idea of computing the volume of cylinders of different sizes stacked one on top of the other. The volume of a simple cylinder is just the area of a base times the height. If a radiologist outlines a tumor on each scan in a session by designating the boundary points then the area of an outlined scan can be computed by methods of either multivariate calculus or computational geometry as

$$A = \frac{1}{2}\sum_{i=1}^{n-1}\left(y_{i+1}x_i - x_{i+1}y_i\right),$$

where the summation is taken over boundary points taken in order. Each of these areas can be multiplied by the distance between scans and those volumes summed to generate an estimate of the volume.

The two outlined tumor scans in Figs. 7 and 8 indicate how crude an estimate this can be however. There is clearly a large discontinuity between the two outlined tumors that requires some form of a continuation or smooth mapping in between. Some current work is concentrated on developing functional form models using B-splines for each of the outlines and then using a mapping technique called a homotopy to map one boundary function to the other and thus generating a reasonable transition model of the volume of the tumor between the two scans.

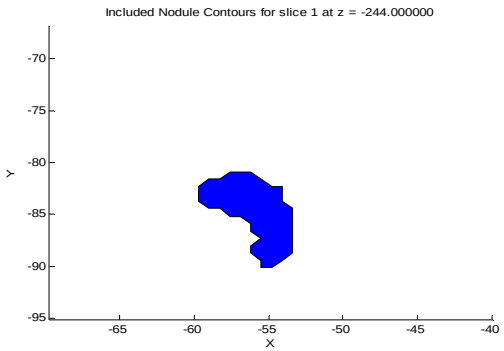


Figure 7. *An outline of a tumor on one scan*

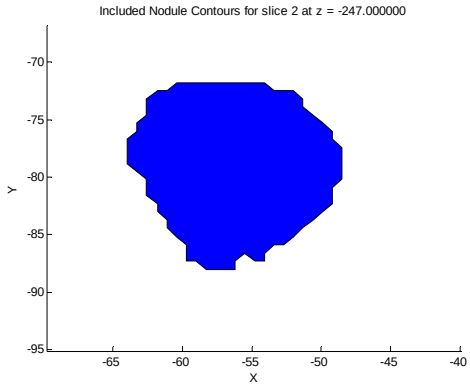


Figure 8. *Outline of the same tumor in a scan 3 mm below the previous.*

For a sufficiently dense set of points on the surface methods of computational geometry based on tetrahedralization can be used to estimate volume. The so-called Power Crust method is under study. One of these modeled data sets is shown below.

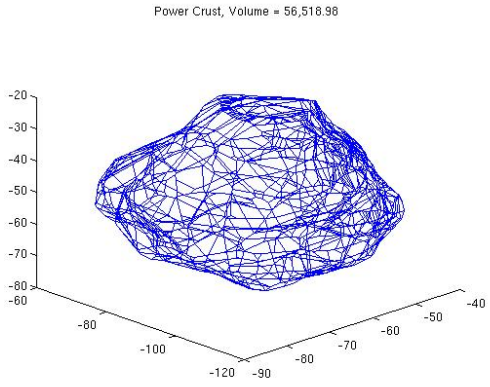


Figure 9. A *triangulated surface of a lung cancer tumor data set.*

Associated ITL Program: *Information Discovery, Use and Sharing.*
Funding: *Biomedical Imaging ACI Initiative (2007).*

Computational Biology and Cell Imaging

Adele Peskin
Javier Bernal
David Gilsinn
Alden Dima (NIST ITL)
John Lu (NIST ITL)
Jim Filliben (NIST ITL)
Anne Plant (NIST CSTL)
John Elliot (NIST CSTL)
Michael Halter (NIST CSTL)

Cell images are obtained by phase-contrast microscopy over extended periods of time. The resulting database of image frames can typically produce gigabytes of data and tens of thousands of images. This makes some form of automated image tracking and analysis critical. There are two parallel efforts within this recently initiated project.

The first effort involves identifying precise boundaries of static images of cells that have been made to fluoresce with a red intensity with one treatment and can then be treated subsequently in order to identify specific cell nuclei. This work primarily will involve development and application of image segmentation algorithms.

In a separate effort, there is work involved to identify the distribution of cell image intensities as the cells begin to split. This involves identifying cell motions throughout many images taken over hours. Not only does this effort involve time series analysis of intensity data, but it involves studying diffusion models of cell motion. There has been recent work in the literature using level set methods to model cell motion. At the moment a review of the current state of cell motion modeling is underway.

Associated ITL Program: *Information Discovery, Use and Sharing*.

Mathematical Applications: Information Technology

Automated Combinatorial Testing for Software Systems

Raghu Kacker
Jeff Yu Lei
James Lawrence
Michael Forbes
Richard Kuhn (NIST ITL)
Vincent Hu (NIST ITL)
Richard Rivello (NIST ITL)
Tao Xie (North Carolina State University)
Renee Bryce (University of Nevada Las Vegas)
Sreedevi Sampath (UMBC)
Sagar Chaki (Carnegie Mellon University)
Arie Gurfinkle (Carnegie Mellon University)

See feature article, page 38.

Funding: *Cyber Security ACI Initiative (2007).*

Foundations of Measurement Science for Information Systems

Ronald F. Boisvert
Isabel Beichl
Brian Cloteaux
John Hagedorn
Fern Y. Hunt
Raghu Kacker
James Lawrence
Roldan Pozo
Judith Terrill
Vladimir Marbukh (NIST ITL)
Anoop Singhal (NIST ITL)

Modern information systems are astounding in their complexity. Software applications are built from thousands of interacting components. Computer networks interconnect millions of independently operating nodes. Large-scale network-based applications provide the basis for critical services of national scope, such as financial transactions and electrical power distribution. In spite of our increasing reliance on such systems, our ability to build far outpaces our ability to secure. As a result, much of our cyber infrastructure is in peril. Communication protocols controlling the behavior of individual nodes lead to unexpected macroscopic behavior, such as waves of network congestion. Local anomalies in power grids propagate in unexpected ways leading to large-scale outages. Vulnerabilities in individual computer systems are exploited in viral attacks resulting in widespread loss of data and system availability. The cost of such events has already led to loss of productivity in the billions of dollars. The long term stability of our critical infrastructure and its resilience to attack is simply unknown.

The critical nature of this problem is widely recognized, and much research is being devoted to the hardening of existing computer infrastructure. However, most of this work is focused on the near-term, providing much needed tools to prevent malicious exploitation of vulnerabilities discovered in current systems. We will continue to remain vulnerable to cyber attack until we can develop a fundamental understanding of the nature of the information systems which we construct. We know, for example, that there are limits to what computers can do: there are functions that are simply not computable. What are the fundamental limits of security? What are the inherent properties of large-scale computing systems? How do these influence security and reliability? How can these be measured? Can such measurements be exploited to ensure the stability of our cyber infrastructure? In short, we are badly in need of an underlying measurement science for complex information systems.

Measurement science has long provided a basis for the understanding and control of physical systems. For example, the Carnot cycle relates elementary measurable properties of gas molecules to a fundamental limit of the efficiency of heat engines independent of their complexity. Similar types of deep understanding and insight are simply lacking for complex information systems. We have no way to relate macro-scale phenomena to measurable properties of the components from which large-scale information systems are built. How do information systems grow? What macro-scale phenomena can they give rise to? Under what circumstances are they stable? When and how do they exhibit phase transitions? The basis for such understanding is beginning to emerge from the study of abstract models of information system structure and dynamics. If such study can lead to the identification of fundamental measurement quantities, then improved methods of design, monitoring, and, ultimately, control can be considered.

In this program, initiated this year, we will begin to develop the theoretical foundations needed for the emergence of a true measurement science for complex information systems. The basis for such work will be found in the fields of discrete mathematics, theoretical computer science, as well as in models and approaches utilized in the physical and biological sciences. We will develop and analyze abstract mathematical models of information system structure

and information flow. In this regard, the study of combinatorial structures (e.g., abstract networks and graphs), probability, information theory, dynamical systems, and control theory will be central. Relevant models will be studied using the emerging theory of discrete random processes, graph theory, queuing theory, and modern Monte Carlo based computational approaches. Such fundamental work will be applied to the study of complex information systems, such as computer networks and distributed systems. An important goal of this work will be to identify and characterize fundamental measurable properties of complex information systems that are indicators of the inherent level of security (i.e., resilience to threats both known and unknown). In this regard, we will study emergent behavior in large-scale networks, network reliability theory, theories of trust, and the analysis and development of self-healing (homeostatic) systems.

To help frame the possibilities for this new research program, a one-day invitational workshop for senior leaders in the field was held on May 29, 2007. The purpose of the workshop was to assess the potential for a mathematical research program in this area, to suggest intermediate-level technical goals, and to identify external research programs with which collaboration would be appropriate. A panel of eight external experts was assembled to participate in a wide-ranging discussion with some 15 NIST staff members who also participated. A workshop report has been issued as a NIST Internal Report [1].

Several projects within MCSD were developed for the initial year of the program. These include the following.

- Methods for Characterizing Massive Networks
  I. Beichl and B. Cloteaux
  See description below.
- Analysis of a Distributed Protocol for Network Control
  F. Hunt and V. Marbukh
  See description below.
- Standard Reference Data for Complex Network research
  R. Pozo
  See description below.
- Visualization of Network Dynamics
  J. Hagedorn and J. Terrill
  See description page 65.
- Automated Combinatorial Testing for Software Systems
  R. Kacker and J. Lawrence
  See description above.

In addition, a project in the ITL Computer Security Division was also supported under this program:

- Modeling Network Vulnerabilities Using Attack Graphs
  Anoop Singhal

  Today's computer systems face sophisticated attackers who combine multiple vulnerabilities to penetrate networks with devastating impact. The overall security of a network cannot be determined by simply counting the number of vulnerabilities. To accurately assess the security of networked systems, one must understand how vulnerabilities can be *combined* to stage an attack. We model such composition of vulnerabilities through *attack graphs*. By simulating incremental network penetration, and propagating attack likelihoods, we measure the overall security of a networked system. From this, we score risk mitigation options in terms of maximizing security and minimizing cost. We populate our attack graph models from live network scans and databases that have knowledge about properties such as vulnerability likelihood, impact, severity, and ease of exploitation. Our flexible model can be used to quantify overall security of networked systems, and to study cost/benefit tradeoffs for analyzing return on security investment [2].

[1] I. Beichl and R. Boisvert, "Mathematical Foundations of Measurement Science for Information Systems: Report of a Planning Workshop," NISTIR 7465, October 24, 2007.

[2] S. Noel, L. Wang, A. Singhal, and S. Jajodia, Measuring Security Risk of Networks Using Attack Graphs, submitted.

Associated ITL Program: *Complex Systems.*
Funding: *Cyber Security ACI Initiative (2007).*

Methods for Characterizing Massive Networks

Isabel Beichl
Brian Cloteaux
Francis Sullivan

A type of data of increasing importance in information technology is one that is based not on floating point values but rather on connections between objects which can be modeled as graphs or networks. Such complex systems arise in many areas such as the power grid, the Internet, communications, and transportation networks. In all these cases, resources are delivered through restricted channels and reliability is critical. We have begun an effort to develop measurement techniques and tools for characterizing such objects based on sound mathematical methods. In most cases, the size of these objects makes exact measurement impossible.

We are in the process of building novel computational tools to generate meaningful statistics about real networks to yield fundamental measures of reliability.

We have developed Monte Carlo methods based on sequential importance sampling to count the number of spanning trees of a graph along with all subforests with k edges for each k. There already exists a method to count all spanning trees of a graph, which uses the determinant of a particular matrix. This is impractical for the size of network we are measuring. However, it has allowed us to test our methods on smaller data. We have also tested our method on real-world examples for which the determinant method cannot be used, e.g., representations of the Internet at the autonomous systems level. This is a graph with 24,566 nodes and 102,946 edges obtained from a database at UCLA.

In the process of this investigation, methods to compute with numbers whose size is such that they can only be represented in practice by their exponent have been developed.

We have used this information to compare the results with a random graph having the same degree sequence as the original data. The random graph with a given degree sequence was generated by a program we developed based on the Blitzstein-Diaconis method, enhanced to include the requirement that the resulting graph be connected. The object was to determine the validity of results obtained by generating random graphs that are "similar" to actual data. The random graph approach to modeling real data is widely used by some research groups. Our preliminary conclusion is the unsurprising but important one—random graphs don't capture important details contained in real data.

Associated ITL Program: *Complex Systems.*
Funding: *Cyber Security ACI Initiative (2007).*

Analysis of a Distributed Protocol for Network Control

Fern Y. Hunt
Vladimir. Marbukh (NIST ITL)

Congestion control is critical to the (smooth) functioning of modern communication networks in particular and the Internet in general. The goal of congestion protocols is to allow many users to share network resources without causing congestion collapse. In addition it is desirable to use as much of the available network capacity as possible and ensure some fairness among users. In was shown by Frank Kelly and later many others that achievement of these objectives can be represented as a global optimization problem of system utility where some notion of fairness could be expressed by a choice of utility function. Congestion control can then be seen as a distributed iterative solution to this problem (Lin and Shroff). Recently Vladmir Marbukh considered the problem of appropriately routing traffic networks within the TCP/AQM protocol and he proposed the introduction of controls that randomly assign routes to traffic in a way that minimizes the mean link load. Simulations of his ideas show that by adjusting the degree of randomness one can optimize network performance and avoid the "route flapping" phenomenon where traffic switches between one or more links with the minimum load at a given point in time. A straightforward representation of the corresponding optimization problem is not convex. This year we reformulated it as a convex optimization problem and we are now in the process of examining the stability of the optimal points as a function of the randomness parameter. We hope to be able to extend the analysis to the dynamics of the protocol examining its behavior when the number of users if large. We also hope to formulate an appropriate fluid limit description in this setting.

[1] V. Marbukh, "Decentralized Control of Large-Scale Networks as a Game with Local Interactions: Cross-layer TCP/IP Optimization (preprint 2007).

Associated ITL Program: *Complex Systems.*
Funding: *NIST Innovations in Measurement Science. Cyber Security ACI Initiative (2007).*

Standard Reference Data for Complex Network Research

Roldan Pozo

The study of complex or "self-organizing" networks is a relatively young, yet active research area. Much of the focus is to determine how graphs from "real" phenomena (such as protein interactions, social networks, structure of the World Wide Web, etc.) differ from synthetic or "engineered" graphs that are constructed by design. Clearly they are different, but how? Three key metrics (clustering coefficient, graph diameter, and degree distribution) are commonly used in the literature, and these techniques have served as encouraging first steps, but further research is needed to yield more meaningful distinctions and truly understand the nature of these complex networks.

One fundamental question to ask, however, is what are the precise "real" graphs that are being used to drive these theories? Clearly, the conclusions being reached can be only as good as the quality and availability of the original data. Most papers provide little information about this, making any kind of comparison or verification of approaches difficult.

For example, a commonly studied collaboration network in the literature is the Movie Actor graph, where two movie actors are connected if they have worked together on the same movie. This information is usually culled from the Internet Movie Database (imdb.com). However, there is no single Movie Actor graph because this database is changing snapshot of the entertainment industry at any given time. Furthermore, what one chooses to count in their particular search (foreign films, documentaries, TV movies, specific epochs or genres) will generate different graphs. In fact, there are several of these Movie Actor graphs floating around in the literature, so a citation is not really meaningful unless it contains a timestamp and documentation of the specific queries used in its generation. A similar argument can be made for other association networks, such as co-author or citation graphs.

Another example is networks based on the structure of the World Wide Web. In this case, each web page is a node and two pages are connected if there a hypertext link from one to the other. These are typically generated by web crawlers and can vary greatly depending on where they crawl and what decisions they make about what constitutes a "link." For example: Do different anchors to the same page count as separate links? How are non-static pages handled? What file types are processed? How robust it is with broken links and sloppy HTML pages? This list can be quite long, and each decision path will yield a different graph. On top of this, the Web is, of course, not a static entity. Many of the Web graphs out there provide little of these details and usually present only the topological structure (i.e., URLs removed) so it is difficult to confirm or validate them independently.

Similar concerns exist for graphs taken from other application domains. Often the data is incomplete, poorly documented, or missing pieces crucial to its description. There are ongoing debates about whether particular networks are really scale-free or not, or whether the conclusions of some papers are actually valid. Usually these concerns lie at the heart of how data was collected and analyzed.

Our motivation for developing a standard reference data set for complex networks is to provide a collection of public and well-documented "real-world" graphs that serve as a test bed of reasonably validated data which researchers can utilize to (1) verify and compare the algorithms and analysis of various approaches in the literature, (2) to ensure that everyone is using the same reference data in their studies, (3) to have one convenient location where one can browse and search for network graphs of various characteristics, (4) to provide a focal point for the research community to contribute and exchange network graphs from various application domains, and (5) to provide a testbed for development of software analysis tools (e.g., graphical viewers, graph partitioners, clustering algorithms) that will aid in the further development and research of complex networks. This effort was begun in mid FY 2007.

Associated ITL Program: *Complex Systems.*
Funding: *NIST Innovations in Measurement Science. Cyber Security ACI Initiative (2007).*