Summary of Awards

Fiscal Year 1996
NSF 97-139

Division of Computer and Computation Research
CISE Directorate

National Science Foundation

ABOUT THE NATIONAL SCIENCE FOUNDATION

National Science Foundation provides awards for research and education in the sciences and engineering. The awardee is wholly responsible for the conduct of such research and the preparation of the results for publication. The Foundation, therefore, does not assume responsibility for the research findings or their interpretation.

The Foundation welcomes proposals from all qualified scientists and engineers and strongly encourages women, minorities, and persons with disabilities to compete fully in any of the research and education related programs described here. In accordance with federal statutes, regulations, and NSF policies, no person on grounds of race, color, age, sex, national origin, or disability shall be excluded from participation in, be denied benefits of, or be subject to discrimination under any program or activity receiving financial assistance from the National Science Foundation.

Facilitation Awards for Scientists and Engineers with Disabilities (FASED) provide funding for special assistance or equipment to enable persons with disabilities (investigators and other staff, including student research assistants) to work on NSF projects. See the program announcement or contact the program coordinator at (703)-306-1636.

National Science Foundation has TDD (Telephonic Device for the Deaf) capability, which enables individuals with hearing impairment to communicate with the Foundation about NSF programs, employment, or general information. To access NSF TDD dial (703)-306-0090; for FIRS (Federal Information Relay Service) 1-800-877-8339.

PREFACE

The purpose of this report is to provide the scientific community with a summary of those grants awarded by the Division of Computer and Computation Research in Fiscal Year 1996.

Similar areas of research are grouped together for reader convenience. The reader is cautioned, however, not to assume that these categories represent either the totality of interests of each program or the total scope of each grant. Projects may bridge several programs or deal with topics not explicitly mentioned herein. Thus, these categories have been assigned administratively and for the purpose of this report only.

In this document, grantee institutions and principal investigators are identified first. Award identification numbers, award amounts, and award durations are enumerated after the individual project titles. Within each category, awards are listed alphabetically by principal investigators. This report lists only those projects that were supported using Fiscal Year 1996 funds, and hence does not list fully funded multi-year awards initiated prior to Fiscal Year 1996.

Readers wishing further information on any particular project described in this report are advised to contact principal investigators directly.

Richard B. Kieburtz
Division Director
Computer and Computation Research

Table of Contents

Division of
Computer and Computation Research
1995–96

Web URL – http://www.cise.nsf.gov/ccr/

Division Director	Richard B. Kieburtz
Deputy Division Director	Caroline Wardle
Administrative Officer	Jan Gatton
Program Directors	
Theory of Computation	Dana May Latch
Numeric, Symbolic and Geometric Computation	S. Kamal Abdali
Computer Systems	Zeke Zalcstein
Operating Systems and Systems Software	Anand Tripathi
Programming Languages and Compilers	D. Helen Gill Anand Tripathi
Software Engineering	D. Helen Gill
Program Staff	
Division Secretary	Deborah J. White
Senior Program Assistant	Benita Bazemore
	Joyce Muir
Program Assistants	Astride Jakobsons
	Gloria Strothers

Division of
Computer and Computation Research
1996–97

Web URL – http://www.cise.nsf.gov/ccr/

Division Director	Richard B. Kieburtz	(703) 306-1910	*rkieburt@nsf.gov*
Deputy Division Director	Bruce H. Barnes	(703) 306-1910	*bbarnes@nsf.gov*
Administrative Officer	Jan Gatton	(703) 306-1910	*jgatton@nsf.gov*
Program Directors			
Theory of Computation	Zeke Zalcstein	(703) 306-1914	*zzalcste@nsf.gov*
Computer Systems, Architecture and Software	Anand Tripathi	(703) 306-1912	*atripath@nsf.gov*
Experimental Software Systems	William Agresti	(703) 306-1911	*wagresti@nsf.gov*
Numeric, Symbolic and Geometric Computation	S. Kamal Abdali	(703) 306-1912	*kabdali@nsf.gov*
Software Engineering	Frank D. Anger	(703) 306-1911	*fanger@nsf.gov*
Program Staff			
Division Secretary	Deborah J. White-Wilkins	(703) 306-1910	*djwhite@nsf.gov*
Senior Program Assistant	Benita Bazemore	(703) 306-1914	*bbazemor@nsf.gov*
	Joyce Muir	(703) 306-1918	*jmuir@nsf.gov*
Program Assistants	Astride Jakobsons	(703) 306-1916	*ajakobso@nsf.gov*
	Gloria Strothers	(703) 306-1918	*gstrothe@nsf.gov*

INTRODUCTION

This report provides summaries of awards made by the National Science Foundation (NSF) through the Division of Computer and Computation Research in Fiscal Year 1996. The Division of Computer and Computation Research activities were carried out through 6 programs:

Theory of Computing

Basic research aimed at improving the current practice of computing and enhancing our understanding of the fundamental issues of computation.

Numeric, Symbolic and Geometric Computation

Basic and applied research on algorithms and software for computational science and engineering with special emphases on (i) symbolic methods, (ii) numeric/symbolic/geometric interfaces, and (iii) areas of intersection among scientific computing and other parts of computer science.

Computer Systems

Research in foundational studies, new architectures, system design, parallel computation, interconnection networks, load balancing, performance evaluation, fault-tolerance, and graphics.

Operating Systems and Systems Software

Research on operating systems, systems tools, and libraries for all levels of computers and networked resources, with emphasis on systems software for high performance environments.

Programming Languages and Compilers

Programming Languages for new or improved means of software creation, development and evaluation of operating systems, and validation of correctness and compliance.

Software Engineering

Creation of models and tools to improve software development, empirical studies of software quality, and the software development process.

THEORY OF COMPUTING

Algorithms: Sequential, Combinatorial and Applied

University of Southern California; Leonard Adleman; ***Computational Complexity and its Relationship to Number Theory*****; (CCR-9403662 A002); $283,734; 12 months. (Joint support with the ARPA-COMPUTATIONAL COMPLEXITY Program)**

The primary focus of this research is the number theoretic aspects of computational problems. In recent years there has been a resurgence of interest in this area. Ancient problems have been solved (e.g. primality testing), new subdisciplines have been spawned (e.g. public-key cryptography) and deep mathematical tools have emerged (e.g. Abelian varieties, class field theory). One area of particular interest is the new "number field sieve'. This collection of algorithms, developed by many researchers, is the most significant advance in integer factoring in the last quarter century. This work continues to investigate improvements and generalizations of this method, in particular, to the function field setting. The new "function field sieve' is used for the calculations of discrete logarithms over finite fields and shows promise for supplanting existing algorithms for finite fields of appropriate form. A second area of interest is the application of recent results to public-key cryptography. The United States Government has recently proposed a national "Digital Signature Standard' (DSS). The proposed scheme relies for its security on the difficulty of computing discrete logarithms. The DSS is controversial and this research investigates the potential of several methods, including the number field sieve and the function field sieve, for breaking the proposed system or for introducing "trapdoors'. Another area of investigation involves the development of new mathematical tools. Finally, the relationships between biological systems and computation is explored. One aspect of this relationship concerns computer viruses.

University of North Carolina – Chapel Hill; James H Anderson and Kevin Jeffay; ***Object-Sharing Technology for Real-Time Systems*****; (CCR-9510156 , A001); $110,329; 18 months. (Joint support with the Operating Systems and Systems Software Program)**

The objective of this project is to develop an object-based framework for synchronization and interprocess communication in real-time computer systems. This framework will integrate work on scheduling theory with work on delay-resilient shared objects. The term "scheduling" refers to the problem of sequencing tasks on a shared execution unit such that it is possible to determine a priori if the timing requirements of tasks are respected. A variety of scheduling policies will be considered in this project, spanning the spectrum from hard to soft real-time systems. By focusing on delay-resilient shared objects, this project represents a significant point of departure from past work on real-time synchronization. Delay-resilient objects are distinguished by the fact that they are not locked. The notion of a delay-resilient shared object builds upon recent promising work on wait-free and lock-free synchronization in shared-memory multiprocessors. Although such work has been ongoing for several years, this project represents the first comprehensive effort to integrate such objects within an object-sharing framework for real-time computing. From a real-time perspective, the key advantage of delay-resilient objects is that they do not give rise to priority inversions. Traditional object-sharing approaches for real-time systems rely on significant operating system support to deal with this problem. The proposed approach does not require such support, and thus entails lower synchronization overhead than traditional approaches. Another advantage of the proposed approach is that it can be applied across a variety of diverse platforms. Target platforms for this project include uniprocessors, shared-memory and message-passing multiprocessors, and workstation clusters.

Princeton University; Sanjeev Arora; ***CAREER: Research into the Hardness of Approximation Probabilistically Checkable Proofs, and Their Connection to Other Areas*****; (CCR-9502747 A001); $48,769; 12 months.**

The classical theory of NP-completeness shows that many optimization problems of practical interest are hard to solve: i.e., NP-hard problems that have no polynomial-time algorithms (assuming P is not equal to NP). A popular method for dealing with such an NP-complete problem is to try to compute approximate solutions: i.e., solutions whose cost is within some multiplicative factor of the cost of the exact solution. The focus of this project is to demonstrate that computing approximate solutions for NP-hard problems is no easier than computing exact solutions. The underlying idea which is exploited, is that of a new probabilistic definition of the NP complexity class, a definition based on probabilistically checkable proofs (PCP's). Some of the major goals of the project include: (1) To better understand the approximability of NP-hard optimization problems (including cataloging the hardness of approximation of many problems, improving existing results about the hardness of approximation, and understanding the limitation of these techniques); (2) To develop further connections between the theory of PCP's and cryptography; (3) To improve existing ways of representing data by error-correcting codes; (4) To simplify the techniques used in the proof of this new definition of NP. The Educational Component of this CAREER Grant includes: (a) Development of CORE undergraduate computer science courses (including Theory of Computation and Applied Discrete Mathematics); (b) Development of a graduate computer science course and a corresponding text dealing with PCP's and (approximate) complexity theory.

Yale University; James Aspnes; *Research Initiation Award (RIA): The Competitive Analysis of Distributed Algorithms*; (CCR-9410228 A002); $24,675; 12 months.

This research applies the technique of competitive analysis, originally developed to study on-line algorithms, to the design of algorithms for distributed systems with unpredictable failures. Competitive analysis compares the performance of an algorithm with that of an optimal "off-line" algorithm designer in effect to predict the future. Using it, an algorithm designer can avoid the perils of designing for the worst case (yielding algorithms too inefficient under milder conditions to be practical) or for a normal case that may turn out to have little correspondence with reality. (1) The first task of the project is to formulate performance measures for distributed algorithms that are amenable to competitive analysis (the surprising difficulty of this task may explain why previous applications of competitive analysis to distributed problems have largely concentrated on the on-line aspects of those problems). (2) The second is to construct distributed algorithms that perform well by the competitive measures. By implementing these algorithms it is possible to test that the measures do in fact predict good real-world performance. This work will yield both insights into the theory of distributed algorithms and better practical algorithms for the distributed system builder.

Mount Sinai School of Medicine; Gary E Benson; *CAREER: Tandem Repeats: Sequence Comparison and Search Algorithms*; (CCR-9623532); $75,000; 18 months. (Joint support with the Computational Biology Activities Program)

This project focuses on expanding the capabilities of the present (DNA) sequence analysis algorithms with respect to comparing sequences and detecting features. In particular, the project concentrates on:(1) developing methods for quantifying the relatedness of biological sequences under a broader model of mutational events than the currently used substitution-indel model, and (2) improving methods for detecting tandem repeats in sequences. The Integrated Educational Plan of this CAREER Grant includes(a)developing a curriculum for graduate study in computational and mathematical biology, (b) developing courses and course materials, including lecture notes, exercises, programming projects and extensive bibliographies, that cover combinatorial and algorithmic topics in sequence alignment, DNA sequencing, DNA physical mapping, prediction of protein folding and RNA folding, and phylogenetic tree construction, (c) instituting a summer training course for new researchers in computational biology.

Massachusetts Institute of Technology; Bonnie A Berger; *Career: Probabilistic Methods and Algorithms to Solve Protein Folding Problems*; (CCR-9501997 A001); $41,049; 12 months. (Joint support with the Computational Biology Activities Program)

The project is in the area of computational molecular biology; specifically it addresses problems in protein folding and the self-assembly of virus shells. The 'grand challenge' problem associated with protein folding is to determine how a protein will fold in 3-dimensions when given only its amino acid sequence. An important first step in tackling the protein folding problem is a solution to the motif recognition problem (i.e., given a known 3D structure, or motif, determining whether this motif occurs in an unknown amino acid sequence). The goal of this effort is to develop new algorithms, based solely on the right pairwise correlation probabilities, which outperform existing motif recognition algorithms. An implementation of these algorithms is being programmed. In fact, the package is aiding in the discovery of the coiled-coil motif in such viruses as Mollney murine leukemia, influenza hemagglutinin, and HIV. The coiled-coil motif is thought to be the mechanism by which proteins fuse with or bind to the cell membrane. This continuing research deals with recognizing motifs when less data is available, predicting inter-strand interactions in motifs, and finding new insights into the 3-dimensional structure of proteins. The second part of the project involves studying the self-assembly of virus shells. In particular, by computationally modeling how these shells are built, it is hopefully possible to suggest ways of interfering with the growth of these viruses and of causing their deformity. The educational component of this CAREER Grant includes: (a) Development of interdisciplinary (computer science, mathematics, and biology) undergraduate and graduate courses and curriculum in computational molecular biology, discrete mathematics and the theory of (biological) algorithms; (b) Development of both undergraduate and graduate seminars in computational molecular biology. Participation of undergraduate students in developing the extended system project and the Virus Shell Assembly project. (This is part of the Research Experiences for Undergraduates Program at NSF and Undergraduate Research Opportunities Program of MIT.)

Carnegie Mellon University; Avrim L Blum; ***NSF Young Investigator (NYI) Award: Understanding Basic Issues in Machine Learning, On-line Planning, and Approximate Algorithms*****; (CCR-9357793 A003); $25,000; 12 months. (Joint support with the Knowledge Models and Cognitive Systems Program)**

This research investigates application of theoretical tools to analyze fundamental issues in machine learning, in planning from incomplete information, and in approximation algorithms. In the area of machine learning, one direction is to provide new algorithmic techniques for fast learning when there is much information available about objects but only a small amount is truly relevant. These techniques are potentially useful both for speeding up learning and for automating the process of selecting what information is given to a learning algorithm. Also new algorithms for learning using neural networks are investigated, especially in the difficult case of high dimensional input spaces. Other fundamental research issues examined include relationships between learning and cryptography, and various ways a learner can usefully employ more active approaches. In the domain of on-line planning past work has been on strategies for traveling without a map in scenes with simple types of obstacles, whose performance can be guaranteed to be not too much worse than the best possible, if one had a map of the region. This is expanded to more varied situations, including those where some kinds of partial maps are available. One interesting case is where partial maps are created by previous explorations. Here one can combine ideas of on-line planning and learning.

University of California – Riverside; Marek Chrobak; ***On-Line Competitive Algorithms*****; (CCR-9503498 A001); $47,217; 12 months.**

This project focuses on a number of topics related to on-line problems. On-line algorithms work with only partial information about the input data. At each time step, such an algorithm receives one unit of data, and has to produce partial results before seeing the remaining data. On-line competitive algorithms are algorithms which return a solution that is not worse than a constant times the optimal one. One of the goals is to solve the k-server problem. An algorithm for k servers that is believed to be optimal has already been developed and shown to be 2-competitive for two servers. A closed form function, which is a good candidate for the potential function for k = 3, is examined. A second goal is to attack a number of other on-line problems for which optimal competitiveness in unknown, including: randomized list update, multi-processor scheduling, and page migration; and to develop a general theory of on-line problems solutions. A third goal is to show hoe the techniques, developed for the solution of the k-server problem can be applied to other on-line problems. In particular, applications of work functions, forgiveness and stable distributions to general on-line problems are developed. Note: This is a companion award to CCR-9503441 (PI: Larmore) of UNLV.

University of Pennsylvania; Fan R Chung, Herbert S Wilf, and Joan P Hutchinson; ***A Conference in Combinatorics and Graph Theory; June 12-15, 1996; Philadelphia, PA*****; (DMS-9612387); $10,000; 12 months. (Joint support with the Algebra & Number Theory Program)**

This award provides partial support for speakers and participants in "A Combinatorics and Graph Theory Conference on the Occasion of Herb Wilf's 65th Birthday," to be held at the University of Pennsylvania on June 12-15, 1996. Topics will include combinatorial identities and enumeration, combinatorial structures, and graph theory, with roughly one day devoted to each topic. This conference is in the general area of Combinatorics. One of the goals of Combinatorics is to find efficient methods of studying how discrete collections of objects can be arranged. The behavior of discrete systems is extremely important to modern communications. For example, the design of large networks,such as those occurring in telephone systems, and the design of algorithms in computer science deal with discrete sets of objects, and this makes use of combinatorial research.

University of California – Berkeley; Edith Cohen; ***Design of Efficient Algorithms*****; (CCR-9616139); $64,559; 12 months.**

(A) This project focuses on the design and analysis of efficient algorithms for network and combinatorial optimization problems. In particular, the project concentrates on (1) theoretical analysis and experimental evaluation of algorithms and (2) algorithmic challenges that arise in applications such as telecommunication networks and information management. Specifically at AT&T, development of new and enhanced communication and information services necessitates algorithmic research and formulation of new models and problems. (B) A principal component of the project is the design of fast algorithms for fundamental network problems (such as shortest path flow computations, and transitive closure). For large problem instances where a quadratic time exact algorithm is infeasible, approximate linear time algorithms are investigated, by considering these well known problems in new contexts. For example, availability of multi-processor systems warrants development of parallel and distributed algorithms, and large interactive databases necessitate the design of dynamic and online algorithms. (C) This award enables the PI to visit the Department of Computer Science, University of California, Berkeley for six months, so that fruitful collaborations with faculty and graduate students can be initiated. In particular, her activities include (3) the development and teaching of a graduate course in network algorithms and combinatorial optimization, and (4) an outreach program to women and other underrepresented groups. It is anticipated that this university/industrial collaborative effort will be continued at both Berkeley and AT&T Laboratories.

New York University; Richard Cole; ***Networks, Pattern Matching and Parallel Algorithms*****; (CCR-9503309 A001); $81,608; 12 months.**

This project involves a number of fundamental problems in several areas of algorithm design and analysis. The first area is pattern matching and in particular, one-dimensional string matching. A major goal is to obtain tight bounds on the performance of string matching algorithms (measured in comparisons). The second is the amortization and the performance of splay trees. This effort is directed towards making further progress on conjectures relating to splay trees, conjectures such as the Split Conjecture, the Traversal Conjecture and the Optimality Conjecture. The third is the area of parallel algorithms. The goal is to obtain work-optimal algorithms which run as fast as possible. Future work on two problems, graph connectivity and Voronoi Diagrams is described.

University of Wisconsin – Madison; Anne Condon, Lloyd M Smith, Robert M Corn, and Max G Lagally; ***DNA Computing and Surface-Based Chemistry*****; (CCR-9628814); $48,961; 12 months. (Joint support with the ARPA-DNA COMP & SURFACE BASED Program)**

This multi-disciplinary SGER project involves collaborative research in the areas of biochemistry, surface chemistry, and algorithm design. The main goal of the project is to develop an approach to computation and information storage, using DNA sequences. Although this approach is both "novel" and "high-risk", there is a high potential that this new DNA/biomolecular computing approach will yield "solutions" to intractable/NP-complete problems. To this end, the project concentrates on: (1) Developing and characterizing complex mixtures of DNA molecules attached to surfaces, by characterizing, for the absorbed DNA molecules,(a) the attachment chemistry, (b) the hybridization chemistry, and (c)the enzymatic activity, using a variety of spectroscopic and biochemical methods; (2) Optimizing the use of the manipulations of DNA sequences in developing molecular computing strategies for solving NP-complete problems; (3)Developing new algorithms for solving the NP-complete problems; and (4) Investigating surface-based methods for high density storage and retrieval of information by DNA hybridization chemistry.

University of Washington; Anne Condon; ***VPW: DNA Algorithms and Approximation Algorithms for NP-Hard Problems*****; (HRD-9627241); $123,061; 12 months. (Joint support with the Advanced Scientific Computing Centers Program and the Computational Biology Activities Program)**

This research involves foundational research on algorithms for NP- hard problems. The primary research area is DNA computing. A careful study of models and algorithms for DNA computation provides a basis for assessing whether DNA computing is viable. This research will provide new, massively parallel algorithms for NP- hard problems that may be useful on other computing models. Another contribution will be analysis of error-prone computation. Research will also be carried out in the area of approximation algorithms. The goal is to extend the range of applicability of the semidefinite programming relaxation method of Goemans and Williamson to an NP-hard timetabling problem that typically arises in high schools. The interactive activities to be undertaken include: teaching a senior-level undergraduate course in theoretical computer science, giving several research presentations in departmental seminars, doing collaborative research with faculty in Computer Science and Engineering and Molecular Biotechnology, and serving as an advisor for an NSF-funded project on development of mentor training packages.

University of California – Berkeley; Nicholas R Cozzarelli; ***Program in Mathematics and Molecular Biology*****; (DMS-9406348 A002); $525,000; 12 months. (Joint support with the Computational Mathematics Program, the Statistics Program, the Topology Program, the Applied Mathematics Program, the Geometric Analysis Program, the Infrastructure Program Program, the Genetics and Nucleic Acids Program Program, and the Computational Biology Activities Program)**

The Program in Mathematics and Molecular Biology (PMMB) promotes the discovery and expansion of the applications of mathematics, including statistics and computer science, to molecular biology because of its precise measurement techniques, molecular biology appears now in the position that physics was a century ago, ready for an explosion of productivity and new insights arising from interactions with mathematics. Although administered through the University of California at Berkeley, this is a Program without walls, a group unified by a common goal. PMMB enables the collaboration of experimental biologists with mathematicians, professions often viewed as disparate. PMMB members apply state-of the-art mathematics and computational techniques to the design and analysis of biology experiments. PMMB provides a forum for mathematicians and biologists with interesting problems and a commitment to the application of mathematics to molecular biology. The Program uses three major mechanisms to achieve this interdisciplinary collaboration. (1) An active fellowship program and well-attended annual scientific meetings draw particularly the younger scientists into the area of study and provide support for interdisciplinary training. (2) The Program actively promotes the education of the general mathematical and biological audiences on the potentialities of the field through presentations and tutorials at national meetings focused small meetings, individual seminars, courses, and didactic articles. (3) The Program provides modest direct support for interdisciplinary studies by the members, their students and postdoctoral associates, fostering a network of interactions among the members for productive collaborations.

University of South Carolina – Columbia; Karel Culik; *Finite Automata Methods in Image Compression and Manipulation*; (CCR-9417384 A001); $55,000; 12 months. (Joint support with the Numeric, Symbolic & Geometric Computation Program)

A new method for image description, compression, and manipulation based on finite automata is studied. A Weighted Finite Automaton (WFA) computes a real function on 0,1, .., n. For n = 2, such a function can be interpreted as the grayscale function specifying an image. The most important application of WFA is the new image- compression method based on the WFA inference algorithm. This method can be classified as a fractal compression method, but it is different and performs considerably better than the previous fractal methods for image compression. It produces an excellent compression size to quality ratio for the widest variety of images. The current image-data compression software performs well; however, it does not reach the full potential of WFA for image compression. One of the main goals of this project is to optimize the compression algorithm to make it faster and to produce better quality regenerated images for the same compression rate. Weighted Finite Transducers (WFT) are a tool for the description and implementation of image transformations. The project will investigate the WFT and their applications, their efficient implementation, and the design of inference algorithms for WFT.

Rice University; Hristo N Djidjev; *Research Initiation Award (RIA): Efficient Algorithms for Special Classes of Graphs*; (CCR-9409181 A002); $24,340; 12 months.

This project focuses on the development of efficient algorithms for graphs that have small separators. The goal is to explore the particular topology of the graphs in order to improve the efficiency of the algorithms. One important class of graphs that is investigated is the class of planar graphs, which arise naturally in applications dealing with regions in the plane. The following types of problems are addressed: (1) Finding shortest paths and distances in graphs: This is a fundamental and well studied problem in computer science with many applications. The problems considered here include the design of dynamic and on-line algorithm for the single-source and the all-pairs shortest paths problems for graphs with weights on the edges; (2) Locating the center and computing the diameter of an outer planar graph: Such problems find important applications in the design and analysis of communication networks; (3) Graph separation problems: Graph separator theorems and the corresponding algorithms are important tools in the efficient solution of various combinatorial problems. Algorithms for finding separators of small cost for special classes of graphs are constructed, where the cost of the separator depends on some previously specified cost function on the edges of the graph.

University of Minnesota – Twin Cities; Ding-Zhu Du; *Approximation of Steiner Minimum Trees and Applications*; (CCR-9530306); $124,989; 36 months. (Joint support with the Algebra & Number Theory Program)

This project's goal is the study of approximations for Steiner minimum trees (SMTs) and related problems. This study extends the PI's research, supported under CCR-92- 08913, on the analysis and the designs of polynomial-time approximations for SMTs. The project includes finding new techniques to solve or make significant progress in both the analysis and construction/implementation of the following problems in the area of SMTs: (1) Open problems on the Steiner ratio (such as, Chung-Gilbert's conjecture, Graham-Hwang's conjecture, and Smith's sausage conjecture); (2) Determination of the performance ratios for various approximations (such as, Chang's approximations, Smith-Lee-Liebman's approximation, and Kahng-Robin's approximation); (3) New approximation algorithms using the variable metric method; (4) Investigation of a conjecture about SMTs approximations and hardness of finding approximate SMTs; (5) Establishment of significant lower bounds for the performance ratio of polynomial-time approximations; (6) Improvement of the analysis of Zelikovsky's algorithm; and (7) Construction of good approximation algorithms for related problems.

University of California – Irvine; David A Eppstein; *NSF Young Investigator (NYI) Award: Algorithms for Molecular Biology Optimal Triangulation, Minimum Spanning Trees, and Geometric Optimization*; (CCR-9258355 A004); $62,500; 12 months.

This project addresses research in the design and analysis of algorithms, motivated by applications in many other areas of computer science. Very similar problems tend to arise in very different applications. As an example, the "least weight subsequence problem" which is studied in connection with certain biological sequence comparison problems has also appeared in page layout algorithms, VLSI design theory, and computational geometry. Also work in motion planning for parts handlers may be couched in the language of automata theory, and is closely related to some problems in coding theory. As a consequence, theoretical computer science research might be most appropriately grouped by solution style and technique, rather than by the problems that it solves. Along with traditional sequential algorithms, there is work with parallel (PRAM model) algorithms. Algorithms in molecular biology have become increasingly important, as the quantity of DNA sequence information known has been growing more quickly than the ability of computers to process it. Until recently, most workers in biological algorithms were mathematicians and biologists. Their research identified and formalized many problems in this area, but the algorithms they developed tended to be simple applications of dynamic programming. This work applies more sophisticated techniques to these problems, resulting in speedups of one or two orders of magnitude. There will also be ongoing research on problems of computing minimum spanning trees, both for graphs and for planar point sets. The main focus is on dynamic versions of such problems as the minimum spanning tree. This has led to a more general study of off-line algorithms, and to data structures for querying the effects on the MST of potential future update.

Rutgers University; Martin Farach; ***Career: Distance Based Methods in Phylogeny Reconstruction: Trees and Strings*****; (CCR-9501942 A001); $41,011; 12 months. (Joint support with the Computational Biology Activities Program)**

This project is in the area of computational biology; specifically, problems which deal with how to construct evolutionary trees, or phylogeny, (i.e., models of the evolutionary history of various species). There are three standard criteria on which tree construction algorithms are based: distance methods, maximum likelihood, and parsimony. The first two methods are used by the biologists and are dealt with in the statistical literature. The traditional algorithms literature for tree construction is focused on the parsimony methods; notably the Perfect Phylogeny Problem. Although this problem has significant combinatorial appeal, there is a gap between the results for this algorithmic theory and the theory for distance based tree construction methods. The project focuses on problems in the distance based phylogeny, including: (1) Methods for computing distances from strings; (2) Methods for comparing different trees for consensus; (3) Methods for fitting distance data to metric models. The second part of the project involves developing empirical methods for testing the statistical robustness of heuristic methods currently in use. The Educational Component of this CAREER Grant includes: (a) Development and implementation of a post doctoral training program in computational molecular biology at DIMACS; (b) Modification of the graduate seminar in biocomputing; (c) Development and modification of the Rutgers University undergraduate program in biomathematics which focuses on combinatorics and molecular biology.

Iowa State University; David Fernandez-Baca; ***Algorithms in Parametric Optimization*****; (CCR-9520946 A001); $34,357; 12 months. (Joint support with the Science and Technology Centers Program)**

Two topics in the design of combinatorial algorithms are addressed: (1) Parametric problems on graphs with vertex and/or edge weights, with particular emphasis on Megiddo's method of parametric search; (2) The construction of evolutionary trees for sets of species, focusing on methods where species are described by the characteristics they exhibit. First, parametric computing deals with problems where the input is a continuous function of one or more parameters; it has applications in: (1a) Sensitivity analysis (that is, studying how changes in costs affect the choice of optimum solution to a problem); (1b) Lagrangian relaxation (a powerful heuristic for coping with difficult constraints in optimization problems); (1c) computational geometry; and (1d) computational biology. The primary goal of this project is to study the circumstances under which parametric problems can be solved within the same time bound as their underlying non-parametric problems. This leads to an examination of the roles of parallelism, dynamic algorithms, and graph and space decomposition techniques in parametric search. A secondary goal is to obtain bounds on the number of times the optimum solution to certain parametric problems changes as the parameter is varied across its range. Second, an evolutionary tree or phylogeny describes how a set of species evolved from a common ancestor. The research on phylogeny construction deals primarily (2a) with finding faster algorithms for important special cases, and (2b) with developing dynamic algorithms for updating existing trees. Dynamic algorithms are particularly useful in practice, as they permit scientists to analyze different scenarios for evolution and to modify existing trees to reflect increased knowledge about sets of species. It appears that the development of efficient dynamic algorithms requires clean characterizations of sets of species admitting certain kinds of phylogenies. Such characterizations may be of interest in their own right.

University of Chicago; Lance Fortnow; ***Presidential Faculty Fellow*****; (CCR-9253582 A004); $100,000; 12 months.**

Work in the structural theory of computation, especially nondeterminism, probabilistic complexity, interactive proof systems, counting classes and program verification is continued. Newly developed techniques of algebraic structure of complexity classes and Gap-P counting functions to further understanding of these areas are used. A theory of program verification to enable an implementation of simple verification schemes is enhanced.

Purdue University ; Greg N Frederickson; ***Graph Algorithms and Data Structures*****; (CCR-9322501 A002); $67,854; 12 months.**

This project focuses on problems in two interrelated areas: (1) algorithms for graph problems, and (2) the design of data structures. The common theme is the investigation of efficient computation in these areas. The project is to generate improved techniques, and characterize significant complexity relationships. A central issue is the efficient coordination of the acquisition of information during a complex computational task.

Columbia University; Zvi Galil; ***Design and Analysis of Computer Algorithms*****; (CCR-9316209 A002); $102,107; 12 months.**

This work has four parts that deal with four areas of algorithmic research. The first part, algorithms on strings, has a number of specific problems for which improved algorithms will need new algorithmic techniques. In the second part, the study of cryptographic protocols, concerns new protocols, and exploring tradeoffs of security and communication for useful tasks. The third part, problems in distributed computing, includes developing precise models and improved algorithms. The fourth part, problems in dynamic graphs, involves the development of algorithmic techniques for graph problems whose underlying graphs may change. This research develops new general tools and techniques as well as attacking the specific problems of special interest. It includes the design of sequential, parallel, and distributed algorithms and protocols. Theoretical considerations as well as practical ones are taken into account.

Massachusetts Institute of Technology; Michel X Goemans; ***Career: Approximation Algorithms: Methodology and Applications*****; (CCR-9623859); $100,000; 24 months.**

This project focuses on designing and analyzing fast approximation algorithms for NP-hard problems, with good performance guarantees. Despite the recent strong results on the hardness of approximations, the PI and his Ph.D. Student have been instrumental in introducing new methodologies for designing and analyzing approximation algorithms, and in improving the best known performance guarantees for a wide range of problems. In particular, the goals of this project include: (1) Development and refinement of methodologies to design and analyze approximation algorithms; (2) The deviation of approximation algorithms with improved performance guarantees for a variety of combinatorial problems;(3) The actual solution of important practical problems, arising in electrical engineering and computer science, in computational molecular biology, and in material science. The Integrated Educational Plan of this CAREER Grant includes: (1) Involvement of undergraduates in developing efficient software codes based on approximation algorithms; (2) Development of undergraduate and graduate courses which address the changing job markets; (3) Development of a multimedia textbook on optimization algorithms.

University of Virginia ; John T Hackett, Maria Bykhovskaia, and Mary Kate Worden; ***CRI: Statistical Analysis and Computer Simulations of Persistent Neurosecretory Activity*****; (IBN-9634400); $99,703; 24 months. (Joint support with the Neuroscience Program)**

Neurons (nerve cells) in our brains communicate with each other at contact points (synapses) which excite the following neurons by means of chemicals known as transmitters. When things go right, this process of signaling continues repeatedly in an organized way that results in integration of information coming into and leaving the brain. This process also is the basis of memory and learning. This research has as its long range goal clarifying the process by which neurotransmitter substances are released from neurons. At an early stage, the neurotransmitter are packaged in a stored form from which they must be mobilized to dock at the membrane of the release site. To help explain the mobilization process, mathematical calculations will test various possible outcomes that will be compared with experimentally obtain results. Of particular interest is the finding that following repetitive stimulation more and more neurotransmitter substance is released by a process known as facilitation. Understanding how facilitation comes about will be key to our understanding of more complicated forms of memory and learning.

University of California – Santa Cruz; David Haussler and Richard Hughey; ***Generalized Hidden Markov Models for Protein Sequence Analysis*****; (DBI-9408579 A001); $175,138; 24 months. (Joint support with the Computational Biology Activities Program)**

This award will support research in the development of statistical methods, algorithms and software for the analysis of protein and RNA sequence data based on generalization of the hidden Markov model. Specifically the investigators will use and develop Hidden Markov Models, Stochastic Graph Grammars, and Markov Random Fields. These methods would have applications to inverse folding, multiple alignment and database searching in molecular and structural biology. This award is supported by programs in Computational Biology and the Theory of Computing.

State University of New York – Binghamton; Thomas J Head and Dennis G Pixton; *Computational Systems with Biological Roots II*; (CCR-9509831 A001); $27,000; 12 months. (Joint support with the ARPA-COMP SYS WITH BIOLOG ROOT Program)

The objective of this 'applied algorithm' project is to continue the establishment and study of the interrelationships between formal language theory and the structure and behavior of the informational macromolecules. The formal representation of recombinant behaviors exhibited by double stranded DNA in the presence of specified sets of enzymatic activities is of crucial interest. This project includes the continued investigation of splicing systems and the languages they define. The fact that DNA occurs naturally in both linear and circular forms, which interact through recombination, has motivated the initiation of a unified theory of interacting linear and circular words. This project includes further development of this theory with emphasis on splicing. A treatment of languages of bi-infinite words and semi-infinite words is made, which will initiate the integration of these theories with the theories of languages of circular and finite words. This project directs special attention to the sofic shifts, which are a class of languages of bi-infinite words that play a major role in dynamical systems. This project is significant in that it establishes and clarifies relations between formal language theory and molecular biology. It develops an additional source of suggestive problems and imagery for formal language theory. Also, over a time scale of decades, it is expected to suggest questions and contribute precision to biochemical thought about informational macromolecules and their behavior. Additionally, the project provides new problems and approaches to the study of symbolic dynamics. This project combines theoretical computer science, biology and dynamical systems in a mutually stimulating and reinforcing way.

University of California – Irvine; Sandra Irani; *Competitive Analysis of Problems in Resource Allocation*; (CCR-9625844); $166,809; 36 months. (Joint support with the Operating Systems and Systems Software Program)

This project focuses on algorithms that work with limited information. This research investigates on-line algorithms in the context of various resource allocation problems in operating systems. This project is taking a twofold approach to each problem. The first approach is to perform a theoretical study using competitive analysis. The competitive measure is used to evaluate on-line algorithms by measuring their performance in relation to the optimal off-line algorithm which has available to it complete information about the problem instance at hand. This effort plans to perform empirical tests to evaluate the performance of various on-line algorithms in a realistic setting. The first two problems addressed in this project are related to memory management. The first of these addresses ways in which information about the access pattern of a program available at compile time can be used to guide page replacement and prefetching policies. The investigator will develop methods which can take a program and automatically generate appropriate memory management instructions to be inserted in the code. The second problem addresses allocating memory resources among users in a multiprocessing environment. This work addresses this problem for general multiprocessing environments as well as methods specifically designed for database systems. The third problem examines preemptive load balancing strategies over local area networks under a probabilistic model for process duration. The last problem is an investigation of disk scheduling strategies. This work is motivated by the fact that the availability of larger cheap memory has made it possible to have long queues of disk updates which means that more sophisticated scheduling policies can have a greater impact on disk utilization.

Kestrel Institute; Richard Jullig; *International Workshop on Constructive Programming: Rio de Janeiro, Brazil, November 4-8, 1996*; (CCR-9617781); $37,200; 6 months. (Joint support with the Software Engineering Program, the Programming Languages and Compilers Program, and the Numeric, Symbolic & Geometric Computation Program)

This award provided funds for an international workshop on Constructive Programming held in Rio de Janeiro, Brazil to explore future directions in formal methods, programming logics, and programming languages. The workshop took place November 4 to 8, 1996 and combined 3 days of technical exchange with 2 days of discussions to produce recommendations to the science funding agencies of the U.S. and Brazil: NSF, CNPq, and FINEP. In recent years Brazilian researchers have made significant contributions in constructive programming and related areas, establishing a number of contacts with U.S. researchers. This workshop will promote further collaborations and provide information and recommendations for future research efforts and funding initiatives.

Carnegie Mellon University; Ravindran Kannan; *Optimization and Learning Over Convex Sets*; (CCR-9528973); $240,000; 36 months.

This is a companion project with Software Capitalization Program Award CCR-95-28215, " Fast Randomized Algorithms for Optimization and Other Applications of Geometric Random Walks". Both of these linked awards are part of the PI's continuing long term project, started under CCR-92-08597, "Random Walks, Parametric Integer Programming". Under this previous grant, CCR-92-08597, random walk algorithms to sample from a wide class of multivariate probability distributions were developed. The goals of this project include: (a) The study of some algorithmic applications of the random walk technique, including (a1) producing and using random samples given a convex set, (a2) investigating the rapid generation and use of contingency tables, and (a3) exploring geometric and efficiency questions arising in the estimation of volumes of convex sets, the original motivation for random walks; (b) The study of some fundamental questions that arise in computational learning theory related to convex sets, including (b1)"synthesizing" a convex set given random samples, (b2) generating algorithms and determining the complexity of certain multi-dimensional convex sets given samples from them, both when the learning is "distribution-free" (in the sense of Valiant) and when the samples are drawn from an uniform distribution on the convex set.

University of Pennsylvania; Sampath Kannan; *Models, Methods, and Criteria for Phylogeny Construction*; (CCR-9612829); $50,000; 12 months. (Joint support with the Systematics and Population Biology Program Program and the Computational Biology Activities Program)

This Small Grant for Exploratory Research (SGER), jointly funded by the Theory of Computing (TOC) Program, CCR, by Computational Biology Activity, BIR and by the Systemic Biology Program, DEB, addresses problems in the construction of phylogenies or evolutionary trees. The formulation and effective solution of this problem requires a collaborative multidisciplinary effort from biologists, statisticians and computer scientists. The ideal methodology for solving this problem would include the following steps:(a) Observe data on the species that exist today; (b) Identify a biological model (such as the Jukes-Cantor model or Kimura two parameter model); (c) Based on the model from Step (b), design an objective function and efficient optimization methods for this function so that the tree that optimizes this objective function is the tree that best fits the model. Unfortunately, this ideal program is impossible to realize, because of roadblocks at every step: (i) Data is subject to experimental error and to errors due to its interpretation and use in phylogeny construction methods;(ii) It seems difficult to identify a precise biological model for evolution; (iii) Given the stochastic model of evolution, one candidate for the optimizing tree is the "most likely tree". Given the uncertainty of what is actually the "best" model, a most likely tree under one model should still be a very likely tree under a slightly different model. Demonstrating this has proven to be a very difficult problem. The goals of this SGER project include:(1) Modification of the ideal methodology so that Difficulties(i) – (iii) are removed; (2) Testing of various models using r-RNA and tufA sequence data supplied by a molecular biologist; (3) Using the experimental results, design of general methods for inferring phylogeny;(4) Comparison of these new models with existing models for the data. In addition, one of the goals of this SGER award is to initiate a multidisciplinary effort between the biologists, statisticians, and computer scientists at the University of Pennsylvania.

Georgia Technology Research Corporation – Georgia Institute of Technology; Howard J Karloff; *Randomization and Derandomization*; (CCR-9319106 A002); $57,058; 12 months.

This project studies randomized algorithms and derandomization techniques; specifically, randomized algorithms for on-line problems, randomized algorithms for graph connectivity questions, techniques for "derandomizing" (making deterministic) algorithms for on-line and connectivity problems, and general techniques for derandomizing large classes of randomized algorithms. An on-line problem is one in which the input is revealed to the algorithm only one piece at a time, and the algorithm is required to respond without knowledge of the future. An example is the problem of paging in a computer system: the operating system does not know which pages will be requested in the future, and therefore does not know which pages put in fast memory and which to leave in slow memory. Recent research has shown that randomized algorithms can be dramatically better than deterministic algorithms for some on-line motion planning, scheduling, and graph coloring. The (edge) connectivity of a connected graph is the minimum number of edges one has to remove to disconnect the graph. Determining the connectivity of a graph is an important problem in the area of network reliability. Recent advances have provided the key to faster algorithms to determine the edge connectivity of a graph. But the new algorithms are not likely to be the fastest possible algorithms. This research builds faster connectivity algorithms, randomized if necessary, deterministic if possible. Much research over the past few years has dealt with derandomizing randomized algorithms. This work continues this effort in the hope of derandomizing algorithms for certain on-line and connectivity problems, and also providing new general techniques for derandomizing randomized algorithms.

University of Maryland – College Park; Samir Khuller; *CAREER: Approximation Algorithms for Graph-Theoretic Problems*; (CCR-9501355 A001); $40,919; 12 months.

Graphs are powerful tools that can be used to model objects and relationships between objects. Indeed, many distinct problems can be cast as optimization problems in graph theoretic terms. Specifically, graph theory provides an excellent way to model problems related to transportation, network design, scheduling, robotics and many other areas. It is used to develop and illustrate the power of different algorithmic tools. There are three broad areas from which the particular problems considered in this project are abstracted: (1) Network Design: How can a network that connects a given set of sites be built? How can the network be made reliable, and also provide high bandwidth between sites (while keeping the cost low)? Such fundamental questions are addressed here. (2) Motion Planning: How does an agent navigate in an environment? What are good navigation strategies for navigating in unknown environments? In this case the agent discovers the environment by actually touching obstacles. How does an agent record an approximate map of the environment, when working with limited memory? (3) Scheduling: Precedence based scheduling problems arise in the context of manufacturing and compilation for parallel machines. How can effective heuristics for such problems be designed? For most of these problems it is very difficult to obtain even an approximate solution in polynomial time. Education Plan: Most computer scientists are constantly faced with the task of designing algorithms to solve problems. Although these arise in many different contexts, some basic principles underlie the design of efficient algorithms and data structures. As a part of the Education Program of this CAREER award, the PI is planning: (a) to undertake the training of students in designing algorithms, and (b) to teach them the importance of re-examining existing algorithms so as to make them simpler and more efficient. An algorithms lab where students undertake projects, implement algorithms, and act as consultants to students in other research groups is a significant step in this direction.

Johns Hopkins University; S. Rao Kosaraju; ***Topics in Algorithm Design*****; (CCR-9508545 A001, A002); $123,567; 12 months. (Joint support with the Computational Biology Activities Program)**

This research project examines algorithm design issues in the areas of (1) Pattern Matching, (2) DNA Sequence Assembly, (3) Data Structures, and (4) Computation of n-body Potential Fields. The proposed problems in pattern matching deal with efficient methods for the construction of suffix trees of strings and trees, treatment of ambiguous symbols with the help of suffix trees, external memory algorithms for suffix trees, and several issues related to computation of convolutions of strings. These investigations are of great significance to the broad areas of Digital Libraries and Computational Biology. With the emergence of faster DNA sequencing methods, there is an immediate need to develop better algorithms for assembling large numbers of DNA fragments. Several problems that arise in such sequence assembly are investigated. Recently, the PI developed an attractive procedure for transforming data-structure-based algorithms which have good amortized speed into ones that have matching worst-case speed characteristics. Another goal of this project is to explore the applicability of this technique to problems related to persistence of data structures and dynamic maintenance of minimum spanning trees. In addition, the PI developed the concept of a well-separated pair decomposition of points in d-dimensional space. For a set P of n-points, this consists of a binary tree whose leaves are in P, with internal nodes corresponding to subsets of P in the natural way, and a list of pairs of nodes, such that the sets corresponding to each node are geometrically separated by appropriate distance, and each distinct pair of points is 'covered' by exactly one of the pairs of the nodes. The PI had made use of this decomposition in designing efficient algorithms for n-body potential field computations and several related problems. These techniques are extended to other problems.

University of Wisconsin – Madison; Max G Lagally; ***DNA Computation on Surfaces*****; (CCR-9613799); $300,000; 12 months. (Joint support with the Experimental Systems Program Program, the Computational Biology Activities Program, and the ARPA-COMPUTATIONAL PROTOTYPING Program)**

(1) This multi-disciplinary project concentrates on developing computation at the molecular level, via manipulation of DNA strands on surfaces. An experimental demonstration of DNA-based computation by Adleman(a) has changed the view of what computation is, (b) offers the potential for unprecedented computing power at the molecular level, and (c) raises fundamentally new research problems in chemistry, computer science, and material science. This project represents a concerted attack on these problems which involves close collaboration between chemists, material scientists, and computer scientists.(2) This research will help answer two major questions on DNAcomputation. First, does computation at the molecular level have the potential to provide computing power that is orders of magnitude greater than foreseeable extrapolations of current technology? Second, is this type of computation suited to solving NP-hard problems (problems that are well beyond the limits of current technology). (3) The potential for huge computing power rests on the use of DNA strands to encode information and the manipulation of these strands in a massively parallel fashion, involving as many as $2^7 0$ distinct strands. The premise of this project is that surface-based chemistry is a critical technology in approaching this scale. With this approach, DNA strands are immobilized on a surface, thus allowing a much greater degree of control in chemical processes that manipulate the DNA than is achievable via the test tube based methodology of Adleman. While surface based chemistry is the basis for recent strides in combinatorial chemistry, this project is the FIRST to fully exploit surface-based manipulation of DNA strands for the purposes of DNA computation.(4) Scaling up current surface chemistry to approach the requirements of DNA-based computation requires high-quality research at the interface of material science and chemistry. Improvements in the nanoscale morphology and chemical makeup of the surface is key to ensuring that a high density of information can be obtained and reliable chemical manipulations performed. Good surface attachment chemistry and control of chemical "operations" or enzymatic processes are also extensively developed. This project should result in significant advances in the state of-the-art in surface chemistry and should

provide a solid basis for predicting the limits of surface-based DNA computation.(5) An eventual goal of DNA-based computation is to perform massively parallel searches for optimal solutions of NP-hard problems. However, the differences between DNA-based computation and conventional technology require that radically new algorithms be developed for this paradigm. Tempering the potential for unprecedented parallelism (up to 2 to the 70th power simultaneous operations) is extremely slow and error prone nature of the operations themselves. Novel strategies for generating and searching solution spaces for NP-hard problems are investigated. These strategies embody sound algorithmic principles that can be applied to a range of problems. Using a combination of analysis and simulations, the effectiveness of these strategies are tested on selected applications. This work provides new ways of attacking NP-hard problems that, while designed for a hypothetical DNA based computer, are valuable for massively parallel computing paradigms, other than the DNA-based paradigm studied here.(6) Thus this comprehensive program will provide a deep understanding of the potential and limitations of DNA-based computation, from both the chemical and algorithmic viewpoints.

University of Nevada Las Vegas; Lawrence L Larmore; ***On-line Competitive Algorithms*****; (CCR-9503441 A001); $51,695; 12 months.**

This project focuses on a number of topics related to on-line problems. On-line algorithms work with only partial information about the input data. At each time step, such an algorithm receives one unit of data, and has to produce partial results before seeing the remaining data. On-line competitive algorithms are algorithms which return a solution that is not worse than a constant times the optimal one. One of the goals is to solve the k-server problem. An algorithm for k servers that is believed to be optimal has already been developed shown to be 2-competitive for two servers. A closed form function, which is a good candidate for the potential function for k = 3, is examined. A second goal is to attack a number of other on-line problems for which optimal competitiveness is unknown, including: randomized list update, multi-processor scheduling, and page migration; and to develop a general theory of on-line problem solutions. A third goal is to show how the techniques, developed for the solution of the k-server problem, can be applied to other on-line problems. In particular, applications of work functions, forgiveness and stable distributions to general on-line problems are developed. Note: This is a companion award to CCR-9503498 (PI: Chrobak) of University of California, Riverside.

Princeton University; Richard J Lipton; ***SGER: Proposal for Research on DNA Based Computation*****; (CCR-9633103); $49,889; 12 months. (Joint support with the ARPA-ITO Program)**

This Small Grant for Exploratory Research (SGER), funded by the Informations Technologies Office, DARPA, addresses the following problems in the new area of Biomolecular/DNA Based Computation:(a) Determining correct models for biomolecular/DNA computations;(b) Developing techniques that make biomolecular/DNA computations error tolerant;(c) Determining a set of "practical" solution techniques for certain NP-hard problems (such as, breaking DES cypher and "accurate" pattern matchings for the human genome project) that exploit the massively parallel biomolecular/DNA computations. One of the goals of this SGER project is to initiate and develop collaborative efforts with the Biochemistry Department at Princeton University.

Yale University; Laszlo Lovasz; ***Algorithms in Convex Geometry and their Applications in Combinatorial Optimization*****; (CCR-9402916 A002); $79,697; 12 months. (Joint support with the Numeric, Symbolic & Geometric Computation Program and the Geometric Analysis Program)**

This project investigates the problems of volume computation, sampling from geometric domains, integer programming, and some basic, but not yet well understood, problems such as containment between convex bodies. The goals are to develop new algorithms, to test implementations and, parallel to this, to explore theoretical bounds on the efficiency of algorithms. The main tool in the design of these algorithms is randomization, in particular random walks and Markov chains.

International Computer Science Institute; Michael Luby; ***Workshop at ICSI: On Randomized Algorithms and Computation December 17-22, l995, Berkeley, California*****; (CCR-9531792); $22,697; 6 months. (Joint support with the Probability Program)**

This project funds speaker and participant costs for a workshop, titled "Randomized Algorithms and Computation", at ICSI, December 17 – 22, 1995. The Workshop emphasizes applications to approximation algorithms for optimization and counting problems, coding theory with network and multimedia applications, cryptography and security, distributed computation, and derandomization of algorithms. The Workshop: (1) Brings together leading researchers from Europe, Israel, Canada and the U.S.; (2) Is being combined with the Weizmann Institute Workshop on "Randomness and Computation" (normally held in Israel in January); (3) Is organized in cooperation with the ESPIRT RAND Working Group. This project is a cooperative funding effort with support for European participants provided by the ESPRIT RAND-REC Project.

Iowa State University; Jack Lutz; ***Presidential Young Inves-***

***tigator (PYI) Award: The Internal Quantitative Structure of Complexity Classes*; (CCR-9157382 A005); $28,000; 12 months.**

New extensions of Lebesgue measure theory are used to investigate the internal, quantitative structure of complexity classes. The relationships among intrinsically pseudorandom objects, pseudorandom number generators, probabilistic and interactive complexity classes, and completeness phenomena are studied. Another focus is the measure structure of exponential time complexity classes.

Carnegie Mellon University; Bruce M Maggs; *NYI: Routing Protocols and Mechanisms for Tolerating Failures in High-Speed Networks*; (CCR-9457766 A002); $25,000; 12 months. (Joint support with the Computer Systems Architecture Program)

This project studies protocols for admitting and routing calls in distributed networks, and for tolerating failures in these networks. Its goal is to understand how to best use high-speed networking technology, such as ATM, to support parallel supercomputing, and to advance the NII (National Information Infrastructure) initiative. In the supercomputing domain, it seeks to develop network structures, routing algorithms, and mechanisms for tolerating faults that are both practical and provably effective. In the realm of the NII, it is analyzing the behavior of simple algorithms, including greedy algorithms, for admitting calls in wide-area integrated-service networks.

Old Dominion University; Stephan Olariu; *On the Linear Structure of Graphs - II*; (CCR-9522093); $58,640; 24 months.

Many families of graphs including interval graphs, permutation graphs, trapezoid graphs, and co-comparability graphs demonstrate a type of linear ordering on their vertex sets. It is precisely this linear order that is exploited, in one form or another, in a search for efficient algorithms for these classes of graphs. The classes mentioned are known to have wide-ranging practical applications. In addition, they are all subfamilies of the class of graphs called the asteroidal triple-free graphs (AT-free graphs, for short). This project extends the research completed under CCR-94-07180: the PI has shown that every connected AT-free graph contains a dominating pair (i.e., a pair (x,y) of vertices with the property that all (x,y)-paths represent dominant sets), which can be found in linear time. In spite of these encouraging first steps in the study of AT-free graphs, the structure of the AT-free graphs is not well understood. In particular, no efficient recognition algorithm is available. The research effort investigates the ATfree graphs using the powerful tool of graph decomposition, with the intention to reveal new structural properties of AT-free graphs that can be exploited to obtain efficient sequential and parallel algorithms. The results provide a unifying look and a common algorithmic base from which to investigate the four classes of graphs mentioned above.

University of California – Berkeley; Christos H Papadimitriou; *Research in Algorithms Complexity and Database Theory*; (CCR-9626361); $253,710; 36 months. (Joint support with the Database & Expert Systems Program)

The PI and his collaborators have been involved in: (a) some of the subfields of theory that have witnessed impressive recent advances (approximation algorithms, on-line algorithms), (b) a subfield that has been very successful and relevant, but is facing new challenges (database theory), and(c) "exporting" complexity-theoretic ideas to other disciplines (e.g., Game Theory, Economics, and Artificial Intelligence). The goals of this project include: (1) The deepening and broadening of recent advances in approximation algorithms and on-line algorithms, by attacking fundamental problems such as the Euclidean traveling salesman problem and the k-server conjecture; (2) The study of alternative models and metrics, for example, in relation to on-line algorithms for storage allocation, that are closer to reality and to computational practice; (3) The introduction of approximation and on-line problems that in a small way contribute to our understanding of the computational requirements of a global information environment; (4) The study of database-theoretic problems that are posed by novel media and heterogeneity; (5) The continuation of PI's research program that considers problems from other disciplines (such as, Game Theory, Genetics, and Artificial Intelligence) from a complexity-theoretic point of view.

Stanford University; Serge A Plotkin; *Design of Efficient Algorithms for Multicommodity Flow and Related Combinatorial Optimization Problems*; (CCR-9304971 A002); $71,635; 12 months.

The project studies combinatorial optimization, with emphasis on problems related to commodity flow and minimum-ratio cuts. Deeper understanding of the combinatorial structure of these problems will result in improved algorithms for a variety of applications, including graph bisection, small area VLSI layout, scheduling, and routing. All of the currently known algorithms for finding exact solutions to multicommodity flow are very slow. These algorithms are based on general linear programming (LP) techniques and do not take advantage of the underlying combinatorial structure of the problem. Applying general LP techniques to solving such problems leads to even slower algorithms. The main goal of this project is to find alternative approaches to solving multicommodity flow and related problems. The goal is to improve both sequential and parallel complexity. In many applications, solving a multicommodity flow is only a first step in approximately solving an NP-complete problem; in the majority of such cases there is no need to have an exact solution of the multicommodity flow problem. One of the focuses of the project is to develop the design of efficient approximation algorithms. Another natural application of the multicommodity flow methods is in the area of routing in distributed communication networks. For example, problems related to managing available bandwidth can be stated as variants of multicommodity flow, where the information rate corresponds to flow and the bandwidth of every link corresponds to its capacity. The distributed, real-time nature of communication networks presents several challenging obstacles that do not exist in an off-line, sequential setting. Among them is the fact that not all the information is known in advance, which means that the algorithms have to work on-line. Second, there is usually no 'manager' node that has a complete information about the network and that makes all the decisions. Therefore, the algorithms should work with partial information.

Stanford University; Serge A Plotkin; *Research into Network Algorithms and Related Problems*; (CCR-9307045 A003); $65,659; 12 months.

This project studies fundamental network problems, including maximum, minimum-cost, and multicommodity flow problems, shortest paths problems, minimum cut problems, and the assignment problem. These are classical combinatorial optimization problems that have numerous applications. Designing efficient algorithms and understanding combinatorial structure of these problems is important from a theoretical point of view because these problems are very basic to the field, and from a practical point of view because instances of these problems need to be solved. There is interest in extending the results of this work to related problems, such as linear programming. The research has three interrelated components: (1) the first involves a better understanding of combinatorial structure of these problems; (2) the second involves development of new algorithms that improve currently known bounds on the problem complexity; and (3) the third involves experimental evaluation of the currently existing and newly developed algorithms. The worst-case theoretical efficiency does not always correspond to practical efficiency. Experimental evaluation of existing and newly developed algorithms for some of the problems is of interest. The experimental research also motivates theoretical research by identifying subproblems and data structures needed by the implementations and by suggesting variations of algorithms which may be more efficient than the original variants originating in theoretical research.

University of Texas – Dallas; Balaji Raghavachari; *Research Initiation Award (RIA): Approximation Algorithms for Hard Problems in Discrete Optimization*; (CCR-9409625 A002); $22,646; 12 months.

The objective of this research is to develop new algorithmic techniques to compute efficiently near optimal solutions to a number of NP-hard graph problems arising in network-design. A partial list of problems studied include: (a) problems in Euclidean setting including the traveling salesman problem; (b) designing low-congestion networks for multicommodity flow; (c) computing broadcast-trees for fast dissemination of information in networks; (d) finding low-cost networks of specified connectivity. The goal is to develop algorithms which not only generate good solutions, but are also practically feasible. Algorithms developed are both implemented and tested.

Carnegie Mellon University; Ramamoorthi Ravi; *CAREER: Approximation algorithms for NP-hard problems in networks and biology*; (CCR-9625297); $90,000; 22 months. (Joint support with the Computational Biology Activities Program)

This project focuses on the design of approximation algorithms for NP-hard problems arising in two important application areas: (1) design of networks and (2) computational molecular biology. The work involves the development of general techniques to solve classes of problems in both areas, applying and extending tools from the area of approximation algorithms, such as the primal- dual schema, bicriteria formulations and approximations, and using graph-theoretic arguments in the proof of performance guarantee. In particular, the issues addressed in the area of network design include improved algorithm for minimum-cost survivable network design by extending the primal-dual method, and a generalization of the Steiner tree problem with covering requirements. In the area of computational molecular biology, the areas of investigation include the application of bicriteria approximation methods to the physical map assembly problem, and fundamental questions in multiple sequence alignments – local, global and tree based. The Integrated Educational Plan of this CAREER Grant includes (a) participating as Resident Scientist in the DIMACS High School Teacher Education Program in Computational Biology, (b) developing a"more modern" undergraduate cross-disciplinary computational biology at CMU, (c) establishing and running a graduate research seminar for doctoral students in the Algorithms, Combinatorics and Optimization Doctoral Program at CMU, and (d) developing new graduate courses in network design and approximation algorithms.

University of Illinois – Urbana; Edward M Reingold; *Decision Trees and Discrete Search Problems*; (CCR-9530297); $54,252; 12 months.

This project explores two aspects of discrete search:(a) The search for the majority in the set of objects;(b) The search for critical points in an unknown function.In both cases the main tools are detailed analyses of binary decision trees and Kraft's inequality. The goals of the project include the design of, analysis of and the proof of lower bounds for search algorithms.

University of Illinois – Urbana; Edward M Reingold and Nachum Dershowitz; *U.S.-France Cooperative Research: Specifying and Finding Solutions to Constraint and Search Problems (with INRIA)*; (INT-9507248); $15,000; 36 months. (Joint support with the Western Europe Program Program and the Programming Languages and Compilers Program)

This three-year award will support U.S.-France cooperative research in theoretical computing and programming languages between research groups at the University of Illinois and at INRIA (the French National Institute for Research in Computer Science and Applied Mathematics). The U.S. investigators, Edward Reingold and Nachum Dershowitz, collaborate with Claude Kirchner and Michael Rusinowitz at INRIA/Lorraine and with other researchers at the University of Nancy and the French Center for Research in Informatics at Nancy. Their research deals with problems in term rewriting and search techniques and in solutions of constraint and search algorithms. The U.S. investigators bring to this collaboration expertise in algorithms and equational reasoning. This is complemented by the French investigators' expertise in programming languages, constraint-based reasoning, and term rewriting techniques. The problems under study have important applications in image processing, programming languages, and automated deduction.

Cornell University; Ronitt A Rubinfeld; *CAREER: Algorithms for Self-testing/Correcting Program and Learning*; (CCR-9624552); $100,000; 24 months.

This project focuses on two lines of research: (1) program correctness and (2) the problem of learning deterministic and probabilistic finite automata:(1) Because the simplest of programs can be full of elusive errors, the study of program checkers, self-testing programs and self correcting programs, was introduced, in order to permit the use of a program without trusting that it works correctly. In this area, the goal of this project is to develop a core of algorithmic techniques for writing fast and simple checkers, self-correctors, and self-testers;(2) The goal in the area of learning deterministic and probabilistic finite automata is to study the applications of automata learning algorithms to on-line page replacement strategies, reinforcement learning, game theory, part-of-speech tagging, phoneme modeling, and parsing. The Integrated Educational Plan of this CAREER Grant includes designing new projects to facilitate undergraduate and graduate student involvement in the research process and mentoring undergraduates.

Stanford University; Jeanette Schmidt; *VPW: Detection and Analysis of Reptitions in Biosequences*; (HRD-9627109); $160,565; 12 months. (Joint support with the Visiting Professorships for Women Program, the Science and Technology Centers Program, the Advanced Scientific Computing Centers Program, the Computational Biology Activities Program, and the ARPA-ITO Program)

This research focuses on the development of computational tools for analyzing and comparing sequences. An important aspect of this work is the collaboration with scientists in the biological sciences. Particular emphasis will be placed on the development of computational tools for the detection of multiple repeated regions and tandem repeats in both DNA and protein sequences. These regions are very important as they have been associated with regulatory, stabilizing, and evolutionary sites. Many such repeated regions are difficult to detect because the repetitions are inexact, sometimes exhibiting a low percentage of identical symbols. An analysis will be conducted to gain more insight into the difficult problem of determining the threshold for statistical significance of such repeats. The efficient identification of multiple repeats requires a subtle similarity measure. This will require the identification and representation of optimal as well as near optimal alignments between subsequences. An empirical comparative study will be conducted to assess the applicability and significance of the new algorithms. Interactive activities include teaching a seminar to encourage a dialogue between students from different disciplines; visiting local high schools; organizing group meetings with undergraduate interns; and interacting with the graduate students; participating in journal clubs, colloquium and seminars; and engaging in joint research with other members of the Biochemistry, Medical Informatics and Genome Center at Stanford University.

University of Houston; L. Ridgway Scott; *NSF Grand Challenge Application Group for Computational Biomolecular Design*; (DMS-9217374 A005); $600,000; 12 months. (Joint support with the Computational Mathematics Program, the Molecular Biophysics Program, and the Computational Biology Activities Program)

The Grand Challenge Application Groups competition provides one mechanism for the support of multidisciplinary teams of scientists and engineers to meet the goals of the High Performance Computing and Communications (HPCC) Initiative in Fiscal Year 1992. The ideal proposal provided not only excellence in science: focussed problem with potential for substantial impact in a critical area of science and engineering) but also significant interactions between scientific and computational activities, usually involving mathematical, computer or computational scientists, that would have impact in high-performance computational activity beyond the specific scientific or engineering problem area(s) or discipline being studied. In the award to Scott, McCammon, Glowinski, and Pettitt, the research will resolve several issues that will enable high-performance computing to solve critical problems in biomolecular design. The most obvious problem relates to the difficulty in using advanced computer architectures. Secondly, new algorithms (mathematical and computer science-based) and methods (physical science-based) are needed to solve problems with reasonable efficiency. Finally, detailed understanding of what types of problems can be treated with vastly improved computing resources is needed to guide and provoke the development of both algorithms and software. The grand challenge problems motivating the computational research include thermodynamic cycle perturbation theory applied to ligand-binding proteins and Poisson-Boltzman/Brownian dynamics applied to protein design. The award will support postdoctoral associates, graduate students, and the purchase of a high performance computing system.

Cornell University; David B Shmoys; ***Near-Optimal Solutions for Combinatorial Problems: Algorithms and Complexity*; (CCR-9307391 A002); $46,404; 12 months.**

The study of approximation algorithms for combinatorial optimization problems dates back to the 1960's, and in spite of significant attention paid to this area, there has not been much progress until the past few years. Recently, there have been a number of important breakthroughs that have introduced new techniques, both in the design and analysis of approximation algorithms, and more significantly, in techniques for providing evidence that problems do not have efficient algorithms that guarantee near-optimal solutions for particular performance guarantees. The problems studied include: (a) the vertex cover problem; (b) the maximum acyclic subgraph problem; (c) the minimum-cost satisfiability problem; (d) the Steiner tree problem; (e) the traveling salesman problem; (f) the bin-packing problem; (g) the graph partitioning problem; and (h) several scheduling problems. In the search for improved approximation algorithms for these problems, the primary focus is on methods that rely on the use of linear programming as a tool both in the design and the analysis of the algorithm. Also these new techniques for proving nonapproximability results are applied in a variety of settings; in particular: (1) proving lower bounds on the absolute error of approximation algorithms (as opposed to the relative error); and (2) strengthening lower bounds (where traditional methods were able to prove lower bounds that did not match the performance of the best known approximation algorithms).

New York University; Alan R Siegel; ***Analyzing Algorithms: Advanced Methods for Recurrence Equations, Probabilistic Estimates, and Operator Methods*; (CCR-9503793 A001); $60,680; 12 months.**

The goal of this continuing research project is to develop more systematic theorems and methods for analyzing difficult algorithms. These methods are briefly characterized as follows: (1) Perturbation theory for recurrence equations in computer science. Perturbation theory is a mathematical technique which was developed for analyzing difficult differential equation problems in mathematical physics. The research objective is to develop a tool kit of theorems that can apply to recurrences ranging from advanced issues in hashing to the elementary concerns of basic irregular divide-and-conquer algorithms; (2) Replacing operator methods in performance analysis. Operator methods also arise in the analysis of physical methods, and have been used to establish conservation laws, and compute the time evolution of fundamental quantities such as energy in complicated physical processes. A variation on these methods has been used to analyze several complicated processes in computer science, including coalesced hashing. However, recently it has been shown that coalesced hashing can be analyzed more completely and more easily by standard analytic methods. The goal in this part is to capture algebraic properties of operator methods, to translate them into analytic formulations, and to use these translated formulations to analyze difficult algorithms; (3) Probabilistic estimates. A growing number of computer science results use large deviation bounds, including fault tolerance analysis, the analysis of throughput in probabilistic situations, and the analysis of probabilistic algorithms. Many computer science applications introduce varieties of extra constraints, which render the standard theorems inapplicable. The final goal is to continue research on probabilistic estimation methods.

State University of New York – Stony Brook; Steven Skiena; ***Interactive Sequencing by Hybridization*; (CCR-9625669); $116,000; 20 months. (Joint support with the Computational Biology Activities Program)**

This project explores a new approach to sequencing by hybridization (SBH), which uses interaction to dramatically reduce the number of oligonucleotides for de novo sequencing of large DNA fragments, while preserving the parallelism which is the primary advantage of SBH. The goals of the project include: (1) Development of efficient algorithms for the interactive design of sequencing chips; and (2) Utilization of simulation results to establish that 64kb fragments can be sequenced (obtained using approximately ten rounds of standard-sized 64k oligonucleotide arrays). The project includes the following theoretical and laboratory research in interactive SBH: (a) In a collaborative effort with biochemists at Oxford University, England, algorithmic issues in the design and manufacture SBH arrays, under a variety of fabrication technologies, are examined; (b) Algorithmic issues inherent in reconstructing DNA sequences from hybridization data, under complications such as non-classical chip designs and hybridization errors, are studied; and (c) Algorithmic issues inherent in other novel variations of SBH, particularly photolithography-based interactive SBH and positional SBH, are investigated. This project is jointly funded with the Computational Biological Activity.

Dartmouth College; Clifford S Stein; *CAREER: Efficient Algorithms in Scheduling, Network Algorithms and Biology: Theory and Practice*; (CCR-9624828); $99,885; 24 months.

9624828 Stein Combinatorial optimization problems arise in fields as diverse as transportation,communication, large-scale computing, and biology. This project focuses on three fundamental areas in combinatorial optimization:(1) Scheduling, including good average-case scheduling algorithms and algorithms for scheduling a network of processors;(2) Flows and Cuts, including multicommodity flows, minimum cuts, and shortest paths;(3) Computational Biology, including problems that arise as part of the human genome project, algorithms for DNA folding, and techniques for studying evolution. The Integrated Educational Plan of this CAREER Grant includes curriculum reform, new course development, training of teachers in new technologies, and outreach to high school students.

Brown University; Roberto Tamassia; *Graph Drawing*; (CCR-9423847 A001); $75,036; 12 months.

Graph drawing addresses the problem of constructing geometric representations of graphs and networks. It is an emerging area of research that has received increasing attention both in the theory and systems community. From the theoretical viewpoint, graph drawing combines flavors of topological graph theory and computational geometry, and offers many algorithmic challenges. From the practical viewpoint, the automatic generation of drawings of graphs has important applications in key computer technologies such as software engineering, computer-aided-design, database systems, and visual interfaces. The research project addresses both theoretical and applied issues in graph drawing. Specific problems studied include: (1) Trade-offs between various properties of drawings, such as area, aspect-ratio, and angular resolution; (2) Upward planarity testing for directed graphs; (3) Dynamic planarity testing and incremental layout maintenance; (4) Three-dimensional drawing; (5) Declarative approaches to graph drawing and related constraint-satisfaction issues; (6) A graph visualization system and its applications.

Princeton University; Robert E Tarjan; *Design and Analysis of Data Structures and Combinatorial Algorithms*; (CCR-9626862); $240,000; 36 months.

This project is a continuation of a multi-year effort to discover and understand efficient data structures, combinatorial algorithms, and the relationships between the two. The emphasis is placed on studying problems that are fundamental, long-standing, and/or of significant practical importance. Sought are not only specific algorithmic results, but also general principles of design and analysis. In particular, the goals of the project include: (1) to devise specific algorithms and data structures that result in more efficient solutions of important combinatorial problems; (2) to study the worst-case and typical-case performance of new and older methods, to allow prediction of the best method for a given situation; (3) to develop general tools and principles for design and analysis, and apply these tools and principles systematically in the construction of new algorithms and data structures. Asymptotic efficiency is not the only goal; methods are sought that display simplicity and even elegance in addition to efficiency. Success is achieved when a method is found that is both theoretically efficient and usable in practice. The objective is not only to devise new methods, but to gain greater insights into old ones, so that all solutions can be stripped down to their essence, achieving both greater efficiency and greater simplicity.

University of North Texas; Stephen R Tate; *Research Initiation Award (RIA): On-Line and Dynamic Algorithms: Computing with an Uncertain Future*; (CCR-9409945 A002); $20,313; 12 months.

This project involves the study of both on-line and dynamic algorithms—types of algorithms with the unifying theme of uncertainty about future requests to the algorithm. New methods of average-case analysis are introduce to the study of on-line algorithms, drawing from experience in information theory and data compression (and, in particular, from stochastic complexity). In addition, the computational complexity of making optimal on-line decisions when the input requests are drawn from a known distribution, is investigated. In the area of dynamic algorithms and data structures, techniques are developed for designing dynamic algorithms by examining the structure of existing (or perhaps new) parallel algorithms. Initial results show how to use easily computable properties of algebraic circuits in order to design dynamic algorithms, and these results are extended in this project to a wider class of problems involving more complex incremental operations on the inputs. In addition, dynamic algorithms for parallel machines are developed, which are important for applications which require very fast response to dynamically changing data.

Cornell University; Leslie E Trotter; *MDC: Computational Optimization research Center*; (DMS-9527124 A001); $300,000; 12 months. (Joint support with the Computational Mathematics Program)

This award covers a broad research program in the area of computational optimization. The central thrust of the research addresses real-life, large-scale mixed-inter programming problems arising from important industrial application areas, such as manufacturing, telecommunications, and distribution. These problems are of vast importance towards the nation s present and future economic and technological development. The problems are also extremely intractable, both from the viewpoint of practical computation, as well as in terms of theoretical understanding, and they will only become larger and more difficult in future applications. The computational and methodological technology for solving integer programming models has made steady advances over the last thirty years, often following improvements in computing machinery as much as new theoretical insights. We are now at the starting point of a major quantum improvement in the technology. Not only are there several new and promising methodological venues available, but we are witnessin g a revolution in the computing world, in particular involving parallel computing and operating systems, with profound implications for science and engineering. It is almost certain that within a few years all computers will have multiple CPUs, and that all major scientific and engineering computing will be done on clusters of multiple-CPU machines. As supercomputing becomes redefined, a push to parallelize algorithms is starting now, and the end result will be software of extraordinary complexity by today s standards. However, it is fair to say that the field of parallel integer programming is, at best, in its infancy. The research seeks to propel computational integer programming into the realm of parallel computing, with the objective of effectively handling difficult real-world problems. The objective will require a multidisciplinary approach. Our work will seek to extend and develop solution methodologies for integer programming, in particular involving cutting planes, column generation, and branch-and-cut techniques. This will be integrated with computational theory techniques, especially the design of efficient data structures for parallel integer programming. Finally, there will be a large component of software engineering, necessary for the implementation of efficient parallel programs. This work combines the scientific and computational resources of three participating organizations, Columbia University, Cornell University, and the IBM T.J. Watson Research Center, to create a collaborative research environment to attain the goals outlined above. Among the many important advantages of such a close relationship between university computing research groups and that of a major hardware producer is the focus on practical ends provided by the industrial component.

Georgia Technology Research Corporation – Georgia Institute of Technology; Vijay V Vazirani; ***Two Themes in Approximation Algorithms: Use of the Primal- Dual Schema, and Problems in Network Design*****; (CCR-9627308); $240,000; 36 months.**

This projects explores problems in the area of approximation algorithms, with emphasis on two themes: (1) Primal-dual schema based on approximation algorithms; and (2) Approximation algorithms for problems in network design. First, several problems are investigated that appear to have "optimal" approximate solutions using primal-dual schema, including generalized Steiner network problem, integral multicommodity flow and multicut in trees, the set cover problem, and the k-minimal Steiner tree problem. The second goal of the project is to provide "good" approximate solutions to problems, identified by AT&T Bell Labs and Bellcore, in the area of implementation and maintenance of low-cost broadband networks.

Kent State University; Paul S Wang; ***Efficient Algorithms and System Interface for Scientific Computation*****; (CCR-9503650 A001); $14,924; 6 months.**

This project focuses on three related areas: (1) No-missing-term p-adic lifting — Establishing a faster algorithm for recovering final results in multivariate polynomial operations such as GCD (Greatest Common Divisor) and factoring. The lifting method is most effective when the polynomials involved are sparse, containing just a few of all possible terms. (2) Function Evaluation Optimization and Applications — Speeding up the repeated evaluation of functions in closed- form formulas through the application of recurrence relations. Nested recurrences called "chain of recurrences" will be derived by symbolic computation and applied to increase the evaluation speed of formulas at regular intervals. Application in interactive graphical visualization will be made. (3) Interfacing Distributed Scientific Systems — Seeking to establish an efficient and standardized protocol to connect different autonomous systems for scientific computation. Such a protocol is necessary to build distributed "problem solving environments". By exploiting sparseness and solving systems of linear equations, the p-adic lifting procedure of item (1) will be investigated. Feasibility will be studied. Algorithms will be specified, implemented and applied. For area (2), automatic generation of chain of recurrences (CR) for functions of one or more variables will be implemented, and tested for speed and effectiveness. The numeric properties of CRs will be studied. The technique will be applied to speed up interactive graphical visualization of mathematical functions. For area (3), an efficient protocol (called MP) will be designed. Libraries in C will support the protocol.

Polytechnic University of New York; Joel Wein; ***Off-Line and On-Line Techniques for Scheduling and Resource Allocation*****; (CCR-9626831); $88,836; 36 months.**

This project focuses on developing algorithms for a variety of problems in the theory and practice of scheduling and resource allocation. The project concentrates on: (i) developing improved approximation algorithms for problems in combinatorial scheduling theory; (ii) developing on-line algorithms for resource allocation; and (iii) implementing and testing these algorithms. In particular, the first goal of the project is to develop general techniques for the design of approximation algorithms for NP-hard scheduling problems. These techniques are based on several classes of integer programming formulations, and they have already been applied to yield approximation algorithms for average weighted completion time for a number of constrained scheduling problems. The second goal in this area is to better understand, from the prospective of worst-case analysis, the strengths of these general techniques. The third goal is to investigate approximation algorithms for multi-criteria scheduling problems. The fourth goal is to study on-line scheduling and resource allocation questions, focusing initially on algorithms that provide good average service to a stream of jobs, of different types and priorities, that arrive over time. The outcome of this study is expected to be: (a) a better theoretical understanding of a number of models and optimality criteria and (b) algorithms for a network of workstations and modern parallel computers (such as the IBM SP2).

MA Fellowships; David B Wilson; ***Mathematical Sciences Postdoctoral Research Fellowships*****; (DMS-9627760); $75,000; 48 months. (Joint support with the Infrastructure Program Program)**

The PI will receive his doctoral degree from the Massachusetts Institute of Technology and will pursue research under the guidance of Alistair J. Sinclair at the University of California at Berkeley. The research is on the development and analysis of algorithms which use the theory of Markov chains to produce random samples from large, finite sets. %%% The Mathematical Sciences Postdoctoral Research Fellowships are awards to recent recipients of doctoral degrees in the mathematical sciences. These awards are a means of contributing to the future vitality of the scientific effort of the nation. As researchers in the mathematical sciences expand their interactions with other disciplines, and as the interplay increases between the various areas of mathematics itself, opportunities for postdoctoral research and training are becoming increasingly important. The Fellowships are designed to permit awardees to choose research environments that will have maximal impact on their future scientific development. The fellowship is designed to provide 24 months of support divided into 18 academic-year months and three periods of two summer months. The recipient may choose (1) the Research Fellowship option which allows for full time support for any 18 academic year months in a three year period, in intervals not shorter than 3 consecutive months or (2) the Research Instructorship option which allows the 18 months of academic year support to be taken as nine months of full time support and 18 months of half time support. Not more than two months of summer support from this Fellowship may be received in any calendar year.

Xerox – Palo Alto Research Center; Frances F Yao, Marshall W Bern, Daniel H Greene, and John R Gilbert; ***Design and Implementation of Approximate Algorithms*****; (CCR-9404113 A002); $35,000; 12 months.**

This award gives partial support for a postdoctoral researcher to work in the theory group at Xerox Palo Alto Research Center (PARC). Postdoctoral researchers would collaborate closely with the principal investigators, freely choosing and mixing topics from current theory group projects and their own areas of interest. The two largest projects are mesh generation and approximate searching. Mesh generation requires the partitioning of a complicated geometric domain into a complex of small simple cells. Approximate searching requires the efficient retrieval of data objects close to a given query point in a high-dimensional space (either combinatorial or geometric). Each of these major projects covers a number of specific algorithmic questions.

Rice University; Willy Zwaenepoel, Alan Cox, and Alejandro A Schaffer; ***Parallel Software for Genetics*****; (DBI-9408503 A002); $81,909; 12 months. (Joint support with the New Technologies Program Program and the Computational Biology Activities Program)**

As information about DNA sequences accumulates, it is becoming increasingly important to identify the location of genes or segments of genes. Genetic linkage analysis is a statistical technique used to map genes and find their location in relation to other known segments of DNA. The most popular software for doing this is called LINKAGE, but the computational time required to solve problems is often prohibitive. FASTLINK, developed by this principle investigator, is an improved version of LINKAGE. This award will support the parallelization of the FASTLINK code to substantially improve the computational speed, and to improve the shared distributed memory system named TreadMarks, which manages the algorithms on a network of workstations.

Parallel, Distributed and Probabilistic Computation

Rutgers University; Jozsef Beck; ***A Probabilistic Theory of Positional Games*****; (DMS-9626151); $40,000; 12 months. (Joint support with the Algebra & Number Theory Program and the Probability Program)**

The investigator is pursuing a new, probabilistic approach to positional games (i.e. board-games, or combinatorial games). The straightforward way to analyze a position is to examine all of its options and all the options of these options and all the options of the options of these options and all the.... The obvious difficulty is that this exhausting search through the game-tree takes an enormous amount of time even for very simple games (exponential running time in terms of the board-size). Note that the game-matrix approach, the set-up of traditional game theory, is much worse: usually requiring double exponential time. A desperate attempt to make up for the lack of time is to study the random walk on the game-tree. Now the basic idea of the probabilistic theory of the investigator is that the probabilistic analysis of the random walk (i.e. the random game) is often a tractable problem, and also what one has learned from this analysis can often be converted, via potential arguments (like resource counting with loss-probabilities), into deterministic optimal strategies. The probabilistic theory is basically a game-theoretic adaptation of the so-called Probabilistic Method in Combinatorics applied to hopelessly complicated games where algebraic and other exact methods fail to work. The surprising thing is that this desperate attempt turns out to be very successful for large, interesting classes of positional games. The main research topics are: the foundations of positional games (a surprisingly non-trivial task), game-theoretic Ramsey theory, game-theoretic random graph theory, game-theoretic Lovasz Local Lemma, and applications in algorithms and complexity. An extremely exciting aspect of studying positional games is that it might give a better understanding of how human intelligence works.

Yale University; Richard Beigel and James Aspnes; *Parallel Fault Diagnosis*; (CCR-9415410 A001); $60,000; 12 months. (Joint support with the Computer Systems Architecture Program)

Computer systems of the future may consist of thousands or even millions of processors. Inevitably some of the processors will fail. Faulty processors need to be repaired so that the system will continue functioning. The goal of this project is to investigate improvements to the existing algorithms for parallel fault diagnosis. Second, and more importantly, the project will develop dynamic models for fault diagnosis that are more realistic than previous models.

Truman State University; Jay Belanger; *Research at Undergraduate Institution (RUI): Average-case Hierarchies*; (CCR-9503601 A001); $18,257; 12 months.

This RUI research project is a companion project to CCR-9424154 (PI: Wang) of UNC at Greensboro. The research project centers around the relations among average-case complexity classes. Complexity theory began with the study of hierarchies among worst-case complexity classes, and these have been thoroughly investigated. Recently, a reasonable amount of work has been done investigating average-case completeness, however very little work has been done looking into separation and hierarchy results for average-case complexity classes. What little work has been done has either used an unsatisfactory definition of average-case complexity or has produced less than optimal results. The goal of this research project is to more thoroughly investigate average-case complexity classes and hierarchies among them.

University of Wisconsin – Madison; Anne Condon; *NSF Young Investigator (NYI) Award: Randomness and Nondeterminism in Game-Like Models of Computation*; (CCR-9257241 A004); $62,500; 12 months.

This award supports on-going research efforts on interactive computational models which incorporate both randomness and nondeterminism in novel ways. A fundamental goal of the work is to address what is probably the major open problem in computer science: how randomness and nondeterminism help in computation. Representative problems in this area are stochastic games, automata theory and random walks on graphs. Stochastic games model a large class of problems in optimization called decision problems under uncertainty. Finite state automata are studied with both nondeterministic and random states in order to understand the difference between such models and their deterministic counterparts. New questions about random walks have been formulated on graphs that are partly controlled by an adversary; in our scenario edges of the graph are colored and the adversary can control the walk by specifying which color edge is followed at a given step. Further, this award supports ongoing research regarding the need for a more realistic theory of parallel computation. The design and implementation of algorithms on parallel machines is combined with analysis of these algorithms on realistic models. Two algorithmic paradigms will be restudied: dynamic programming and branch and bound. Emphasis will be on applications of dynamic programming to problems in DNA sequence analysis and of branch and bound algorithms to problems in combinatorial optimization. Also a new look is taken at what it means for a problem to be "inherently sequential".

Carnegie Mellon University; Alan Frieze; *Probabilistic Considerations in the Analysis of Algorithms*; (CCR-9530974); $54,970; 12 months. (Joint support with the Probability Program)

This is jointly funded project (between the Theory of Computing Program, CCR and Probability Program, DMS), that started under CCR-92-25008, "Probabilisitic Considerations in the Analysis of Algorithms". Probalilistic considerations arise in the analysis of algorithms in at least two important ways: (1) In a randomized algorithm the outcomes of random events are used to determine the progress of the algorithm. Randomization is now a standard tool of the computational scientist; (2) A second area of consideration is when the problem instances come from some probability distribution, and one wants to understand the average performance of a particular algorithm, which is often far better than its worst case. This project explores problems in both areas. In (1) the area of randomized algorithms, the problems examined include: (1a) random walks on graphs; (1b) approximation algorithms; and (1c) learning and routing problems. In (2) the area of probabilistic analysis, the problems studied include: (2a) greedy algorithms; (2b) in computational biology, sequencing by hybridization and physical mapping reconstruction for DNA; and (2c) some geometric problems.

Brown University; Maurice Herlihy; ***CONACyT: Topological Methods in Distributed and Concurrent Computation*****; (CCR-9613785); $191,885; 18 months. (Joint support with the CISE Institutional Infrastructure Program)**

This NSF-CONACyT Collaborative Renewal Award continues to apply techniques from the abstract mathematical area of simplicial algebraic topology (developed under CCR-9505949) to the problem of coordinating concurrent processes in a distributed or parallel computing system. Coordination problems arise at all scales of modern distributed and concurrent systems, way from synchronizing access to shared data objects in tightly-coupled multiprocessors, to allocating data paths in ATM networks. Coordination is difficult because modern parallel and distributed systems are inherently asynchronous (that is, processes may be delayed without warning for a variety of reasons, including interrupts, pre-emption, cache misses, communication delays, or failures). These delays can vary enormously in scale: (a) a cache miss might delay a processor for fewer than ten instructions; (b) a page fault for a few million instructions; and (c) operating system pre-emption for hundreds of millions of instructions. Coordination protocols that do not take such delays into account run the risk that if one processor is unexpectedly delayed, then the remaining processes may be unable to make progress. Such problems become increasingly severe as systems scale. In the past few years, a number of researchers have developed a range of new techniques, based on classical algebraic topology, for analyzing distributed and concurrent protocols and data structures. These techniques have been applied successfully, under the US-Mexico collaborative effort supported by CCR-9505949, to a range of computational problems. This renewal collaborative project extends these algebraic topological techniques to a range of more difficult problems, not just for distributed and parallel computing, but for any problem area that requires reasoning about uncertainty and partial information.

Carnegie Mellon University; Ravindran Kannan; ***Fast Randomized Algorithms for Optimization and Other Applications of Geometric Random Walks*****; (CCR-9528215); $50,000; 12 months.**

This is a companion project with CCR-95-28973 (entitled "Optimization and Learning over Convex Sets") and is an award under the Software Capitalization Program. The project's goal is to upgrade and update software that had been produced. These upgrades and updates are based on algorithms developed under support of CCR-92-08597. The software is for executing random walks that produce a sample from certain multivariate probability distributions. An update of the software involves simply changing the "number of steps" parameter in the random walk. Applications of this discrete probabilistic software (i) to contingency tables and (ii) to the membership oracles for optimization over convex sets, are investigated. With good planned dissemination (for example, on the world wide-web) and well thought-out testing, other applications of this software are anticipated.

Massachusetts Institute of Technology; David R Karger; ***CAREER: Randomization in Graph Optimization Problems*****; (CCR-9624239); $100,000; 24 months.**

This project focuses on the uses of randomization, with particular emphasis on graph algorithms. In particular, the problems to be "tackled"(but not necessarily exactly solved) by the PI's randomized graphical optimization techniques include: (1) The maximum flow and minimum cut problems can be solved exactly, with the major objective being the design of faster, simpler and more practical algorithms for arriving at a correct answer; (2) Graph coloring and bisection, analysis of network reliability, and construction of low-cost high-connectivity networks, are extremely hard to solve exactly, with the major objective being to find efficient algorithms that give answers as close as possible to the elusive optimum. Implementation and testing of the algorithmic solutions is an integrated component of this project. The Integrated Educational Plan of this CAREER Grant includes the teaching, improvement and development of theoretical computer science courses, including: (1) Undergraduate algorithms course; (2) Undergraduate discrete mathematics course; (3) Graduate randomized algorithms course; (4) Undergraduate information retrieval and text databases.

Brown University; Philip N Klein; ***Presidential Young Investigator (PYI) Award: Computational Problems in Network Design*****; (CCR-9157620 A006); $75,000; 0 months.**

Computational problems in network design and analysis are addressed. Since standard approaches to finding exact solutions to these problems are time consuming, the focus is on the development of new approximation and parallel algorithms. Multicommodity flow is used as a tool to formulate and solve network problems.

Massachusetts Institute of Technology; Nancy Lynch and Nir Shavit; *The Topological Approach to Asynchronous Computability*; (CCR-9520298); $76,034; 12 months. (Joint support with the Software Engineering Program)

The choice of primitive coordination operations, whether reading and writing a shared memory, message-passing, or more complex operations, has become a crucial design issue. To make such choices, one must analyze the behavior of multiprocess executions, reasoning about the multitude of possible combinations of concurrent operations. This is an unusually delicate and laborious task. The goal of this research is to replace this need to argue about concurrent executions with combinatorial conditions on the geometric representation of problems. This should allow researchers and designers, instead of analyzing behaviors, to apply powerful mathematical tools to recognize when certain protocols are impossible, to evaluate the power of alternative synchronization primitives, or to make explicit the assumptions needed to make a given problem solvable. This project pursues the problem of characterizing and classifying the power of synchronization primitives in different models. Prior research has developed powerful new tools based on classical topology for analyzing fault-tolerant concurrent aloglrithms and data structures in a variety of models and architectures. This research has shown that a large class of synchronization problems can be associated with a high-dimensional geometric structure called a simplicial complex, and that the computational complexity of solving the problem using a given underlying architecture is captured by certain topological properties of the complex. These properties have been used to give a complete characterization of the synchronization problems solvable in asynchronous read/write memory, and to provide impossiblity results and lower bounds for several known open problems in both the asynchronous model and in the synchronous message-passing model. These techniques are being applied toward a comprehensive theory of multiprocessor synchronization.

Stanford University; Rajeev Motwani; *NSF Young Investigator (NYI): The Complexity of Probablistic Algorithms*; (CCR-9357849 A003); $62,500; 12 months.

The probabilistic analysis of NP-hard problems has demonstrated that they are all not equally hard to solve in practice; for example, Hamiltonian paths are easy to find on the average, while random graph coloring still resists attack. This project considers the class DNP (the class of NP problems with input distributions) defined by Levin, and examine the issue of whether DNP = NP? While several DNP-complete problems have been identified, none is a natural problem under a natural distribution. This work proves a result of the type: satisfiability or graph coloring under a uniform distribution is DNP-hard. In particular, it demonstrates that random 3-SAT formulas exhibit a sharp threshold phenomena with respect to the ratio of the number of clauses to the number of variables and more significantly, to show that the random instances of this problem are provably intractable when the ratio is at the threshold value. The general question demonstrates that problems which admit a sharp threshold behavior are hard to solve in the immediate neighborhood of the threshold value. Another problem considered is that of relating DNP-hardness to NP-hardness, or perhaps some weaker intractability assumption. For example, for some class of problems, does an efficient average-case solution imply an unlikely outcome such as P = NP?

University of Texas – Austin; C. Greg Plaxton; *Theory of Parallel and Distributed Computation*; (CCR-9504145 A001); $64,000; 12 months.

The goals of this continuing research project are to identify primitive operations in parallel and distributed computation, and to establish tight bounds on their complexity. Within this general framework, emphasis is placed on: (i) The design and analysis of algorithms for REALIZABLE models of parallel and distributed computation; (ii) the development of techniques that are applicable for a wide range of computational models and problem variations; (iii) the analysis of 'simple' classes of algorithms which are easy to implement. This project concentrates on the following areas: (a) normal hypercube algorithms; (b) sorting networks; (c) list ranking; (d) algorithms for distributed networks, in particular, for the dynamic asynchronous network model; and (e) multiple-resource scheduling.

Duke University; John H Reif; *An Exploratory Study of Parallel Molecular Computation*; (CCR-9633567); $50,000; 12 months. (Joint support with the ARPA-ITO Program)

This Small Grant for Exploratory Research (SGER), jointly funded by the Theory of Computing (TOC) Program, CCR and by the Informations Technologies Office, DARPA, addresses problems in the new area of Biomolecular/DNA Based Computation; in particular, in the area of parallel molecular computation (PMC), massively parallel computation at the molecular scale. The goals of this SGER project include:(1) Exploring possible implementations of PMC;(2) Assessing the feasibility of the PMC implementations; (3) Empirically determining the parameters of performance (e.g., time, volume, etc.) of PMC for some set of applications;(4) Doing trial recombinant DNA implementations of such key operations as: (a) parallel memory read and write, and (b) parallel arithmetic on small integers;(5) Planning a study of the implementation of a biological application of PMC with small or at most moderate I/O and volume requirements. In addition, one of the goals of this SGER award is to facilitate a multidisciplinary effort between the Computer Science Department, the Duke Center for Molecular Diversity and the Microbiology Department of Duke University.

Lehigh University; Eunice E Santos; ***CAREER: Determining Parallel Complexity of Numerical Computation Problems via Dependency Graphs*****; (CCR-9624721); $100,000; 24 months. (Joint support with the Numeric, Symbolic & Geometric Computation Program)**

The research objective of this project is to determine the time complexity of solving various numerical computation problems on parallel distributed-memory machines in general by abstracting the common properties of these problems via dependency graphs. To that effect, two important topics are considered: (1) the design and analysis of parallel numerical algorithms and (2) the derivation of lower bounds on complexity. Working within LogP, a recently-introduced model for parallel computation, parallel complexity will be analyzed by taking into account three important factors: (i) initial data placement, (ii) communication scheduling, and (iii) local operations. Algorithms which are efficient on LogP are portable from one parallel machine to another and will be efficient on any parallel machine. Initially several different numerical computation problems will be considered and the parallel complexity for each specific problem will be determined. The overall results will be used to help extract the important criteria of dependency graphs and thereby classify numerical computation problems via these criteria. Furthermore, the parallel complexity for each classification of numerical computation problems will be determined. The importance of the results obtained lies in the ability to pinpoint what types of data placements and communication schedulings are needed in order to achieve efficient or optimal running times for large classes of numerical problems. Clearly, this research potentially impacts a wide variety of applications and computational problems from various domains which are currently difficult to address. In order to successfully accomplish this research and apply it to these domains, there must be education and training of people in the parallel and numerical computing field. To that effect, a program in parallel and numerical computing will be developed beginning with course preparation at the undergraduate level and culminating in a research program at the doctoral level.

University of California – Berkeley; Alistair Sinclair; ***A Proposal for Research on Random Processes and Algorithms*****; (CCR-9505448 A001); $105,072; 12 months. (Joint support with the Probability Program)**

During the past decade, the quantitative analysis of random processes has led to dramatic advances in the design of efficient algorithms for fundamental computation problems. The PI is continuing research on the application of these ideas, as well as beginning to investigate new paradigms based on nonlinear and cooperative processes. Specifically, the project focuses on the following directions: (1) the analysis of mixing rates of Markov chains, with applications to efficient algorithms for problems in statistical physics and combinatorics; (2) the study of computational complexity of counting problems, and their classification with respect to efficient approximability; Algorithmic applications of nonlinear random processes, in particular quadratic dynamical systems; (3) the theoretical and experimental study of general randomized search heuristics in combinatorial optimization, including Metropolis algorithm, simulated annealing, and cooperative processes.

University of Massachusetts – Amherst; Ramesh Sitaraman; ***Research Initiation Award (RIA): Communication and Fault Tolerance in Parallel Computers*****; (CCR-9410077 A002); $30,979; 12 months. (Joint support with the Computer Systems Architecture Program)**

The project focuses on two important inter-related issues in the design of large scale parallel computers: inter-processor communication and fault-tolerance. The nature of the research is one of synthesizing theoretical and practical approaches to parallel computing. Efficient inter-processor communication is at the heart of efficient parallel computing. With some exceptions, much of the theoretical work has focused on provably optimal algorithms that are often too complex to implement. On the other hand, much of the implemented routing hardware have simpler designs but lack theoretical guarantees as to their performance. This project will attempt to bridge this gap and focus on simple routing schemes that are provably good (though not always optimal). Fault tolerance is becoming ever-increasingly important in the world of parallel computers with a large number of fallible components. The study of fault-tolerant routing schemes focuses on algorithms that can route packets efficiently even in the presence of faulty switches. The two research directions represent very important but largely uncharted territory. The first concerns the question of intrinsic limitations on the fault tolerance of networks. The second research direction concerns reconfiguring in increasingly important class of parallel computers known as heterogeneous parallel computers.

Rutgers University; Endre Szemeredi; ***Probablistic Methods in Computer Science*****; (CCR-9503254 A001); $46,501; 12 months.**

Probabilistic methods in Computer Science range from design of algorithms to analysis of typical behavior of algorithms to proving lower bounds in complexity theory. This research concentrates on a number of specific problems where application of the probabilistic method seems promising. The work includes: (1) graph algorithms (finding approximation algorithms for the coloring problem and the independent set size problem); (2) parallel sorting (in the so-called PRAM model); (3) algebraic questions (diameter of Cayley graphs); (4) problems in circuit complexity (non-linear lower bounds); and (5) complexity of on-line algorithms. Based on past experience, the asymptotic combinatorial tools are expected to be applicable more broadly.

University of New Hampshire Durham; Pilar de la Torre; ***Research Initiation Award (RIA): Towards a Unified Calculus of Efficient Parallel Computation*****; (CCR-9410592 A002, A003); $30,928; 18 months.**

Is there a fundamental set of rules by which we can design and analyze portable algorithms for locality and/or efficiency? This project continues work towards a model of parallel computation that captures locality for wide classes of parallel computers. Theoretical tools are developed that can be applied to gain a better understanding of the following: (1) Parallel Algorithms — how to design and analyze architecture independent generic algorithms capable of exploiting abstract-locality for wide classes of parallel models; (2) Calculus of Architectures — how to develop a calculus that, given such a generic algorithm, identifies necessary and/or sufficient characteristics of the parallel models for which the problem can be solved with prescribed levels of abstract-locality and/or efficiency; (3) Parallel-Performance Limitations — how to discover those levels of abstract-locality, speedup, and efficiency which cannot be, or are highly unlikely to be, achieved by any such parallel algorithms for a given problem. These tools are tested by applying them to design and analyze portable algorithms, and to prove parallelism limitations, for fundamental combinatorial problems.

Cornell University; Sam Toueg; ***Unreliable Failure Detectors for Reliable Distributed Systems*****; (CCR-9402896 A001); $70,000; 12 months. (Joint support with the Operating Systems and Systems Software Program and the Computer Systems Architecture Program)**

The starting point for this research is a fundamental problem in fault-tolerant distributed computing: reaching agreement among processes in a system that is subject to failures. It is well-known that this problem, called Consensus, has no deterministic solution in asynchronous systems, even if it is assumed that communication is reliable and no more than one process may fail. The impossibility of solving Consensus (and other related problems such as Atomic Broadcast) is one of the most severe obstacles to implementing fault-tolerant applications in asynchronous systems. In recent work the PI has introduced a novel approach to circumvent such impossibility results: he showed that unreliable failure detectors can be used to solve Consensus (and Atomic Broadcast), even if the information that they provide about failures is highly inaccurate, e.g., even if they make an infinite number of mistakes. Since such failure detectors can be implemented in realistic distributed systems, and since various considerations make the asynchronous models especially attractive, this work suggests an approach to fault-tolerance that is viable in practice. The objectives of this research are to broaden the applicability of this approach by removing the limitations of the earlier work, and to explore in more concrete terms its practicability. Specific goals include: (1) tolerating communication failures, including network partitions (the earlier work assumed reliable links); (2) tolerating process failures of various types (the earlier work dealt with crash failures only); (3) considering shared-memory systems (the earlier work dealt with message-passing systems); and (4) solving other problems that are central to fault-tolerant distributed computing, including Group Membership and Group Multicasts (the earlier work solved Consensus and Atomic Broadcast). In order to assess the cost and benefit of using unreliable failure detectors complexity questions are also explored. Finally, the practicality of this approach is validated by implementation on an experimental platform.

University of Maryland – College Park; Uzi Vishkin; ***Parallel Algorithmics: Some Current Changes*****; (CCR-9416890 A001); $71,317; 12 months.**

The project addresses major application domains of parallel computing including parallel data bases, and addresses the fundamental issues of the evolving technology of parallel algorithms. Various properties of families of parallel computer systems are studied with the intention of moving towards more ubiquitous general purpose parallel computing. Specifically the role of interprocessor communication as part of the theoretical foundation of parallel computer systems is studied.

Duke University; Jeffrey S Vitter; ***Efficent I/O Communication for High-Performance Computing*****; (CCR-9522047); $90,000; 12 months.**

The project focuses on the following crucial areas of large-scale computing and communication, all of which revolve around issues of Input/Output (I/O) communication and locality: (a) External memory and I/O performance: The relative bottleneck in many large scale applications is the I/O communication between main memory, and external memory and/or other levels of memory. The goal is to develop an understanding of the limits of I/O systems and to construct external-memory algorithms that are simultaneously optimal in terms of internal computation and I/O. (b) TPIE: a Transparent Parallel I/O programming Environment: A system called TPIE is proposed, consisting of an easy-to-use object-oriented interface linked to powerful disk and memory management primitives, to facilitate implementation and evaluation of external- memory algorithms. (c) Prediction and applications to caching: By contrast, the problem of cache replacement involves designing a utility (which can be part of an operating system) to manage the cache memory so that the data requested during the application algorithm are most often local in the cache, thus avoiding costly page fetches.

University of Texas – Austin; David I Zuckerman; *NSF Young Investigator (NYI) Award: Randomness in Computation*; (CCR-9457799 A002); $60,000; 12 months.

The primary focus of this research is the role of randomness in computation. Randomization has proved extremely useful in almost all areas of computer science, such as Monte Carlo simulations, cryptography, distributed computing, and network constructions. These uses of randomness seem wonderful, until one realizes that computers don't have truly random bits available to them. Most computers get their random bits by using pseudo-random generators. Usually, these pseudo-random generators work well in practice. Nevertheless, it is not at all unusual to find instances where they don't. It is therefore important to find generators that are provably good, if such generators exist. Can we always remove the randomness from an efficient randomized algorithm, and be left with an efficient deterministic algorithm that produces the correct answer? This is the general problem, which has different theoretical formulations, depending on the notion of efficiency used. Since polynomial time is the most common measure of efficiency, the most common theoretical formulation is as follows: does randomized polynomial time (BPP) equal deterministic polynomial time (P)? Unfortunately, the only direct progress on the BPP vs. P question has relied on unproven assumptions. This work focuses on two more specific questions. First, can BPP be simulated using a general weak random source that outputs n bits with bounded entropy? Second, does RSPACE(S) = DSPACE(S) or, more realistically with five years, can randomized SPACE(S) algorithms that use more that poly(S) random bits be simulated deterministically in SPACE(S)?

Foundations of Computing

Rutgers University; Eric Allender; *Computational Complexity Theory and Circuit Complexity*; (CCR-9509603 A001); $70,000; 12 months.

The goals of this project include: (1) Continued investigation of complexity classes related to the computation of the determinant and the rank of integer matrices. Such classes include probabilistic logspace and the class #L of nondeterministic logspace counting problems(for circuits); (2) Continued exploration of resource-bounded measure, particularly as it can be applied to small complexity classes such as P.

University of Washington; Paul W Beame; *Computational Complexity Lower Bounds*; (CCR-9303017 A002); $68,163; 12 months.

This research project examines computational complexity in a variety of settings. In particular, this research focuses on obtaining lower bounds on the complexity of solving specific problems in these domains. The main goal is to extend the body of techniques in circuit complexity, communication complexity, and resource tradeoff lower bounds that have been developed over the last decade and find new applications of these techniques for proving lower bounds. Some specific areas of interest are the study of graph connectivity, the use of multiparty communication complexity in lower bounds, the power of branching programs, and the extension of circuit complexity techniques into the domain of the complexity of proofs in propositional logic — a subject that is fundamental to our understanding of the relationship between NP and co-NP.

Yale University; Richard Beigel; *Small-depth Circuit Complexity*; (CCR-9522084); $17,327; 12 months.

Circuits are a fundamental model of computations. Communication complexity and polynomial approximation have yielded many advances in the last decade; yet many important theoretical and practical problems are unsolved. This investigates the relative power of threshold, mod, and binary gates as computational devices. This project provides support postdoctoral associates to work with the PI to address these problems.

University of California – San Diego; Mihir Bellare; *Career: Cryptography, Proof Checking and Approximation*; (CCR-9624439); $100,000; 24 months. (Joint support with the Algebra & Number Theory Program)

This CAREER award is a companion project to the CAREER award CCR-96-24560, "Practice-Oriented Provable Security", University of California, Davis. This project focuses on the areas of (1) provable security and probabilistic proof checking with orientation towards developing practical, but proven secure protocols and (2) probabilistic proof checking (PCP) with application to the derivation of nonapproximability results for NP-hard optimization problems. The problems examined cannot be solved by some kind of "technology transfer" from theory to practice; rather the theoretical principles have to be used to develop new models and solve new problems. Applications of the techniques of practice-oriented provable security are applied to such problems as: session key distribution, use of block cyphers to provably achieve basic cryptographic goals, and balancing between individual privacy and law enforcement, via key escrow or other means. In the area of PCP, the problems studied include determining tight results for the approximability of Max Clique, and improved results for Max SNP problem. The Integrated Educational Plan of this CAREER grant includes developing: (a) new undergraduate and graduate courses in principles of sound protocol design, with joint university/industrial training projects; and (b) an undergraduate algorithm design course in which proof checking tehniques for determining good approximate solutions are taught and applied to NP-hard optimization problems.

University of Michigan – Ann Arbor; Andreas R Blass; ***Topics in Logic and Set Theory*****; (DMS-9505118 A001); $38,300; 12 months. (Joint support with the Foundations Program)**

This research addresses problems in four areas of set theory: (1) Cardinal characteristics of the continuum, including both the general theory of such characteristics (involving Galois-Tukey connections) and the detailed study of particular characteristics, especially the recently introduced evasion and prediction numbers; (2) Applications of set theory, especially in the theory of abelian groups; (3) Semantics of linear logic, especially game semantics and its variants; (4) Logic of topoi, particularly the logical properties of geometric morphisms. This research project also concerns interpreting general recursive definitions, especially in non-determinisitic and concurrent contexts. Specific questions include: (1) Finding a universal class of structures with a recursion operation, so that every structure will be contained in one from this class; (2) Comparing proposals for the semantics of concurrent programming languages; (3) Seeking connections with fixpoint logics used in finite model theory; (4) Extending the current theory to allow more detailed modelings of parameter and message passing between subprograms.

University of California – San Diego; Samuel R Buss; ***Proof Theory and Computational Complexity*****; (DMS-9503247 A001); $32,000; 12 months. (Joint support with the Foundations Program)**

The goal of this project is to investigate axiomatics and conditional separations of bounded arithmetic theories; relative proof complexity of propositional proof systems including Frege, extended Frege, constant-depth Frege, Hilbert Nullstellensatz and cutting plane proof systems; problems in designing computerized theorem provers based on extensions of the sequent calculus; and various problems in low-level computational complexity. This research concerns: (i) formal systems of mathematical logic and the complexity of proofs in such systems; and (ii) computational complexity and feasible complexity. A central goal for this research is to relate the strength of proof systems to the complexity or difficulty of computations. This work centers on proof systems corresponding to functions which are feasibly computable, especially on proof systems for polynomial time computation. This project will also investigate the relative effectiveness of various proof systems for propositional and first-order logic. This research is fundamental to understanding the power of computation in mathematics.

Boston College; Peter G Clote; ***Propositional Logic, Invariance Groups for Boolean Functions, and Parallel Higher Type Functionals*****; (CCR-9408090 A002); $45,960; 12 months.**

This project concentrates on propositional proof systems, Boolean complexity, and parallel/sequential computable functionals of higher types. The project specifically concerns: (1) upper and lower bounds for the number of symbols in proofs of various combinatorial statements for an extension of resolution called cutting planes, originating in operations research, as well as for related propositional logics (logics of approximate reasoning or threshold logics, logics with the bi-conditional, with modular counting gates, etc.); (2) the relation between the parallel complexity of a language L and the algebraic structure of so-called invariance groups and extended invariance groups corresponding to L (structure of invariance groups of regular and context free languages, structure of extended invariance groups of fast parallel computable languages, invariance groups of monotonic versus non-monotonic boolean functions, relation to Krohn-Rhodes semi-group theory); and (3) the study of variants of the scheme of bounded primitive recursion applied to higher type functionals (functions whose arguments may be functions), and the study of parallel computable real valued functions. This research applies techniques from complexity theory, proof theory (mathematical logic), combinatorics and finite group theory. Initial experimental studies of extended invariance groups will be done using a parallel computer. The goal of this research is to increase our understanding of low level parallel complexity and associated logics with long term possible applications in logic programming, database update methods, VLSI design, and programming language design.

University of Southern Maine; Stephen A Fenner; ***Career: Properties of Typical Computable Languages*****; (CCR-9501794 A001); $31,250; 12 months.**

The main research goals of this project are to study new analytical techniques for elucidating the internal structure of complexity classes. These techniques adapt the theories of Lebesgue measure and Baire category to computational complexity. They provide quantitative methods to differentiate 'typical' from 'exceptional' properties of languages in a particular class, shedding light on the relationships between the structure and intrinsic difficulty of computational problems. These concepts are also closely related to the study of pseudorandom and 'pseudogeneric' sequences. Improvements to some of these analytic techniques are explored, and new types of complexity-theoretic Baire category and their applications are investigated. The Educational Component of this CAREER Grant includes the goals of: (a) Establishing theoretical computer science as an integral part of the undergraduate curriculum at the University of Southern Maine, expanding the theory offerings and more closely tying theory with applications and with other subjects; (b) Implementing innovative ways to attract nonmajors to the computer science department, especially those who are interested in computer science, but may not be ready for a traditional first programming course. The routes to these goals run through teaching, advising, mentoring, and curriculum development.

Washington University; Sally Goldman; *NSF Young Investigator (NYI) Award: New Directions in Computational Learning Theory*; (CCR-9357707 A003); $62,500; 12 months.

Building machines that learn from experience is an important research goal of artificial intelligence. Recently, considerable research attention has been devoted to the theoretical study of machine learning. This project focuses on several new directions designed to enhance the current learning models to more accurately model real-life learning situations. Most work in the area of concept learning assumes there is a well-defined border that divides all objects into those that are instances of the concept and those that are not. In reality, though, categorization of objects is often not so clear cut: an algorithm designed to read handwritten cheques will likely encounter many handwritten characters that look somewhat like a '4,' and somewhat like a '9.' In these situations, one possible goal for the learner is to determine which objects are unclassifiable (i.e., the classification is not clear cut) as well as determining the classifications of objects which are classifiable. Another possibility is to require only that the learner determine a categorization so that no object is incorrectly categorized. Thus, for this second goal, the learner can arbitrarily categorize those objects that can go either way. This project defines and studies learning models for both possibilities. Also this project continues the PI's work initiated on developing and studying formal models of teaching to understand how a teacher with knowledge of the target concept and the learner can reduce the training time needed by the learner. As well as being of theoretical interest, there are potential applications of the research to improving automated manufacturing environments.

Carnegie Mellon University; Robert B Griffiths; *Consistent Histories and Quantum Computation*; (CCR-9633102); $50,000; 12 months. (Joint support with the Theoretical Physics Program and the ARPA-ITO Program)

This Small Grant for Exploratory Research (SGER), funded by the Theory of Computing (TOC) Program, CCR, by the Theoretical Physics Program, PHY and by the Informations Technologies Office, DARPA, explores the application of the consistent history formulation of quantum mechanics to some central problems in the new field of quantum computation. These applications include improving algorithms for quantum computation (1) by carrying out measurements at itermediate times, and (2) by error detection and correction. Quantum computation has recently attracted a lot of attention, because Peter Shor showed that certain tasks, including factoring large numbers, could be carried out in polynomial time on a quantum computer, whereas no polynomial-time algorithm for factoring on a "classical" computer is presently known. Clearly, if it could be built, such a computer would have many important applications in many areas of computational science, including in communications and information processing. Actually constructing such a quantum computer represents a daunting task, because of severe technical difficulties. One of the goals of this SGER project is to initiate a collaborative effort between computer scientists, theoretical physicists, and experimental physicists at Carnegie Mellon University, in order to explore the applications of the consistent histories framework to determining the feasibility of such a quantum computing device. The consistent histories approach makes it possible to think about the computational process in more realistic terms (where atoms, photons, etc. are actually doing something over time) than is possible in the usual "textbook" approach, that views the quantum system as a "great smokey dragon" up until the point in time at which a final measurement is made, and the wave function of the system "collapses" into something which is physically understandable.

University of Michigan – Ann Arbor; Yuri Gurevich; *On Foundational Issues in Computer Science*; (CCR-9504375 A001); $84,995; 12 months.

This project concentrates on three areas: (1) Average-case complexity: Tuning up the average-case reduction theory, finding more average-case complete problems, and proving average-case hardness of real- life problems; (2) Finite model theory: Investigating whether semantical properties (such as monotonicity) have any syntactic characterizations on finite structures, determining whether parts of polynomial time properties can be captured by polynomial-time-bounded logics, and studying finite structures with weights in infinite structures (such as arithmetic); (3) Evolving algebras: Examining a new computation model (evolving algebras) in order to use these algebras to specify and verify real-life algorithms.

Northwestern University; Lisa Hellerstein; *CAREER: Structural Properties and Irrelevant Attributes: Implications for Learning and Complexity*; (CCR-9501660 A001); $42,350; 12 months.

The research component of this project addresses fundamental questions in computational learning theory and complexity. It also addresses specific learning problems that are motivated by practical applications. Computational learning theory provides a framework for the formal study of problems in machine learning. The first goal of this project is to determine necessary and sufficient conditions for learning, prove lower bounds, and design learning algorithms. A focus is on query models of learning, which represent situations in which a learner can ask questions of a teacher or expert, or perform experiments. Emphasis is placed on the effect of forbidden projections on the learnability of investigating Boolean function classes. The project also aims to determine structural properties of languages, in order to prove complexity results. It is directed particularly at resolving open questions concerning automata with powers of both randomness and nondeterminism. Another major goal of the project is to develop algorithms for learning in the presence of irrelevant data. A particular problem is to develop PAC (Probably Approximately Correct) algorithms that are robust in the presence of irrelevant attributes in the following way: they depend on only a few more attributes than the concept they are learning. This problem is theoretical but motivated by practical applications, such as medical diagnosis. In these applications, output hypotheses are used for future prediction (diagnosis). Determining the value of an attribute is expensive perhaps involving a clinical test, and therefore it is desirable to limit the number of attributes whose values must be determined. The educational component of this project includes goals in both undergraduate and graduate education. At the undergraduate level a basic goal is to make theoretical courses accessible and intellectually exciting to a wide range of students. Methods for achieving this goal include the following: preparing handouts, introducing historical and biographical material, presenting algorithms to students using algorithm animation, and using computer lab 'algorithm experiments.' At the graduate level, the project includes teaching a graduate course in computational complexity theory with a significant research component. In addition, a reading group in computational learning theory, started last year, will be continued.

University of Rochester; Lane A Hemaspaandra; *Structural Complexity Theory*; (CCR-9322513 A002); $57,416; 12 months.

The structure of feasible computations determines whether one can efficiently solve the optimization problems posed by nature and the modern world. The long-term goal of this research is to understand which problems have efficient solutions, via learning why certain problems may lack computationally feasible algorithms. Of particular interest are the structural relationships between complexity classes, and the internal structure of complexity classes. Though the project intends to broadly investigate this area, focus will be given to the research via two themes: (a) dealing with unpredictable execution order; and (b) the properties of sparse sets. On the first topic, this project investigates the class of languages computable via a large number of short subtasks, each of which is given the (small) result of the subtask that executed before it. This has been studied by Cai and Furst and others for the case where the scheduling of the subtasks is trivial and fixed. However, the project studies the case in which the order in which the subtasks execute is unpredictable. On the second topic, several issues dealing with sparse sets (which are a classic examples of 'sets of low information content') are examined.

University of Rochester; Lane A Hemaspaandra and Edith Hemaspaandra; *U.S.-Germany Cooperative Research on Structure in Complexity Theory*; (INT-9513368); $16,020; 24 months. (Joint support with the Western Europe Program Program)

This award supports the PIs to collaborate in computer science with researchers of the Friedrich Schiller University of Jena, Germany. Together they will study several topics of current interest in computational complexity theory These include (1) rigorous studies of the effect of query order on complexity (i.e., whether and when the order of access to multiple information sources matters) and (2) certificate complexity (calculating the conditions that control the size of classes of sets which always have acceptance certificates that can be easily computed.) The U.S. and German participants in this collaboration bring to bear complementary strengths in the creative theoretical aspects and the rigorous mathematical aspects of computational complexity. The joint effort of these talented and prolific researchers should produce significant contributions to the study of structural aspects of complexity theory. Results will have implications in many aspects of computer and information science.

State University of New York – Albany; Harry B Hunt and Richard E Stearns; *Topics on the Complexity of Decision, Optimization, and Approximation Problems*; (CCR-9406611 A002); $105,878; 12 months.

The objective of the proposed research is to understand the interrelationships among (a) the complexity of problems, (b) the algebraic, combinatorial, or graphical structure of problem instances, and (c) the methods used to describe problem instances. Further given this understanding, this research will develop when possible more efficient algorithms for the problems considered. Research will be carried out on several different fronts as follows: Work on succinctly-described objects is carried out to better understand how the structure of the description can be used to efficiently solve the problem instances described. Very efficient structure-preserving reductions are used to identify problem similarities not adequately captured by polynomial time or log space reductions. Work on efficient approximation algorithms and scheme is carried out for hard problems including PSPACE- and EXPTIME-hard problems. Finally, work is initiated on the average case complexity of several hard natural decision, optimization, and approximate optimization problems.

University of Massachusetts – Amherst; Neil Immerman; *Descriptive Complexity*; (CCR-9505446 A001); $65,486; 12 months.

Descriptive complexity is an approach to complexity theory that measures the richness of a language or sentence needed to describe a given property. The language studied are variants of first-order logic and their targets are finite, ordered structures. In this setting, there is a non-trivial relationship between the traditional computational complexity of a problem and the descriptive complexity of the problem. Indeed, all major complexity classes have been shown to have natural descriptive characterizations. This project continues this research into descriptive complexity by concentrating on the following four research areas: (1) Order-independent languages; (2) First-order interpretations; (3) Applications to databases and dynamic complexity; (4) Trade-offs in descriptive complexity.

University of California – San Diego; Russell Impagliazzo; *NSF Young Investigator (NYI) Award: Small Depth Boolean Circuits and Complexity - Theoretic Cryptography*; (CCR-9257979 A004); $25,000; 12 months.

This award funds continuing investigations into two long-term interests: the complexity of small depth Boolean circuits, and the computational complexity foundations of cryptography. Computational complexity as a field strives to provide a mathematical foundation for computer science by developing techniques to evaluate the inherent difficulty of computational problems. The Boolean circuit seems the most natural and robust concrete model of computational complexity. Much work has gone into developing the theory of circuit complexity in the hope of resolving such questions as P vs. NP. There are strong connections between bounds on circuit depth and issues in the theory of parallel computation, logic (mostly proof theory), learning theory, and the theory of neural nets. Continuing projects in circuit complexity are: establishing links between circuit bounds and Frege proofs; examining the behavior of formulas under random restrictions, (also known as neural nets). Of course, circuit complexity is a dynamic area, and specific projects may change with new results. Work continues on the foundations of cryptography. Recent work in developing the structural complexity of cryptography has been so successful that perhaps all the major, tractable questions in the area have been resolved. Indeed, there seem to be some theoretical roadblocks to further progress in this direction. However, there are several important open questions regarding the relationship between average-case complexity and cryptography, and that between oblivious transfer and secret agreement. More importantly, a large gap remains between theory and practice. Another goal of the project is to develop theoretical tools that will actually lead to implementable cryptosystems, and to handle such questions as virus protection, secure electronic transfer of funds, and such issues as privacy vs. security for electronic mail.

University of Washington; Richard M Karp, Walter J Ruzzo, and Martin Tompa; *Computational Problems in Physical Mapping and Sequencing*; (DBI-9601046); $225,000; 12 months. (Joint support with the Computational Biology Activities Program)

This project is an investigation into some computational problems that arise in the physical mapping and sequencing of chromosomal DNA. The goals of the project are (1) to investigate new strategies for certain parts of these computations, aimed at improving speed and/or reducing the amount of expert intervention required, while preserving solution quality, and (2) to add new capabilities to existing mapping and sequencing systems. A pervasive theme throughout the project is to take experimental error into account in the computations. The investigation into physical mapping concentrates largely on one method in use at the University of Washington's Department of Molecular Biotechnology, namely Multiple Complete Digest mapping. Within the map assembly stage of this method, new algorithms are being investigated for identifying the clone fragment containing the cloning vector, determining which clone fragments arise from identical fragments on the target DNA being mapped, determining the probability of pairwise clone overlap, and other related problems. In addition, the project includes investigations into a general placement problem in physical mapping, a method of sequencing without mapping, and the effects of experimental error on molecular computation. The principal investigators will use their experience in algorithm design and analysis to help establish a rigorous foundation for the computational aspects of these biological problems. This may in turn result in direct impact on the experimental methods used.

Boston University; Assaf J Kfoury, Pawel Urzyczyn, and Jerzy Tiuryn; ***Combinatorial Problems in Typed Lambda-Calculi*****; (CCR-9417382 A001); $100,000; 12 months. (Joint support with the Programming Languages and Compilers Program)**

This award supports research in the area of typed lambda-calculi, with special emphasis on the combinatorics, complexity, and logic of type inference. Specific problems concern type-reconstruction, type-checking, and related aspects, and are motivated by design and implementation issues in programming languages. Particular attention is given to algebraic characterizations of these problems. Such characterizations give rise to new and more general forms of unification, in particular semi unification, often with applications in other parts of computer science (term-rewriting systems, database theory, computational linguistics). Based on some of the same techniques, the project explores problems in two research areas, both with important applications: (1) higher-order matching (for automatic theorem proving) and (2) extensions of ML-style polymorphism (for polymorphically typed programming languages).

University of California – Santa Cruz; Phokion G Kolaitis; ***Aspects of Computation Theory*****; (CCR-9307758 A002); $53,248; 12 months.**

This work investigates a broad spectrum of problems dealing with logic in computer science that span finite model theory, database theory, complexity theory, and optimization theory. The question of whether or not certain important algorithmic problems are expressible in various logical query languages, such as Datalog and extensions of Datalog with negation, are examined. Further, the relationship between first- order logic, fixpoint logic, and polynomial-time computability on arbitrary classes of finite structures are explored. In the area of logical definability and asymptotic probabilities, 0-1 laws are investigated for certain powerful extensions of fixpoint logic and an attempt is made to find necessary and sufficient conditions for the existence of 0-1 laws under variable probability measures. In a different line of investigation, the descriptive complexity of optimization problems continues to be developed. An attempt is made to determine if there is an effective syntax for the approximable NP optimization problems and to seek structural explanations for the diverse behavior of polynomial-time optimization problems with respect to parallel approximation.

Cornell University; Dexter C Kozen; ***Topics in the Theory of Computation*****; (CCR-9317320 A002); $85,957; 12 months.**

This research addresses problems that span several distinct but related areas of theoretical computer science: algorithms in programming languages (type inference and type checking for type systems with subtypes and recursive types, solving systems of set constraints), computational algebra (decomposition of algebraic functions, calculations of Newtonian graphs of algebraic functions), program logics and semantics (deductive systems for the propositional - calculus, typed Kleene algebra).

Rutgers University; Paul J Krasucki; ***Research at Undergraduate Institution (RUI): Knowledge and Concurrency*****; (CCR-9528919); $93,938; 24 months. (Joint support with the Software Engineering Program and the Programming Languages and Compilers Program)**

This (i) Research at Undergraduate Institutions (RUI) and (ii) cooperative US-India research project investigates distributed systems, using reasoning about knowledge: During a run of the distributed system, processes acquire knowledge about the state of the system through messages. The goals in this area are to (a) characterize changes in the state of knowledge of processes due to communications and (b) relate knowledge structure of the system to the temporal structure on event occurrences in the system. The third research goal of this project is to (c) attempt to modify a model, created by economists to reason about rational agents communicating values of certain agreed-upon function, so that it can be used to design and analyze protocols leading to distributed consensus.

University of Massachusetts – Amherst; Jonathan Machta; ***Statistical Physics of Complex and Disordered Systems*****; (DMR-9632898); $75,000; 12 months. (Joint support with the Materials Theory Program)**

This grant supports the work of a mid-career materials theorist who is getting involved in the interface between conventional materials properties and the analysis and development of complex computational algorithms. In particular, this grant supports further development and applications of the invaded cluster algorithm, devised earlier by the PI. The other two projects are: (1) investigation of computational complexity in collaboration with a computer science colleague and (2) a collaboration with experimental colleagues on the flow properties of fluids in porous media. This grant supports work at the interface of computer science and materials theory.

University of California – Los Angeles; Donald A Martin, John R Steel, and Yiannis N Moschovakis; ***Recursion Theory and Set Theory*****; (DMS-9505153 A001); $80,000; 12 months. (Joint support with the Foundations Program)**

The PI will work in set theory and descriptive set theory, especially determinacy hypotheses, and his first goal is to complete his long-awaited research monograph on proofs of determinacy from large cardinal axioms. Steel will also work generally in set theory and descriptive set theory, and more specifically on the construction and study of canonical, inner models of ZFC that satisfy specified large cardinal or determinacy assumptions. Moschovakis will work on fixpoint recursion and its applications to the theory of computation, including the mathematical modeling of recursive algorithms, the logic of recursive definitions, and the construction of canonical solutions for recursive equations in non-deterministic and concurrent settings. The PI and the co-PI will continue working on this fundamental program, seeking to extend it and to delineate its limitations. Moschovakis changed interests about ten years ago, and his current research program aims to provide a rigorous, mathematical foundation for computer science. One of the main, ultimate aims of his work is to construct a "logic of programs" that will be useful in the development and verification of computing systems.

University of Pittsburgh; Toniann Pitassi; *NSF Young Investigator (NYI) Award: Logic and Complexity Theory*; (CCR-9457782 A002); $25,000; 12 months. (Joint support with the Foundations Program)

The study of the complexity of proof systems is fundamental to logic and has a broad range of applications in computer science as well. First, the efficiency of any deduction system is tied to the speed of logical reasoning (e.g., Prolog, Datalog, and many inferencing systems in artificial intelligence). Secondly, the complexity of propositional proof systems is intimately connected to the central question in complexity theory of whether NP equal coNP. Finally, there are links between particular propositional proof systems, systems of arithmetic, and complexity classes. Exploiting these connections, one can apply a whole new range of tools from logic to attack difficult complexity questions (such as P = NP). The basic question in propositional logic to be considered is: How long does a proof of a propositional tautology have to be, as a function of the size of the tautology? The best upper bound for all known proof systems is exponential in the size of the tautology and it is conjectured that a similar lower bound holds. This question is equivalent to whether or not NP = coNP, and while it has been studied extensively for the last 25 years, it is still far from solved. Propositional proof systems can be divided into classes that correspond directly to complexity classes. An approach toward proving NP = coNP is to establish superpolynomial lower bounds for specific proof systems of greater and greater proof strength. Much of this research is aimed at attacking this difficult problem. Another line of research concerns feasible arithmetic (also known as bounded arithmetic) and its connections to propositional proofs and complexity theory. The concept of a feasible proof has been the subject of much research in the last ten years.

Indiana University – Bloomington; Paul W Purdom; *Speed of Constraint Satisfaction Algorithms*; (CCR-9402780 A002); $59,592; 12 months.

This research investigates finding fast algorithms for constraint satisfaction problems. A constraint satisfaction problem is one in which a series of constraints is imposed on a set of discrete variables. The task is to find a set of values for all the variables that satisfies all the constraints simultaneously. For example, when scheduling traffic lights each light may be in one of only three discrete states. Furthermore, there are constraints on the states of the traffic lights that require that no two cross-streets get green lights at the same time. A constraint satisfaction problem may have thousands of variables and thousands of constraints. Constraint satisfaction problems arise in diverse areas. These include generating tests of correctness for electronic circuits, understanding line drawings by computer, solving combinatorial problems, and predicting the folding of RNA molecules. The set of constraint satisfaction problems is one of the NP complete problem sets, so it is unlikely that a fast algorithm for all constraint satisfaction problems will ever be found. Indeed, brute force algorithms for constraint satisfaction problems may be extremely slow. Nonetheless, algorithms have been found that are fast for most problems in certain NP complete problem sets. This research concerns finding and characterizing the algorithms and problem sets for which constraint satisfaction problems can be solved quickly, with the aim of finding general algorithms that are fast over as wide a range of problem sets as possible.

University of California – Davis; Phillip Rogaway; *CAREER: Practice-Oriented Provable Security*; (CCR-9624560); $100,000; 24 months.

This CAREER award is a companion project to the CAREER award CCR-96-24439, “Cryptography, Proof Checking and Approximation”, University of California, San Diego. This project focuses on the new area of provable security. Provable security is achieved for a problem, when one provides (i) a definition of security for the problem, (ii) a protocol, and (iii) a proof that the protocol meets its definition, assuming, for example, the intractability of factoring. Instead of just producing triples (definition, protocol, proof), the goal of the project is to utilize reductions as a technique for making precise statements about the security of concrete cryptographic schemes; for example, “if you can break authentication protocol P with resources X, then you can break the Data Encryption Standard with resources Y”. In particular, applications of these techniques have already lead to a treatment of asymmetric encryption, message authentication, entity authentication, and key distribution. Protocols designed employing practice oriented provable security have resulted in protocols used in products of IBM and Microsoft, and in a proposed International Standard. The Integrated Educational Plan of this CAREER Grant includes developing (a) a new graduate course in modern cryptography; (b) undergraduate security projects which are difficult, open-ended, and design oriented; (c) course and course materials for a new undergraduate course in graphs, games and coloring.

University of California – Los Angeles; Bruce L Rothschild and Alexei P Stolboushkin; ***Finite Model Theory and Applications to Databases*****; (CCR-9403809 A002); $30,365; 12 months.**

This research examines certain definability and decidability problems for finite models, motivated by the theory of relational databases. Among them are problems of existence of recursive characterizations for specific classes of queries (formulas), such as, safe, monotone, or distributive queries. In addition, this project investigates the expressive power of different logical languages for finite models, particularly in the absence of linear order. Among the targeted languages are those of first order logic, different fixpoint logics, implicit definability, and of fragments of infinitary logics. The main goal here is to expand numerous definability results for ordered models to wider classes of models, for example, finitely axiomatizable classes of rigid models.

Syracuse University; James Royer; ***Higher-Order Computation and its Computational Complexity*****; (CCR-9522987 A001); $68,425; 12 months.**

One of the problems with programming using higher-order procedures is the absence of formal tools for analyzing the time and space complexity of such procedures. This absence is partially due to the fact that the current mathematical models of computation for higher types are very divorced from the complexity-theoretic concerns. The aim of this project is to work toward constructing a usable, realistic complexity theory for higher-order computation. This research model is partially based on Cook and Kapron's basic feasible functionals (BFFs), which are a class of finite-type functionals that are intended to be the finite-type analogue of PF, the polynomial-time computable functions. At type-2 there is a reasonable body of work which appears to confirm that type-2 BFFs are a reasonable type-2 analog of PF. Natural non-sequential extensions of the type-2 BFFs are investigated and a theory of non-sequential 'polynomial-time' functionals is constructed. This extended theory of type-2 BFFs is also used to develop a theory of computational complexity for machine learning. Unlike the situation for type-2, for type-3 (and above), there is not the weight of evidence that the BFFs are a reasonable analog of PF. This research tests several models (partly based on the type-2 complexity) at type-3, by analyzing some higher-order algorithms involving monads, continuations, and other higher-order constuctions. An intermediate goal is to use the experience gained from this work on the complexity of finite-type functionals, to develop 'complexity theoretic' solutions to domain equations. Such solutions have the potential of providing settings in which one could examine both semantic and complexity-theoretic features of languages such as ML.

SRI International; John Rushby, Patrick D Lincoln, and Natarajan Shankar; ***An Integrated Framework for Mechanical Verification of Fault-Tolerant Systems*****; (CCR-9509931); $345,147; 36 months. (Joint support with the Software Engineering Program, the Operating Systems and Systems Software Program, the Computer Systems Architecture Program, and the Numeric, Symbolic & Geometric Computation Program)**

Formal verification is a technique for reducing certain questions about the requirements, designs, or implementations of computer software and hardware to a mathematical form that can examined by methods having the character of calculation. Similar to the way in which questions about the properties of an aerofoil can be reduced to calculations that can be mechanized by computational fluid dynamics, mechanized formal verification uses automated theorem proving and related methods (e.g., model checking) to perform logical calculations about the properties of computer systems. Whereas testing can examine only some of the behaviors of a complex system, mechanized formal verification can examine all the behaviors of a model of the system (again, the situation is analogous to flight test vs. computational fluid dynamics). Using mechanized formal verification, suitably realistic models can provide strong evidence for the correctness of computer systems, and simplified models are often useful for detecting bugs that testing does not encounter. Mechanized formal verification is most effective when applied to the hardest and most critical elements of design. An area where it holds great promise is in the analysis of fault-tolerant systems, where the presence of faults can produce a complexity of behavior that defies other forms of analysis. The increased assurance provided by mechanized formal verification can be of particular value for fault- tolerant systems because these are generally found in applications that are critical to safety (such as aircraft and other vehicles) or national infrastructure (such as the telephone, power-distribution, and banking systems). The objectives of this research are to reduce mechanized formal verification of fault-tolerant distributed algorithms from tours-de-force to routine practice, and to develop a framework for mechanical verification of the architectures of complete fault-tolerant systems composed from such algorithms. The research focuses on developing systematic methods for the formal specification of classes of fault- tolerant algorithms, and mechanized strategies for their exploration and verification. These strategies will build on techniques that combine decision procedures and model checking within an interactive framework. The algorithms considered will include those for distributed consensus, synchronization, diagnosis, and group membership. The work is expected to yield improved understanding of the assumptions, constraints, and fault models underlying the algorithms studied, more refined arguments for their correctness, and coherent arguments for their synthesis into an overall architecture. In addition, the research is expected to yield new results in the construction and application of formal verification systems, and improved methods for the formal representation and analysis of fault- tolerant algorithms.

Kent State University; Meera Sitharam; *Research Initiation Award (RIA): Proving Circuit Complexity Bounds Using Classical Analytic Methods*; (CCR-9409809 A002); $20,219; 12 months.

This project examines problems in circuit complexity, using classical analytic methods. The project is centered around three basic questions, the first and last of which are specifically geared towards basic, long-unresolved issues that have hindered progress in the area. (1) Do simple operations (such as shifts, or products) on hard functions preserve hardness? (2) In what ways can hardness be used for learning functions in a complexity class? (3) What are the analytic properties that are common to constant depth circuits with various sets of symmetric gates? Already partial answers have yielded the following: (a) a non-trivial, complexity bounds involving constant depth circuits and more significantly, the analytic patterns underlying such bounds; (b) general methods for using hard functions to obtain large classes of pseudorandom generators and training sets for learning; and (c) conversion of lower bounds on circuit complexity into upper bounds on the complexity of learning, and vice versa. The analytic techniques used, are partly multivariate generalizations of univariate approximation methods, partly from coding theory and partly from spectral analysis. The techniques developed are of independent interest to approximation theorists, and solve, in particular, a number of problems involving general Boolean functions, and combinatorics over the multidimensional unit-cube.

Kent State University; Meera Sitharam; *Foundations and Mathematical Aspects of Computer Science (An AMS session) to be held at Kent State University November,l995*; (CCR-9529950); $7,001; 12 months. (Joint support with the Computational Mathematics Program and the Algebra & Number Theory Program)

This project funds the speaker and participant costs for a special session, titled "Foundations and Mathematical Aspects of Computer Science", of the Central Section of the American Mathematical Society, Kent State University, November 3 – 4, 1995. The purpose of this special session is to encourage interaction between mathematicians and the more theoretical computer scientists. Such interaction is currently rare, and would benefit both communities: the process of formalizing and solving many challenging problems that arise in computer science theory not only benefits from the mathematical formulation but often requires the development of new and deep mathematics. The aim of this session is to create awareness of these special connections and problems, and to open communication channels with mathematicians attending the AMS Meeting. In fact, both the formal session (with talks that outline problems in such areas as algorithms, complexity theory, computational learning theory, computational geometry, parallel computing models, probabilistic and randomized computation,and computer security and cryptography), and the planned open problem sessions (for an interchange of solution techniques and problems) should facilitate interdisciplinary collaborations between mathematicians and computer scientists. There are three NSF programs which are split funding this project: Theory of Computing Program of CCR, Computational Mathematics Program of DMS, and Algebra and Number Theory Program of DMS.

University of Maryland – College Park; Carl H Smith and William Gasarch; *Learning, Complexity, & Testing*; (CCR-9301339 A003, A004); $78,017; 18 months.

Various aspects of machine learning have been under empirical investigation for quite some time. And, more recently, theoretical studies have become popular. This research contributes towards the goal of understanding how a computer can be programmed to learn by isolating features of incremental learning algorithms that theoretically enhance their learning potential. In addition to pursuing lines of research that have already been established with a firm base, investigations of other topics in machine learning are initiated; for example, investigating learning algorithms with a limited ability to remember observed data and other information. Inductive inference is in some sense the dual of program testing. For the inference problem, one must generalize from a finite set of examples to an entire function or concept. The testing problem requires the formulation of a finite test set that distinguishes a given function from all the others in a given class. The program testing problem is more difficult than the program equivalence problem, hence it is highly unsolvable. Consequently, practitioners are doomed to ad hoc techniques. However, it is possible to discuss testing strategies that give an indication of how confident we are that the program being tested is reliable. In another vein, the complexity of certain problems is studied. Some of the earlier work in recursive graph theory has now been applied to the complexity of NP-hard problems. This work is extended to problems in P. The complexity of Boolean functions is also investigated.

University of Nebraska – Lincoln; Douglas R Stinson; *Combinatorial Cryptography*; (CCR-9402141 A002); $31,956; 12 months.

This research addresses various areas of cryptography that have a significant interaction with combinatorics. The three primary areas of research are: (1) secret sharing schemes; (2) authentication codes; and (3) resilient functions. Secret-sharing schemes allow shared computation of a secret key by certain specified groups of people; they have applications in any scenario where shared access or shared control is important. codes are a fundamental cryptographical tool used to guarantee that information has not been altered, and that it originated with the intended sender. Resilient functions are a type of cryptographic function that have application to different problems, including designing of stream and block ciphers and key exchange protocols. All of these areas involve combinatorial theory in an essential way. Combinatorial designs, graph theory, error-correcting codes, and linear programming all play an important role in this work.

Harvard University; L. G Valiant; *Computational Rationality*; (CCR-9504436 A001); $97,000; 12 months. (Joint support with the Knowledge Models and Cognitive Systems Program)

In computational studies of intelligent behavior several phenomena have been isolated, among the most important of which are learning, reasoning, and planning. Obstacles that have been found pervasive are that the formalization of the various phenomena are often inconsistent with each other, and, even more often, that they are computationally intractable. The main thrust of this research project is to develop a computational framework and some algorithms that would comprise a unified approach to this problem area. The proposed starting point is the idea that the concept of probably approximately correct (PAC) learning provides a quantitative criterion on the behavioral performance of a system. These analytic methods used in computational learning theory, are extended to a wider class of phenomena, including various notions of reasoning and planning.

University of North Carolina – Greensboro; Jie Wang; *Research at Undergraduate Institution (RUI): Structural Aspects of Average-Case NP-Completeness*; (CCR-9424164 A001); $27,221; 12 months.

This RUI project is a companion project to the RUI project CCR-9503601 (PI: Belanger) of Northeast Missouri State University. (1) The first part of this research project aims to better understand the properties and structures of average-case NP-complete problems under randomizing reductions. Randomizing reducibility is a powerful tool used to learn about average-case completeness. In particular, the existence of isomorphisms among complete problems and the structure of complete problems under randomizing reductions are studied. Polynomial isomorphisms are explored in the setting of randomizing many-to-one reductions based on recent results on the NP-isomorphism problem with respect to random instances, which have explicit support for this goal. (2) The second research topic is rankable distributions. P-rankable distributions were proposed recently to capture the ranking of probability distributions, but they are not directly comparable to p-time computable distributions used in the literature. The goal here is to investigate whether p-rankable distributions can provide harder instances than p-time computable distributions for randomized NP decision problems and randomized NP search problems.

University of Pennsylvania; Tandy J Warnow; *NSF Young Investigator (F) Award: Computational Problems in Evolutionary Tree Construction*; (CCR-9457800 A002); $62,500; 12 months.

The main focus of this work is in the area of algorithms for reconstructing evolutionary history. The evolutionary history for a set of species S is described by a rooted tree called an evolutionary tree (also called phylogenetic tree or phylogeny) in which the leaves represent the species in S and the internal nodes represent the common ancestors. There is a wide range of methods for constructing phylogenetic trees given in the literature, and most of these are based upon optimization criteria. That is, an optimization criterion is given with which to evaluate a given phylogeny, and the objective is to find a phylogeny with an optimal score with respect to that criterion. Almost all of the optimization problems have been shown to be NP-hard score with respect to that criterion. Despite this, a plethora of heuristics have been developed for constructing phylogenetic trees. The standard modus operandi is to analyze the data set using many different methods for tree construction, and compare the outcomes of these methods in order to select a phylogenetic tree. This research effort with respect to phylogeny construction has three basic components: (1) the development of a comprehensive mathematical framework by which current methodologies can be compared; (2) the development of efficient algorithms for phylogeny construction with meaningful output; and (3) application of these techniques to problems in biology and linguistics.

University of Pennsylvania; Scott Weinstein and Maria L Bonet-Carbonell; *Logical Methods Applied to Complexity Theory*; (CCR-9403447 A001); $139,457; 12 months.

This project's goal is the development of a logical approach to the study of computational complexity. The approach will exploit two of the main traditions in mathematical logic, namely, (a) model theory and (b) proof theory. Model theory studies the relation between logical languages and their interpretations and has developed powerful techniques to answer questions about the expressive power of a variety of logical languages. Proof theory studies the notion of logical inference and elevates mathematical proof itself to the position of an object of mathematical study. In so doing, proof theory makes it possible to undertake a rigorous examination of the complexity of establishing conclusions within a given system of inference. A significant feature of this logical approach is that it forgoes, to a large extent, the use of machine models of computation in the study of complexity. Much valuable work in complexity theory has involved careful analysis of the resource requirements of various algorithms on the specified abstract machines. This project seeks to obtain new insights into the complexity of combinatorial problems by analyzing these problems in machine-independent, logical terms. Among the questions considered are whether the polynomial time decidable properties of finite graphs are recursively enumerable and whether NP = co-NP. Negative answers to either of these questions would imply that P is not equal to NP. Though these implications are certainly significant, the focus of the project is to achieve a deeper understanding of the problems themselves by the use of logical techniques.

Princeton University; Andrew C Yao; ***Complexity Studies in Communications and Quantum Computations*****; (CCR-9627819); $63,000; 12 months. (Joint support with the ARPA-STUDIES IN COMM & QUANTUM Program)**

This project studies the computational complexity of communication problems and quantum computing. Topics explored include: (1) complexity theories for (a) multiparty communications, (b) probabilistic communications, (c) quantum communications, and (d) quantum circuits; and (2) complexity questions for algebraic decision trees. The mathematical techniques employed are drawn from such areas as: (I) combinatorics; (ii) probability theory; and (iii) topology.

NUMERIC, SYMBOLIC, AND GEOMETRIC COMPUTATION

Numerical Methods

Pennsylvania State University – University Park; Jesse L Barlow; ***Numerical Solution of Eigenvalue Problems and Related Least Squares Problems*****; (CCR-9424435 A001); $49,859; 12 months.**

This research seeks to improve the accuracy of methods to solve some particular eigenvalue problems. The first is the classical symmetric eigenvalue problem. Recent results have shown that there is a large class of symmetric eigenvalue problems for which small structured perturbations result in small changes in the eigenvalues in the relative sense. This class of matrices is called "well-behaved". The problem of finding a necessary and sufficient condition for a matrix to be "well-behaved" is being investigated. A number of related issues have to be considered, including: analysis of methods for updating indefinite eigenvalue problems; investigation of methods for updating the ULV and ULLV approximations to the singular value decomposition and generalized singular value decomposition; and development of analysis of methods for computing the sparse generalized singular value decomposition. The last two issues are extremely important in the solution of ill-posed least squares problems and total least squares problems. The implications of derived results for these problems are being investigated.

University of Kansas; Ralph Byers; ***Numerical Linear Algebra and Control*****; (CCR-9404425 A001); $39,233; 12 months.**

This project studies existence, uniqueness and methods for calculating a smoothly varying singular value decomposition of a smooth matrix valued function. The smooth singular value decomposition has applications in differential algebraic equations including time dependent optimal control problems for descriptor systems. Preliminary work with real analytic functions is to be extended to a larger class of smooth functions and more efficient numerical methods for tracking all of or part of a smoothly varying singular value decomposition are being investigated. This project is also developing numerical methods for calculating the controllability radius, regularity radius, and other distances of interest in computational control. It is developing efficient, provably reliable, numerically stable methods. Success here will generalize to a means of estimating the distance from a generic matrix pencil to an algebraic variety of non-generic pencils and to general condition estimators.

Society For Industrial and Applied Mathematics (SIAM); James M Crowley; ***SIAM Workshop on Computational Differentiation Feb. 12-15, l996, Santa Fe, New Mexico*****; (CCR-9627456); $10,000; 6 months.**

This workshop on computational differentiation was conducted by SIAM with the cooperation of the special interest group on numerical computing (SIGNUM) of the Association of Computing Machinery (ACM). The workshop brought together researchers and students working on constructive and accurate methods for differentiation and their applications. The workshop attendees discussed recent advances in theory, implementation, and application of computational differentiation. The resulting close interaction between developers of tools and their users in applied mathematics, computation, science, and engineering will accelerate progress.

Massachusetts Institute of Technology; Alan S Edelman; ***Numerical Methods for the Nonlinear Eigenvalue Problems in High Performance Materials Design*****; (CCR-9404326 A001); $55,594; 12 months.**

The objective of this project is to develop adaptive numerical methods for the first principles prediction of the properties of technologically important materials. These materials may be characterized as having complex structures and compositions, large numbers of particles (> 100) and little or no symmetry. The methods being investigated include a suite of iterative, acceleration, and minimization schemes for the nonlinear eigenvalue problem and the accompanying elliptic problem arising in first principles prediction of material properties. This suite includes the eigenvalue iterative methods of Rayleigh-Ritz and Longsine and McCormick; minimization methods such us conjugate gradient, steepest descent, and the trace minimization method by Sameh and Wisniewski; and multilevel methods such as multigrid and fast multipole method. These methods will be used with adaptive grid to exploit the underlying dynamical, localized structure inherent in electronic wave functions. Modern parallel architectures will be used to reduce the computational time and memory costs. The various components of this project will be developed through the close collaboration of researchers from chemistry, physics, computer science, and numerical analysis.

University of Colorado – Boulder; Charbel Farhat, Xiao-Chuan Cai, and Jan Mandel; ***10th International Conference on***

***Domain Decomposition Methods in Sciences and Engineering. University of Colorado at Boulder, Summer1997*; (ECS-9628513); $30,000; 18 months. (Joint support with the New Technologies Program and the Computational Engineering Program)**

This conference is designed to promote the interaction between mathematicians, numerical analysts, physical scientists, engineers, and computer scientists. For this tenth meeting, the focus theme will be High Performance Computing. The conference organizers plan to devote a significant amount of effort on attracting the participation scientists not only from academia and government research laboratories, but also from the computing, simulation software, and key engineering industries. Support will be provided for the active participation of graduate students who will benefit from exposing their research to a diversified audience of potential employers, and from keeping up-to-date with the latest developments in this field.

University of Colorado – Boulder; Carlos A Felippa; *High-Performance Computational Methods for Coupled Field Problems*; (ECS-9217394 A005); $1,231,813; 12 months. (Joint support with the Magnetospheric Physics Program Program, the Network Infrastructure Program, the Physical Oceanography Program, the Solar-Terrestrial Program, the Computational Engineering Program, the Materials Engineering Program, the Solid Mechanics Program, and the Computational Mathematics Program)

This project is a collaboration between applied mathematics, astrophysics, computer and computational science, engineering, and geophysical sciences, to develop parallel algorithms for problems relating to Coupled Fields and Coherent Structures and Geophysical and Astrophysical Fluid Dynamics (GAFD) Turbulence. The common denominator of the various application areas are performance measures, massive data sets, and the visualization of multi-terabyte data sets.

Stanford University; Gene H Golub; *Computational Methods in Linear Algebra Using the Theory of Moments*; (CCR-9505393 A001); $92,408; 12 months. (Joint support with the Computational Mathematics Program)

Numerical methods are developed for solving large scale matrix problems arising in applications. These applications are quite diverse and include such areas as statistical modeling and quantum chromodynamics (QCD). The basic tool for these computations comes from the classical theory of moments and numerical quadrature. These problems are equivalent to estimating upper and lower bounds for a quadratic form. One area of particular interest is computing the generalized cross validation parameter. This technique is used in solving ill-posed least squares problems by regularization methods. A numerical procedure is developed which enables one to analyze very large data sets. Also, a method for computing a parameter which is of importance in solving ill-conditioned sparse least squares problems is developed. The basic objective of this research is to develop new and robust algorithms for solving very large scale linear problems.

Brown University; David Gottlieb; *High Order Methods for Time Dependent Equations*; (DMS-9500814 A001); $50,000; 12 months. (Joint support with the Computational Mathematics Program)

The project entails research in long term integrations of genuinely time dependent partial differential equations. It involves the development and application of high order finite differences and finite element schemes as well as Spectral methods. The methods to be developed and applied are suitable for problems in several space dimensions with complicated geometries. A major consideration in developing and assessing high order methods for long term integrations is their scalability. Based on the experience gained already in parallelizing ENO codes, a further effort to parallelize spectral and finite element based ENO codes is proposed. The research has two components, theory and applications, which are linked together. The theory is motivated by the type of applications under investigation and the applications use the theoretical developments. It has been shown that in order to numerically simulate complicated flows that change in time, high order methods are necessary. However those methods, because of their high accuracy, are less robust. The proposed research deals with the issue of the successful implementations of high order methods. This involved theoretical issues (as solving the 100 years old Gibbs Phenomenon) as well as more applied issues. On the applied side, the problem of mixing enhancement by interactions of Shock waves and hydrogen jets will be studied. This entails the simulation of an air shock passing through an hydrogen jet in high Mach numbers. High order methods are mandatory here since the process of interest is the mixing and combustion inside the jets. The current codes, Spectral and ENO are to be modified to handle more general cases. Parallel versions of ENO and spectral methods are being applied now and will be further modified.

University of North Carolina – Chapel Hill; John H Halton; *Sequential Monte Carlo Techniques for the Solution of Large Linear and Non-Linear Systems*; (CCR-9322102 A002); $71,275; 12 months.

Almost all mathematical problems in science and engineering are approximated by linear and non-linear systems of integral and/or partial differential equations, which vary widely in number, form, and character. Since analytical solutions do not exist for most problems of interest, numerical approximation techniques, such as finite element and finite difference methods, are widely used, ultimately resulting in a system of linear or non-linear algebraic equations, which must then be solved. In the realistic representation of practical applications over large and/or domains, such systems may contain m=104 to 107 unknowns. For non-linear or transient problems, the system requires the repetitive solution of corresponding linearized problems for changing parameters. Direct solution methods often cannot be used because of excessive storage and time requirements (the latter are O(m^3)). Iterative methods (such as multigrid, conjugate gradients, SSOR, orthomin, and GMRES), which take time O(m^2) per iteration, are almost universally applied. Another candidate for solving very large linear systems is the Monte Carlo method, which takes time O(m) per iteration; but the high variances of the estimates and the very slow convergence has tended to eliminate it from consideration. The sequential Monte Carlo method is a promising modification which shows guaranteed convergence. This method is applied to broad classes of large systems of linear and non-linear equations. Preliminary computations show promisingly large improvements in speed and efficiency over both plain Monte Carlo and deterministic methods.

University of Houston; Chenyi Hu; *Research at Undergraduate Institution (RUI): Finding All Numerical Solutions for Large-Scale Nonlinear Systems of Equations In Parallel and Reliably in a Given Domain*; (CCR-9503757); $88,686; 36 months.

This project investigates efficient parallel algorithms to find all numerical solutions for large-scale nonlinear systems of equations in a given domain with mathematical certainty, even in the presence of uncertainty in the data, roundoff error, and nonlinearities by finite digit computations. The basic algorithm being used is based on the interval Newton/Generalized bisection method. Parallel algorithms will be implemented on vector processors, shared-memory multiprocessors, and distributed-memory multicomputers. On a vector processor, new data structures will be designed to take full advantage of completely pipelined functional units of vector processors. To speed memory accessing, techniques will be developed to increase cache hit ratio on machines with cache such as the IBM 3090, and to reduce bank conflict on interleaved memory systems such as the CRAY C90. On a shared memory multiprocessor, optimal partitioning schemes will be studied for given nonlinear systems based on two protocols. By overlapping communication and computation, in addition to reducing communication and contention, large speedup is expected. On distributed-memory multicomputers, algorithms on coarse grained SPMP schemes and fine-grained MPMP schemes will be implemented on realistic machines. General sparsity and scalability will be considered in all implementations.

North Carolina State University – Raleigh; Ilse Ipsen; *Relative Perturbation Techniques for Eigenvalue and Singular Value Decompositions*; (CCR-9400921 A002); $79,658; 12 months. (Joint support with the Computational Mathematics Program)

This project develops perturbation techniques for deriving bounds on the relative error of singular values and eigenvalues, as well as bounds on the error angles of singular vectors and eigenvectors in terms of a relative gap. Deflation and convergence criteria based on these bounds have the potential to lead to highly accurate and efficient QR-type and Jacobi algorithms for computing eigenvalue, singular value and URV decompositions. In the context of the singular value decomposition of a matrix B, a class of perturbations is investigated which represents all possible that do not change the rank of B. In particular, it includes componentwise relative perturbations of bidiagonal and biacyclic matrices, and perturbations that annihilate an off-diagonal block in a block triangular matrix. So far, many existing relative perturbations bounds have been derived for singular values and vectors of bidiagonal matrices from general results for this class of perturbations. Also, new perturbation results have been proved for singular values and vectors of biacyclic, triangular and shifted triangular matrices. These results are extended to the generalized eigenvalue problem, and to the eigenvalue problems for structured and non-Hermitian matrices.

Stanford University; Thomas Kailath; *Fast and Numerically Accurate Algorithms for Matrices with Displacement Structure*; (CCR-9628117); $79,180; 24 months.

Numerous applications in sciences and engineering, as well as in computational, applied and pure mathematics give rise to problems involving matrices with Toeplitz, Hankel, Toeplitz-plus-Hankel, Vandermonde structures, controllability, observability, Cauchy, Loewner, Pick, Schur-Cohn, Routh-Hurwitz matrices, Bezoutians, along with many other patterns of structure. Standard mathematical software tools (e.g., MATLAB, MathCad, Maple V, NAG library, LAPACK), are based on standard (structure-ignoring) methods, and therefore their use is often not appropriate for obtaining a satisfactory solution. Here are two major reasons. First, ignoring the structure artificially squares the size of the data, which requires unnecessary storage and an extremely large amount of CPU time. Secondly, many structured matrices, e.g. Hankel, Pick, Hilbert, Vandermonde, are extremely ill-conditioned, which means that all available standard methods often fail to produce even one correct digit in the computed solution. The theme of this project is the study of several theoretical and computational problems related to Cauchy and Cauchy-like matrices, which arise in many applied areas, including signal processing, system theory H-infinity control. Use of the concept of displacement structure suggests that algorithms for Cauchy-like matrices can be used to design efficient practical algorithms for many other important kinds of structured matrices, such as Toeplitz, Hankel, Toeplitz-plus-Hankel, Vandermonde, Chebyshev-Vandermonde matrices, etc. In fact, all these matrices can be transformed into Cauchy-like matrices just by applying a few fast Fourier, cosine of sine transforms. It is planned to extend the results to develop fast and numerically accurate algorithms for new classes of structured matrices.

University of Colorado – Boulder; Thomas A Manteuffel, Stephen McCormick, and Thomas F Russell; ***Multilevel and Algebraic Iterative Methods in Large-Scale Computation*****; (DMS-9312752 A002); $150,000; 12 months. (Joint support with the Computational Mathematics Program and the Applied Mathematics Program)**

The project continues to develop numerical methods for solving problems of flows in porous media and to develop and analyze fast algorithms for partial differential equations, for computational fluid dynamics, and for several linear algebra problems. Specifically, the problems are: (a) develop fast parallel numerical methods for the solution of the Boltzmann equation used to model the transport of neutral and charged particles in 1 and 3 spatial dimensions; (b) develop multilevel zonal techniques for large scale computational fluid dynamics applications on parallel computers; (c) study discretization and solution techniques for fluid flows in porous media that involve complex geometry with heterogeneous material properties; (d) develop a new methodology for the solution of systems of partial differential equations by introducing new dependent variables and forming a least-squares functional. In each case, the work exploits and extends multi-level computational techniques. These projects are focused on important applications, such as calculating radiation doses for cancer treatment, ion implantation for semiconductor development, global weather modeling, fluid flow over airfoils and through rocket nozzles, oil recovery and contaminant transport in subsurface water supplies. They reflect the growing recognition that effective numerical modeling of physical phenomena involves several stages — formulation of the mathematical model, transforming the model into a discrete problem, and developing algorithms for the numerical solution of the discrete problem on advanced computers — none of which should be attempted without consideration for the others. Each project examines all components of the process, reformulating the model and discretization when necessary to yield discrete problems that can be efficiently solved numerically on parallel computers.

North Carolina State University - Raleigh; Carl D Meyer and William J Stewart; ***Computational Methods in Markov Chains*****; (CCR-9413309 A001); $64,758; 12 months.**

This project emphasizes practical as well as theoretical aspects surrounding the application and implementation of algorithms for computing stationary probabilities of discrete Markov chains. Special attention is devoted to applications in which discrete models are used to understand and analyze the dynamics of large systems comprised of a collection of loosely coupled subsystems such as in the study of queuing models and networks, telecommunications, computer performance evaluation, stochastic automata networks, and economic modeling and forecasting. Algorithmic facets concern the development and analysis of parallel multilevel methods based on aggregation/disaggregation techniques—both direct and iterative—and some of these techniques are applied to problems involving stochastic automata networks. An architectural aspect involves the identification of features of vector and parallel computers that are best suited for implementing Markov chain algorithms. Theoretical components involve studying properties of Markov chains. These ideas are then used to generate error analyses aimed at understanding the convergence and stability properties of multilevel aggregation/disaggregation algorithms.

Carnegie Mellon University; Gary L Miller; ***Algorithm Design and Implementation for Parallel Scientific Computation*****; (CCR-9505472 A001); $99,602; 12 months. (Joint support with the New Technologies Program)**

This project applies parallel algorithm design theory to problems arising from scientific problems governed by partial differential equations. Most of the effort is restricted to solving algorithmic problems including mesh generation and partitioning. This project develops algorithms and code for the solution of scientific problems that are highly unstructured and whose solution may be rapidly changing with time. Thus, the emphasis is on parallel mesh generation algorithms that will run on parallel machines in conjunction with the system solvers. The project also continues to improve the parallel efficiency of linear system solvers and will extend the work on combinatorial approaches to the construction of good preconditioners for linear system solvers. These are used to speedup the convergence rate of iterative methods. A central approach taken by this project for the design of new algorithms for problems such as mesh generation and preconditioner construction is to use the fact that the graphs for these problems have small separators that can be quickly found. Under prior projects partially supported by NSF, algorithms were discovered and developed to partition, for instance, finite element meshes into roughly two equal size pieces by removing a small number of elements. This work is referred to as the geometric approach to separators. The geometric approach quickly produces provably good partitions for finite element meshes. This project intends to show how the role of mesh generation methods and separator technology can be reversed. In particular, good mesh generation methods are being developed which use the separator technology. This technology will also be used in the construction of good iterative methods for solving these underlying systems. Further improvements are being made in the geometric separator algorithms via experimental and analytic methods. Relationships between the geometric method and other known algorithms and methods, such as spectral, are being investigated.

Baylor University; Ronald B Morgan; ***Implicitly Restarted GMRES and Arnoldi Methods for Nonsymmetric Systems of Equations*****; (CCR-9522612); $56,189; 36 months.**

This project investigates improvements of the restarted GMRES method for solving large nonsymmetric systems of linear equations. The goal is to develop a method that is competitive with short recurrence Lanczos methods such as QMR, but maintains the desirable properties of GMRES. These properties include extraction of the minimum residual solution from the subspace, generally steady convergence, and numerical stability. It is also desired to not significantly raise cost per iteration and storage requirements. The convergence of GMRES can be improved by including approximate eigenvectors in the subspace. Once these eigenvalues are essentially deflated from the spectrum of the matrix, removing even just a few small eigenvalues from the spectrum can significantly improve the convergence. This project studies a new and efficient way of including eigenvectors. Sorensen's implicit-QR Arnoldi method for eigenvalues is combined with an interior eigenvalue version of Arnoldi. The surprising result is that the approximate eigenvectors become part of a new Krylov subspace. This Krylov subspace can be used with GMRES. The expense per iteration is even less than that of regular GMRES for the case where matrix-vector products are expensive, and the number of iterations needed is sometimes greatly reduced. The resulting implicitly restarted GMRES method will be studied and compared with other methods.

Texas A&M University; Paul Nelson, Marvin L Adams, and G. Donald Allen; ***Novel Computational Approaches to the Transport Equation*****; (CCR-9302782 A002); $106,197; 12 months. (Joint support with the Computational Engineering Program and the Computational Mathematics Program)**

This project is directed toward computational solution of the linear particle transport equation, which usually is termed simply the "transport equation." This equation historically has provided a major challenge to the state-of-the-art computer hardware available at any given time. More specifically, the project is directed toward spatial approximations to this equation, because new approaches and better basic understanding in this area seem essential to permit the solution of the transport equation in the complex material arrangements underlying some of the emerging applications of this equation (e.g., identification of underground waste locations for environmental restoration). A certain abstract model (CLOF methods) of one-dimensional approximations in particle transport is being extended and applied, as follows: a massively parallel code is being developed for a general class of one-dimensional spatial approximations; the model is being extended to multiple (two or more) spatial dimensions; a code that permits computational determination of the order of accuracy of an arbitrary two-dimensional CLOF is being developed; the CLOF methods that have the very desirable property of satisfying the diffusion limit are being characterized. The corner-balance method, which is a novel spatial approximation that seems to enjoy several properties that are highly desirable for multidimensional calculations, is being extended to the three-dimensional setting, initially for parallelepiped cells, and subsequently for meshes composed of cells that are general (convex) polyhedra.

University of Maryland – College Park; Dianne P O'Leary and Gilbert W Stewart; ***Numerical Methods for Ill-Posed Problems and Markov Chains*****; (CCR-9503126 A001); $74,509; 12 months. (Joint support with the Computational Mathematics Program and the Applied Mathematics Program)**

This project concerns two topics in numerical linear algebra: the numerical treatment of Markov chains and regularization of ill-posed linear systems. The topics share common features: singular or nearly singular matrices; use of iterative methods for large-scale problems; problems in matrix perturbation; and connections with infinite dimensional problems. The computation of the steady-state vector of a Markov chain as well as other important quantities are being investigated. In particular, algorithms are being developed and analyzed for calculating mean first passage times and for computing the recurrence matrix of M/G/1 type queues. In addition aggregation methods with overlapping blocks and the behavior of chains with sluggish transients are being considered. When ill-posed problems are discretized they result in ill-conditioned linear systems which must be regularized to yield accurate solutions. There are three main goals in this work. The first is to prove the folk theorem that if the components of the data vector with respect to the singular vectors decay sufficiently fast then the conjugate gradient iteration will produce a regularizing set of solutions. The second is to treat problems with errors in the matrix by a regularized total-least-squares approach. The third is to further preconditioning strategies for iterative solution methods.

University of California – San Diego; Maria E Ong; ***Numerical Methods for the Nonlinear Eigenvalue Problems in High Performance Materials Design*****; (CCR-9403864 A001); $42,038; 12 months.**

The objective of this project is to develop adaptive methods for the first principles prediction of the properties technologically important materials. These materials may be characterized as having complex structures and compositions, large numbers of particles (> 100) and little or no symmetry. The methods being investigated include a suite of iterative, acceleration, and minimization schemes for the nonlinear eigenvalue problem and the accompanying elliptic problem arising in first principles prediction of material properties. This suite includes the eigenvalue iterative methods of Rayleigh-Ritz and Longsine and McCormick; minimization methods such us conjugate gradient, steepest descent, and the trace minimization method by Sameh and Wisniewski; and multilevel methods such as multigrid and fast multipole method. These methods are being used with adaptive grid to exploit the underlying dynamical, localized structure inherent in electronic wave functions. Modern parallel architectures are being used to reduce the computational time and memory costs. The various components of this project are being developed through the close collaboration of researchers from chemistry, physics, computer science, and numerical analysis.

University of California – Los Angeles; Stanley J Osher and Bjorn Engquist; ***Development, Analysis, and Applications for Numerical Methods for Nonlinear Partial Differential Equations*****; (DMS-9404942 A002); $125,000; 12 months. (Joint support with the Computational Mathematics Program and the Applied Mathematics Program)**

The project studies the development, analysis, and applications of numerical methods for nonlinear partial differential equations. The three main topics addressed are: (1) high resolution methods, including shock capturing, front capturing, numerical homogenization, and particle methods, (2) multiscale analysis applied to these and other related problems, and (3) fast methods for linear evolution equations based on multiscale analysis and high frequency asymptotics. Engineering and physical applications studied include combustion, fluid dynamics, crystal growth, Stefan problems, microwave scattering, and reacting gases. These innovative numerical methods are used to simulate real world problems in the areas of aeronautics, oil recovery, materials science, and environmental science. Transition to industrial and military application is made in the above areas, as well as in low observables, semiconductor device modeling, and shape recognition. These methods are applicable wherever the phenomena to be studied have sharp changes of state variables or are beyond the resolving capabilities of standard methods.

University of Minnesota – Twin Cities; Haesun Park and J. Ben Rosen; ***Solution of Structured Total Least Norm and Parameter Estimation Problems*****; (CCR-9509085 A001); $70,932; 12 months. (Joint support with the Computational Mathematics Program)**

A solution to an over determined linear system, where there may be errors in both the data matrix and the target vector, is required in many important applications in sciences and engineering. Recently, a new technique called Structured Total Least Norm (STLN) has been developed for obtaining the solution that preserves any affine structure, after the data are perturbed to account for the errors. It also permits the error to be minimized in various norms. The STLN algorithm is being investigated and developed into an efficient and practical solution method for many applications. Its computational performance is being studied and a more complete convergence theory is under development. Significant algorithmic and improvements are being made in order to efficiently handle large-scale problems and special structures. A particular study is being conducted to apply the STLN method to the model reduction by Hankel norm approximation problem. The effect of different norms are being explored. The STLN algorithm is being extended to solve more general problems where the structured matrix is a differentiable function of parameters to be estimated, and its performance is to be compared to other parameter estimation methods such as Prony's method.

Wake Forest University; Robert J Plemmons and Todd C Torgersen; ***Adaptive and Nonlinear Methods in Signals and Imaging*****; (CCR-9623356); $127,623; 24 months. (Joint support with the Science and Technology Centers Program and the New Technologies Program)**

This project investigates adaptive and nonlinear methods for the solution of certain large and often ill-posed inverse problems, where the solution may not depend continuously on the data. The particular applications addressed concern large scale adaptive filtering computations, such as in active noise control, and nonlinear and adaptive image enhancement computations. The project is using a 400 node IBM SP2 at the Maui High Performance Computing Center. This particular facility specializes in image and signal processing applications. The project includes the following topics:

- Fast and stable adaptive algorithms for estimation and filtering, along with sensitivity analysis, with applications in signal processing.
- Nonlinear and adaptive image enhancement methods, including Total Variation minimization involving partial differential equations, adaptive multi-level segmentation-based methods, and a combination of these two approaches.
- Coupling numeric and symbolic computations.
- Parallelization of the algorithms. The project has strong educational and REU research components, and includes participation by underrepresented minorities.

Old Dominion University; Alex Pothen; ***Parallel Sparse Matrix Computations*****; (CCR-9412698 A001); $40,000; 12 months.**

This project investigates the design, implementation, analysis of algorithms for several problems that arise in the solution of large, sparse systems of equations and least squares problems on high-performance parallel computers. The irregular nature of the memory access required by sparse matrix algorithms makes good performance difficult to achieve. Innovative algorithmic approaches capable of better concurrency than the traditional sequential approaches, and implementations that exploit the memory hierarchy are essential to obtain good performance. The following problems are being investigated: (1) An algebraic approach to the partitioning problem that arises in partitioning data and computation on parallel computers, and in computing nested dissection orderings, (2) The problem of reordering sparse matrices to minimize with applications to computing incomplete factorization (3) The product-form-inverse approach to the solution of sparse triangular systems on parallel computers, (4) The solution of rank-deficient least-squares problems on sequential and parallel computers.

University of Texas – Austin; Gregory J Rodin, J. Tinsley Oden, Robert A van de Geijn, and James C Browne; ***Application of High Performance Computational Methods to Micromechanical Characteristics of Composite Material*****; (ECS-9422707 A001); $438,523; 12 months. (Joint support with the New Technologies Program, the Computational Engineering Program, and the Solid Mechanics Program)**

Design of composite materials to have desired properties is a task of great engineering significance. Presently this is usually accomplished through trial and error laboratory synthesis and ad hoc analysis with the result that non-optimal design is commonplace. The dramatic increase in computational power available for mathematical modeling and simulation raises the possibility that computationally-based design can play a significant and complementary role to laboratory development of composite materials, as has been the case in other engineering design problems. The computational problem is to determine the overall properties (not necessarily linear elastic) of the composite material from the constituent properties. The goal of this research is to create a toolbox of computational components with which engineers can design and analyze composite materials directly form the microstructural level. The approach taken is to adapt the emerging methods for high performance parallel computation, being developed for computational modeling in other domains, to micromechanical analysis of composite materials. The application of large-scale computation to this calls of problems has been impeded by the fact that neither available computation power nor available computational methods were adequate for realistic micromechanical models. The methods to be adapted in this project are: (i) parallel adaptive finite element and boundary element methods for computational fluid dynamics and acoustics, and (ii) parallel fast multipole methods emerging form many- particle dynamics. The most significant properties of these methods are: Superalgebraic convergence rates, permitting effective simulations with models requiring a minimum computational effort to achieve acceptable accuracy's. Amenability to common infrastructure for efficient parallel implementation. Natural applicability to micromechanical modeling of composite materials since they perform similar roles in their current domains. The methods are complementary in their domains of application to composite materials. The development of the methods and of the toolbox will be empirical and example driven.

University of Minnesota – Twin Cities; Youcef Saad, Harry G Perros, and William J Stewart; ***US-France (INRIA) Cooperative Research: Numerical Solution of High Speed Network Models*****; (INT-9600422); $36,000; 36 months. (Joint support with the Western Europe Program)**

The objective of this US-France cooperative research is to evaluate the performance of crucial aspects of high-speed communication networks. The project evaluates the effectiveness of Stochastic Automata Networks (SANs) as a modeling technology for the analysis of telecommunications systems and development and deployment of high-speed networks. The approach will be applied to one aspect of modeling high-speed networks, namely performance analysis of packet delay and loss in a high-speed multiplexer. This multiplexer has been developed to provide flexibility in integrating multi-media traffic. The US investigators bring to this collaboration expertise in numerical modeling, networking, and parallel processing. This is complemented by the French investigators' expertise in Stochastic Automata Networks.

University of California – Berkeley; James A Sethian and Alexandre J Chorin; ***Computational Techniques from Geometry and Statistical Physics Applied to Fluid Mechanics and Interface Problems*****; (DMS-9504950 A001); $150,000; 12 months. (Joint support with the Computational Mathematics Program and the Applied Mathematics Program)**

This project studies the motion of singularities in solutions of differential equations, in particular, the motion of interfaces as described by level-set methods and applied to problems in interface motion, and the motion of vortices in three space dimensions, as described by vortex methods and other Lagrangian techniques. The interface techniques exploit Hamilton-Jacobi methodology coupled to upwind schemes for hyperbolic solvers; new work focuses on developing fast methods, methods for multiple interfaces, and the incorporation of effects of visibility, surface diffusion, and reflection/re-emission in surface profile motion. The vortex motion through Lagrangian methods develops machinery from renormalization methods and polymer analysis from statistical physics applied to fluid phenomena through linkage with Shenoy-Williams methodology, and Kosterlitz-Thouless formulations. Methods are being developed to follow the motion of interfaces and vortices. Such problems arise in flows of materials, for instance, and whenever the places where objects or materials meet or end change with time. These are common situations throughout science and engineering. The work is applied to a number of areas. In etching, deposition, and lithography in the manufacture of microelectronic components, the work is used to analyze process steps leading to higher yield and efficiency of electronic devices. In image enhancement, segmentation, and recognition applied to medical imaging, the work leads towards automatic software for recognizing cardiac data in CT and MRI scans. In grid generation, the techniques are used to develop new techniques for building grids around complex bodies, as required in aircraft and engine design. The Lagrangian and vortex particle work focuses on generation of large turbulent structures in wakes, motions of polymers in superfluids, and fundamentals of superconductivity.

Rensselaer Polytechnic Institute; Robert L Spilker, Mark H Holmes, Boleslaw K Szymanski, Mark S Shephard, and Joseph E Flaherty; ***Understanding Human Joint Mechanics through Advanced Computational Models*****; (DMS-9318184 A004); $500,000; 12 months. (Joint support with the Bioengineering Program, the New Technologies Program, and the Computational Mathematics Program)**

The objective of this research is to develop and demonstrate the capability for predicting, by computer simulation the ground motion of large basins during strong earthquakes, and to use this capability to study the seismic response of the Greater Los Angeles Basin. The research seeks to: (1) Develop three-dimensional models of large-scale, heterogeneous basins that take into account earthquake source, propagation path, and site conditions; (2) Develop nonlinear models for sedimentary basins that experience sufficiently strong ground motion; (3) Develop unstructured mesh methods and associated fast parallel solvers, enabling the study of much larger basins; (4) Develop software tools for the automatic mapping of the computations associated with large unstructured mesh problems on parallel computers; (5) Characterize the computation and communication requirements of unstructured mesh problems, and make a set of recommendations for the design of future parallel systems. While the work is motivated by an interest in gaining a better understanding of strong seismic motion in large basins, the algorithms and software tools developed will be applicable to a wide range of applications that require unstructured meshes.

University of Utah; Frank Stenger and Kris Sikorski; ***Sinc Algorithm Development and Studies*****; (CCR-9307602 A002); $70,583; 12 months.**

This project investigates SINC methods for ODES. The following tasks are undertaken; (1) develop a SINC ODE-IVP (sequential and/or parallel) program package for solving ordinary differential equation initial value problems, (2) develop a (sequential and/or parallel) program package to evaluate multidimensional convolution integrals, (3) write a program package which uses the Maple program to reduce a differential or integral equation to a system of algebraic equations, based on SINC derivative and/or SINC convolution collocation, (4) write program packages, based on SINC derivative and/or SINC convolution collocation, to solve problems which are important for ongoing research at the University of Utah (the forward problem Maxwell equations, with variable permeability and permittivity, but with conductivity; the solution, based on SINC convolution collocation, of the integral equation formulation of a scalar conservation law problem in two and three dimensions), and (5) comparison studies of complexity and stability of SINC and classical methods, especially in areas in which one can expect SINC methods to have advantages over methods, such as for problems with singularities, problems over unbounded regions, boundary layer problems, stiff problems, and intensive problems such as the solution of the above cited equations.

Columbia University; Joseph F Traub; ***Average Case and Probabilistic Setting of Information-Based Complexity*****; (CCR-9420543 A001); $134,796; 12 months.**

Problems with only partial or contaminated information arise in many disciplines: computer science, physics and chemistry, control theory, statistics, prediction and estimation, scientific and engineering computation, geophysics, decision theory, and finance. Furthermore, this information is often expensive to obtain. The goal of information-based complexity is to create a computational complexity theory for problems with partial, contaminated, and priced information, and to apply the results to solving specific problems in varied disciplines. The prime foci of this project are the following: (1) Breaking Intractability: Many problems in science, engineering, and even finance, deal with functions that have a large number, d, of variables. However, typical multivariate problems are intractable in the worst case deterministic setting. The only way to break intractability is to weaken the assurance by switching to another setting. Randomized, average case, and probabilistic settings continue to be studied. In particular, it has been proven that multivariate integration and approximation are both strongly tractable on the average. The proof is non-constructive; the problem of how to sample in d dimensions to achieve strong tractability is being addressed. (2) Software Development and Testing: Theoretical investigations have led to algorithms for important problems such as high dimensional integration. To make these methods widely available, software is being developed and tested. Since these methods contain a large amount of inherent parallelism, the implementation is being done on a workstation network. Testing is also planned on parallel computers. Real-world problems are used for testing. (3) Probabilistic Complexity of Piece-Wise Smooth Functions: Piece-wise smooth functions arise in diverse areas of science and technology such as computer vision, image and signal processing, scientific computation, meteorology, seismology, and tomography. Because of negative results of the worst-case setting these problems are being studied in the probabilistic setting. New algorithms for integration, approximation, and detection of singular points are being studied and implemented.

Cornell University; Lloyd N Trefethen; ***Non-Normal Matrices and Operators: Analysis, Computations, Applications*****; (DMS-9500975 A001); $90,000; 12 months. (Joint support with the Computational Mathematics Program and the Applied Mathematics Program)**

The project concerns problems involving highly non-normal matrices and operators. Approximately speaking, this means matrices and operators whose eigenvectors or eigenfunctions are far from orthogonal. Most matrices and operators that arise in applications are normal or close to normal, but the highly non-normal cases constitute an important minority, one that has been poorly handled by traditional methods based on eigenvalues. The emphasis of this project is on methods that avoid the use of eigenvalues. Some of these make use instead of certain sets in the complex plane known as pseudospectra, defined in terms of perturbations of the spectrum or equivalently as regions bounded by level curves of the norm of the resolvent. The applications investigated fall in three areas: numerical solution of non-self-adjoint partial differential equations; iterative solution of large-scale matrix problems; and fluid dynamics, specifically hydrodynamic stability and transition to turbulence. This work is connected with some major ongoing developments in science and engineering. One is a basic change in the prevailing view of how fluid flows become unstable and eventually turbulent, with potential applications across many branches of mechanical, aeronautical, and civil engineering as well as in meteorology, geophysics, and other scientific fields. A second, less accessible to the public but still of great importance, is the great shift underway in computational science from "direct" to faster "iterative" methods for solving large-scale linear algebra problems.

Cornell University; L. B Wahlbin; ***Algorithms and Numerical Analysis for Partial Differential and Integral Equations*****; (DMS-9311406 A002); $105,000; 12 months. (Joint support with the Computational Mathematics Program and the Applied Mathematics Program)**

The project studies new techniques and methods of analysis of behavior in existing approximation methods, and introduce and analyze new algorithms for the numerical approximation of various problems. Specific areas to be considered include: (1) preconditioned iterative methods for fast solution of resulting algebraic equations, (2) finite element methods in domains with curved boundaries and interfaces, (3) local behavior of finite element methods in differential and integral equation problems with singularities, and (4) superconvergence phenomena in differential and integral equations. Numerical experimentation and examples play important roles in the discovery and elucidation of various numerical methods. Of course, practical application is a motivating factor for the problems studied. One of the aims of the work is to develop a deeper understanding of many existing methods, improve them or develop new ways of treating old and new problems. This kind of understanding is central to the design of accurate and efficient algorithms for the use in present and future computer environments. Various questions concerning the practical solution of problems of interest in engineering and science are pursued. One aim is to construct methods that are faster and more accurate than existing ones. A second aim is to invent and study techniques for the solution of problems heretofore beyond the scope of even the most powerful computers. Areas of application in which the studies are important are structural mechanics, fluid dynamics, groundwater flow and electromagnetic theory.

New York University; Olof B Widlund; ***Iterative Substructuring Methods for Elliptic Problems and Related Algorithms*****; (CCR-9503408 A001); $82,000; 12 months.**

The development of numerical methods for large systems is central in the development of efficient codes for fluid dynamics, elasticity, and other core problems of continuum mechanics. Many other tasks in such codes parallelize relatively easily. The importance of the algebraic system solvers is therefore increasing with the appearance of parallel and distributed computing systems, with a substantial number of fast processors, each with relatively large memory. The development of improved methods will, together with more powerful computer systems, make it possible to carry out simulations in three dimensions, with quite high resolution, relatively easily. Work will continue in developing iterative substructuring and other domain decomposition methods for increasingly difficult elliptic problems. Domain decomposition algorithms are iterative methods, often of conjugate gradient type, for the parallel solution of the large linear, or nonlinear, systems of algebraic equations that arise when partial differential equations are discretized by finite elements, finite differences, or spectral methods. In each iteration step, local problems representing the restriction of the original problem to a potentially large number of subregions are solved approximately. The subregions, which can be allocated to individual processors of a parallel computer, form a decomposition of the entire domain of the problem. In addition, the inclusion of a coarse subproblem substantially increases the efficiency of the preconditioner. In this study, which combines mathematical analysis with the design and numerical testing of algorithms, a special emphasis is on the study of elements and other high order finite element methods, on nonconforming methods such as the mortar finite elements, and on the difficulties which are related to distorted subregions or elements.

University of Texas – Austin; David M Young and David R Kincaid; ***Iterative Algorithms and Software for Parallel Architectures for Partial Differential Equations*****; (CCR-9504954 A001); $62,769; 12 months.**

The objectives of this project are (1) to carry out research on iterative algorithms for solving large sparse systems of linear algebraic equations with emphasis on systems arising from problems involving partial differential equations and on the use of parallel computers, (2) to develop research-oriented prototype software based on the algorithms, and (3) to test the algorithms and software on a range of problems. Among the algorithms to be considered are parallel alternating-type methods, periodically generated iterative methods, rational iterative methods, and iterative methods for solving large sparse nonsymmetric linear systems.

University of Texas – San Antonio; Xiaodong Zhang; ***Research at Undergraduate Institution (RUI): High Performance Parallel/Distributed Computing for Numerical Algorithms, Software and Scientific Applications*****; (CCR-9400719 A001); $17,000; 6 months. (Joint support with the East Asia and Pacific Program Program)**

This project focuses on three important problem areas high performance parallel/distributed computing: 1) designing, analyzing, implementing and testing several groups of structured methods for solutions of large sparse nonlinear systems of equations on parallel and systems; 2) applying numerical algorithms to solve large real-world problems; and 3) studying execution patterns of scientific application programs for effective parallel program and architecture design.

Symbolic and Algebraic Computation

University of Wisconsin – Madison; Eric Bach; ***Algorithmic Number Theory*****; (CCR-9510244); $80,534; 12 months. (Joint support with the Computational Mathematics Program and the Algebra & Number Theory Program)**

The goal of this project is to study the computational complexity of number-theoretic problems. Algorithms for solving these problems are useful in the design of systems for random number generation, cryptography, and computer algebra. In addition, many of the classic questions in complexity theory (such as whether randomization helps in efficient computation) are motivated by computational number theory. The investigations being undertaken fall into three categories: a) Computations and explicit bounds in analytic number theory. b) The design and analysis of number-theoretic algorithms. c) Constructive techniques in algebraic geometry and commutative algebra.

City University of New York – City College; Gilbert Baumslag; ***Computational Group Theory*****; (DMS-9626231); $135,165; 12 months. (Joint support with the Computational Mathematics Program and the Algebra & Number Theory Program)**

This project is to develop software to study finitely presented groups. Many problems that arise in diverse mathematical disciplines can be distilled into related problems about finitely presented groups. One of the major developments of the decade beginning in the mid 1950s has been the proof that almost all questions about such finitely presented groups are algorithmically undecidable. Nevertheless, the very importance of these groups makes it essential to try to obtain at least partial information about them. In particular, although problems about finitely presented groups are unsolvable in general, it may be possible in a number of specific instances to solve such problems. In order to gain some insight into these problems, it is often necessary to try to carry out some experiments and computations with finitely presented groups. Such experiments and computations are very time-consuming and usually impossible to carry out by hand. This project aims to make such experiments and computations possible by using very fast computers, enabling one to gather enough information together about many important problems to facilitate their solution. The project creates and implements a package of group-theoretical algorithms and procedures, which can be run by means of a graphical user interface and which is easily accessible by a wide variety of users.

The ubiquitousness and usefulness of groups make the creation of a software package that will lead to a deeper understanding of many of them an important enterprise.

University of Illinois – Chicago; Daniel J Bernstein; ***Algorithmic Problems in Number Theory*****; (DMS-9600083); $59,000; 36 months. (Joint support with the Algebra & Number Theory Program)**

The number field sieve is the latest algorithm for factoring integers into primes. This project continues research into the speed of the number field sieve. The plan is first, to study the precise distribution of certain vectors constructed in the number field sieve; second, to develop tools to analyze the probability that vectors chosen from a non-uniform distribution will be independent; third, combine these results to sharpen the known heuristics for the time taken by the number field sieve. The project will also consider various problems in multiprecision arithmetic from a geometric viewpoint.

University of California – Berkeley; John F Canny; *A Toolkit for Non-linear Algebra*; (CCR-9320588 A002); $108,548; 12 months. (Joint support with the Computational Mathematics Program and the Algebra & Number Theory Program)

This project undertakes the implementation of an algebraic-geometric toolkit, written in C, which can solve systems of equations over the complex numbers, and inequalities over the reals. There are many applications of such a toolkit, since problems from many branches of science and engineering can be formulated using systems of polynomial equations and inequalities. By several well-known theorems, in particular, Tarski's theorem for the reals, it is possible to solve such problems in principle. But the worst case bounds are exponential or doubly-exponential in the number of variables, and no practical system has appeared that can deal with more than a few special cases. On the other hand, systems of polynomials may have a special structure, a kind of sparseness, that implies a complexity (measured by the algebraic degree) much lower than the worst case. Indeed, most of the important applications are of this type. Very recently, algorithms have been developed that can exploit this structure. This realization is a strong motivation for developing the toolkit at this time. In addition, there have been improvements in algorithms for sign determination and symbolic- numeric computation, and it is felt that these methods are now advanced enough to warrant implementation.

University of California – Berkeley; Andrew J Casson and Robion C Kirby; *Representations of Three-Manifold Groups*; (DMS-9505053); $130,800; 12 months. (Joint support with the Computational Mathematics Program and the Topology Program)

This project will study the Geometrization Conjecture for 3-dimensional manifolds, proposed by Thurston in 1982. A proof of this conjecture would complete the topological classification of closed 3-manifolds, provide insight into topological properties of 3-manifolds, and lead to efficient methods to determine whether two given 3-manifolds are topologically equivalent. The project will explore a very direct approach to Thurston's conjecture; given a topological description of a 3-manifold M, an attempt will be made to find triangulations of M that are closely related to a geometric structure on M. The project will continue the work in Topological Quantum Field Theory. It will try to explain the Kontsevich and Vassiliev invariants, using intersections of certain codimension-two manifolds in a compactification of products of the given 3-manifold with diagonals deleted. It will also continue to study the formulae for the Witten-Reshetikhin-Turaev invariants, looking for topological interpretations and applications. Since the space that we live in is three-dimensional, and all our lives we perfect our three-dimensional geometric intuition simply by going about our everyday business, it is always surprising to the layman to learn that mathematicians are frequently more capable of answering questions about higher dimensional geometric objects than about the analogous objects in dimension three. This is, however, very much the case. The most famous outstanding conjecture in topology is the three-dimensional Poincare conjecture, made at the turn of the century. On the other hand, its higher dimensional analogues have all been settled. The research of this project is directly addressed to these notoriously recalcitrant problems.

Northeastern University; Gene Cooperman and Larry A Finkelstein; *Applications of Group Theory to Large Scale Computation*; (CCR-9509783 A001); $79,044; 12 months. (Joint support with the Algebra & Number Theory Program)

This project addresses computational issues for finite groups that are currently beyond the scope of current algorithms and existing technology. The groups are represented either as permutation groups or as matrix groups defined over finite fields. The questions of interest concern both group theoretic structure and applications of the group theory. A central motivation for the research is the construction of representations for matrix groups defined over finite fields. This is an important part of a large-scale international effort to develop a suite of computational tools for performing matrix group computations. The tools to be developed have application to problems in the physical sciences, engineering, and those parts of mathematics dealing with high dimensional matrix representations of groups. A second research area is computational techniques for solving the minimum word problem for finite groups. This problem arises in many contexts, such as routing in interconnection networks and coding theory, and is known to be computationally difficult. The research will lead to new algorithms which use group automorphisms to prune the search tree of possible solutions. A third research area is the development of a software architecture to bring parallel computers to bear on large scale problems in computational group theory, such as the construction of very large matrix and permutation representations. Such as integrated, interactive environment will enable researchers to quickly overcome the initial learning barrier for effective use of parallelism. They will be able to combine readily available networks of workstations at their home site (such as student labs), with an already familiar language, such as GAP or LISP.

University of South Carolina – Columbia; Karel Culik; ***Finite Automata Methods in Image Compression and Manipulation;*** **(CCR-9417384 A001); $55,000; 12 months. (Joint support with the Theory of Computing Program)**

A new method for image description, compression, and manipulation based on finite automata is studied. A Weighted Finite Automaton (WFA) computes a real function on 0,1, .., n. For n = 2, such a function can be interpreted as the grayscale function specifying an image. The most important application of WFA is the new image-compression method based on the WFA inference algorithm. This method can be classified as a fractal compression method, but it is different and performs considerably better than the previous fractal methods for image compression. It produces an excellent compression size to quality ratio for the widest variety of images. The current image-data compression software performs well; however, it does not reach the full potential of WFA for image compression. One of the main goals of this project is to optimize the compression algorithm to make it faster and to produce better quality regenerated images for the same compression rate. Weighted Finite Transducers (WFT) are a tool for the description and implementation of image transformations. The project will investigate the WFT and their applications, their efficient implementation, and the design of inference algorithms for WFT.

IBM Thomas J Watson Research Center; Samuel S Dooley and Barry M Trager; ***East Coast Computer Day, Yorktown Heights, NY April 13, l996;*** **(CCR-9626881); $8,301; 6 months.**

The purpose of this series of meetings is to stimulate interest in and enhance understanding of the technical aspects of computer algebra, by providing a way for those in the region who are active or interested in becoming more active in computer algebra to meet with each other in an inexpensive and easily accessible way. This meeting is more informal than the annual ISSAC meetings, with ample time for unstructured interaction. Participation by the region's graduate students is particularly encouraged through special invitation to contribute to the poster sessions, and through travel support. This meeting has become a regular series for the region, with different academic and industrial institutions hosting the meeting in rotation. This year's meeting was the third in the series: the first meeting was held in 1994 at Drexel University in Philadelphia, and the second meeting in April 1995 at the University of Delaware. The meetings have had the endorsement of ACM's Special Interest Group for Symbolic and Algebraic Manipulation (SIGSAM) and of the Society for Industrial and Applied Mathematics (SIAM).

University of Illinois – Urbana; Daniel R Grayson; ***A Software System for Algebraic Geometry Research;*** **(DMS-9622608); $46,536; 12 months. (Joint support with the Computational Mathematics Program and the Algebra & Number Theory Program)**

This award supports the development of a computer software system enabling large scale research computing projects dealing with systems of polynomial equations in many variables. The result of the project are a usable and efficient computer program, which executes the Groebner basis algorithm for computing standard bases and syzygies of ideals and modules over polynomial rings. Improvements over the existing program, Macaulay, include the addition of a good programming language for the user, the ability to work over more general coefficient rings, and increased speed of execution. This project is in the general area of commutative algebra and the computer-aided computational aspects of this field. There is a growing interest in using computers to answer theoretical questions in algebra and conversely, algebra is quite useful for the development of algorithms. This project makes use of both the theory and practice of symbolic calculations to solve real problems in commutative algebra.

Colorado School of Mines; Willy A Hereman; ***Development of Symbolic Software for Nonlinear Partial Differential Equations;*** **(CCR-9300978 A001); $5,000; 0 months.**

This Research Experiences for Undergraduates (REU) project deals with the development and implementation of new symbolic algorithms for the computation of Lie symmetries, the investigation of integrability, and the calculation of exact solutions of systems of nonlinear partial differential equations (PDEs). The first symbolic program will assist in computing classical and nonclassical Lie-point symmetries, as well as generalized symmetries, for large systems of differential equations. The second software package will automatically carry out the Painleve integrability test for the existence of soliton solutions. Via a unified bilinear transformation method, it will create "solitons" for a large class of integrable PDEs. The development of novel mathematical techniques, in addition to the refinement and generalization of existing methods, is an essential part of this research project. The goal is to provide high quality symbolic programs, in Macsyma and Mathematica, to researchers working on soliton theory, dynamical systems, mathematical physics; specifically wave phenomena in meteorology, biosciences, fluid dynamics, plasma and particle physics, and nonlinear optics.

Colorado School of Mines; Willy A Hereman; ***Symbolic Software for the Investigation of Nonlinear Partial Differential Equations;*** **(CCR-9625421); $87,000; 24 months. (Joint support with the Computational Mathematics Program)**

This project continues the development and implementation of new powerful symbolic algorithms for testing the integrability and the calculation of conservation laws. The construction of closed-form solutions for systems of nonlinear partial differential equations (PDEs) will also be investigated. In particular, for systems of nonlinear evolution equations, a first symbolic software package will automatically generate polynomial-type conservation laws. Based on scaling properties (symmetries) of the equations, the algorithm provides an elegant way to construct conserved densities and their associated fluxes. Based on homogenization techniques, a second software package will allow the user to compute the analytic form of particular solutions for large classes of PDEs. For integrable equations, the technique leads to soliton solutions. For non-integrable equations, closed-form expressions of homoclinic and hetereoclinic orbits can be computed efficiently with the method. The symbolic software to be developed will complement and build on earlier algorithms for the computation of Lie-point and non-classical symmetries, to find soliton solutions via Hirota's direct method, and to test the integrability of differential equations via the Painleve test. The programs will be written in Macsyma and Mathematica syntax. They will be entirely symbolic and produce analytical output. The development of novel mathematical techniques, in addition to the refinement and generalization of existing methods and algorithms, is also a part of this research project. The aim is to provide high-quality symbolic packages to researchers working on soliton theory, dynamical systems, mathematical physics; specifically nonlinear wave phenomena in networks, elastic media, chemical kinetics, biosciences, fluid dynamics, plasma physics, and nonlinear optics.

University of Southern California; Ming-Deh Huang; *Computational Number Theory and Computational Algebraic Geometry*; (CCR-9412383 A001); $67,190; 12 months.

This research explores the rich and potentially fruitful interplay between number theory and algebraic geometry to investigate computational problems in number theory, with emphasis on: discrete logarithms over various groups, which is a fundamental problem of cryptographic significance; polynomial congruences modulo primes, including the univariate and multivariate cases; small nonresidues in finite fields; efficient algorithms for computing with curves over finite fields, with emphasis on fundamental arithmetical problems on curves which are important for applications in computational number theory, cryptography and coding theory; applying the results obtained to public-key cryptography, coding theory, and other areas as well.

Rensselaer Polytechnic Institute; Erich L Kaltofen; *Efficient Computer Algorithms for Symbolic Mathematics*; (CCR-9319776 A002); $74,169; 12 months. (Joint support with the Algebra & Number Theory Program)

This project continues research in design, analysis, implementation of efficient algorithms for several basic computer algebra problems. The parallel time complexity of solving linear systems with a minimum number of processors is further investigated. New sequential and parallel randomized algorithms for sparse linear systems are explored, and polynomial-time algorithms that make irreducible polynomials factorizable via substitution of the variables by monomials are sought. Another research direction is how to handle imprecise data in the course of symbolic computation. The allowance of floating point coefficients in a symbolic (i.e., parameterized model) is considered crucial for a symbolic approach to real-world problem solving. In particular, the question of computing in polynomial-time the nearest singular Toeplitz matrix for a given Toeplitz matrix, and the nearest pair of polynomials with a common divisor for a given pair of relatively prime polynomials, is a starting point for the theory of precision-tolerant symbolic computation. Ultimately, the problem of factoring over the complex numbers a minimally perturbed version of an absolutely irreducible curve is attacked. Algorithm implementation is done using the DSC system for distributing a symbolic computation over a network of workstations.

University of Oregon – Eugene; William M Kantor; *Groups, Algorithms, and Geometries*; (DMS-9301308 A003); $135,600; 12 months. (Joint support with the Computational Mathematics Program and the Algebra & Number Theory Program)

This project addresses groups, algorithms and geometries. Probabilistic and asymptotic properties of finite groups are studied: the nature of random elements, or of random subgroups, of a given permutation group. Some results obtained are applied to algorithmic questions concerning permutation groups. The study of polynomial-time and parallel (complexity class NC) group-theoretic algorithms continues. Additional new practical group-theoretic algorithms are obtained, in some cases based on methods developed in polynomial-time situations. All of this work makes detailed use of the classification and properties of finite groups. Some of these algorithms depend on geometric methods. Other geometric projects are continued including investigations into the number of planes, designs, generalized quadrangles, or other geometries, based upon special properties of their automorphism groups. This project is in the general area of finite group theory and the computer-aided computational aspects of this field. There is a growing interest in using computers to answer theoretical questions in algebra and conversely, algebra is quite useful for the development of algorithms. This project makes use of both the theory and practice of symbolic calculations to solve real problems in group theory.

State University of New York – Albany; Deepak Kapur; *In-*

***vestigation of the Dixon Resultants*; (CCR-9622860); $123,573; 24 months.**

Dixon's formulation for simultaneously eliminating several variables from a system of multivariate polynomial equations is being investigated. Solving polynomials equations and identifying conditions on parameters under which the equations have common solutions are fundamental problems in symbolic computations which arise almost everywhere. Polynomials are used to model many phenomena — particularly in robotics, kinematics, computer vision, solid modeling, graphics, chemical equilibrium and other engineering, CAD-CAM design, etc. Dixon's formulation directly generalizes Bezout-Cayley's formulation of a resultant for eliminating a single variable. Dixon resultants have not been extensively investigated apparently because they presumably got overshadowed by Macaulay's formulation of resultants for simultaneously eliminating several variables. Implementations of Dixon resultants, Macaulay resultants and sparse resultants have revealed that Dixon resultants are more efficient to compute in practice. Dixon resultants will be analyzed with a particular emphasis on studying the relationship of the size and structure of the Dixon matrix to the input polynomial system.

University of Akron; Kathy J Liszka; *Parallel Algorithms for Generating Large Tables of Primes*; (CCR-9634559); $15,843; 18 months.

The search for efficient algorithms to produce primes is motivated largely by the needs of cryptologists and number theorists. Many open problems in these areas rely on very large tables of primes for their solution. The Sieve of Erathosthenes is a well known method for generating tables of primes up to a given natural number, N. Significant work has been published that improves on the original sieve to generate primes, some in sublinear time. This project will investigate six variations on the sieve to generate large tables of primes efficiently using parallel architectures. Initial testing and implementation of the algorithms will be based on selected mapping and communication structures that are determined to be optimal. The most effective architecture will be identified for each algorithm. Preliminary data will give insight into the scope of memory limitations that must be addressed in order to generate primes to N = 10^18. Break even points between the various algorithms will be provided because it is important to determine how feasible it is to reach a certain point given a particular algorithm.

University of North Carolina – Chapel Hill; Dinesh Manocha; *Numeric and Symbolic Methods for Polynomial Manipulation*; (CCR-9319957 A002); $66,104; 12 months.

Algorithms for symbolic and numeric manipulation of systems are needed for basic symbolic, numeric and geometric computations and their application to problems in graphics, geometric modeling, robotics, vision and engineering. This research takes a three-fold approach to these problems using multipolynomial resultants: (1) Develop better algorithms for symbolic resultant computation and solving nonlinear polynomial equations using resultants and matrix computations. Algorithms are being developed in the context of exact arithmetic as well as floating point arithmetic. It involves better formulations of resultants in terms of matrices and determinants and use of symbolic and numeric algorithms making use of the matrix formulation. (2) Specialize these algorithms to applications in computer graphics, geometric modeling and robotics by making use of the specific polynomial systems arising in these applications. (3) Develop a library of routines, ELIMPACK, for symbolic resultant computation and finding roots of polynomial equations (in exact and floating point arithmetic). This package is of great utility to symbolic computation, numerical computation, geometric applications and the engineering community. The results also help in understanding the complexity of geometric problems described in terms of polynomial equations.

City University of New York – Herbert H Lehman College; Victor Y Pan; *Polynomial and Matrix Computations*; (CCR-9625344); $77,895; 12 months.

This project continues the investigation of effective algorithms for some major problems of polynomial and matrix computations. One of the main subjects of the current project is the study of approximating polynomial zeros, which has recently resulted in the design of new solution algorithms running in optimal time (up to polylog factors). The optimality has been reached under both Boolean and arithmetic and both sequential and parallel models of computing. The project includes the work on extension of the latter results to other major problems of algebraic and numerical computing, such as computing approximate common divisors of pairs of univariate polynomials, approximation of matrix eigenvalues, solving systems of polynomial equations, and algebraic optimization problems. Another major direction of the project is to continue and extend the work on computations with structured matrices, which have applications to numerous major practical and theoretical computational problems, including the areas of speech and signal processing, Markov chains, fundamental algebraic computations with general matrices, and the solution of partial differential and integral equations.

University of Georgia; Carl B Pomerance, Andrew J Granville, and Helmut Maier; *Topics in Analytic and Algorithmic Number Theory*; (DMS-9206784 A004); $92,000; 12 months. (Joint support with the Computational Mathematics Program and the Algebra & Number Theory Program)

This research addresses problems in analytic, computational, and additive number theory. Work continues on integer factoring algorithms, in particular the number field sieve. Size of the coefficients of cyclotomic polynomials and irregularities in the distribution of primes is addressed. Work also occurs on the distributions of the primes as well as sieve theory. This is research in the field of number theory. Number theory starts with the whole numbers and questions such as the divisibility of one whole number by another. It is among the oldest fields of mathematics and it was originally pursued for purely aesthetic reasons. Within the last half century, it has become an essential tool in developing new algorithms for computer science and new error correcting codes for electronics.

Ohio State University; Akos Seress; *Efficient Computation in Finite Groups*; (CCR-9503430 A001); $28,568; 12 months. (Joint support with the Algebra & Number Theory Program)

This project continues research in the area of manipulation of finite groups and estimation of their parameters. Algorithms for finite groups and their associated Cayley graphs have a wide range of applications in mathematics and computer science. These applications range from problems related to the classification of finite simple groups to the mixing rate of Markov chains, the design of interconnection networks for large interacting arrays of CPU's, the graph isomorphism problem (of relevance to chemical documentation), and group-based cryptography (such as DES). The principal consumers of the expected results will be the areas of computational group theory and the complexity theory of algorithms; but applications arise in the other areas mentioned as well.

Marquette University; Michael Slattery; *Second Magma Conference on Computational Algebra; May 12-16, 1996; Milwaukee, WI*; (DMS-9626327); $10,000; 12 months. (Joint support with the Computational Mathematics Program and the Algebra & Number Theory Program)

In the past few years a number of significant advances have taken place in several different branches of computational algebra and number theory. Because of the growing interdependence between the different branches, an algorithmic advance in one area often has a major impact on the performance of algorithms in other areas. Consequently, this meeting brought together workers from some of the key areas, including researchers working on software development as well as those using these tools. The conference served to inform algorithm designers and users of recent developments in cognate areas of computational algebra, number theory, and geometry and inform key theoretical mathematicians of the possibilities offered by the computational algebra tools that are now becoming available. It also helped identify desirable algorithm and software developments that would result in significant new applications in the near future.

Butler University; Jonathan P Sorenson; *Research at Undergraduate Institution (RUI): Design and Analysis of Number Theoretic Algorithms*; (CCR-9626877); $56,150; 24 months. (Joint support with the Computational Mathematics Program)

This project investigates the following problems: (1) Determine explicit bounds involving primitive roots and Artin's conjecture that are useful in the analysis of several number-theoretic algorithms under the assumption of the extended Riemann hypothesis. (2) Study two problems involving GCD algorithms. The first is to design and implement genetic algorithms for the extended GCD problem. The second is to analyze the left- and right-shift k-ary GCD algorithms, including determining average-case behavior and studying the so-called "spurious factors" introduce during the main loops of the algorithms. (3) Design, analyze, and implement algorithms for computing the Jacobi symbol based on Lehmer's Euclidean GCD algorithm. Undergraduate students will be involved in this research, in particular in the design and implementation of algorithms in the last two problems above.

University of New Mexico; Stanly L Steinberg; *Symbolic Computing and Dimensionally Reduced Models of Fluid Flow*; (CCR-9531828); $41,778; 36 months. (Joint support with the Eastern Europe Program Program)

This project develops new methods for numerically modeling incompressible fluid flows of the type that appear in global environmental modeling problems. The models are hyperbolic and dispersive vertically-averaged models of stratified incompressible flow which are derived from the Navier-Stokes equations. These models reduce three-dimensional models to a system of complex two-dimensional models, which promise greater numerical efficiency. However, the averaged models are not well posed for some interesting ranges of their parameters. Consequently, their instability and regularization must be investigated. This analysis, along with the discretization of the continuum equations, the accuracy and stability analysis of the discrete problems, and the coding of the discrete equations are all algebraically complex, so the use of computer algebra will play a critical role in this project. The results of the project can be used for the global modeling of the ocean, the atmosphere, and other complex environments. The current codes implementing detailed models of such environments require computational resources that tax even the largest computers. The goal of this project is to produce new models and numerical methods for their solution which would represent a significant improvement in efficiency and accuracy over existing models for some important applications.

Cornell University; Michael E Stillman; ***A Software System for Algebraic Geometry Research*****; (DMS-9623232); $49,290; 12 months. (Joint support with the Computational Mathematics Program and the Algebra & Number Theory Program)**

This awards supports continued development of Macaulay 2, a system to enable large scale computations dealing with systems of polynomial equations in many variables. The project includes work on the relationship of syzygies of a projective variety to the geometry of the variety. This research is in the field of algebraic geometry, one of the oldest parts of modern mathematics, but one which blossomed to the point where it has, in the past 10 years, solved problems that have stood for centuries. Originally, it treated figures defined in the plane by the simplest of equations, namely polynomials. Today, the field uses methods not only from algebra, but also from analysis and topology, and conversely it is extensively used in those fields. Moreover it has proved itself useful in fields as diverse as physics, theoretical computer science, cryptography, coding theory and robotics.

University of Illinois – Urbana; Alexander F Vakakis and Cetin Cetinkaya; ***Research Experiences for Undergraduates (REU) Site: Undergraduate Symbolic Computations in Engineering and Science*** **; (CDA-9531571); $48,900; 12 months. (Joint support with the CISE Institutional Infrastructure Program)**

This Research Experiences for Undergraduates (REU) project is supporting ten students per year in programs carried out during the summer, for three years. The project promotes the emerging field of symbolic computational science as an approach to computational modeling and scientific computing with problems drawn primarily from mechanical and industrial engineering. Specific goals of the project are (a) to introduce the students to modern computational and symbolic methodologies for solving challenging problems, (b) to increase understanding of the qualitative and quantitative nature of these problems, (c) to develop the students' ability to derive and interpret mathematical models, and (d) to provide an integrated educational experience to a selected group of undergraduate students with special emphasis on underrepresented groups.

Rutgers University; Wolmer V Vasconcelos; ***Studies in Commutative Algebra and Its Applications to Computational Algebra*****; (DMS-9500786 A001); $60,000; 12 months. (Joint support with the Computational Mathematics Program and the Algebra & Number Theory Program)**

This project concerns the study of Cohen-Macaulay structures, in algebra, combinatorics, and geometry, and is driven by their wide presence and the opportunities offered in issues of computational complexity. The project investigates invariants of blowup algebras for their role in deciding the property of Cohen-Macaulay; contributes to the ongoing process of algebraization of tangent and secant algebras; studies the role of various measures of complexity in the processing of basic operations of computational algebra; and develops methods and algorithms that facilitate the efficient processing in large scale computations in geometry. This project is in the general area of computational algebra and the computer-aided computational aspects of this field. There is a growing interest in using computers to answer theoretical questions in algebra and conversely, algebra is quite useful for the development of algorithms. This research will not only aid in the development of computer programs in commutative algebra but will influence the development of computer programs in other areas. One aspect of this project concerns the development of a computer program which will solve systems of polynomial equations. This is significant in manufacturing both for modeling and for the development of robots.

Kent State University; Paul S Wang; ***Efficient Algorithms and System Interface for Scientific Computation*****; (CCR-9503650 A001); $14,924; 6 months. (Joint support with the Theory of Computing Program)**

This project focuses on three related areas: (1) No-missing-term p-adic lifting — Establishing a faster algorithm for recovering final results in multivariate polynomial operations such as GCD (Greatest Common Divisor) and factoring. The lifting method is most effective when the polynomials involved are sparse, containing just a few of all possible terms. (2) Function Evaluation Optimization and Applications — Speeding up the repeated evaluation of functions in closed-form formulas through the application of recurrence relations. Nested recurrences called "chain of recurrences" will be derived by symbolic computation and applied to increase the evaluation speed of formulas at regular intervals. Application in interactive graphical visualization will be made. (3) Interfacing Distributed Scientific Systems — Seeking to establish an efficient and standardized protocol to connect different autonomous systems for scientific computation. Such a protocol is necessary to build distributed "problem solving environments". By exploiting sparseness and solving systems of linear equations, the p-adic lifting procedure of item (1) will be investigated. For area (2), automatic generation of chain of recurrences (CR) for functions of one or more variables will be implemented, and tested for speed and effectiveness. The numeric properties of CRs will be studied. The technique will be applied to speed up interactive graphical visualization of mathematical functions. For area (3), an efficient and protocol (called MP) will be designed. Libraries in C will support the protocol.

Temple University; Doron Zeilberger; *Combinatorics, Special Functions, and Computer Algebra*; (DMS-9500646 A001); $40,000; 12 months. (Joint support with the Computational Mathematics Program and the Algebra & Number Theory Program)

This project investigates problems in the areas of algebraic combinatorics, the theory of special functions, and computer algebra. One direction of this project is a continuation of the research program on the use of computer algebra for computerized proofs of identities in combinatorics and the theory of special functions. Other goals of this project include the development of a general theory of permutation statistics, the study of holonomic aspects of combinatorial statistical mechanics, and the extension of methods developed in the principal investigator's recent proof of the alternating sign matrix conjecture to other, still open, problems in this area. Because both special functions and combinatorics are so prevalent in the natural sciences, this research has many potential applications to both science and technology, as well as to pure mathematics. This research is in the general area of combinatorics. Combinatorics attempts to find efficient methods to study how discrete collections of objects can be arranged. The behavior of discrete systems is extremely important to modern communications. For example, the design of large networks, such as those occurring in telephone systems, and the design of algorithms in computer science deal with discrete sets of objects, and this makes use of combinatorial research.

Computational Geometry

Duke University; Pankaj Agarwal; *NSF Young Investigator (NYI) Award: Geometric Algorithms and Their Applications*; (CCR-9357814 A004); $50,000; 12 months.

Computational geometry has provided efficient algorithms for geometric problems in several applied areas including graphics, robotics, CAD, and geographic data processing. The project continues studying these problems and explores new areas where computational geometry can play an important role. Work on problems involving arrangements of surfaces, dynamic data structures, geometric optimization, interaction detection, and multi-dimensional searching is pursued. There is a special emphasis on higher dimensional problems.

Texas A&M University; Nancy M Amato; *CAREER: Building and Searching Data Structures for Spatial Environments*; (CCR-9624315); $100,000; 24 months.

In recent years geometric modeling and simulation techniques have become essential tools for applications involving physical objects. Although much progress has been made, there are important applications that cannot be realized with current technology. Examples include controlling autonomous robots operating in dynamic situations, training an employee for a dangerous task in a virtual reality environment, and helping an engineer interactively analyze design choices using a three-dimensional model of a complex product. The general objective of this research is to design and implement practical algorithms for such applications. Toward this goal, the applicability of techniques developed in the field of computational geometry to the following target applications will be studied: performing robotic path planning and planning manipulation tasks with contact, and analyzing product designs for manufacturability, assembly, and disassembly. The research will be performed in collaboration with an ongoing research effort at Texas A&M in dexterous manipulation and with researchers at GE R&D analyzing product designs for maintainability.

Two educational activities will be undertaken. First, mentoring and general support programs will be initiated for undergraduate and graduate women in the Computer Science Department at Texas A&M. The goal of these programs is to encourage more women to pursue graduate studies and careers in computer science and engineering. Second, group learning methods will be incorporated into undergraduate and graduate courses in theoretical computer science. The change from the standard lecture format is motivated by research that has shown that group learning can increase student interest and help them to understand concepts more fully.

University of Notre Dame; Danny Ziyi Chen; *CAREER: Theoretical and Practical Solutions for Geometric Path Planning and Related Problems*; (CCR-9623585); $100,000; 24 months.

This project will study the design, analysis, and implementation of algorithmic techniques for solving geometric shortest path problems, their generalizations, and applications. Of interest are not only theoretically efficient algorithms, but also practically efficient ones. The research has three main interests: (a) Develop efficient algorithmic solutions for several fundamental geometric shortest path problems that are still outstanding (e.g., geometric shortest path queries), (b) investigate new approaches to computing approximate geometric shortest paths, and (c) design effective and practically efficient paradigms for planning robotic shortest paths in the plane and in higher dimensional spaces. Several general frameworks for processing exact and approximate geometric shortest path queries will be investigated. These frameworks offer the promise of achieving new efficient algorithmic techniques and data structures for geometric shortest path queries and for other related problems. Also, a paradigm is being studied for obtaining practical solutions to planning shortest obstacle-avoiding paths for robot motion in planar and higher dimensional environments.

This paradigm is based on new data structures called framed-quadtrees and framed-octrees, and has led to new practical robotic path planning algorithms. In addition, algorithmic solutions for shortest paths and their generalizations are applied to practical applications such as data compression, computer vision, image processing, and VLSI design. New approaches for solving application problems based on algorithms for shortest path problems and their generalizations are studied. The possibility of finding new methods for geometric shortest paths that yield efficient implementation performance on existing Massively Parallel Processing (MPP) systems is also explored. The research also includes an important experimental component.

The education plan is to develop a new environment for teaching and experimenting with Massively Parallel Processing (MPP) systems by utilizing the new EXECUBE-based MPP architectures. The goal is to develop an inexpensive, but very robust, MPP system for upper level undergraduate and entry level graduate students to study and gain "hands on" experience with parallelism. The curriculum materials to be developed include concise programming tutorials, lecture notes, sample programs, and projects with sample solutions. This work could provide an integral part of not just electives in parallelism, but virtually of all the upper level computer science and engineering curriculum, and could even provide a basis to spill over into other engineering and scientific disciplines.

Princeton University; David P Dobkin and Bernard Chazelle; *Design and Experimentation in Geometric Computing*; (CCR-9623768 , A001); $241,045; 24 months.

The project is to pursue ongoing investigations into core areas of computational geometry, in particular, cell decomposition, mesh generation, multidimensional searching, and geometric optimization. The research spans the spectrum from the theoretical (e.g., lower bounds, geometric sampling techniques, derandomization) to the practical (implementation and distribution of geometric software). Progress in the field has been hampered by the difficulty in testing out new geometric ideas on a computer. As a result, experimentation has often taken a backseat to theoretical research. Essential to this project is the building of software tools (e.g., algorithm animation, visual debugging, test data generation) that facilitate and encourage the use of experimentation.

University of Michigan – Ann Arbor; Rida Farouki; *Computational Methods for Pythagorean-hodograph Curves in CAD/CAM Applications*; (CCR-9530741); $121,030; 24 months.

Pythagorean-hodograph (PH) curves offer many important computational advantages over the "ordinary" Bezier/B-spline representations commonly used in CAD/CAM applications. Among the attributes that distinguish PH curves from the latter are: exact rectification in terms of a polynomial arc-length function; exact rational representations of their offsets; a closed-form expression, suitable for purposes of optimization, for their elastic bending energies; and a propensity to yield fairer loci — exhibiting a smoother curve distribution — in the interpolation of discrete data. Moreover, existing CAD systems can readily take advantage of these properties, since the PH curves are entirely compatible with the Bezier/B-spline representation. This project investigates outstanding theoretical and algorithmic issues concerning the construction and manipulation of PH curves. Their solution will permit the full potential of PH curves to be realized in applications such as rapid prototyping and NC milling. These issues encompass trimming procedures for PH curve offsets; real-time CNC interpolators for PH curves; optimal shape design with PH curves; algorithms for efficient construction of C2 PH splines; and alternate algebraic formulations for PH space curves.

City University of New York – City College; Jacob E Goodman; *Conference on Discrete and Computational Geometry July 13-19, 1996, Mount Holyoke College, South Hadley, MA.*; (CCR-9530506); $8,000; 6 months.

This meeting brought together researchers in computational and discrete geometry. The purpose of the meeting was to foster interaction, centered on geometric questions, between computer scientists and mathematicians. Such interaction started mainly by a similar conference held in Santa Cruz in 1986. During the ten years since that meeting, researchers in each area have made substantial contribution to the other. Combinatorial and discrete geometers have gotten more and more interested in algorithmic and complexity-related questions, while the computational geometers have used their techniques successfully to make inroads into problems that had previously seemed intractable to mathematicians. Both fields have grown as a result of this interaction. The new meeting surveyed much of the progress made during the past decade, and delineated the areas in which progress is likely to take place during the coming years.

Johns Hopkins University; Michael T Goodrich; *Application-Motivated Geometric Algorithm Design*; (CCR-9625289); $107,398; 24 months.

This project investigates basic research in computational geometry that is well-motivated by applications in geometric computing. Specific focus is given to the construction and maintenance of geometric structures, as well as methods for efficiently searching in such structures, with principle motivating applications coming from computational metrology, computational graph drawing, and geometric information retrieval. Emphasis is placed upon the design of algorithms for single processors as well as methods for exploiting coarse- grain parallelism such as exists in systems with parallel disk arrays or clusters of workstations.

Stanford University; Leonidas Guibas; ***Computational Geometry in the Physical World*****; (CCR-9623851); $123,693; 24 months.**

As the application of computers to the physical world widens, there is an ever increasing need for computational tools to deal with the physical attributes of objects, such as their geometry. The field of Computational Geometry has made impressive strides in the last twenty years in providing an algorithmic foundation for the solution to many geometric problems in computer graphics, computer vision, robotics and manufacturing, and many other application areas. The purpose of this project is to develop a set of novel geometric representations and algorithmic techniques that will make computational geometry more useful in modeling physical-world objects and processes. The aim is not to obtain universal, domain-independent solutions — such as a complete solution to robustness concerns in geometric computing. Instead the focus will be on certain key issues, such as scale-sensitivity, shape variability, and motion, that arise in many of the important uses of geometry in modeling physical objects and systems. Geometric models that properly capture certain key aspects of the uncertainty or variability in models of physical objects and processes will be investigated, and efficient algorithms for manipulating these models effectively will be developed. It is expected that this set of research directions will generate new mathematical and algorithmic problems that will be interesting and worthy of study in their own right, in addition to producing geometric tools that are well-tailored to physical world applications.

Southwest Texas State University; Carol Hazlewood; ***A Prototype Object-Oriented Geometry Software Library*****; (CCR-9623854); $13,278; 12 months.**

A user-callable geometry library in C++ is being developed. The software package will include documented C++ libraries, a user manual, sample drivers, and sample data sets. It will be distributed on the Internet and diskettes, and announced through conferences and publications. Implementation in C++ will make the software accessible to the computational geometry community and to an increasing number of programmers in scientific computing. The library can advance science and engineering by reducing software development time for its users. The library is small enough to be completed quickly, yet provide a basis for future expansion. The project will encompass the four layers identified for geometry libraries, namely, geometric primitives, algorithms, support environment, and applications. The support environment will include sample drivers, sample data, and an interface to Geomview, the visualization software from the Geometry Center at the University of Minnesota.

Purdue University; Christoph Hoffmann and Elisha Sacks; ***Spatial Geometric Constraint Solving Using Constructive Geometry and Homotopy Continuation*****; (CCR-9505745 A001, A002); $115,250; 12 months.**

The project develops the theory and implementation of a practical spatial geometric constraint solver that integrates symbolic methods from constructive geometry with numerical methods from homotopy continuation theory. Geometric constraint solvers are a key component in modern Mechanical Computer-Aided Design systems, and offer a critical technology for increased productivity, rapid response manufacturing, and agile manufacturing. Effective solvers exist for planar constraints, but spatial problems of moderate size exceed the capabilities of many research solvers. The solver being developed will cover the broad class of problems specified by distance, angle, and incidence constraints among points, lines, and planes. It will solve small problems interactively and problems of moderate size in a few seconds. Using a constraint graph whose nodes represent elements and whose edges represent constraints, a constraint analysis will decompose the constraint problem into a series of small construction problems that can be solved efficiently using preprocessing. Elementary constructions will be carried out algebraically, more complex construction steps by homotopy continuation. The hybrid algorithm combines the robustness and completeness of symbolic methods with the uniformity and efficiency of numerical methods.

University of North Carolina – Chapel Hill; Dinesh Manocha; ***ACM Workshop on Applied Computational Geometry, May 27-28 1996, Philadelphia, PA.*****; (CCR-9622493); $8,000; 6 months.**

This workshop was mainly designed to bring together practitioners and theoreticians in computational geometry and other related applied fields to discuss the current research and future directions in these communities, to facilitate technology transfer from theory to practice, and to identity future joint programs and collaboration.

University of North Carolina – Chapel Hill; Dinesh Manocha; ***CAREER: Algebraic and Geometric Techniques for Interface Detection in Static and Dynamic Environments*****; (CCR-9625217); $99,962; 24 months.**

The problems of interference detection are fundamental in computational geometry, solid modeling, robotics, manufacturing, computer graphics and computer-simulated environments. These include collision detection between geometric models undergoing motion, surface intersection between linear and curved surfaces, collision avoidance for robot motion planning, visibility computations, hidden surface removal and boundary computations. This research takes a three-fold approach to these problems using algebraic, numeric and geometric techniques: 1. Develop fast and robust algorithms for interference detection between geometric models in static and dynamic environments. The set of models include polyhedra, splines, algebraic surfaces and deformable objects. 2. Specialize these algorithms to interactive collision detection, hidden surface removal, surface intersection and boundary computation with special emphasis on robustness. 3. Develop efficient software libraries for interference detection between moving polyhedral objects, solid models defined using algebraic surfaces and boundary evaluation. The resulting libraries will be distributed as public domain code and demonstrated on different applications.

The research efforts will be complemented with the following educational goals: 1. Development of undergraduate and graduate courses on geometric, symbolic and scientific computing. 2. Develop novel teaching methods and electronic course notes using the World-Wide Web (WWW) technology. 3. Use of graphics, geometry and animation tools to develop an algorithm animation and programming laboratory for undergraduate courses. 4. Serving as a mentor to high school and undergraduate students to work on graphics and modeling related projects in our laboratory.

State University of New York – Stony Brook; Joseph S B Mitchell and Esther M Arkin; *Algorithmic Studies in Applied Geometry*; (CCR-9504192 A001); $100,000; 12 months.

This investigation has three major objectives: (1) to add new algorithms and methodologies to the repertoire of geometry techniques, including new approximation methods and algorithms that are simpler and easier to implement; (2) to formulate precise algorithmic versions of a large body of geometric problems from allied fields, in such a way that the results of computational geometry can be applied successfully and so that the community of largely theoretical computational geometers can be properly exposed to the more practical applications (allied fields include cartography, computer vision, design, graphics, manufacturing, etc.); and (3) to disseminate methodologies of computational geometry to applied researchers who work in allied fields but who have been largely unaware of progress in computational geometry, particularly since it has been construed as a primarily theoretical instrument. The third objective requires discussions of research directions with contacts in research labs and through publishing papers for conferences and in journals whose readership includes a breadth of engineering disciplines.

Smith College; Joseph O'Rourke and Ileana Streinu; *Research at Undergraduate Institution (RUI): Visibility, 3D, and Other Topics in Computational Geometry*; (CCR-9421670 A001); $53,665; 12 months.

The focus of computational geometry is designing for solving geometric problems. These arise in a number of contexts, such as graphics, robotics, medical imaging, and manufacturing. The project's concentrations touch on all these application areas. A number of visibility questions are being examined, including covering regions with floodlights, and the structure of visibility graphs. Several of the problems being explored involve three dimensions: reconstruction from parallel slices (and its application to graphics morphing), computing diameters and centerpoints of sets (with applications to pattern recognition), and motion planning configuration spaces (application to robotics). Fundamental questions in the theory of arrangements are being addressed. Several problems arising from the manufacturing community are being tackled, including that of scheduling a punching tool to create a hole in a sheet of metal of a prescribed shape. Female undergraduate students are engaged in all aspects of the research, and several useful software packages are being produced and distributed.

Purdue University; Jorg Peters; *NSF Young Investigator (NYI) Award: Surface Splines Over Irregular Meshes*; (CCR-9457806 A003); $62,500; 12 months.

Models of die casts in manufacturing, level sets of functions, deposition fronts in simulation, pleasing shapes in animation, fast prototyping, and visualization of interactions in a virtual world have one thing in common: the need for efficient representations of smooth surfaces. At present, some of our understanding and use of scientific results is handicapped by the fact that the standard representation used for the task, tensor-product B-splines, can only model a small subset of these surfaces naturally and efficiently, namely surfaces that allow for a regular, checker-board arrangement of its pieces. Free-form surface splines extend the spline paradigm and techniques to cover also irregular arrangements. In extending, they preserve the intuitive generation of the surface from the control mesh by a process of cutting with hyperplanes, also known as corner cutting or subdivision. So far only first-order, tangent continuous surface splines exist. Extension to higher-order continuity is both of fundamental and practical interest. The fundamental aspect consists of a new theory that arises when surfaces from pieces using massive computing power in contrast to analyzing surfaces as in classical Gaussian differential geometry. The goal of this research is to develop a general framework for surface splines and to use the representation to optimize the shape of surfaces with respect to various cost functions. The immediate goal is the derivation of surface splines and to find tools to adjust the variation of curvature, an important quality criterion for surfaces. Since the challenge is considerably greater than the definition of first-order surface splines over irregular meshes, progress requires heavy use of symbolic computation and graphics for quality assessment.

New York University; Richard Pollack, Janos Pach, and Micha Sharir; ***Studies of Geometric Algorithms and Their Applications*****; (CCR-9424398 A001); $112,158; 12 months.**

The relatively new field of computational geometry has developed at an accelerating rate in the past few years. The problems that it encompasses cover a wide range of applications, such as motion planning and computer graphics, and their solution often involves the use of sophisticated techniques drawn from many branches of mathematics and theoretical computer science. Prior work includes extensive contributions to many basic and applied problems in the area, including motion planning, hidden surface removal and related visibility problems, combinatorial and algebraic analysis of arrangements of curves and algebraic surfaces, randomized algorithms for linear programming and other optimization problems, etc. The purpose of this research is to continue the work in these fields and to extend it into new core areas in computational geometry. A major portion of the research is devoted to the study of arrangements of curves and surfaces. Specifically, the following problems are being studied: 1. Combinatorial and algorithmic problems related to substructures (lower envelopes, single cells, zones, etc.) in arrangements of surfaces in higher dimensions. 2. Related algorithms in real algebraic geometry for computing cells and road maps in semi-algebraic sets. 3. Combinatorial and algorithmic problems involving planar arrangements of segments or curves (also known as "geometric graphs". A theme that is manifest in this work is the rich cross-fertilization between basic research in computational geometry and the various application areas, where problems in one area motivate the study of new basic problems whose solution in turn finds applications in many other areas. Another theme is the strong connection between the combinatorial analysis of substructures in arrangements and the design of corresponding efficient algorithms for constructing and utilizing these structures. Often the efficiency of such an algorithm crucially depends on the size of the structure to be computed, and most of the algorithm design has to be devoted to the combinatorial analysis of the complexity of that structure. These considerations are reflected in the work being done.

Lehigh University; Eunice E Santos; ***CAREER: Determining Parallel Complexity of Numerical Computation Problems via Dependency Graphs*****; (CCR-9624721); $100,000; 24 months. (Joint support with the Theory of Computing Program)**

The research objective of this project is to determine the time complexity of solving various numerical computation problems on parallel distributed-memory machines in general by abstracting the common properties of these problems via dependency graphs. To that effect, two important topics are considered: (1) the design and analysis of parallel numerical algorithms and (2) the derivation of lower bounds on complexity. Working within LogP, a recently-introduced model for parallel computation, parallel complexity will be analyzed by taking into account three important factors: (i) initial data placement, (ii) communication scheduling, and (iii) local operations. Algorithms which are efficient on LogP are portable from one parallel machine to another and will be efficient on any parallel machine. Initially several different numerical computation problems will be considered and the parallel complexity for each specific problem will be determined. The overall results will be used to help extract the important criteria of dependency graphs and thereby classify numerical computation problems via these criteria. Furthermore, the parallel complexity for each classification of numerical computation problems will be determined. The importance of the results obtained lies in the ability to pinpoint what types of data placements and communication schedulings are needed in order to achieve efficient or optimal running times for large classes of numerical problems. Clearly, this research potentially impacts a wide variety of applications and computational problems from various domains which are currently difficult to address.

In order to successfully accomplish this research and apply it to these domains, there must be education and training of people in the parallel and numerical computing field. To that effect, a program in parallel and numerical computing will be developed beginning with course preparation at the undergraduate level and culminating in a research program at the doctoral level.

Brigham Young University; Thomas W Sederberg; ***Implicitization of Rational Surfaces with Base Points*****; (CCR-9622768); $60,000; 12 months. (Joint support with the Computer Systems Architecture Program)**

The objective of this project is to investigate a new technique for finding the implicit equation of a rational surface. Standard implicitization techniques for rational surfaces are based on resultants. But resultants vanish in the presence of base points. Therefore, when base points are present, the implicit equation can be recovered using resultants only by carefully perturbing the parametric equations to eliminate the base points. But such perturbations generally introduce extraneous factors, which then need to be painstakingly removed. Thus base points seriously complicate implicitization using resultants. On the other hand, base points simplify the implicit equation of a rational surface by lowering its degree. Preliminary work has roughed out a new implicitization technique based on algebraic "moving surfaces" that "follow" a rational surface. Not only is this new implicitization technique impervious to base points, it also simplifies in the presence of base points. Moreover this new method generates a more compact matrix representation for the implicit equation than methods based on resultants. Currently this new method is substantiated only by several Mathematica examples and some preliminary theorems. It is planned to develop the theoretical foundations of the new method and then to apply the method to solve a variety of practical problems in computer graphics, geometric modeling, and symbolic computation. A wide variety of additional problems in elimination theory might also be attacked by these new techniques including, but not limited to, inverting a rational map, computing the singular locus of a rational surface, intersecting rational curves and surfaces, and intersecting algebraic curves and surfaces.

University of North Texas; Farhad Shahrokhi; *Crossing Number Problems in Geometric Drawings of Graphs*; (CCR-9528228); $48,098; 24 months.

Geometric drawings give rise to many interesting and important problems which are the extensions, generalizations, or the dual forms of the well known planar crossing problem. A k-crossing in a geometric drawing is defined to be a set of k edges which pairwise cross in the drawing. An interesting problem here is to generate geometric drawings with small number of k-crossings, for a given graph and a given k>1. Another important problem is the crossing clique problem, in which the objective is to generate geometric drawings so that the largest clique of crossings (the largest k so that the drawing has a k-crossing) is small. This problem is closely related to the page number problem and has nice applications to the theory of VLSI. The dual concept to a k-crossing is a k-disjoint set: Given a geometric drawing of the G, a k-disjoint set is a set of k edges which are pairwise non-crossing in the drawing. The concept of k-disjoint sets gives rise to problems which are similar to those of k-crossings, in a dual setting. It is the goal of this investigation to study k-crossings, k-disjoint sets, and the crossing clique problem. The aim is to develop algorithms which can generate solutions whose quality can be guaranteed and/or to develop general structural results in a form which could be applicable to a broad class of graphs.

New York University; Micha Sharir and Jacob T Schwartz; *Implementation and Experimentation with New Geometric Algorithms for Motion Planning*; (CCR-9311127 A002); $84,111; 12 months.

Exact combinatorial solutions to the motion planning problem in robotics, which use precise computational geometry techniques, especially for systems with few degrees of freedom, tend to be more robust than heuristic or approximating solutions; they can achieve performance that is theoretically faster than that of naive solutions by an order of magnitude or more, and it is therefore desirable to test their pragmatic feasibility in real systems context, both to complete the somewhat abstract algorithm variants typical of theoretical developments in computational geometry, and to show that they indeed perform well in practice. This project undertakes the development of an 'industrial-strength' motion planning system for bodies translating in 3-dimensional polyhedral environments. This system utilizes efficient and precise geometric algorithms obtained through ongoing theoretical work, whose worst-case complexity attains the best theoretical bounds known for the problem, and also turn out to be very efficient in practice.

Washington University; Subhash Suri; *Approximation Algorithms in Computational Geometry*; (CCR-9501494 A001); $35,203; 12 months.

This project investigates the topic of approximation algorithms in computational geometry, with focus on the two problems: polyhedral surface modeling and Euclidean shortest paths. In surface modeling, one tries to construct a polyhedral model of a complex surface, based on a set of unordered sample points. The complexity of constructing and manipulating the model depends on its number of vertices and faces, and therefore a natural optimization goal is to minimize these features in the model without exceeding a pre-specified error tolerance. The project studies various surface-modeling problems within the formal framework of computational geometry, and analyzes the issues of computational complexity. The problem of computing shortest paths is central to robotics and autonomous navigation. The research investigates approximation algorithms for shortest path problems in two and three dimensions. In two dimensions, the query version of the shortest path problem is studied: preprocess a set of polygonal obstacles so that given two points p and q, the distance between them can be determined in polylogarithmic time. In three dimensions, the goal is to design simpler and more practical alternatives to the exact algorithms for shortest paths among polyhedral solids or on a polyhedral surface.

Optimization and Mathematical Programming

Carnegie Mellon University; Egon Balas and Gerard Cornuejols; *Polyhedral Methods in Integer and Combinatorial Optimization*; (DMI-9424348 A001); $132,001; 12 months. (Joint support with the Operation Research Program, the Computational Mathematics Program, and the Applied Mathematics Program)

This project continues earlier research in integer and combinatorial optimization problems. The earlier research developed a new approach to solving mixed integer programming problems based on a framework described as 'lift-and-project cuts' within the context of the branch and bound technique. The procedure has been founde to be robust and more efficient than any of the approaches currently used in solving mixed integer programs. The current research is aimed at investigating some other unresolved questions associated with the improved technique. Included in the questions to be resolved are matters related to the characterization of the computational behavior of the algorithm, development of some other new algorithms, and the design and implementation of these algorithms to solve a broader class of problems. The lift-and-project cut approach to mixed integer programming has applications to several areas of design, manufacturing, distribution, facility location, and civil infrastructure development. One such application is the optimal design of multi-product batch plants in chemical engineering, where the optimization is concerned with the choice of unit plants sizes for various stages of the process and the batch sizes for the products. Another example is the design of a cost-effective fiber optics network that protects services against the consequences of equipment failure. In the area of civil infrastructure development, one example is the design of a minimum weight truss consisting of a number of specified bars with fixed modal locations, under specified conditions on stresses, bar elongations, and nodal displacement. Another class of problems studied in this research is the generalized traveling salesman problem where the development technique can be used to construct optimal rolling cycles i n a hot strip mill. This research has the potential to significantly enhance the advancement and the understanding of integer and combinatorial optimization problems. Furthermore, several sectors of the economy stand to benefit from the results.

Massachusetts Institute of Technology; Dimitri P Bertsekas; *Laboratory for Information and Decision Systems*; (DMI-9300494 A002, A003); $92,500; 12 months. (Joint support with the Operation Research Program)

Many operations in design, production planning, and resource management encounter optimization problems, that is, mathematical problems of choosing optimally several factors that affect a productivity index. These are problems of considerable computational difficulty that have been extensively investigated. This research focuses on several such problems that arise in practice and have a large scale structure, such as a network or a dynamic system structure. Methodological improvements for the solution of these problems has significant practical impact. The research aims at the understanding of the performance of existing network optimization, linear programming, and nonlinear programming algorithms, and develops algorithms with improved performance for both a serial and a parallel computing environment.

Rice University; Robert E Bixby; *Investigations in Mixed Integer Programming*; (CCR-9407142 A001); $70,000; 12 months.

While recent years have seen remarkable progress in programming, both theoretically and practically, most of that progress has come in the solution of pure 0/1 problems. On the other hand, mixed problems, problems with both continuous and discrete variables—usually 0/1 variables—dominate in practice; moreover, practitioners regularly produce instances of these models that cannot be solved using currently available codes. This combination of success in solving pure 0/1 problems and the existing demand for methods to solve difficult real-world mixed problems has received considerable attention in the research community. Nevertheless, general-purpose mixed-integer solvers remain primarily solvers, and while they often employ rather sophisticated "heuristic" devices, they still use few of the algorithmic ideas implicit in recent theoretical developments. This research will attempt to change that situation by developing a research code including both the standard methods that have been successful in practice, as well as incorporating ideas based upon recent theoretical developments. The focus will be on actually solving real-world instances that arise as general mixed integer programs but cannot currently be solved.

University of Minnesota – Twin Cities; Shantanu Dutt and Vipin Kumar; *US-France Cooperative Research: Highly Parallel Branch-and-Bound Algorithms for Solving Optimization Problems (with INRIA)*; (INT-9512014); $27,000; 36 months. (Joint support with the Western Europe Program Program)

The objectives of this US-France cooperative research are to develop highly parallel branch-and-bound algorithms for solving combinatorial optimization problems (COPs) and software for optimization computations on parallel computers. Branch-and-bound algorithms are used widely in solving combinatorial optimization problems that appear in a variety of applications like computer vision, robot motion planning, and integer programming and computer- aided design of very large scale integrated circuits. These problems take a long time to solve. The US investigators will develop time and memory efficient techniques for solving these problems using parallel processing. The US investigators bring to this research their expertise in parallel computing and techniques for solving COPs using distributed memory machines and workstation clusters. This is complemented by the French investigators' expertise in operations research and the work on branch-and-bound techniques suitable for shared-memory multiprocessors.

University of Wisconsin – Madison; Michael Ferris; ***Presidential Young Investigator (PYI) Award: Development of Efficient Serial and Parallel Algorithms Based on Mathematical Programming Theory*****; (CCR-9157632 A005); $37,500; 12 months.**

Efficient serial and parallel algorithms, backed by a mathematical programming theory, are developed. Emphasis is placed on effective distribution of constraints among processors as well as line-search stabilization techniques, nonlinear complementarity and other problems. Finite termination of algorithms via weak sharp minima and related concepts are being investigated.

University of Minnesota – Twin Cities; Vipin Kumar; ***Highly Parallel Direct Solvers for Sparse Linear Systems*****; (CCR-9423082 A001); $60,555; 12 months.**

Direct methods for solving sparse linear systems are important because of their generality and robustness. For linear systems arising in certain applications, such as linear programming and some engineering applications, they are the only feasible methods for numerical factorization. This project is developing a scalable parallel direct solver for solving sparse symmetric-positive definite (SPD) systems of linear equations, and is investigating its utility and effectiveness in the context of linear programming problems and least square problems. A highly parallel formulation of the factorization phase of the direct solver for SPD systems has been developed. This is being extended to the other three phases (ordering, symbolic factorization and triangular system solving) of a complete scalable parallel direct solver. Parallel formulations are being investigated on a variety of hardware platforms, including traditional MPPs such as CM5 and Cray T3D, as well as clusters of DEC and IBM workstations. The parallel algorithms are being implemented in a portable form using a programming library that supports collective communication, and will be made available to researchers and practitioners elsewhere.

Yale University; Laszlo Lovasz; ***Algorithms in Convex Geometry and their Applications in Combinatorial Optimization*****; (CCR-9402916 A002); $79,697; 12 months. (Joint support with the Theory of Computing Program and the Geometric Analysis Program)**

This project investigates the problems of volume computation, sampling from geometric domains, integer programming, and some basic, but not yet well understood, problems such as containment between convex bodies. The goals are to develop new algorithms, to test implementations and, parallel to this, to explore theoretical bounds on the efficiency of algorithms. The main tool in the design of these algorithms is randomization, in particular random walks and Markov chains.

University of Wisconsin – Madison; Olvi L Mangasarian; ***Applications, Algorithms and Theory of Mathematical Programming*****; (CCR-9322479 A002); $156,270; 24 months.**

Highly promising mathematical programming techniques applied to medical diagnosis and prognosis as well as to machine learning in general. Five years of experience with a highly accurate system for breast cancer diagnosis, in current use at University of Wisconsin Hospitals, is drawn upon to develop a linear-programming-based prognostic system. Parallel algorithms for the solution of large-scale constrained optimization problems are developed, by distributing ingredients of the problem (constraints, gradients, or/and variables) among parallel processors. Each processor has complete responsibility for varying its own problem ingredients, while allowing the other ingredients to vary in a restricted fashion. The processors share new information computed, then a fast synchronization is performed and the process is repeated. Preliminary algorithm prototypes have been tested successfully with some high parallelization efficiency on publicly available test problems and significant real-world applications. Error bounds are developed for possibly inconsistent systems of inequalities, programs and complementarity problems. The novel idea here is that the system may be unsolvable, and the bounds are meaningful whether the system is solvable or not. In the latter case, it bounds the distance between the point under consideration, and the set of least error solutions of the system. New mathematical programming approaches to some machine learning problems are studied. In particular a new quadratic programming model for a cascade architecture in neural networks is being studied, as is another one for the inherently difficult problem of minimizing the number of misclassified points by a separating plane.

Stanford University; Walter Murray and Michael Saunders; ***Large-Scale Constrained Optimization*****; (DMI-9500668 A001); $100,000; 12 months. (Joint support with the Operation Research Program)**

The objective of this research is to develop, study, and implement algorithms for solving large-scale, nonlinear optimization problems. A large-scale optimization problem is characterized by a large number of variables and constraints. Two solution approaches form the focus of the research: the interior point (barrier) method and the sequential quadratic programming (SQP) method. The interest in the interior point method stems from their success on linear programming problems, and some interesting properties that may be of value when certain difficult constraints are encountered. The interior point method can be used to eliminate constraints on the eigenvalues of a matrix. On the other hand, interest on the use of SQP is prompted by the success of the approach on small and medium-sized problems, and the experiences of the investigators on the performance of the technique on two classes of large, practical problems that differ substantially in character. If successful with these approaches, the developed algorithms will be coded into a software to facilitate their applications by practitioners. These are several engineering, manufacturing, scientific, and business problems that can be described by large-scale nonlinear optimization models. To date, finding a general solution algorithm to large scale optimization community. If successful, the algorithms being investigated in this research could open the window to a better understanding of this class of optimization problems. Many problems in various areas of human endeavor stand to benefit from the results of this work.

Northwestern University; Jorge Nocedal; ***Nonlinear Optimization*****; (CCR-9625613); $135,951; 24 months.**

Three research projects in nonlinear optimization will be undertaken; they range from theoretical studies of constrained optimization algorithms to the development of a network-based optimization system capable of solving problems automatically over the Internet. The first project is devoted to the design of limited memory approaches for large-scale nonlinear programming. Limited memory methods have proved to be quite useful for solving many classes of large unconstrained problems, but their economy and simplicity have not been fully exploited in the nonlinearly constrained case. The basic guiding principle in this work is to employ only algorithms with modest computational requirements and to represent limited memory matrices in compact form, so that the overall cost of the iteration is only a fraction of that required by the standard Sequential Quadratic Programming iteration. The new algorithms will allow the user to choose between direct and iterative methods for solving the linear subproblems arising in the iteration. The second project is devoted to the design and analysis of interior-point methods for nonlinearly constrained (non-convex) optimization problems. This is a long term research effort that first aims at identifying the crucial ingredients of the algorithms, such as the choice of model, the function, and the numerical techniques for solving the (ill-conditioned) subproblems. The new methods will be formulated in a trust region framework, which allows a uniform treatment of the convex and non-convex cases. The research will be guided by a combination of numerical testing and convergence analysis. An Internet-based optimization system called NEOS has recently been developed by a group of investigators at Northwestern and Argonne. One of the objectives of NEOS is to allow users to solve optimization problems remotely and with minimal effort using modern web browsers. The third project considers extending the capabilities of NEOS so that it can interact with modern modeling languages. A new constrained optimization server will also be developed.

New York University; Michael L Overton; ***Numerical Methods for Non-Smooth Optimization*****; (CCR-9625955); $170,000; 24 months. (Joint support with the Operation Research Program, the Computational Mathematics Program, and the Applied Mathematics Program)**

Nonsmooth optimization problems involve objectives or constraints which are not differentiable everywhere. The term structured nonsmooth optimization means that the specific nonsmooth nature of the problem is known. One especially interesting class repeatedly arises in many applications: problems involving eigenvalues of matrices. The work primarily focuses on various aspects of eigenvalue optimization, including semidefinite programming, quasiconvex eigenvalue optimization, and non-Lipschitz eigenvalue optimization. The goal is fourfold: the development of fast, robust numerical algorithms; analysis of theoretical questions concerning optimality conditions and algorithm convergence; development of software which can be used by the general scientific community; application to the solution of important interesting problems which arise in practice. Application areas of particular interest include combinatorial optimization, robust control and structural analysis.

Johns Hopkins University; Jong-Shi Pang; ***Computational Research on Complementarity Problems and Variational Inequalities*****; (CCR-9624018); $127,000; 24 months. (Joint support with the Operation Research Program and the Computational Mathematics Program)**

The continued growth and expanded applications in engineering, economics, and sciences have greatly increased the need for advanced computational methods for solving some highly complex nonlinear complementary problems and variational inequalities. This investigation focuses on several major aspects: (a) foundational—convergence properties and posteriori error analysis of new and existing solution methods, (b) numerical—computer implementation and testing and software development of new algorithms, and (c) practical—the efficient solution of important classes of engineering and economics problems. The class of mathematical programs with equilibrium constraints is also being investigated.

Yale University; Marios Papaefthymiou; ***CAREER: Parallel Integer Programming for Architectural-Level VLSI Design*****; (CCR-9624587); $100,000; 24 months.**

The research component of this project is to design and develop portable parallel software for solving architectural-level optimization problems that arise in the design of VLSI systems. Several of these problems can be formulated as very large scale integer programs that can be solved in polynomial time using iterative relaxation techniques. The practical running times of these algorithms on serial computers can be prohibitively long, however, especially for interactive design environments. The scaleable parallel software developed in this project will be able to execute orders of magnitude faster than conventional serial code. It will thus enable VLSI designers to interactively explore their design space at an early stage in the design process and to efficiently consider several architectural alternatives before committing to any particular implementation.

The educational part of this project aims at the seamless integration of parallel computation in the undergraduate and graduate curriculum. To that effect, two new courses will be developed at Yale. The first course will introduce undergraduates to the basics of serial and parallel computer systems organization. The second course will cover algorithms for computer-aided design of VLSI systems with a research focus on parallel electronic design automation.

University of Florida; Panos M Pardalos, William W Hager, and Donald W Hearn; ***Conference on Network Optimization: State of the Art*****; (DMS-9522573); $15,000; 12 months. (Joint support with the Operation Research Program and the Applied Mathematics Program)**

This conference brought together leading researchers working on many different aspects of network optimization and on diverse applications in fields such as engineering, computer science, operations research, transportation, telecommunications, and manufacturing. It also provided a unique opportunity for cross-disciplinary exchange of research. The conference received endorsements from the Mathematical Programming Society and the Institute for Operations Research and Management Science, and the Society for Industrial and Applied Mathematics.

Networks are pervasive in modern life, whether real, as in the case of highways, pipelines, electrical systems, etc., or more abstract network models which represent interconnections among components of a complex system. Advances in data structures and computer technology and the development of new algorithms have made it possible to solve classes of network optimization problems that were formerly intractable. Among these are problems in airline scheduling, transportation planning, satellite communications, water resource allocation, environmental modeling and VLSI chip design. A selection of the papers presented at the conference will be placed in the public domain by the publication of special issues of refereed scholarly journals.

Purdue University; Nagabhushana Prabhu; ***Complexity of the Simplex Method*****; (CCR-9625425); $25,000; 12 months.**

This Speculative Grant for Exploratory Research (SGER) investigates the celebrated Hirsch conjecture using the new theoretical framework that was developed recently for resolving the conjecture. In contrast to all of the previous attacks on the conjecture which were inherently combinatorial, the new strategy brings to bear on the conjecture the powerful machinery of continuous analysis and representation theory of continuous groups. Preliminary results from the new approach include (i) the discovery of a new formulation of the Hirsch conjecture with connections to Gaussian elimination, and (ii) the emergence of the most compelling theoretical and empirical evidence to date about the validity of not just the Hirsch conjecture but a considerably stronger version of it. The project aims at proving the strong Hirsch conjecture by employing powerful techniques from differential geometry and cohomology theory. The strong Hirsch conjecture, if proved, would show the existence of exponentially many paths of linear length between the initial and optimal vertices of a polyhedron, and would hence revitalize the research for polynomial-time Simplex algorithms.

University of Michigan – Ann Arbor; Romesh Saigal; ***Studies on Affine Scaling Method*****; (CCR-9321550 A001, A002); $62,933; 12 months.**

Affine scaling methods form an important class of point algorithms for solving a linear programming problem, and there have been some major developments in the convergence and convergence rate theory of these methods. In particular, when the problem may be degenerate, if the step size is less than equal to two thirds of the distance to the boundary, the primal converges to the interior of the optimal face, while the dual to the analytic center of the optimal dual face. Also, by a step selection strategy, these methods have been shown to converge superlinearly. More efficient variants have also been developed. Under this project, four different aspects of these methods are being studied: 1) study of their convergence properties, convergence rates and complexity; 2) study of stable implementations to achieve these higher rates of convergence in computer implementations; 3) study of their connection to Kalman filter and their application to solving stochastic linear programming problems; and, 4) study of their efficient implementations in vector/parallel/distributed computing environments.

Cornell University; Michael J Todd, Yuying Li, Leslie E Trotter, and Robert G Bland; ***Computational and Mathematical Investigations in Optimization*****; (DMS-9505155 A001); $164,000; 12 months. (Joint support with the Operation Research Program and the Computational Mathematics Program)**

This project undertakes broad studies of optimization algorithms and the mathematical theory underlying them. The project covers continuous and discrete optimization algorithms, both exact and approximate, their computational complexity, practical large-scale implementation, performance guarantees (in the case of approximate algorithms), and applications in the physical sciences such as elementary particle tracking, as well as more conventional applications in manufacturing and logistics. Particular investigations include: (1) a continuation of large-scale computational testing of implementations of minimum-cost network flow algorithms; (2) development and testing of algorithms for large-scale nonlinear minimization problems with bound constraints; (3) performance guarantees for efficient approximation algorithms for machine-scheduling problems with budget constraints and for other combinatorial optimization problems; (4) development of more efficient interior-point methods for convex programming and for linear programming problems where no initial interior feasible solution is known; (5) studies of the polyhedral structure of combinatorial problems in graph theory and partially ordered sets, as well as in applications such as crew scheduling — computational implementation and testing have been an integral aspect of these studies; and (6) further development of complexity theory for continuous nonconvex problems.

These problems range from modeling through algorithm development and analysis to implementation and testing. Optimization addresses questions like: (1) How do you schedule the production of hundreds of items with varying demands over several months to minimize setup and holding costs, subject to resource limitations and capacities? (2) How do you determine the geometric conformation of a protein (how it "folds") given an approximate energy functional for any position?

Indeed, complex systems arising in the engineering, physical, and biological sciences, and in the manufacturing and service industries, involving hundreds to millions of interacting variables, frequently lead to optimization problems when one wishes to control or predict their behavior. The project investigates a number of specific issues related to modeling and algorithm development for a variety of applications, ranging from traditional settings like resource allocation, planning, logistics, and distribution problems to new application areas such as airline crew-scheduling, elementary particle tracking, and shop scheduling. A common thread is the search for efficient methods to solve exactly large-scale instances of such problems, or, if that is not likely to be possible, for efficient algorithms that are guaranteed to obtain a solution that is close to optimal in a precise sense. These algorithms are tested computationally on representative instances of such problems using serial and parallel computers.

Princeton University; Robert J Vanderbei; ***Interior-Point Methods for Convex and Semidefinite Programming*****; (CCR-9403789 A001); $43,676; 12 months.**

Interior-point methods have proven to be efficient for solution of large, sparse linear and quadratic programming problems. These methods are being extended to cover more general classes of problems. In particular, interior-point technology are being used to create efficient algorithms for the solution of convex programming problems and semidefinite programming problems. The first class is important for general nonlinear optimization and the second has important applications to integer programming.

Computational Logic and Theorem Proving

Carnegie Mellon University; Peter B Andrews; ***Automated Theorem Proving in Type Theory*****; (CCR-9624683); $131,350; 24 months.**

This research is concerned with automated theorem proving in type theory, which is also known as higher-order logic. The basic purpose of this research is to automate and facilitate the use of rigorous logic. Rigorous reasoning plays (or should play) an important role in a wide variety of intellectual endeavors, and computerized systems which facilitate the use of logic have many important potential applications. The focus of this research is on proving theorems of a formulation of higher-order logic known as the typed lambda-calculus. This formal language includes first-order logic, but in a practical sense it has greater expressive power, and it is particularly well suited to the formalization of mathematics and other disciplines and to specifying and verifying hardware and software. Part of this research involves continued development of an existing computerized theorem proving system called TPS, which can be used to construct and check formal proofs (in natural deduction style) interactively, semi-automatically, and automatically. In automatic mode, TPS first searches for an expansion proof, which expresses in a nonredundant way the fundamental logical structure of proofs of the theorem in a variety of styles, and then transforms this into a proof in natural deduction style. The interactive commands for applying rules of inference are available in a related program called ETPS (Educational Theorem Proving System), which is used interactively by students in logic courses to construct natural deduction proofs. The possibility of using TPS in a mixture of automatic and interactive modes makes it an attractive tool for working on complex logical problems in a variety of disciplines. Research will continue on methods of searching for expansion proofs, methods of translating back and forth between expansion proofs and natural deduction proofs, on higher- order unification, on enhancement of TPS as a useful logical tool, and on related problems and questions.

State University of New York – Stony Brook; Leo Bachmair, R C Sekar, and I. V Ramakrishnan; ***Enhancing the Power and***

***Performance of Equational Systems*; (CCR-9510072); $159,183; 24 months.**

This project continues and expands the research in the design, development, and implementation of techniques for enhancing the power and performance of equational systems. The three key areas for this research effort are rewrite operations, equational programming, and equational deduction, though the results will likely also be applicable to other areas such as (constraint) logic programming. Research on rewrite operations will focus on variants of pattern matching algorithms that are specifically tailored to lazy evaluation strategies for functional programming and to build the semantics of certain functions into the matching algorithms for equational deduction. The investigations in equational deduction will center on the PI's recently proposed basic completion method, in which no inferences are applied at those subexpressions in formulas that are built up from substitutions generated in previous inference steps. The equational programming component of the project will focus on the further development of Equals — a fast parallel implementation of a lazy language and the Integral system which uses symbolic constraint solving techniques to glue together program analyses across different paradigms.

San Jose State University Foundation; Michael J Beeson; *Research at Undergraduate Institution (RUI): Logic and Computation: Automating Proofs in Calculus and Analysis in Teacher Preparation*; (CCR-9528913); $85,035; 24 months.

At present there are large computer programs capable of representing mathematical proofs (such as NuPrl) and large computer programs capable of symbolic computation (such as Macsyma), and many theorem-proving programs that try to find proofs. But these three capabilities are nowhere present in the same system. In this project, a single system smoothly integrating symbolic computation and representation of mathematical proofs will be built, and supplemented with proof-finding algorithms. A prototype version has already been able to generate automatically an epsilon-delta proof of the continuity of f(x) = x^3. The system to be built will be able to generate many other epsilon-delta proofs and other more complicated proofs in analysis.

Wichita State University; Shang-Ching Chou; *Automated Generation of Human-Readable Proofs in Geometry*; (CCR-9420857 A001); $66,517; 12 months.

This project continues investigation of automatic generation of proofs of theorems in geometry. The emphasis is on short and human readable proofs. For this, the newly discovered area method and new methods based on other geometrical concepts, such as full angles and trigonometric functions, are being developed. Extension of these techniques are being sought to geometric inequalities as well as to non-Euclidean and higher-dimensional geometries. The techniques are being implemented in a system with a sophisticated graphics display.

Kestrel Institute; Richard Jullig; *International Workshop on Constructive Programming: Rio de Janeiro, Brazil, November 4-8, 1996*; (CCR-9617781); $37,200; 6 months. (Joint support with the Software Engineering Program, the Programming Languages and Compilers Program, and the Theory of Computing Program)

This international workshop on Constructive Programming was held to explore future directions in formal methods, programming logics, and programming languages. The workshop combined 3 days of technical exchange with 2 days of discussions to produce recommendations to the science funding agencies of the US and Brazil: NSF, CNPq, and FINEP. In recent years Brazilian researchers have made significant contributions in constructive programming and related areas, establishing a number of contacts with US researchers. This workshop was intended to promote further collaborations and provide information and recommendations for future research efforts and funding initiatives.

Duke University; Donald W Loveland; *Workshop on Future Directions of Automated Deduction March 2-3, l996, Chicago, IL*; (CCR-9625544); $19,655; 12 months.

The purpose of this workshop is to assess the achievements, opportunities and goals of the field of automated deduction. Some key researchers of the field and a sample of users of the technology were assembled to address in particular the impact of automated deduction on other fields, present and future. The workshop will produce reports to NSF and the larger scientific community, identifying the major accomplishments, and discussing future directions, for the automated deduction field.

Duke University; Donald W Loveland and Owen L Astrachan; *US-Germany Cooperative Research to Enhance the Performance of the Model Elimination Proof Procedure*; (INT-9514375); $13,080; 24 months. (Joint support with the Western Europe Program Program)

This grant supports US-Germany cooperative research to study the Model Elimination proof procedure, a general-purpose theorem-proving technique of interest to both research groups. The procedure is derived from the logic programming language Prolog, and can use the sophisticated implementation architecture developed for Prolog. There is a significant amount of research being performed on Model Elimination in Germany as part of a five year multi-university research program on deduction that is lead by the TUM group. A fundamental issue being studied by both the German and US groups is the intelligent use of lemmas that are automatically generated by the Model Elimination system for use in accelerating the subsequent proof finding. The collaborating US and German research groups have experience respectively with global and local lemma techniques. These are quite different approaches, and combining them is expected to lead to enhancements of the lemma techniques which will expand the usefulness of the Model Elimination procedure for proving more complex theorems.

State University of New York – Albany; Neil V Murray; *Implementation and Analysis of Inference Techniques for Classical and Multiple-Valued Logics*; (CCR-9404338 A001); $46,302; 12 months.

This project continues a research program in logic. The studied areas of research stem directly from an analysis of the structure of formulas in negation normal form and from investigations into multiple-valued logics. The former work led to the development of the deduction rule called path dissolution. The study of this rule continues, including the study of its properties as an inference mechanism and the development of a portable C++ implementation. For some time it has appeared that dissolution would be a useful tool in the design of prime implicant/implicate algorithms. A new negation normal form (NNF) algorithm, PI, has been developed that can be combined with dissolution for this purpose, and formula classes have been discovered that are hard for CNF/DNF-based methods but that are easy for PI/dissolution. A prototype system has been implemented and is being further developed. These techniques are being extended to multiple- valued logics (MVL's). The work in MVL's also includes the study of annotated logic programming and SLD-style proof procedures, and inference methods for fuzzy logics.

University of North Carolina – Chapel Hill; David A Plaisted; *Instance-Based Theorem Proving with Semantics and Equality*; (CCR-9627316); $89,204; 24 months.

First-order logic is a common formalism for representing mathematical and other formal knowledge. Many traditional techniques for proving first-order theorems on a computer suffer from severe inefficiencies on certain kinds of problems such as near-propositional highly non-Horn problems. Under previous grants, theorem proving strategies have been developed that avoid many of these propositional inefficiencies. These include hyper-linking and hyper-linking with semantics. The latter method is able to make use of realistic models, or semantics, of the problem. However, the kinds of models it can use are limited to those involving linear inequalities and those with finite domains. This approach is now being extended to include a much more general class of models, those whose functions and predicates are computable. First-order logic with equality is a more powerful formalism that is also often applicable. For this formalism, methods based on term-rewriting systems are commonly applied. The above-mentioned semantic techniques will be extended to first-order logic with equality. Other topics related to equality will also be studied, such as rigid E-unification and specialized algorithms for handling permutations that arise from equational systems. A general framework—recently developed for studying the search efficiency of theorem proving strategies, and applied in a largely propositional context—will be extended to first-order logic. Complexity issues related to term-rewriting proofs will also be studied.

University of New Haven; Erik Rosenthal; *Research at Undergraduate Institution (RUI): Applications of Classical Inference Techniques to Multiple-Valued Logics and to Prime Implicate Algorithms*; (CCR-9504349 A001); $49,650; 12 months.

This project is for research in computational logic. The areas of research stem directly from prior analysis of the structure formulas in negation normal form. That work raised many questions and led to explorations in several directions. Substantial results at the propositional level, preliminary experimental at the first order level, and certain theoretical results that the most important of these may be path dissolution, a rule of inference that is strongly complete at the ground level. This project continues exploration of this and related inference mechanisms, largely through experimentation. One major thrust is to further the implementation of the techniques developed earlier. The current first order system ("Dissolver") is a solid platform on which the following techniques may be tested: link selection; computing prime implicants; backtracking; theory links and dissolution; and star chains. While abstract proof-theoretic questions are of interest in their own right, enhancing the performance of Dissolver is an important motivation for the theoretical work in this project. The investigations into multiple-valued logics largely fit the category; the study of the following issues contributes to the development of Dissolver: dissolution and multiple- valued logics; dissolution, analytic tableaux, and distributive law; quantifier duplication, proof length, and cycles; algorithms for computing prime implicants; and induction and equality.

SRI International; John Rushby, Patrick D Lincoln, and Natarajan Shankar; *An Integrated Framework for Mechanical Verification of Fault-Tolerant Systems*; (CCR-9509931); $345,147; 36 months. (Joint support with the Software Engineering Program, the Operating Systems and Systems Software Program, the Computer Systems Architecture Program, and the Theory of Computing Program)

Formal verification is a technique for reducing certain questions about the requirements, designs, or implementations of computer software and hardware to a mathematical form that can be examined by methods having the character of calculation. Similar to the way in which questions about the properties of an aerofoil can be reduced to calculations that can be mechanized by computational fluid dynamics, mechanized formal verification uses automated theorem proving and related methods (e.g., model checking) to perform logical calculations about the properties of computer systems. Whereas testing can examine only some of the behaviors of a complex system, mechanized formal verification can examine all the behaviors of a model of the system (again, the situation is analogous to flight test vs. computational fluid dynamics). Using mechanized formal verification, suitably realistic models can provide strong evidence for the correctness of computer systems, and simplified models are often useful for detecting bugs that testing does not encounter. Mechanized formal verification is most effective when applied to the hardest and most critical elements of design. An area where it holds great promise is in the analysis of fault-tolerant systems, where the presence of faults can produce a complexity of behavior that defies other forms of analysis. The increased assurance provided by mechanized formal verification can be of particular value for fault-tolerant systems because these are generally found in applications that are critical to safety (such as aircraft and other vehicles) or national infrastructure (such as the telephone, power-distribution, and banking systems). The objectives of this research are to reduce mechanized formal verification of fault-tolerant distributed algorithms from tours-de-force to routine practice, and to develop a framework for mechanical verification of the architectures of complete fault-tolerant systems composed from such algorithms. The research focuses on developing systematic methods for the formal specification of classes of fault-tolerant algorithms, and mechanized strategies for their exploration and verification. These strategies will build on techniques that combine decision procedures and model checking within an interactive framework. The algorithms considered will include those for distributed consensus, synchronization, diagnosis, and group membership. The work is expected to yield improved understanding of the assumptions, constraints, and fault models underlying the algorithms studied, more refined arguments for their correctness, and coherent arguments for their synthesis into an overall architecture. In addition, the research is expected to yield new results in the construction and application of formal verification systems, and improved methods for the formal representation and analysis of fault-tolerant algorithms.

SRI International; Mark E Stickel; ***Research on Automated Deduction*****; (CCR-9408630 A001); $45,146; 12 months.**

The model elimination theorem-proving procedure, like top-down reasoning procedures, has the defect of repeatedly solving the same goals. The problem can be ameliorated by lemmas or caching, but the "upside-down meta-interpretation" approach to be investigated offers a more comprehensive solution that entails executing the model elimination procedure by a bottom-up reasoning engine, which also enables more control over search strategy. Theory resolution is a framework for incorporating theories into a resolution theorem-proving program, thereby making it unnecessary to resolve directly upon axioms of the theory and improving efficiency. Although many ways of incorporating theories into a resolution theorem prover can be seen as instances of theory resolution, theory resolution provides little guidance on how to incorporate theories. Predicate- and function-matching rules and a multilateral representation for residues in partial theory resolution are being developed as methodologies for using theory resolution. Numerous previously open problems in the theory of quasigroups have been solved recently by automated reasoning systems such as efficient implementations of the Davis-Putnam procedure. In collaboration with other automated deduction researchers and a mathematician expert on the domain, additional effort will be devoted to obtain new results in the theory of quasigroups.

University of California – Santa Cruz; Allen Van Gelder; ***Practical Algorithms for Problems in Propositional Logic*****; (CCR-9503830 A001); $63,028; 12 months.**

Problems in propositional logic, such as Boolean satisfiability, are studied from the point of view of developing practical algorithms, although the problems are often NP-complete or co-NP-complete. Such problems are of practical importance in VLSI design verification, test generation, related CAD applications, as well as automated verification. Many other applications also occur naturally. E.g., higher order theorem provers require a satisfiability checker as a module. Recent hardware advances bring solution of some of these problems within reach. Algorithms are being implemented and tested for performance in typical applications, such as test pattern generation, sequential circuit verification, and program verification. Model search (usually called the Davis-Putnam method) and resolution are two well known, but somewhat conflicting, strategies for solving propositional satisfiability. Model search "succeeds" by discovering a satisfying assignment, whereas resolution "succeeds" by detecting that no such assignment exists. Some work on combining the two ideas in one algorithm has recently appeared. This project investigates more powerful combinations of these methods. Besides resolution, recognition can be valuable. The goal is to identify contexts in which resolution, equality recognition, and other inferences can be beneficial in reducing the search space. Equally important is to find implementations, usually requiring advanced data structures. Another goal is the development of application-independent methods. Heuristics described in the literature for particular applications are examined to determine whether the application content can be abstracted away, making the heuristic applicable to a logic problem without knowing its origin.

University of Houston; Rakesh M Verma; ***Computational Aspects of Normalization, Matching and Unification*****; (CCR-9303011 A002); $51,261; 12 months.**

Normalization, matching, and unification are three computational steps found in virtually all symbolic computation systems. Applications of these operations include term rewriting, functional, equational and logic programming languages, abstract data type specification and verification, automated deduction, code generation, type inferencing, and computer algebra. This project addresses these problems in computing and the means to design and analyze efficient algorithms for them. Practical algorithms are designed on both sequential and parallel machines. Algorithms which learn efficiently from their history are designed for normalization with respect to several important and useful classes of rewrite systems. Several facets of matching and unification problems are studied from a novel perspective. Algorithms that correlate information from previous match/unify operations are designed for tree pattern matching and its interpreted versions, and subterm unification and its interpreted versions. The practical performance of these algorithms is investigated. This research is useful in enhancing the efficiency of symbolic systems as well as provide understanding of the inherent complexities of these operations.

State University of New York – Stony Brook; David S Warren and I. V Ramakrishnan; ***Enhancing the Power and Performance of Tabled Evaluation of Logic Programs*****; (CCR-9404921 A003); $106,000; 12 months. (Joint support with the Knowledge Models and Cognitive Systems Program and the Programming Languages and Compilers Program)**

This project continues work on the development of XSB system. XSB is a system that integrates logic programming, deductive databases and nonmonotonic reasoning. It uses SLG resolution, a tabled evaluation strategy, to combine Prolog's SLD and deductive database's bottom up evaluation in a computationally effective way. Current benchmarks show that the XSB system is significantly faster than more traditional approaches to deductive database query evaluation, and shows promise of being very efficient for certain types of nonmonotonic reasoning. In the current project, the XSB implementation will be improved by; (1) improving the underlying WAM-based engine to make it more efficient, and (2) developing program transformation techniques to permit XSB programs to be optimized at the source level.

University of Iowa; Hantao Zhang; ***NSF Young Investigator (NYI) Award: High Performance Automated Reasoning and its Applications*****; (CCR-9357851 A003); $25,000; 12 months.**

Reasoning algorithms are used in many application areas of computer science and artificial intelligence, and high performance of reasoning algorithms is indispensable to the successes of these applications. The objectives of this research are to develop high performance reasoning algorithms, and to apply these algorithms in formal verification of hardware and software designs. This research will substantially improve the performance of a software system called RRL (Rewrite Rule Laboratory) by adding new reasoning algorithms into it, and by building new applications on the top of it. The research will focus on redundance control of automated deduction and mathematical induction, and efficient representation of data structures. The topics include restricted inference rules, powerful simplification rules, unnecessary computation detection, induction schemata formulation, induction hypotheses handling, and efficient implementations of all the involved algorithms. Practical verification problems will be chosen from software and hardware design as test problems for RRL and RRL will be enhanced to meet the need of large-scale applications by providing failure-resistant and user-friendly interfaces. Special effort will be made on distributing RRL world-wide to the people who are interested in using RRL as either a research tool or a teaching tool in automated reasoning.

University of Iowa; Hantao Zhang; ***High Performance Automated Reasoning*****; (CCR-9504205 A001); $59,983; 12 months.**

Two software systems, RRL (the Rewrite Rule Laboratory) and SATO (SAtisfiability Test Optimized), developed under the past grant, have been successfully used for solving many problems often considered a challenge for automated theorem provers, including several dozens open problems in quasigroups. The current research will further increase the reasoning power of RRL and SATO. The objectives of this research are to develop new high performance reasoning algorithms, and to apply these algorithms in specification and verification of hardware and software, and in solving open problems in algebra, logic and combinatorics. The research focuses on constraint satisfaction and automated mathematical induction. The topics include efficient constraint propagation, parallel and distributed reasoning, intelligent induction schemata formulation, management of large proofs, and implementation and experimentation of all the involved algorithms. Practical problems will be chosen from software and hardware design verification and combinatorics as test problems for both RRL and SATO. Both systems will be enhanced to meet the need of large-scale applications by user-friendly interfaces. Special effort will be made to distribute RRL and SATO to the automated reasoning community worldwide.

Problem Solving Environments

University of Colorado – Boulder; Elizabeth Bradley; ***NSF Young Investigator (NYI) Award: New Approaches to Engineering Design – Controlled Chaos and Computer Automation;*** **(CCR-9357740 A004); $62,500; 12 months.**

This project explores two new approaches to engineering design. Each of these threads is composed of synergistic subprojects; these subprojects combine analytical mathematics, numerical simulation, symbolic reasoning, and physical experiments. The first approach is to actively use chaos to improve the designs of various engineering systems. One example of this is causing gases to burn more effectively by induction of chaotic mixing in a combustion chamber. A small, well-placed obstruction can interact with the inflow and with the chamber's geometry to cause large effects: chaos's symptomatic "sensitive dependence on initial conditions." These effects are tested both experimentally and with computational fluid dynamic simulations. Another example is the use of chaos to broaden the capture range of a phase-locked loop circuit. The second approach to improving engineering design is through computer automation. The ultimate goal of this endeavor is a suite of programs that explores different high-level analysis and design paradigms. Such tools would not only streamline an engineer's problem-solving task, but actually amplify creativity. The first step is a program that autonomously constructs mathematical models from information about a system, presented in a variety of forms and formats: measurements made directly on the physical device; a user's observations described to the program in varying degrees of precision, symbolically or graphically; partial models and hypotheses suggested by a user; and general mathematical theory that is encoded in the program's knowledge base.

California Institute of Technology; K. Mani Chandy, James C Pool, Daniel I Meiron, John H Seinfeld, and Steven E Koonin; ***MDC: Integration of Symbolic Computing with Framework of Classes and Problem-Solving Archetypes;*** **(CCR-9527130 A001); $600,000; 12 months.**

The goal of this project is to help scientists and engineers carry out computational tasks more effectively by doing research on:

- The integration of symbolic and numeric computation to facilitate problem specification, exploration and result verification.
- Hierarchies of reusable computational abstractions organized in libraries with well-defined navigational structures.
- Exploiting commercial standards and tools for object-request brokers, compound documents, and collaboration infrastructure.

The project takes a novel approach to problem-solving environments (PSEs). Its focus is on the entire process of problem-solving from initial conception, symbolic specification, solution, visualization, comparison with experimental data, and feed back. The project focuses on problem-solving environments, not merely programming environments. This focus leads to research on reusable abstractions at all levels of problem-solving. Important abstractions include the problem-solving method itself, method-specific abstractions for sequential and parallel program design, reasoning and debugging, performance analysis and tuning, and frameworks for class libraries. PSEs are more useful when they are more specific. However, rather than build one problem-solving environment for one class of user, application domain, programming language and machine, this project aims at the basic research that will enable us to develop tools for building specific PSEs. To use a metaphor, rather than doing research on workbenches we want to do research on machine tools that can be used to build workbenches. Traditional PSEs deal with extending a workstation environment whereas this project is concerned with developing structured environments for problem-solving. Scientists and engineers collaborate when they solve problems, and they share much more than code. PSEs based on collaboration also provide object-request brokering and compound multimedia document interfaces. Since this vision is both novel and ambitious, a critical aspect of the project is leveraging methods and tools developed for scientific, commercial and home-computing applications.

University of Minnesota – Twin Cities; Linda Petzold, Robert T Tranquillo, and J. Ben Rosen; ***MDC: A High-Performance Problem-Solving Environment for Optimization and Control of Chemical and Biological Processes;*** **(CCR-9527151 A001); $549,999; 12 months. (Joint support with the New Technologies Program and the Computational Mathematics Program)**

The primary goal of this research is the development of high-performance problem solving environment (PSE) for the optimization and control of chemical and biological processes, with initial emphasis on bioengineering applications. The optimization and control of such processes requires the repetitive solution of time-dependent partial equations (PDEs) in two or three spatial dimensions. The requirements of this problem, which must be solved interactively, can only be met by the use of massively parallel computers. Such a comprehensive and powerful PSE does not currently exist, and its development presents significant computational and computer science challenges. The initial application is the design and optimization of a small diameter bioartificial artery. Over 600,000 surgical procedures for blood vessel replacement are conducted in the US annually, many involving synthetic polymer substitutes for small diameter blood vessels that fail due to lack of biocompatability, which is not an issue for bioartificial arteries. A PDE model based on continuum mechanical theory that describes the distribution of cells, fibers, and stresses in a tissue will be used to simulate the evolving compaction of a bioartificial artery. Optimization of the ultimate properties of the bioartificial artery is sought, in particular the ability to withstand pulsatile arterial pressure. PDE systems of a similar mathematical structure commonly arise as models for chemical processes with a need for their control and optimization. The project includes collaboration with and input from scientists and engineers at several industrial and government laboratories with applications in processing. To optimize or control a process described by PDEs, the original time interval is divided into subintervals in a multiple-shooting type approach. The PDEs on each subinterval are discretized in space via parallel adaptive finite element methods to give a large system of algebraic equations (DAEs). Parameters or control variables are then optimized, subject to state and control variable and continuity constraints (equality and inequality) using large-scale nonlinear programming techniques. Extensive knowledge and research experience in adaptive methods for PDEs, DAEs, large-scale optimization, parallel computing, computing environments, and chemical and biological processes, are required to successfully develop such a PSE. The research team members cover this broad range of expertise, and have a successful record of collaboration with each other.

Ohio State University; Feng Zhao; ***Intelligent Simulation Methods for Dynamical Systems*****; (CCR-9457802 A002); $62,500; 12 months.**

This project investigates new computational methods for prototyping physical systems with desired dynamical behaviors, with expected applications in design and control engineering. It develops a computational environment for visualizing and manipulating dynamics of physical systems through a combination of numerical and symbolic computations. More specifically, this project studies (1) automated characterization of dynamical systems through analytic and geometric analysis, and (2) techniques for interactive synthesis, optimization, and visualization of dynamics, for example, "computational scissors" that cut-and-paste vector fields and shape the desired dynamical properties (e.g. stability regions) for a complex system that is otherwise difficult to design.

COMPUTER SYSTEMS

Computer Architecture

Rutgers University; B. R Badrinath and Tomasz Imielinski; ***Designing Communication Support for Mobile Hosts in Distributed Systems*****; (CCR-9509620 A001); $45,970; 12 months.**

The rapidly expanding technology of cellular communications, wireless LAN, and satellite services will make it possible for users to access information anywhere and at anytime. In this environment, users will be mobile, use low powered computers, and communicate with the network over narrow bandwidth links. Such an environment introduces new challenges in the design of communication mechanisms due to constraints such as bandwidth, energy, cost, and the availability of computing resources due to host mobility. Hence, communication protocols should be designed for adaptability to changing environments brought by mobility. The research project will design and build support for mobile communications applications that explicitly take into account bandwidth limitations, energy limitations, and mobility using the following approaches: (1) The communication between the client and the server will be via an intermediary. Indirect interaction allows for optimization and development of specialized protocols for the wireless link and mobile host without having to change any protocol that is running on the fixed hosts. (2) Design environment aware communication primitives that will have access to the current environment state and also be notified about any changes. Thus, the type of communication to use can be tailored to the prevailing conditions.

University of Washington; Jean-Loup Baer; ***Reducing Memory Latency in High-Performance Systems*****; (CCR-9401689 A002); $99,951; 12 months.**

The focus of the research is the design and evaluation of hardware assets and software techniques to improve the performance of the memory hierarchy of high-performance systems. Processor speed has steadily improved during the last decade but memory latency and bandwidth have been progressing at a slower pace. Several techniques to reduce or tolerate high memory latencies have recently surfaced. Some of these techniques are evaluated using trace-driven and instruction-level simulations of benchmarks. A major thrust of the effort is to devise improvements in the simulation techniques, e.g., by using selective sampling and parallel discrete event simulation, so that performance results can be obtained faster. The experiments in parallel simulation serve as a test bed to customize programs for parallel execution. Additional projects continuing present research include the design of cache coherence protocols with various block sizes for transfer and coherence. These protocols are geared towards a reduction in contention in the interconnect of shared-memory multiprocessors and a decrease in the amount of traffic caused by sharing. Similar goals are the subject of software techniques such as compiler directives for bypassing the cache and placement of data in the right levels of the memory hierarchy.

Princeton University; Douglas W Clark and Kai Li; ***Architectural and Organizational Issues for High-Performance Uniprocessors*****; (CCR-9423123 A001); $69,481; 12 months.**

Caches and write buffers are becoming increasingly important to memory hierarchy designs since they are the key components for bridging the widening performance gap between microprocessors and DRAMs. This research studies the details of memory-referencing behavior in contemporary uniprocessors such as multiple referencing behavior . In particular, the research investigates the fine distinctions that characterize data references with respect to their cache and write-buffer behavior, and also studies possible architectural and organizational enhancements that will benefit writes.

University of Wisconsin – Madison; James R Goodman; ***A Study of Long-Term Architectural Limits to Computing: Galileo*****; (CCR-9509589 A001); $88,555; 12 months.**

Galileo is an investigation into the implications of long range technology advances across the computer field. While projections of narrow scope are difficult and risky because of the rapid progress over a very broad range of technology, it is possible to extrapolate current trends and make intelligent guesses about future computer building blocks. Extrapolating current rates of progress to as little as a decade hence suggesting that incredibly powerful computer systems might not be only possible, but even affordable. It is the goal of this research to identify those development directions that are likely to come into conflict or otherwise be retarded, and to investigate the implications of these effects. A major goal is to identify the relative improvements on different kinds of operations, and to understand how these changes will affect architecture, the programming environment, and the algorithms used. Increasingly the processor is becoming less effective, limited not by its peak processing power, but by the ability of the system to provide work to keep it busy. While processing power is rapidly becoming cheaper, the effective use of memory and communication is becoming the critical challenge in achieving higher performance. Memory costs have declined rapidly and steadily for more than twenty years, and will continue to decline in cost per bit. Access time has not improved dramatically, however, and high-bandwidth, low latency-memory will always be relatively expensive. Limitations in memory access tile and bandwidth are already posing difficult design challenges, and this will only be aggravated by technology trends. This research is investigating how a system can get higher performance from its memory system by applying cheap processing power wherever it can be used effectively. By today's standards, future systems will have enormous processing power, huge amounts of memory, and incredible communications bandwidth. Galileo is investigating how the architecture of such systems should differ from systems of today. Galileo is investigating how the architecture of such systems should differ from systems of today, and how such systems should be programmed.

University of South Florida; Lizy Kurian John; ***CAREER: Improving the Access-Execute Balance in High Performance Processors*****; (CCR-9624378); $110,000; 24 months.**

Program behavior studies reveal that most current high performance processors spend more time computing addresses and accessing data than performing the functional operations required by application programs. This research consists of investigating issues in the design of a balanced multiple issues architecture that improves the access-execute balance, maximized the decoupling between the various functional units, exploits higher concurrency and achieves higher performance. The techniques used to alleviate access- execute imbalance include (I) executing multiple access tasks in parallel, (ii) providing architectural support for complex access primitives, and (iii) providing application- specific access primitives. Also performed will be investigations on improving concurrency between access and execute processes. Concurrency between access and execution has been exploited by queues in many decoupled access-execute architectures, but data dependent control dependencies often prohibit prefetching of data to queues. This research will investigate techniques to facilitate anticipatory loading to queues even in presence of data dependent control dependencies. The research will build on the PI's previous studies on access execute decoupling and other schemes to improve the performance of memory accessing. The PI is incorporating innovative techniques such as active learning, and cooperative learning into the classroom. She wants to emphasize design, design automation, and experimental aspects in her courses. As part of curriculum improvement, a new graduate course on Advanced Computer architecture and an undergraduate course and laboratory on Digital Design using FPGAs are being developed. The PI also plans to participate in outreach activities for high-school students and special programs for women and minorities. The overall objective is to stimulate efficient learning at graduate, undergraduate and pre-college levels of education.

University of Minnesota – Twin Cities; Zhiyuan Li; ***CAREER: A Compiler-Assisted Approach to High Performance Memory Architectures*****; (CCR-9502541 A001); $45,000; 12 months.**

The widening gap between processor speed and memory latency will continue to be a critical issue to the performance of computer systems. The latency problem is even more acute on multiprocessors that share the same space, in which a main memory module is typically remote to all or all but one of the processors. It is therefore important to reduce unnecessary remote memory references and unnecessary network activities, and to achieve high hit ratio at the private, primary caches. The project explores coherent designs of compilers and memory architectures to improve the average memory latency, and the network traffic. The main thrust of this work is to combine compilers' ability to predict program behavior with harrdware's ability to record memory reference patterns at run time. The research focuses on three major issues: (1) The use of a compiler-assisted and hardware controlled coherence mechanism on multiprocessors with private cache memories; (2) The use of sophisticated interprocedural analysis to support the proposed coherence mechanism, via array data flow analysis and static task scheduling; and (3) The effects of program transformations, false sharing and limited directory pointers on the proposed coherence mechanism. This work covers a variety of multiprocessors that share a common address space, including bus-based systems, the Non-Uniform Access (NUMA) multiprocessors and the Cache-Only Memory Access (COMA) multiprocessors. The main objectives of the PI's education plan in this funding period are (1) to redesign the compiler and language design course sequence for the Computer Science Department of Minnesota and to solidify the interfaces of these sources with other language design sources; (2) to augment the main compiler and language design sources by seminar courses which address new issues regarding the interfaces of the compilers with the rest of the computer systems, and (3) to achieve teaching excellence in these courses so as to provide students solid understanding of the principles of programming language design and implementation and compiler techniques which will be important to their scientific and engineering career. The PI will conduct two experiments in this funding period. One is to develop a language design source to be taught after the compiler courses and the other is to add an operating system implementation course to the curriculum. Additional educational activities of the PI include exploration of new avenues for creating industry co-op programs for students in the computer systems area and exploration of new avenues or providing professionals with excellent access to both the basic courses and the seminar courses, e.g. through the university's Distributed Multimedia Learning Center.

Princeton University; Margaret R Martonosi; *CAREER: An Integrated Hardware and Software Performance Monitoring System*; (CCR-9502516 A001); $45,000; 12 months.

Over the past 15 years, processor speeds in modern computers have increased at a much faster rate than main memory speeds. Thus, relative memory latencies in both sequential and parallel computers have become quite large, and continue to increase rapidly. Such trends form the motivation for this research, which attacks the issues of how to monitor, analyze and improve computer performance in the face of a large processor-memory performance gap. Because of this performance gap, it becomes increasingly important to develop performance monitoring systems that allow programmers, compiler writers, and system designers to identify and tune portions of their design where memory overhead is limiting performance. To accomplish these goals, this project focuses on several area: (i) exploring hybrids of compile-time and run-time software-based techniques for application memory performance analysis, (ii) combining these hybrid monitoring techniques with dynamic compilation to implement flexible, low-overhead, detailed performance tools, (iii) designing and implementing novel, software-aware, hardware performance monitoring techniques, and (iv) designing software tools based on this novel hardware support. By studying both hardware and software solutions concurrently, one gains insights into the strong and weak points of each approach. The end result of this research is a suite of performance tools that span a range in the level of monitoring detail provided and the level of hardware support required.

Michigan State University; Philip K McKinley; *Collective Communication in Wormhole-Routed Massively Parallel Computers*; (CCR-9503838 A001); $69,346; 12 months.

The supercomputing industry is now dominated by parallel architectures. In many such systems, memory is physically distributed among processors in order to achieve better scalability, requiring nodes to communicate through a network. In order to reduce communication latency, many systems have the wormhole-routing switching strategy, in which messages are pipelined through and a single destination, or collective. Communication operations may be either point-to-point, which involves a single source and a single destination, or collective, in which more than two processes participate. Collective communication operations are particularly important in scientific computing, where large data arrays are distributed over the local memories of the nodes, but are manipulated through single operations. This research project addresses the design of collective communication operations for wormhole-routed networks. The research is intended to explore architectural and algorithmic alternatives in the design of communications subsystems for next generation wormhole-routed networks. The project can be divided approximately into four subareas. First, both models will be used to develop new algorithms for collective operations in addition to those already under investigation. Target platforms will include both direct networks, as well as wormhole-routed switch-based interconnects. Second, the resulting algorithms will be applied in order to support higher-level parallel programming constructs, such as virtual process topologies, parallel I/O operations and distributed shared memory. Third, the resulting communication services will be used to improve the performance of parallel applications, in particular parallel numerical algorithms. Finally, the use of the new models will be investigated in wormhole-routed systems that serve purposes other than parallel computing, such as high speed local area networks. The methods used will be analysis, simulation, and experimental studies on commercial wormhole-routed systems. The significance of this research is that it will provide new solutions to the communication needs of parallel applications. Besides providing a better understanding of the issues and problems that must be addressed in design and use of collective operations in wormhole-routing systems, this project will produce tangible results in the form of new protocols, prototype implementations of new algorithms, and new applications software, all of which will be made publicly available for use by other researchers.

Southern Methodist University; Dan Moldovan; *Marker-Propagation Networks: A Quantitative Analysis*; (CCR-9406998 A001); $80,000; 12 months.

Marker-propagation networks (MPNs) represent a powerful computational model especially suitable for artificial intelligence and other applications where information is often uncertain and incompletely specified. MPNs refer to networks of active nodes and labeled links. The power and novelty of this model derives from the fact that the processing instruments are programmable data patterns of flexible length called markers. Node processing is activated by the arrival of markers. Marker propagations through the network are not driven by destination addresses, instead, they depend on local conditions in the nodes, network topology and the information programmed in each marker. This is a major difference between MPNs and other parallel processing models, and it is the reason why MPNs are capable of handling weakly or incompletely specified applications. Previous work in the area of marker–propagation processing assumed much simpler models. The model studied in this project operates asynchronously, and is highly programmable. Some of the critical issues related to the implementation of MPNs on existing parallel computers are possible programming paradigms, sources of parallelism, load balancing, design tradeoffs, and others. The approach taken in this project is a quantitative one, meaning that the model is implemented on several parallel computers, performance is measured and the results compared. The MPN model offers scalability and has the potential of being a preferred architecture for future supercomputers dedicated to knowledge processing applications.

University of Southern California; Ulrich Neumann; *CAREER: Closed-Loop Augmented Environment Systems*; (CCR-9502830 A001); $40,222; 12 months.

The research component of this project addresses augmented environment (AE) and virtual environment (VE) tracking system inadequacies through a new 'unified-system' approach to AE system design. Even with the best head-mounted displays (HMDs) and graphic systems available, the shortcomings of tracking systems hinder or preclude the usefulness of many AE/VE applications. Such applications, including on-line surgical visualization, assembly or maintenance assistance, and interactive training simulations, require robust solutions to the following problems. 1) Achieving millimeter (position) and 0.1 degree (orientation) accuracy. 2) Readily achieving precise calibration of tracking systems to the real world. 3) Eliminating (or minimizing) latency in tracking (and graphics) systems to prevent gross misalignment during head motion. This project aims to develop new technology and improved methods for making video-based AE systems usable and productive for commercial and scientific applications. The long term research goal is the accurate alignment of real and virtual objects in an augmented environment under all conditions of user activity including rapid head motion. Achieving such performance required the integration of new image and tracking function components into an AE system. A 2D alignment error metric forms a basis for the compensation of a variety of 3D tracking and calibration errors. Corrections are computed and applied at every frame creating a closed loop system which compensated for both static and dynamic (latency-induced) misalignment. The project will produce increased understanding of closed loop AR systems by 1) implementing a software simulator to facilitate rapid exploration of closed loop AE system design options for feature detection error measurement, and error compensation, 2) designing a new AE system architecture with components which have simple, well-defined interfaces to other components in the system, and 3) developing a prototype of a complete closed-loop AE system and quantifying its performance characteristics. In the education component of this project a two-course Computer Graphics sequence for graduate students is under development. These courses build on the introductory undergraduate course and emphasize project creation. Hands on system-building learning opportunities, at the appropriate levels for students, are a major focus of the educational activities. In addition to syllabus design and class teaching, directed-research and advising promotes student-faculty interactions. A CD-ROM with computer graphics simulations of graphics concepts will be developed for both introductory and advanced levels.

University of Virginia ; Paul F Reynolds; *Delta Cache Coherence Protocols*; (CCR-9503143 A001); $96,658; 12 months.

Finding a good solution to the cache coherence problem is critical to the effective use of medium and large scale multiprocessors. The high cost of remote memory access in such systems makes it important that processes be able to access data locally, but allowing processes to hold local copies of shared writable data requires finding a good way to keep the copies consistent. This project will investigate a new family of cache coherence protocols called delta cache protocols that maintain cache coherence with little or no synchronization. Delta cache protocols are directory, as opposed to snoopy, protocols. The delta cache protocols are based on isotach time systems, time systems implemented by a class of networks called isotach networks. Isotach networks are characterized not by topology - they can be implemented on arbitrary topologies - but by the guarantees they provide about the relative order in which operations appear to be delivered. Isotach systems require custom switching elements, but can be built with off- the-shelf processors and in a variety of configurations ranging from tightly -coupled multiprocessors to clusters of workstations. A simulation study of isotach systems shows that they outperform comparable conventional systems under realistic workloads, in many cases by an order of magnitude or more. In extreme cases, they outperform conventional systems by an arbitrary amount. The project is a three year study of the cache coherence problem with the goals of exploring the design space for cache coherence protocols opened by isotach networks; and evaluating the performance of delta cache protocols. The exploratory work includes protocol and cache design; extending the approach to a variety of topologies and to other levels of the memory hierarchy; and devising ways to reduce the network bandwidth requirements of the protocols and to improve the efficiency of the protocols through data migration. The performance study will simulate isotach and conventional systems with caches and compare their performance under a variety of workloads. The simulation will build on the already completed simulation of isotach systems without caches.

University of Washington; David H Salesin; *Presidential Faculty Fellowships*; (CCR-9553199 A001); $100,000; 12 months.

This Presidential Faculty Fellow award is made in recognition of significant achievements of this principal investigator as a researcher and teacher, with potential for future contributions to this country's scientific and engineering enterprise. The research supported by this award is in the general area of image synthesis in computer graphics. In particular, the following topics will be investigated: (1) Real-time walkthroughs of complex environments (2) Controlled-precision photorealistic image synthesis, in which reliable error metrics are used to improve both speed and accuracy and (3) Computer-generated pen-and-ink illustrations.

Harvard University; Margo I Seltzer; *CAREER: High Performance Storage Systems*; (CCR-9502156 A001); $45,000; 12 months.

The demand for network services is growing at an exceptional rate. This growing demand requires that the necessary infrastructure be developed to facilitate rapid and easy access to vast, distributed information stores. The focus of this research is the essential base technologies required to make distributed information systems generally useful. Specifically, this research includes: high performance file system design and analysis, rotationally optimal disk scheduling, and distributed caching data. The research will result in a high-performance, distributed information server, utilizing new file system techniques, advanced disk scheduling algorithms, and caching based upon network topological location (geographical push-caching). The file system design borrows the essential characteristics of today's log-structured and journaling file systems, combining them with adaptive, rotationally optimized disk scheduling techniques. New caching techniques designed for scalability enhance the file system's feasibility by reducing user-perceived latency for distributed requests. The teaching component emphasizes curriculum development for both computer science concentrators and non-concentrators and training programs for undergraduate and graduate teaching assistants. For concentrators, the focus on undergraduate curriculum complements Harvard's commitment to the establishment of a strong computer systems research group. Undergraduate courses have been and will continue to be modified to incorporate the fundamental concepts in an area, the classic literature, and research activities. This combination of class activities is the cornerstone for producing undergraduates who can fully participate in ongoing research projects. For non-concentrators, the introductory computer science sequence has been modified to emphasize design and problem solving skills over programming mechanics. Future developments will extend this idea still further with the development of an introductory design course. The course is intended to serve both as a common ground for computer science and engineering concentrators as well as a broad introduction for non-concentrators. The focus of the graduate curriculum development is on the application of the scientific method to systems research. Students are asked to critique published work, design experimental evaluation platforms, analyze and critique data sets, and produce quality research. The last aspect of the teaching proposal is the development of teaching assistant training seminars. The PI has been working closely with the Derek Bok Center for Teaching and Learning, participating in their workshops and presenting panels and seminars with her teaching assistants. The orientation procedure she has developed will be enhanced so that it will be suitable for teaching assistants in all compute science classes, and potentially other fields.

University of Wisconsin – Madison; Gurindar S Sohi and James E Smith; *Prototyping Multiscalar Processors*; (MIP-9505853 A001); $309,962; 12 months. (Joint support with the Experimental Systems Program)

This project is evaluating a new architectural paradigm that can extract and exploit the parallelism in sequential code. This new approach uses both software scheduling in the compiler, as in VLIW, and hardware scheduling at run-time, as in superscalar architectures. The compiler segments code into large blocks of instructions that form subgraphs of the control-flow graph, though not necessarily basic blocks. The compiler appends synchronization information to each block that describes which registers must be shared with blocks. The compiler appends synchronization information to each block that describes which registers must be shared with other blocks. The blocks or tasks are passed to separate identical parallel execution units, each of which executes its task sequentially. Tasks are scheduled optimistically, so that some tasks may be executed by a unit when they would not be executed in a sequential system; in such cases, the tasks are "squashed," which means that their results are not written to memory or registers. At the end of each task, it blocks until the system determines that the task will not be squashed; at that time results are committed. The collection of execution units appears logically to be one unit, with a single register file. Shared register values are passed on an inter-execution-unit network, and tasks that need shared values block until the values are produced. During the first two years of this award is devoted to compiler design for the new architecture, to comparison with other architectures using simulation, and to conceptual refinement of the architecture. During later years, work is expected to focus on implementation.

Louisiana State University & Agricultural and Mechanical College; Jerry L Trahan and Ramachandran Vaidyanathan; *Algorithmic Scalability in Reconfigurable Bus-Based Models*; (CCR-9503882 A001); $78,433; 12 months.

For a model to be a workable component of a parallel processing system, it must be scalable. Researchers often design algorithms so that the assumed machine size for running the algorithm grows with the problem size. (This is particularly the case with algorithms for reconfigurable busbased models.) A physical machine size is, of course, fixed. For a scalable model, however, the arbitrary sized machine for running the algorithm can be mapped to the smaller physical machine that actually solves the problem without loss of efficiency. A compiler can perform this mapping, so algorithm designers and programmers can design algorithms for arbitrary size machines and need not design different algorithms or write programs for different sized problems. No reconfigurable bus-based model is known to be scalable, however, unless its computational power is significantly reduced. To better understand the extent to which these models are scalable, the investigator is less restrictive than traditional scalability. This research will investigate scalability of two main types of reconfigurable bus-based models, namely, the Reconfigurable Mesh (R-Mesh) and the Reconfigurable Multiple Bus Machine (RMBM). The RMBM comprises four versions: the B-RMBM, S-RMBM,F-RMBM, and E-RMBM. The investigators have previously established relative scalability results for the S-RMBM and for 'separable' algorithms of the E-RMBM, the most powerful version. The research will seek to investigate scalability of general E- RMBM algorithms and study different scalability approaches for different ranges in the relationship of simulated to available machine sizes. Scalability questions for the R- Mesh will also be investigated. These will include relative scalability and scalability of certain classes of algorithms, topics not yet examined for the RMesh, and extensive attempts at complete scalability of the RMesh, a topic that has received only to narrow attention in the past. Scalability questions will also be studied for directional versions of the R-Mesh and RMBM, which have proven particularly useful for directed graph algorithms.

Duke University; Kishor S Trivedi; *Industry/University Cooperative Research Center for Advanced Computing and Communication*; (EEC-9418765 A002); $174,200; 12 months. (Joint support with the Industry/University Cooperative Research Centers Program and the NSA-CCSP ATM NETWORK Program)

The North Carolina State University Industry/University Cooperative Research Center for Advanced Computing and Communication is becoming a multi-university Center with the addition of a research site at Duke University. Four companies are becoming part of the Center in conjunction with Duke University becoming a research site. The research agenda of the new multi-site Center will address high-speed networking, reliable and fault-tolerant systems, digital communication systems, distributed systems, and algorithms and digital image processing algorithms. Additional funding to augment the ongoing research program and to obtain a membership in the Center is being provided by the U.S. Department of Army and National Security Agency.

Fault Tolerance

Yale University; Richard Beigel and James Aspnes; *Parallel Fault Diagnosis*; (CCR-9415410 A001); $60,000; 12 months. (Joint support with the Theory of Computing Program)

Computer systems of the future may consist of thousands or even millions of processors. Inevitably some of the processors will fail. Faulty processors need to be repaired so that the system on continue functioning. The goal of this project is to investigate improvements to the existing algorithms for parallel fault diagnosis. Second, and more importantly, the project will develop dynamic models for fault diagnosis that are more realistic than previous models.

University of California – Irvine; Douglas M Blough; *Design and Evaluation of System-Level Fault Tolerance Mechanisms for Parallel Computers*; (CCR-9318495 A002); $69,899; 12 months.

Recent developments have dramatically increased the importance of ultra-reliable parallel computing. First, parallel processing has moved from a specialized scientific computation tool to an instrument for general-purpose computing. At the same time, computers have begun to control virtually every aspect of human life and are, therefore, increasingly relied upon to provide continuous service. These trends make the development of general but practical techniques for fault-tolerant multiprocessing a necessity. This project focuses on loosely-coupled parallel computers. These systems are ideally suited for ultra-reliable applications due to their built-in hardware redundancy that can be used to achieve fault tolerance. The project has both basic and applied research components. The basic research component studies new models and algorithms for the problems of multiprocessor system fault diagnosis and fault-tolerant routing. The applied research component involves the development of an experimental testbed for fault-tolerant multicomputer systems. A Transputer-based MIMD multicomputer system provides the testbed hardware and low-level operating system. Special-purpose software is developed to provide a system-level fault tolerance framework, a fault simulator, and a data collection tool for the testbed. The testbed allows experimental evaluation of the system-level fault tolerance mechanisms developed in the basic research component and is also made available to other research groups working on multiprocessor system fault tolerance.

University of Minnesota – Twin Cities; Daniel Boley; *Robust Fault Tolerance for Computations in Linear Algebra and Signal Processing*; (CCR-9628786); $59,862; 24 months. (Joint support with the Programming Languages and Compilers Program)

The objective of this project is to develop and test fault tolerant algorithms for numerical computation in various domains, including sparse matrices, recursive filtering and signal processing. The methods used are a combination of checks made during the computation based on preconditioned checksums and thresholds, and a-posteriori checks made on the final computed results based on backward error assertions (BEAs). In spite of the presence of rounding errors, BEAs allow very tight, cheap, validation of the final results independent of the specifics of the implementation or sources of error, and checksums and thresholds supplement the BEAs by providing checks during the computation. These methods will be extended beyond prototype examples to more realistic large computations, demonstrating their scalability and applicability to large domains.

California Institute of Technology; Jehoshua Bruck; *NSF Young Investigator (NYI): Efficient Fault-Tolerant Parallel and Distributed Computing*; (CCR-9457811 A002); $62,500; 12 months.

This research program addresses a number of fundamental issues related to the design of novel algorithms, protocols and architectures that enable efficient fault-tolerant parallel and distributed computing for scientific and commercial applications. The program is a blend of basic research and experimental systems activities. It includes work on the understanding of new computing paradigms inspired by neural networks as well as designing and prototyping of novel communication schemes for parallel and distributed systems.

Arizona State University; Partha Dasgupta; *Collaborative Research: High Performance Parallel Processing: Fault Tolerant Computing on a Network of Workstations*; (CCR-9505519 A001); $44,091; 12 months.

This project deals with both theoretical and prototyping research in parallel processing to harness the power of workstation clusters. The primary techniques will based on aggressive scheduling, evasive memory and dispersed data management. The formal work expands the previous results of the PIs in asynchronous computing in several directions: (1) Extend the models to models of fine-grained programs. (2) Permit the inclusion of architecture specific characteristics into theoretical frameworks. (3) Techniques to distinguish inconsistencies in memory states as viewed by different processors resulting from time-outs, interrupts and failures (or indefinite postponement of a computation). (4) Techniques to allow non-deterministic executions. (5) Model input/output operations of fault-tolerant programs. The experimental part of the project will build a scaled version to execute parallel programs on clusters of workstations. The prototype will address issues related to the eager scheduling of threads, the implementation of evasive memory and data management. The prototype will attempt to evaluate the effectiveness of the environment in terms of efficiency. This project is conducted jointly by P. Dasgupta of Arizona State University and Z. Kedem, of New York University.

Carnegie Mellon University; Ronald Jr. Bianchini ; *NSF Young Investigator (NYI) Award: Fault Tolerance in Distributed Systems*; (CCR-9257973 A004); $100,000; 12 months.

The first area addressed in this project is fault-tolerance in distributed systems with primary focus on continuous systems operation and collaboration. Collaboration is achieved when multiple distributed entities contribute toward a common goal. For collaboration to be successful in a distributed system, the goal must be achieved in a potentially unstable network environment. The second area is the investigation of telecommunication switching architectures. Specifically, the design and deployment of high-speed multimedia networks.

New York University; Zvi M Kedem; *Collaborative Research: High Performance Parallel Processing: Fault-Tolerant Computing on a Network of Workstations*; (CCR-9411590 A001); $71,233; 12 months. (Joint support with the Experimental Systems Program)

This project deals with both theoretical and prototyping research in parallel processing to harness the power of workstation clusters. The primary techniques will based on aggressive scheduling, evasive memory and dispersed data management. The formal work expands the previous results of the PIS in asynchronous computing in several directions: (1) Extend the models to models to fine-grained programs. (2) Permit the inclusion of architecture specific characteristics into theoretical frameworks. (3) Techniques to distinguish inconsistencies in memory states as viewed by different processors resulting from time-outs, interrupts and failures (or indefinite postponement of a computation). (4) Techniques to allow non-deterministic executions. (5) Model input/output operations of fault-tolerant programs. The experimental part of the project will build a scaled version to execute parallel programs on clusters of workstations. The prototype will address issues related to the eager scheduling of threads, the implementation of evasive memory and data management. The prototype will attempt to evaluate the effectiveness of the environment in terms of efficiency. This project is conducted jointly by P. Dasgupta of Arizona State University and Z. Kedem, of New York University.

Carnegie Mellon University; Bruce M Maggs; *NSF Young Investigator (NYI): Routing Protocols and Mechanisms for Tolerating Failures in High-Speed Networks*; (CCR-9457766 A002); $25,000; 12 months. (Joint support with the Theory of Com-

puting Program)

This project studies protocols for admitting and routing calls in distributed networks, and for tolerating failures in these networks. Its goal is to understand how to best use high-speed networking technology, such as ATM, to support parallel supercomputing, and to advance the NII (National Information Infrastructure) initiative. In the supercomputing domain, it seeks to develop network structures, routing algorithms, and mechanisms for tolerating faults that are both practical and provably effective. In the realm of the NII, it is analyzing the behavior of simple algorithms, including greedy algorithms, for admitting calls in wide-area integrated-service networks.

Harvard University; Michael O Rabin; *Asynchronous Parallel Computing, Fault Tolerance, and Randomization*; (CCR-9313775 A002); $110,524; 12 months.

An important class of practical parallel computers consists of machines comprising many large grained processors. These processors execute asynchronously and memory is assumed to be distributed. Trying to execute a general n-thread parallel program P on such a parallel system poses theoretical and practical challenges. One must provide mechanisms for implementing the synchronization barriers, for load sharing and balancing amongst the processors, and for program executing that respects memory locality. The approach used in this research is to define algorithms and a program transformation (compilation) T so that T(P) executes correctly and efficiently on any P. This simulated execution does not require any specialized hardware or operating system mechanisms in the target machine. Of particular interest is the case that the parallel program is large-grained, i.e. that in each parallel step a substantial block of instructions is executed on some or most of the parallel threads. It is assumed that the parallel program is already optimized so as to make the blocks of code within the parallel step quite large. This project studies the most general case of large grained parallel programs, paying particular attention to memory locality and to memory fault-tolerance, employing randomization. A further goal is to achieve Las Vegas style results, i.e. small expected running time and no errors.

SRI International; John Rushby, Patrick D Lincoln, and Natarajan Shankar; *An Integrated Framework for Mechanical Verification of Fault-Tolerant Systems*; (CCR-9509931); $345,147; 36 months. (Joint support with the Software Engineering Program, the Operating Systems and Systems Software Program, the Numeric, Symbolic & Geometric Computation Program, and the Theory of Computing Program)

Formal verification is a technique for reducing certain questions about the requirements, designs, or implementations of computer software and hardware to a mathematical form that can examined by methods having the character of calculation. Similar to the way in which questions about the properties of an aerofoil can be reduced to calculations that can be mechanized by computational fluid dynamics, mechanized formal verification uses automated theorem proving and related methods (e.g., model checking) to perform logical calculations about the properties of computer systems. Whereas testing can examine only some of the behaviors of a complex system, mechanized formal verification can examine all the behaviors of a model of the system (again, the situation is analogous to flight test vs. computational fluid dynamics). Using mechanized formal verification, suitably realistic models can provide strong evidence for the correctness of computer systems, and simplified models are often useful for detecting bugs that testing does not encounter. Mechanized formal verification is most effective when applied to the hardest and most critical elements of design. An area where it holds great promise is in the analysis of fault-tolerant systems, where the presence of faults can produce a complexity of behavior that defies other forms of analysis. The increased assurance provided by mechanized formal verification can be of particular value for fault- tolerant systems because these are generally found in applications that are critical to safety (such as aircraft and other vehicles) or national infrastructure (such as the telephone, power-distribution, and banking systems). The objectives of this research are to reduce mechanized formal verification of fault-tolerant distributed algorithms from tours-de- force to routine practice, and to develop a framework for mechanical verification of the architectures of complete fault-tolerant systems composed from such algorithms. The research focuses on developing systematic methods for the formal specification of classes of fault- tolerant algorithms, and mechanized strategies for their exploration and verification. These strategies will build on techniques that combine decision procedures and model checking within an interactive framework. The algorithms considered will include those for distributed consensus, synchronization, diagnosis, and group membership. The work is expected to yield improved understanding of the assumptions, constraints, and fault models underlying the algorithms studied, more refined arguments for their correctness, and coherent arguments for their synthesis into an overall architecture. In addition, the research is expected to yield new results in the construction and application of formal verification systems, and improved methods for the formal representation and analysis of fault- tolerant algorithms.

University of Massachusetts – Amherst; Ramesh Sitaraman; *Research Initiation Award ( ): Communication and Fault Tolerance in Parallel Computers*; (CCR-9410077 A002); $30,979; 12 months. (Joint support with the Theory of Computing Program)

The project focuses on two important inter-related issues in the design of large scale parallel computers: inter-processor communication and fault-tolerance. The nature of the research is one of synthesizing theoretical and practical approaches to parallel computing. Efficient inter-processor communication is at the heart of efficient parallel computing. With some exceptions, much of the theoretical work has focused on provably optimal algorithms that are often too complex to implement. On the other hand, much of the implemented routing hardware have simpler designs but lack theoretical guarantees as to their performance. This project will attempt to bridge this gap and focus on simple routing schemes that are provably good (though not always optimal). Fault tolerance is becoming ever-increasingly important in the world of parallel computers with a large number of fallible components. The study of fault-tolerant routing schemes focuses on algorithms that can route packets efficiently even in the presence of faulty switches. The two research directions represent very important but largely uncharted territory. The first concerns the question of intrinsic limitations on the fault tolerance of networks. The second research direction concerns reconfiguring in increasingly important class of parallel computers known as heterogeneous parallel computers.

Cornell University ; Sam Toueg; *Unreliable Failure Detectors for Reliable Distributed Systems*; (CCR-9402896 A001); $70,000; 12 months. (Joint support with the Operating Systems and Systems Software Program and the Theory of Computing Program)

The starting point for this research is a fundamental problem in fault-tolerant distributed computing: reaching agreement among processes in a system that is subject to failures. It is well-known that this problem, called Consensus, has no deterministic solution in asynchronous systems, even if it is assumed that communication is reliable and no more than one process may fail. The impossibility of solving Consensus (and other related problems such as Atomic Broadcast) is one of the most severe obstacles to implementing fault-tolerant applications in asynchronous systems. In recent work the PI has introduced a novel approach to circumvent such impossibility results: he showed that unreliable failure detectors can be used to solve Consensus (and Atomic Broadcast), even if the information that they provide about failures is highly inaccurate, e.g., even if they make an infinite number of mistakes. Since such failure detectors can be implemented in realistic distributed systems, and since various considerations make the asynchronous models especially attractive, this work suggests an approach to fault-tolerance that is viable in practice. The objectives of this research are to broaden the applicability of this approach by removing the limitations of the earlier work, and to explore in more concrete terms its practicability. Specific goals include: (1) tolerating communication failures, including network partitions (the earlier work assumed reliable links); (2) tolerating process failures of various types (the earlier work dealt with crash failures only); (3) considering shared-memory systems (the earlier work dealt with message-passing systems); and (4) solving other problems that are central to fault-tolerant distributed computing, including Group Membership and Group Multicasts (the earlier work solved Consensus and Atomic Broadcast). In order to assess the cost and benefit of using unreliable failure detectors complexity questions are also explored. Finally, the practicality of this approach is validated by implementation on an experimental platform.

Performance Evaluation

State University of New York – Buffalo; Jin-yi Cai and Min-You Wu; *Realistic Uncheatable Benchmarks*; (CCR-9634665); $242,237; 36 months.

The project on the design of realistic uncheatable benchmarks strives to achieve a more sophisticated design methodology of computer performance benchmarks, based on the theory of computational complexity. The goal is to integrate ideas from randomization and cryptography with realistic existing benchmark in everyday use. There has been tremendous progress in theoretical computer science in recent years, especially with regard to the power of randomization and interaction. These ideas are well suited to the design of uncheatable benchmarks. The purpose of benchmark testing is to estimate the performance and reliability of hardware and software systems. By and large existing benchmark design has focused on "typical" programs and data sets for the system. The problem of making benchmarks resistant to tampering has been mostly ignored. The on-going research in this project has demonstrated the feasibility of applying ideas from computational complexity and cryptography to a wide variety of existing benchmarks, and making them more accurate, resistant to tampering and thus more trustworthy.

University of Massachusetts – Amherst; Wei-Bo Gong; *Rational Approximation for Some Computationally Complex Performance Analysis Problems*; (CCR-9506764 A001); $50,177; 12 months.

Performance index functions with integer parameters are important in the evaluation and analysis of computer systems, communication networks and general distributed and parallel processing systems. These functions are often difficult to evaluate for large integers. On the other hand, most of these performance functions have nice properties such as monotonicity, convexity, analyticity, as well as obtainable asymptotic behavior. This research uses such properties to extrapolate the performance functions based on their values at small integers. The objective of this research is to further establish the theoretical foundation and to develop guidelines, algorithms and their software implementations that will be applicable to a wide array of systems. The results of this research will make it possible to express the performance measure function in terms of the system size of dimension as a low-degree rational function for some general classes of systems. The successful completion of this research will provide a framework in which various difficult modeling, control and optimization issues can be attacked.

Texas Technology University; Noe Lopez-Benitez; ***Performance and Reliability/Availability Models for the Analysis of Distributed Systems*****; (CCR-9520226); $137,732; 36 months.**

The purpose of this project is to investigate high level models for the performance and performability analysis of distributed computing systems hosting replicated programs and data files. The high level models studied rely on Stochastic Petri Nets (SPN) used hierarchically to model the behavior of the system in the presence of several types of hardware faults as well as software failures. Systems targeted include real-time systems where critical jobs are deadline constrained. These jobs are partitioned into tasks with precedence relations depicted via a task graph. Although task graphs can be evaluated directly, the evaluation assumes the existence of an unlimited number of processing units. Using Stochastic Petri Nets (SPN) the evaluation of task graphs can include other constraints such as a limited number of processors and existing allocation schemes static or dynamic. At the SPN level, state reduction can be achieved by generating only those states that satisfy functional as well as performance requirements specified in the form of symbolic boolean expressions. A synergistic approach will be investigated to predict the effect of hardware failures on the performance of a system running a single or multiple tasks, that may require access to distributed data files. The combination of SPN models and previously proposed reliability models such as those used to evaluate multiterminal reliability, has the objective of reducing the state space of the underlying Markov model (failure/repair models) even though hardware failures include nodes and links. Once the failure/repair model is generated, performance reward levels can be associated with each structural state of the system. This work will contribute to the systematic design of distributed systems. Design requirements such as delay, network throughput, load, task completion time, cost, etc., will be analyzed in terms of their susceptibility to failures in the hardware support for a particular application. It will also contribute to the evaluation of existing systems in order to detect bottlenecks or weak points and suggest appropriate remedial course of action (to either update or replace existing equipment). The systematic prediction of performance and reliability in terms of selective critical tasks, will aid in the process of updating currently saturated computer systems. In particular, the performance models of partitioned jobs will be useful in predicting the average completion times under user-selected allocation schemes via a single SPN-based model of a given task graph. Thus, giving users a rapid assessment on the suitability of particular schemes in a real-time distributed computing environment.

University of California – Davis; Charles Martel; ***Designing Parallel Algorithms Using Modeling, Simulation and Experiments*****; (CCR-9403651 A002); $64,643; 12 months.**

Parallel computers have the potential to significantly outperform current computer systems. In order to exploit this potential, techniques for creating and evaluating efficient parallel programs are needed. Current parallel computers have significant overheads and latencies for messages and this will continue for at least several years. Thus it is essential that programs for parallel computers be designed with these overheads in mind. Parallel computers are complex systems, so it is often difficult to predict the performance of programs based only on analytic models. Thus, it is important to evaluate parallel programs using both simulations and implementations. Networks of workstations are already widely available and they are likely to become even more prevalent. They provide an easy way to integrate the latest workstation technology into a parallel system, and they provide a cost effective parallel computer. Thus programs developed for a network of workstations have the potential to have a significant impact. This research focuses on developing and evaluating programs for networks of workstations. It results in effective algorithms for both basic problems such as load balancing, median finding, and simple graph algorithms, as well as for more complex applications such as modeling air or fluid flows. It also results in improved network support for algorithms on networks of workstations. As indicated, the performance of these programs is evaluated using simulations, so that the performance can be predicted for a range of settings, and by measurements of actual implications. A second focus of this research is to model cache-coherent shared memory systems, and to develop algorithms with good locality to help improve performance. Since algorithms for networks of workstations often exploit data locality, some of the algorithms developed for the network of workstations settings will also be effective on cache-coherent shared memory machines. Simulations and tests help determine which of the network algorithms perform well in this setting.

Northwestern University; Valerie Taylor; ***NSF Young Investigator (NYI) Award: A Tool for the Improvement of Uniprocessor Performance*****; (CCR-9357781 A003); $67,500; 12 months.**

The goal of this research is to investigate and develop techniques needed to produce a tool that allows the applications programmer and computer designer to work to increase uniprocessor performance. Computational units, such as the floating-point units, are becoming faster and readily available because of the rapid advances in technology and computer-aided design (CAD) tools. The performance, however, is limited by the rate at which data is applied to the units. To achieve good uniprocessor performance, the computational units, memory, and memory ports must be efficiently utilized. This may require that the applications programmer redesign the algorithms to better exploit the unique features of the architecture or the computer designer utilize CAD tools to design special-purpose processors that exploit the unique features of the algorithms or for best results, both. This research investigates and develops: (1) techniques for extracting statistics on the utilization of the aforementioned processor resources, and (2) methods for displaying the performance statistics to benefit both the applications programmer and the computer designer. The results of this research permit the applications programmer and computer designer to interact closely to address the problem of efficiently solving challenging problems such as climate modeling, vehicle dynamics, semiconductor modeling, and quantum chromodynamics, which are beyond the scope of existing computers.

Computer Graphics, Visualization and Computer-Aided Geometric Design

University of Washington; Anthony DeRose; *Subdivision Methods for Computer-Aided Design and Engineering*; (CCR-9502694 A001); $38,926; 12 months.

Representing curved shape is a fundamental problem in geometric modeling and design. A variety of approaches currently exist for modeling curved shape, by far the most popular of which is a boundary representation (B-rep) whose surfaces are trimmed NURBS (non-uniform rational B-splines). Unfortunately, building a modeler based on trimmed NURBS is exceedingly difficult for at least two reasons. First, trimming two NURBS patches to agree along a common edge usually involves complex surface/surface intersection algorithms and substantial approximation. Second, even in the absence of sharp edges, patching is often required since a single NURBS patch is capable of modeling only surfaces topologically equivalent to planes, cylinders, and tori. Modeling surfaces of other topological types (even those as simple as a sphere) require the surface be decomposed into a quiltwork of individual NURBS patches. Complicated continuity conditions must then be imposed to guarantee that adjacent patches join smoothly. Subdivision provides a promising alternative to NURBS. Using subdivision, a polyhedron is mapped to a sequence of denser polyhedra whose limit is a curved object. Each new polyhedron is derived from the previous polyhedron by splitting each of its faces into four new faces and positioning its new vertices using a fixed averaging mask. Different masks yield surface features such as corners, curved edges, or smooth faces in the limit object. The method is simple, requiring only polyhedral modeling. The method is general, producing objects with smooth faces and sharp curved edges without trimming or patching. Finally, subdivision provides a hierarchical representation that allows the design of efficient multi-resolution processing algorithms. This research addresses a combination of theoretical and applied problems whose solution is essential if subdivision is to be of use as a practical method for modeling curved shape. Specifically it develops: Curvature continuous subdivision methods; Subdivision methods for local refinement and introduction of shapes and edges; Functional subdivision methods over irregular triangulations (necessary for FEA) and algorithms for: Conversion to and from other representations such as NURBS, Editing and display of subdivision surfaces, Finite element analysis over subdivision surfaces, Computing the intersection, union, or difference of subdivision surfaces.

Massachusetts Institute of Technology; Julie Dorsey; *CAREER: Computer Graphics Techniques for Modeling and Rendering Weathered Materials*; (CCR-9624172); $100,000; 24 months.

Research Component Photorealistic image synthesis is an important problem for such diverse fields as architecture, lighting and industrial design, scientific visualization, simulation systems, entertainment, and advertising. An important component that has been missing from image synthesis is the effect of weathering. Existing models of materials in computer graphics assume that the materials are in pristine condition. This research program will focus on the development of computer graphics techniques for modeling and rendering weathered materials. The research will proceed along four lines. The first part of the program involves the development of comprehensive material models; and emphasis will be placed on models that can capture the changing appearance of a given material over time. The second part of the program will entail the development of new high-level representations for complex surfaces. The third aspect of the program includes creation of techniques simulating the effects of the environment in both batch and interactive modes. The final stage involves developing techniques to render weathered materials. Educational Component The pursuit of an integrated collection of educational initiatives is an important goal. The education plan proposes the development of interactive pedagogical aids to augment traditional classroom teaching, adaptive hypertext environments for on-line student assistance, and interactive challenge- response tools for student evaluation and certification. Advanced courses in computer graphics are being developed using these tools. Mentoring and training of new researchers will integrate the research and educational activities. The principal investigator's track record of interdisciplinary work, combining expertise in architecture and computer science, will facilitate the work.

Cornell University ; Donald P Greenberg; ***MRA: Physically and Perceptually-Based Parallel Global Illumination Solutions*****; (ASC-9523483); $330,000; 12 months. (Joint support with the New Technologies Program and the Advanced Scientific Computing Centers Program)**

Recognizing that computer graphics and visualization have become an essential part of the mainstream computing environment, this MRA, in partnership with the Cornell Theory Center, will investigate one aspect of the technology, and provide a forum for the community to exchange information/ideas/studies. Global illumination solutions, pioneered by the principal investigator will be further extended for parallel systems, and new more efficient algorithms developed. These will be experimentally tested and validated in a light measurement laboratory. A series of workshops will be held at Cornell University to bring together academic and industry experts to provide feedback and input to the project. These workshops will help determine priorities and formats for providing data and documentation on the World Wide Web.

Stanford University; Patrick M Hanrahan; ***Efficient Light Transport Algorithms for Computer Graphics*****; (CCR-9508579 A001); $89,401; 12 months.**

Many realistic image synthesis algorithms have been devised, but unfortunately they tend to be very inefficient, often requiring hours or days of calculation. As a result, the visual cues provided by these algorithms (shading, shadows, transparency, reflections, etc.) still have not been used routinely by applications requiring visualization, primarily because they are not supported in real-time on current generation graphics workstations. However, both increased performance and greater functionality can be expected in future generation workstation. This research project develops efficient high quality image synthesis algorithms. The project will attack several specific areas. Continuing to develop the theory behind rendering so that it has a sound basis in physics and mathematics continuing the development of fast hierarchical algorithms based on wavelets for solving the rendering equation. Investigating various hybrid algorithms that combine the best features of the finite element and Monte Carlo approaches. Finally investigating various issues that result from the need to handle very large, complex environments. The algorithms will be implemented in an experimental image synthesis system that is being developed and they will be tested on a variety of different scenes and environments.

Washington State University; John C Hart and Patrick J Flynn; ***Recurrent Modeling*****; (CCR-9529809); $206,435; 36 months.**

Current fractal tools lack the level of control necessary to model specific natural structures. For example, many fractal models of trees have been constructed, but given a particular tree, it would be difficult to represent the tree's specific shape using current models of fractal modeling. The recurrent modeling project identifies three aspects to this problem, and attempts to solve each with several new fractal modeling tools. (1) Recurrent Model Representation. Recurrent modeling strives to elevate fractal representation to the level of sophistication that smooth surfaces enjoy in computer-aided geometric design by translating their classical self-referential representation into standard implicit and parametric forms. Recurrent modeling also attempts "procedural geometric instancing," and enhancement of the classic object instancing paradigm that enables it to more efficiently represent the development fractal models currently specified by L-systems. (3) Interactive Recurrent Modeling. Recurrent modeling extends the tools of computer-aided geometric design to fractal geometry. A direct manipulation interface provides real time feedback in the fractal modeling process. The implicit formulation from the previous objective allows the application of standard blending formulations specifically designed for natural modeling. Extending constructive solid geometry to include fractal models supports the construction of complex shapes from primitives. (3) Automatic Recurrent Modeling. Recurrent modeling attacks the inverse problem of fractal geometry: "find the parameters of a fractal model that approximates a given shape" from the domain of model-based computer vision. Recurrent modeling continues this research to develop an automated system for discovering self-affinity in natural structures. The new tools recurrent modeling proposes would impact the fields of computer graphics and forest science. The new tools would represent highly detailed geometries, such as a forest and crops. Such models would have applications in image synthesis, particularly in animation, virtual environments, physically- based modeling and remote sensing. As the smooth representations of computer-aided geometric design aid manufacturers, the efficient representations of detail proposed by recurrent modeling representations would aid the study of the environment.

Carnegie Mellon University; Paul S Heckbert; ***NSF Young Investigator (NYI) Award: Simulating Global Illumination*****; (CCR-9357763 A003); $25,000; 12 months.**

Synthesizing realistic images is a fundamental challenge for computer graphics. One of the most difficult aspects of this task is the simulation of global illumination: the interreflection of light between surfaces. This research applies finite element methods to the simulation of global illumination. The work is divided into three subprojects: (1) error metrics for global illumination, (2) adaptive meshing for radiosity, and (3) generalization to scenes with arbitrary reflectance. Rather than attempting to measure "realism", which is a rather nebulous perceptual quality, simpler, quantitative accuracy measures are being developed. Using these error measures, existing and future image synthesis algorithms can be evaluated and quantitatively ranked. Adaptive meshes for radiosity mesh generation is only partially automated in many radiosity systems, and little is known about the errors introduced by various meshes. Recent experiments on simplified, two-dimensional scenes show that the use of adaptive meshes and higher degree approximations can lead to much faster and/or more accurate solutions than those of existing algorithms. More accurate, better looking pictures result when the radiosity mesh follows discontinuities in the solution function such as those at shadow boundaries. These "discontinuity meshing" techniques have been embraced by researchers. In this research, meshes for radiosity are compared and evaluated using error metrics as an objective criterion. This rigorous analysis of radiosity functions suggests new mesh generation techniques allowing more accurate simulations. Generalization to scenes with arbitrary reflectance, the final sub-project, is the generalization of these methods to scenes with realistic light scattering. The reflectance of real surfaces is not purely diffuse or specular, but is described by a complex bidirectional reflectance distribution function (BRDF). Recent work suggests that the use of spherical harmonics or other appropriate basis functions greatly facilitates the approximation of direction-dependent light functions. Such representations for functions of both position and direction are being analyzed and developed. The ultimate goal is the development of accurate, efficient algorithms for simulating global illumination on complex scenes with arbitrary reflectance and transmittance.

University of Maryland – College Park; Joseph F Ja'Ja'; ***Research on Designing Practical Algorithms for Computing and Image Processing*****; (CCR-9627210); $181,899; 24 months.**

A fundamental problem in parallel computing is to design high-level, architecture independent, algorithms that execute efficiently on general purpose parallel machines. The purpose of this project is to advance our understanding of the main factors required for designing practical parallel algorithms and to develop techniques and data sets for experimentally validating the results. As a byproduct, portable parallel programs and data sets for a number of specific important problems arising in combinatorial computing and image processing will be developed and made available to other researchers through the PI's Web page. There are currently two factors that make this fundamental problem more tractable. The first is the convergence of parallel architectures toward a unifying model and the second is the emergence of standards such as the message passing standard MPI. The technical approach to be followed is based on (1) a simple computation model that incorporates communication costs into the complexity of the algorithms, and (2) a SPMD programming model that makes effective use of communication primitives.

Brigham Young University; Thomas W Sederberg; ***Implicitization of Rational Surfaces with Base Points*****; (CCR-9622768); $60,000; 12 months. (Joint support with the Numeric, Symbolic & Geometric Computation Program)**

The objective of this proposal is to investigate a new technique for finding the implicit equation of a rational surface. Standard implicitization techniques for rational surfaces are based on resultants. But resultants vanish in the presence of base points. Therefore, when base points are present, the implicit equation can be recovered using resultants only by carefully perturbing the parametric equations to eliminate the base points. But such perturbations generally introduce extraneous factors, which then need to be painstakingly removed. Thus base points seriously complicate implicitization using resultants. On the other hand, base points simplify the implicit equation of a rational surface by lowering its degree. Preliminary work has roughed out a new implicitization technique based on algebraic "moving surfaces" that "follow" a rational surface. Not only is this new implicitization technique impervious to base points, it also simplifies in the presence of base points. Moreover this new method generates a more compact matrix representation for the implicit equation than methods based on resultants. Currently this new method is substantiated only by several Mathematica examples and some preliminary theorems. It is proposed to develop the theoretical foundations of the new method and then to apply the method to solve a variety of practical problems in computer graphics, geometric modeling, and symbolic computation. A wide variety of additional problems in elimination theory might also be attacked by these new techniques including, but not limited to, inverting a rational map, computing the singular locus of a rational surface, intersecting rational curves and surfaces, and intersecting algebraic curves and surfaces.

Indiana University – Bloomington; Peter Shirley; ***Realistic Computer Rendering of Complex Scenes Using Quantitative Perceptual Error Metrics*****; (CCR-9401961 A001); $47,343; 12 months.**

This research is concerned with the design of computer graphics that produces realistic images of environments that are complex with respect to geometry, reflection behavior, and light transport. This design will be pursued with the additional goals of real-time performance on general purpose hardware that will exist within one to two decades, and compatibility with display algorithms based on human perception.

University of Washington; Werner Stuetzle, Anthony DeRose, Thomas E Duchamp, and John A McDonald; *3D Scanning: From Physical Objects to Electronic Models*; (DMS-9402734 A002); $120,005; 12 months. (Joint support with the Integration Engineering Program, the Computational Mathematics Program, and the Statistics Program)

The goal of this project is to develop methods that will enable rapid, automatic, and inexpensive "3D scanning"; that is, the use of optical scanning technology to create electronic models from physical objects such as a clay model of a car, or a turbine blade. 3D scanning is at present an important technology, used in a number of critical US industries, such as automobile and aerospace production. However, the expense of current scanner hardware and the need for extensive human intervention has prevented 3D scanning from achieving its full potential. The methods developed by our interdisciplinary research team will make 3D scanning far less expensive, faster, and more automatic. The goal of 3D scanning is the inverse of computer aided manufacturing: given a physical object, such as a clay model of a car, a turbine blade, a chair, or a house, create an electronic model that captures its shape and color. 3D scanning is at present an important technology in a number of critical US industries, such as aerospace and automotive production. However, the expense of current scanner hardware and the need for extensive human intervention in both data collection and modeling has prevented 3D scanning from achieving its full potential. This grant will result in new algorithms and software that will enable new scanner systems that are faster, more automatic, smaller, less expensive, model other visual properties in addition to shape, and produce higher quality and more usable output.

State University of New York – Stony Brook; Amitabh Varshney; *CAREER: Improvement, Simplification, and Clustering in Polygonal Datasets for Virtual Reality Applications*; (CCR-9502239 A001); $38,327; 12 months.

The objective of the research being done is to devise a set of techniques and tools that will allow improvement, simplification, and clustering of polygonal datasets within a unified framework. This is an important first step towards the long-term objective of formulating techniques to enable interactive visualization of complex virtual environments. The techniques being used to solve the above problems are all based on the notion of 'rolling a sphere', formalized by the concept of a-hulls. For improving the existing polygonal datasets that have cracks and small missing polygons the intuitive idea is to roll a sphere that will fill-up these minor holes. Such flaws are common in present-day datasets and no general techniques are currently known that will handle them in a unified manner. Simplification of polygonal objects will be achieved through simplifying the genus of the objects. This provides a natural extension to, and is orthogonal to, the existing techniques for creating topology-preserving multiresolution hierarchies. The research being done will remove topological features that are less than a user-specifiable size. Clustering of polygonal datasets is of considerable use in devising efficient methods for solving global illumination and visibility preprocessing for interactive walk-throughs of synthetic environments. This research will accomplish clustering of polygons in a dataset by using sufficiently large rolling spheres. The goal of the educational activities being done is to expand the current entertainment-oriented view of virtual reality by presenting it as a useful tool for serious tasks in engineering, science, and medicine. The activities include attempts to reach out to the high-school students, undergraduates, and the community at large through slide-and-video shows and hands-on demonstrations of the scientific applications of virtual environments. The activities also include the creation and development of a new graduate-level course on virtual environments. The tools and techniques developed for the research component of this project will be used to impart a hands-on education to inculcate a feel for application-oriented research and development work in virtual environments.

Rice University; Joe Warren; *Subdivision Methods for Computer-Aided Design and Engineering*; (CCR-9500572 A001); $50,066; 12 months.

Representing curved shape is a fundamental problem in geometric modeling and design. A variety of approaches currently exist for modeling curved shape, by far the most popular of which is a boundary representation (B-rep) whose surfaces are trimmed NURBS (non-uniform rational B-splines). Unfortunately, building a modeler based on trimmed NURBS is exceedingly difficult for at least two reasons. First, trimming two NURBS patches to agree along a common edge usually involves complex surface/surface intersection algorithms and substantial approximation. Second, even in the absence of sharp edges, patching is often required since a single NURBS patch is capable of modeling only surfaces topologically equivalent to planes, cylinders, and tori. Modeling surfaces of other topological types (even those as simple as a sphere) require the surface be decomposed into a quiltwork of individual NURBS patches. Complicated continually conditions must then be imposed to guarantee that adjacent patches join smoothly. Subdivision provides a promising alternative to NURBS. Using subdivision, a polyhedron is mapped to a sequence of denser polyhedra whose limit is a curved object. Each new polyhedron is derived from the previous polyhedron by splitting each of its faces into four new faces and positioning its new vertices using a fixed averaging mask. Different masks yield surface features such as corners, curved edges, or smooth faces in the limit object. The method is simple, requiring only polyhedral modeling. The method is general, producing objects with smooth faces and sharp curved edges without trimming or patching. Finally, subdivision provides a hierarchical representation that allows the design of efficient multi-resolution processing algorithms. This research addresses a combination of theoretical and applied problems whose solution is essential if subdivision is to be of use as a practical method for modeling curved shape. Specifically it develops: Curvature continuous subdivision methods; Subdivision methods for local refinement and introduction of shapes and edges; Functional subdivision methods over irregular triangulations (necessary for FEA) and algorithms for: Conversion to and from other representations such as NURBS, Editing and display of subdivision surfaces, Finite element analysis over subdivision surfaces, Computing the intersection, union, or difference of subdivision surfaces.

University of California – Santa Cruz; Jane Wilhelms and Allen Van Gelder; ***Hierarchical Exploration and Visualization of Large and Complex Multi-Dimensional Data Sets*****; (CCR-9503829 A002, A003); $151,112; 18 months.**

This project investigates the use of hierarchical methods for accessing and visualizing large and complex multi-dimensional data sets. Visualization can provide an efficient and powerful means of exploring such data However, most visualization methods are developed for moderately sized non-hierarchical data sets that can be imaged acceptably fast on graphics workstations. In practice, there are many data sets that cannot be effectively explored using these methods. The project is developing and implementing visualization approaches designed for very large, multi-dimensional data sets over a variety of grid types. Such data sets present significant problems in the management of time and space resources, which will be addressed by this research. The unifying theme will be the use of hierarchical data structures to manage both resources effectively. Addressing time efficiency, techniques are being developed for the selective traversal and rapid interactive imaging of the data. For space efficiency, various data models, such as wavelets, are investigated to achieve compression, while being directly amenable to visualization. These hierarchies are being integrated with existing visualization techniques for isosurface extraction, direct volume rendering, and flow visualization, many of which have already been developed at UC Santa Cruz and elsewhere for non-hierarchical data. Techniques for rectilinear, structured, and unstructured grids are studied. Parallelization of these methods of SIMD and MIMD machines is also is being explored. This research is developing new algorithms for visualization of large, complex data sets, as well as making software systems available for use by scientists. This software is modular to facilitate integration with existing visualization packages. While the main objective is visualization, the hierarchical methods developed will also be of use for any data analysis hampered by access problems in dealing with very large data sets.

Distributed and Parallel Computing

University of Texas – Austin; Craig M Chase; ***Research Initiation Award (RIA): Efficient Algorithms for Automatic Parallel Program Decomposition*****; (CCR-9409736 A002); $28,501; 12 months.**

Parallel program decomposition is the process by which a single application is decomposed into multiple processes for execution on parallel hardware. For real programs, the execution speed of a good decomposition can exceed that of a poor decomposition by factors of ten or more. Unfortunately, optimal program decomposition is known to be an NP-hard problem, and efficient exact solutions are unlikely to exist. Hence, the key to finding automatic approaches to the program decomposition lies in discovering efficient algorithms that produce approximate solutions. A novel decomposition algorithm with nearly linear time complexity is explored. The algorithm initially decomposes the program into a sequence of small, parallelizable sections of code, known as partitionables. A locally optimal decomposition is selected for each partitionable based on information contained in an empirically derived database. Dynamic data redistribution is combined with a static technique called decomposition smoothing to produce a good global decomposition. The principal contributions of the research are: (1) to define new methods for program decomposition that go beyond simply minimizing communication and consider aggregate parallel execution performance, (2) to establish the extent to which fast decomposition strategies can produce acceptable solutions for practical applications, and (3) to collect and disseminate a database of performance results from kernels of scientific applications.

University of Illinois – Urbana; Andrew A Chien and Rajesh K Gupta; ***PDS: A Flexible Architecture for Executing Component Software at 100 Teraops*****; (ASC-9634947); $100,000; 12 months. (Joint support with the New Technologies Program and the ARPA-NEW MILLENIUM Program)**

Application software for 100 TeraOps machines be based on the increasing use of component software (libraries, GUIs, problem solving environments) and extensive use of interoperability frameworks (CORBA, OLE, SOM, etc.), producing applications which are complex conglomerates of variegated programs. To support these applications, the machine must present a programmable interface (i.e. efficient shared address spaces and data movement) and scaleable high performance. This project will study a flexible machine architecture that recognizes these technological realities and not only supports advanced software structures, but exploits them to configure its programmable hardware, adapting itself to the application. The programming flexibility can provide a wide range of convenient interfaces, ranging from uniform cache-coherent shared memory, to clusters of smaller shared memory systems, to fully distributed memory. In addition to a range of interfaces, this flexible machine can customize both operations and protocols, exploiting appropriate mechanisms and policies to minimize the critical performance aspect–communication–at all levels of the system. This machine design leverages the wealth of research into optimal architectural mechanisms, customizable cache-coherence, as well as technological advances in interprocedural analysis, programmable hardware, and hardware synthesis technology, to achieve general-purpose high performance.

Johns Hopkins University; Robert Cypher; ***Principles and Mechanisms for Asynchronous Parallel Computing*****; (CCR-9522301); $54,783; 12 months.**

MIMD (Multiple-Instruction-Stream, Multiple-Data-Stream) parallel computers are naturally asynchronous. Many MIMD computers use a separate clock for each processor in order to avoid clock skew problems. Furthermore, even if all the processors use a common clock, the timing of communication events is not known in advance due to factors such as data dependent processing times, processor interrupts and variable network congestion. This asynchrony has a major impact on the design of the machine's communication sub- system the semantics of the programming model, and the creation of parallel algorithms. This research studies principles for the design and use of asynchronous parallel computers. Specifically, three key areas of asynchronous parallel computing are studied. (1) specifying the semantics of message-passing, shared memory, and other, novel, programming models. (2) Creating efficient communication protocols that implement the above programming models, and (3) evaluating metrics and models for the design of algorithms for asynchronous parallel computers. This research will result in more efficient communication in MIMD machines, novel techniques for reasoning about and using asynchronous parallel computers, and a better understanding of how the theoretical community's body of knowledge in parallel algorithms can be applied to real machines.

SRI International; Jose Meseguer; ***System Level Issues for Multiparadigm Computing and SIMD and MIMD/SIMD Architectures*****; (CCR-9505960 A001); $78,574; 12 months.**

Parallel computing applications differ widely in the type of computational paradigm that is best suited to efficiently exploit their parallelism. Current software systems for parallel hardware support only single-paradigm computing. The objective of this research project is to investigate system-level support for efficient multiparadigm parallel computation. Previous work on coordination languages and multiparadigm sequential programming languages form a solid foundation for this research, although significant new ideas are required. Although current high-performance computer architectures are moving toward a combined MIMD/SIMD organization, very little software and system support exists for such hybrid computer architectures. This research will thus also focus on aspects of multiparadigm computing for massively parallel SIMD and MIMD/SIMD machines, including the Rewrite-Rule Machine as a testbed architecture. The research results will enable future systems designers to better appreciate the engineering trade-offs to efficiently support multiparadigm computing in a parallel environment. The results of this research are expected to enhance the future programming ease of high-performance massively parallel computing systems.

California Institute of Technology; Paul Messina; ***Parrallel I/O Methodologies for I/O-Intensive Grand Challenge Applications*****; (CDA-9318145 A003); $640,000; 12 months. (Joint support with the Networking and Communications Research Program, the Advanced Scientific Computing Centers Program, the CISE Institutional Infrastructure Program, and the Computational Engineering Program)**

This multi-disciplinary project will investigate and develop strategies for efficient implementation of I/O intensive applications in computational science and engineering. Scalable parallel I/O approaches will be pursued by a team of computer scientist and applications scientists who will work together to: (1) Characterize the I/O behavior of specific application programs running on large massively parallel computers (2) Abstract and define I/O models (templates) (3) Implement and test application-level I/O tools on large-scale computations The Pablo performance analysis environment will provide the foundation for the performance instrumentation and analysis. The application programs are already fully operational on advanced architecture systems and their authors are all co-investigators in this project. The principal computers used will be the Intel Touchstone Delta and Paragon systems at Caltech, each with over 500 computational nodes. Five application areas will be included: fluid dynamics, chemistry, astronomy, neuroscience, and modeling of materials-processing plasmas. The parallel programs for these applications cover a range of I/O patterns and volume, and the techniques that will be developed in this project will be of relevance to a broad spectrum of engineering and science applications. In addition, by overcoming their current I/O limitations, the specific applications targeted in this award will achieve significant new science and engineering results. By the end of the project, sustained teraFlops computers will become available. The project will devise and implement general methods for scalable I/O using today's advanced computers, immediately apply those methods to carry out unprecedented applications in several fields, and use the methods and experience gained in the first half of the project to tackle the I/O issues in future sustained teraFlops computers. The primary objective of the project is to develop methods and insights that are as generally applicable as possible. This award is being supported by the Advanced Projects Research Agency as well as NSF programs in engineering and computer science.

University of Texas – San Antonio; Kleanthis Psarris; *Research at Undergraduate Institution (RUI): A Novel Dependence Analyzer for Parallelizing Compilers*; (CCR-9528330); $65,027; 24 months.

The development of fast restructuring compilers for the automatic parallelization of sequential programs requires efficient data dependence analysis techniques in order to obtain exact data dependence information. The proposed techniques in data dependence analysis fall into two categories, either efficient and approximate tests or exact and exponential. The proposed research will attempt to show, though, that exact data dependence information can be computed efficiently in practice. The GCD test and the Banerjee inequality are the two tests most widely used to detect data dependencies between pairs of array references inside loop nests. These tests are approximate in the sense that they are necessary but not sufficient conditions for data dependence and, therefore, they may introduce artificial dependences which limit parallelization. Their major advantages is their low computation cost. A recent study provided necessary and sufficient conditions for the accuracy of the GCD and Banerjee tests. This study has also led to the development of the I (Interval) test, an improved dependence analysis test. The I Test is more accurate than a combination of the GCD and Banerjee tests and is able to provide exact data dependence information in practice, unlike the classic tests, at no additional computation cost. The goal of this project is to develop an efficient and accurate dependence analyzer. The research has four objectives. The first objective is to derive conditions for the accuracy of the GCD and Banerjee tests in general trapezoidal regions and perform an empirical study of how often these conditions occur in practice. The second objective is to extend the I test to the general trapezoidal regions and to assess its performance benefits on a suite of benchmark codes. The third objective is an analytical and empirical comparison of the I test with all the other tests, such as the Omega test, Power test, etc. An analytical and experimental evaluation of all the data dependence analysis test's is essential to determining which tests should be performed in practice. The fourth and principal objective is to integrate the above research work and develop an improved dependence analyzer as a part of the Paraphrase-2 parallelizing compiler. It is expected that the completion of the proposed research will improve the understanding of data dependence analysis, demonstrate the effectiveness and practical importance of the proposed techniques, and significantly improve the state of the art in parallelizing compilers. By providing efficient and exact dependence analysis techniques a number of compiler optimizations for high performance computer architectures can be performed in practice.

University of California – Santa Barbara; Klaus Schauser; *CAREER: Efficient Communication Primitives and Realistic Models for Modern Multiprocessors*; (CCR-9502661 A001); $45,000; 12 months.

Efficient communication is central to parallel computing. Modern multiprocessors obtain high bandwidth, low latency user-level communication through special hardware support, often implemented as a dedicated communications coprocessor. Future parallel machines, including high performance networks of workstations, will be built using communications co-processors. The goal of this research is to study how the presence of a dedicated co-processor affects the performance of interprocesor communication in modern parallel machines. In particular, the project focuses on two aspects. First, it studies the impact of communications co-processors on the performance of various communication primitives and determines the architectural requirements for implementing efficiently many-to-one communication such as active messages. Second, it studies how the communications so-processor can be captured by a theoretical performance model. The goal is to derive a performance model that is realistic enough to characterize machines and accurately predict performance, but still simple enough to be used in designing and analyzing parallel algorithms. The project leverages off the work on active messages, the parallel languages Split-C, and the LogP performance model, developed at UC Berkeley. The main experimental platform is a 64-processor Meiko CS-2, which the UCSB Computer Science Department acquired through an NSF CISE infrastructure grant. The CS-2 reflects, a modern high- performance network of workstation enthronement equipped with communications co-pocessors, making it extremely well suited for the study. The principal investigator has already implemented active messages and Split-C on the CS-2. The project develops a realistic model for parallel computation which captures the capabilities of communication co-processors. To create a framework to characterize communication and measure its performance, the project implements a simulator based on this model as well as a Split-C profiler. This framework is used to fine-tune applications and for architectural studies. Besides answering these important research questions, the work also provides infrastructure which is used for teaching purposes as well as by other researchers. The educational objective of this project is to teach students in these fields, to get them excited about computer science and engineering, and to prepare them for a successful professional career.

Santa Clara University; Weijia Shang; *CAREER: Fundamental Problems in Parallel Programming*; (CCR-9502889 A001);

$29,470; 12 months.

Parallel computers are difficult to program. Often, users have to specify the parallel execution, i.e., to specify data distributions to different local memory modules (as in High Performance Fortran), data communications between processors, and processor synchronization. The insufficient programming support limits the applicability and further development of parallel systems. The objectives of this research are (1) to search and find solutions (optimal or feasible) for those fundamental problems in programming parallel systems such as how to distribute/replicate input data to different local memories in the system to minimize inter-processor communications, how to schedule and partition the computations to minimize the overall running time, how many processors should be used and how to choose the right message size for a message vector; (2) to develop new techniques for data alignment, scheduling partitioning and estimating the number of processors to be used for optimal execution of algorithms with nested loops; and (3) to implement these techniques and add them to some existing software tools such as PVM and MPI. Education: The education plans are as follows: (1) Undergraduate computer architecture lab at Santa Clara University has not been changed for six years and is obsolete. It is planned to update the lab and syllabi and redesign the students' projects to include recent advances in computer design, technology and debugging software to have students more motivated and prepared when they graduate. (2) Develop new parallel processing lab projects. One project is to have students run parallel programs in PVM environment so students can see and think those fundamental problems in parallel programming. (3) Be involved in mentoring women and underrepresented undergraduates and graduate students. (4) Do research projects with several senior undergraduate students, especially women engineering students to motivate them to pursue graduate studies in computer engineering at SCU

Harvard University; Michael D Smith; *NSF Young Investigator (NYI): A System Architecture for High Performance Networking*; (CCR-9457779 A003); $62,500; 12 months.

Experimental research is performed to explore and develop sophisticated compiler algorithms and novel computer architectures that meet the application-to-network needs of such data-intensive applications as video-on-demand and real-time image retrieval.

OPERATING SYSTEMS AND SYSTEMS SOFTWARE

Operating Systems

University of California – Berkeley; Thomas E Anderson; *Presidential Faculty Fellows*; (CCR-9453532 A002); $100,000; 12 months.

The project will (1) investigate novel approaches to providing extremely efficient communication between protected domains running on the same machine, (2) develop tools for helping programmers tune the performance of applications limited by cache and virtual memory behavior, (3) develop new approaches to building ATM network switches to make them both easy to implement yet capable of delivering high performance, (4) investigate file system designs to improve scalability and to reduce power consumption, and (5) produce software for teaching undergraduate operating systems.

University of Arizona; John Hartman; *CAREER: A Caching Infrastructure*; (CCR-9624845); $97,330; 24 months.

The focus of this research is to develop an effective caching infrastructure for the design and implementation of caches in modular systems. Caching is a well-known technique for exploiting locality to improve performance; unfortunately, caches are often added to systems in an ad-hoc fashion. The caching infrastructure will improve cache development in the following ways. First, it will define a set of practical and efficient external cache interfaces for integrating caching into modular systems, such as object-oriented operating systems. Second, it will allow the functions of the cache itself to be modularized, defining internal interfaces that allow the functions to be easily modified and replaced. The flexibility this affords will allow the cache infrastructure to serve as a test-bed for caching, enabling different cache algorithms to be easily implemented and studied. Third, the use of a common cache infrastructure will allow global resource allocation decisions to be made among caches that are competing for resources, thus optimizing overall system performance. This resource management functionality will also include mechanisms for efficiently transferring resources, such as data blocks, between caches. Finally, the caching infrastructure will contain a toolkit to enable the cache control functions, such as the cache coherency protocol, to be defined via a high-level specification which can be verified and compiled into an efficient implementation.

Massachusetts Institute of Technology; M. Frans Kaashoek; *NSF Young Investigator (NYI) Award: Operating Systems for Multiscale Computers*; (CCR-9457791 A002); $49,930; 12 months.

Advances in VLSI technology have made distributed systems (e.g., clusters of workstations) suitable for parallel computing. Today's software, however, limits the functionality of distributed systems. The goal of this research is to discover and evaluate techniques for (1) automatic management of locality and work distribution, (2) fault isolation, (3) high-performance communication, and (4) fair sharing in a cost-effective distributed system. The approach is based on shared objects, an exokernel-based operating systems, a user-level communication unit, and gang scheduling. The project builds an experimental distributed system that delivers performance comparable to custom-designed supercomputers, while maintaining fault isolation and fair sharing.

University of Maryland – College Park; Peter Keleher; *CAREER: Sparks: Coherence as an Abstract Type*; (CCR-9624803); $100,000; 24 months.

Software distributed shared memory (DSM) systems provide the abstraction of shared memory to networks of workstations and distributed memory machines, such as the Paragon, CM-5, or SP-2. Unfortunately, the latencies for global operations in either environment are several orders of magnitude more expensive than on tightly-coupled multiprocessors. The result is that current DSMs perform well for only a restricted class of applications. The Sparks project is designed to address this restriction by providing a protocol construction library that will allow DSM performance to be improved to within a few percent of tightly-coupled multiprocessors. Sparks' abstractions allow the design space of high-level synchronization operations to be explored cleanly and systematically, rather than in an ad hoc fashion. The emphasis is more on creating and investigating the abstractions that make a broad variety of optimizations possible, rather than on the individual optimizations themselves. However, the performance gains allowed by the synchronization types created via the Sparks library will be thoroughly quantified. The Sparks interface will be implemented in CVM, a new distributed shared memory system. In addition to providing support services to Sparks, CVM implements novel coherence protocols and multi-threading support.

Parallel and Distributed Systems

University of Illinois – Urbana; Gul A Agha; ***Efficient Implementation of Dynamic Parallel Programs*****; (CCR-9523253); $145,747; 24 months.**

The project is studying the feasibility of using high level programming constructs which directly express the parallelism in an algorithm. Although parallelizing compilers for sequential programming languages provide good results for some numerical codes, namely those with regular communication patterns, such compilers are not effective in extracting parallelism in dynamic programs. The methodology used in this project is to express a computation as a collection of encapsulated, asynchronous fine- grained objects, or actors, which interact by sending each other messages. The project is studying ways of reducing the costs associated with managing fine-grained objects and messages. Moreover, because efficient execution on concurrent computer architectures also requires architecture-specific placement (and migration) strategies, the project is studying methods to allow code for placement to be independently expressed and combined with code for the logical operations in an algorithm. The advantage of such a separation is to allow portability of the algorithmic code across a wide range of concurrent architectures.

University of Illinois – Urbana; Roy H Campbell, Andrew A Chien, Simon Kaplan, and Daniel A Reed; ***Broadband Network Infrastructure and Data Storage for Research and Education*****; (CDA-9401124 A002); $270,127; 12 months. (Joint support with the CISE Institutional Infrastructure Program)**

This award provides support for an experimental, high-speed network infrastructure that would interconnect parallel systems, large-capacity secondary and tertiary storage devices, and muli-media workstations capable of supporting high-quality digital audio and video. The network architecture is based on a broadband ISDN backbone of Asynchronous Transfer Mode (ATM) switches and fiber optics links. A high-bandwidth, gigabit/second HIPPI network would interconnect an Intel Paragon distributed memory parallel system and a tertiary storage system. The research spans three basic areas: system software infrastructure for managing high-speed networks, data prefetching and staging policies for high- latency, high-capacity secondary and tertiary data stores, and information access and multimedia collaboration software.

Duke University; Jeff Chase; ***CAREER: System Support for Collaborative Information Systems*****; (CCR-9624857); $100,000; 24 months.**

This research focuses on operating system support for collaborative information systems, in which multiple users or programs cooperate to maintain a shared information resource on a computer network. The central focus is development and evaluation of a "network virtual store", a virtual memory based storage repository for networked clusters of 64-bit computers. The network virtual store represents a convergence of distributed file systems, software-based distributed shared memory systems, and "object-oriented database systems" supporting transactional access to persistent data structures. The goal of the project is to explore new storage management approaches for group computer- aided design systems, scalable database servers, and other applications that require high-performance access to recoverable state in a distributed environment.

Syracuse University; Alok Choudhary; ***System Software Support for Input-Output on Parallel Computing*****; (CCR-9509143); $67,686; 12 months.**

A large number of applications from information processing as well as science and engineering require processing large amounts of data. Particularly, in the science and engineering domain, the so called "grand challenge applications" fall in this category. Thus, it has become apparent that I/O performance much more than CPU or communication performance may be the limiting factor in future parallel computing systems. Many software and hardware components play a key role in I/O operations and any one of them can cause severe bottlenecks. The disparity between the performance of disk systems and CPU (and interconnect) performance requires the use of large-scale parallelism (and supporting software) to address the scalability and performance problems in I/O subsystems. But parallelism in the I/O systems introduces complexities for the system software including compilers, runtime systems and file systems. This research addresses many system software problems in the parallel I/O domain. To ameliorate the I/O bottlenecks, the following are necessary: 1) techniques for sophisticated runtime systems that incorporate the notion of collective I/O, data reuse and prefetching; 2) compiler analysis to identify various optimizations including data access reorganization, opportunities for overlapping computations and I/O, and data staging to handle limited memory for out-of-core computations; and 3) file systems optimizations to recognize collective access patterns for coordinating accesses of processes that are part of parallel programs, and using these access patterns for prefetching, buffer allocation and replacement strategies. Traditional interfaces for file systems lack the necessary semantics to permit the above optimizations. Thus, it is necessary to develop interfaces between the compiler and the runtime systems, and the runtime system and the file systems, that facilitate the flow of semantic information about access patterns to enable optimizations. This research pursues the design and development of techniques for compiler analysis, mechanisms and algorithms for runtime systems, and interfaces and optimizations for file systems to enable high- performance I/O in parallel computers. These techniques are being evaluated by implementing them on parallel systems.

Rice University; Alan Cox; ***NSF Young Investigator (NYI) Award: High Performance Distributed Shared Memory Multicomputers*****; (CCR-9457770 A002); $62,500; 12 months.**

The goal of this research is to make parallel computing simple and efficient on general-purpose, commodity processors and networks. Recent work has focused on algorithms for the efficient implementation of software distributed shared memory (DSM) on low-cost, high-performance workstation clusters. This work has shown that software DSM can achieve performance close to a hardware implementation of shared memory on a variety of scientific applications that synchronize with moderate frequency. The objective is to improve this work in two ways; (1) to achieve the efficient execution of applications that synchronize with greater frequency and, (2) to support languages ordinarily used to solve symbolic problems. Specifically, the following approaches are evaluated: (a) run-time support for programming high-level synchronization operations, such as task queues, that reduces the amount of communication by the DSM system, (b) compiler support to reduce the amount and/or hide the latency of communication by the DSM system, and (c) garbage collection algorithms integrated with the DSM system to enable the efficient execution of advanced programming languages.

University of Connecticut; Phyllis E Crandall; ***Research Planning Grant (RPG): A Software Support System for Parallel Processing in Multiprocessor Clusters*****; (CCR-9625959); $18,000; 12 months.**

This effort is funded as a Research Planning Grant (RPG). The availability of shared-memory, multiprocessor workstation clusters offers the opportunity for the simultaneous application of shared memory and distributed memory parallel processing techniques to a single problem. However, software systems that facilitate this dual-level parallel processing paradigm and allow the programmer, compiler, and run-time system to make the best use of the computational power offered by this platform are not yet available. The long-term goal of this research is threefold: (1) to provide a software system that furnishes advice regarding the partitioning methods and cluster configuration expected to yield the best performance, (2) to develop a communication library that handles both shared and distributed memory communications with a single syntax, and (3) to incorporate both the advisory system and the communication library into a data-parallel compiler that transparently handles interprocessor communication. During the research planning phase, which is being supported by this grant award, four initial steps toward the long-term research goals will be accomplished. (a) The computation and communication characteristics of the multiprocessor cluster will be characterized. (b) Mathematical models of performance of applications that exhibit stencil data dependencies will be developed. (c) The advisory system will be designed and implemented with the initial mathematical models. (d) The current body of research on shared memory and distributed memory communication libraries and data-parallel compilers will be surveyed to determine the best approach for developing the dual-level communication library and incorporating the library into existing compilers.

Princeton University; Edward W Felten; ***NSF Young Investigator (NYI) Award: Efficient Communication for Distributed Parallel Processing Applications*****; (CCR-9457813 A002); $62,500; 12 months.**

The research has two parts: systems software and applications. In the systems software area, the inherent complexity and cost of communication and synchronization between distributed processors is studied. This work includes advanced implementations of communication software in experimental parallel architectures, including tools for generating specially-tailored communication mechanisms for particular parallel applications. The goal is to understand how best to support communication in future systems. Collaboration with researchers in other scientific fields to develop application programs that effectively use the capabilities of today's high-performance parallel computers is the goal of the applications aspects of this project.

University of Texas – Austin; Vijay K Garg; ***Detection of Global Predicates in Distributed Programs*****; (CCR-9520540 A001); $54,041; 12 months.**

The objective of this project is to study theoretical and practical issues in the evaluation of general global predicates and functions of distributed software. As an application of the techniques of the research, the project will focus on distributed monitoring systems for fault tolerance. An example of detecting a global predicate arises in implementing the most basic command of a monitoring system: 'stop the program when the predicate is true'. To stop the program it is necessary to detect the predicate q, which is not a trivial task if q requires access to the global state. There are three components to the project: theory, algorithms and implementation. There are two questions that will be investigated in terms of theory. First, the class of global predicates that are efficient to detect will be classified. This will also entail studying the limits of efficient detection of global predicates. Second, a predicate logic which is suitable for efficient detection of global predicates will be developed. After developing the logic, efficient decentralized algorithms for detecting occurrence of logic formulas will be designed. This work will be based on the use of a centralized checker process developed by the PI. Finally, the decentralized algorithm will be used to design an experimental distributed monitoring system.

University of Kentucky ; James N Griffioen; ***Research Initiation Award (RIA): A Memory-Based Architecture for High-Performance Distributed File Systems*****; (CCR-9309176 A001); $5,000; 0 months.**

This research addresses the issues involved in the design of scalable high-performance distributed file systems based on volatile storage. The flexibility of such system combined with recent technology trends provides an opportunity for exceptional performance and superior resource utilization. The research has four components. First, a new distributed file system abstraction that incorporates both volatile and stable storage into a single abstraction will be investigated. The abstraction allows users to adjust the persistence guarantees of a file to match the requirements of the application, thereby optimizing performance. Second, a novel memory architecture called loosely-coupled memory that provides a massive aggregate memory space for use by the proposed file system will be introduced. The architecture also incorporates asymmetric compression techniques to make effective use of memory and underutilized processing power, and to improve file system access time. Third, a new abstraction and file system architecture into the VM Xinu operating system to verify our design. VM Xinu provides a high-performance, low-latency communication protocol ideal for support of a loosely-coupled memory system. Finally, the file system will be incorporated into the Mach operating system to demonstrate and evaluate our ideas in the context of a large user population.

University of Illinois – Urbana; Wen-mei Hwu; ***Data Speculation for Instruction-Level Parallel Processing Systems*****; (CCR-9629948); $109,328; 24 months. (Joint support with the Programming Languages and Compilers Program)**

Data speculation refers to the execution of instructions before it has been determined that the data being used is valid. Data speculation depends on dynamic memory disambiguation to determine if the data was valid, and forces recovery code to execute if the data was invalid. A data speculation compiler generates recovery code to handle the situation where such execution violates a memory dependence. This project examines data speculation techniques, and attempts to understand how to utilize them more effectively. The issues addressed are (1) using data speculation with predication, (2) improving the scheduling algorithms used with it, and (3) using static analysis to help determine when it should be applied. This work will be done using the Memory Conflict Buffer (MCB) technique for dynamic disambiguation. The concepts will be developed so they are applicable to other forms of dynamic memory disambiguation.

University of Wisconsin – Madison; James R Larus; ***NSF Young Investigator (NYI) Award: Programming Massively-Parallel Computers*****; (CCR-9357779 A003); $62,500; 12 months.**

The goal of this research is to make massively-parallel computers easy to program. This means that the natural users of these machines; scientists and engineers working on large-scale, state-of-the-art research, as exemplified by the Grand Challenge Problems routinely exploit these computers as a research tool. The problem of programming massively-parallel computers is attacked from three directions: (1) collaborating with computer architects on shared-memory architectures that facilitate programming and compilation, (2) developing new parallel languages that permit more natural expression of parallel algorithms and the techniques to implement them, and (3) building performance evaluation and debugging tools to help programmers to understand and increase their program's performance and correctness.

University of Rochester; Thomas J LeBlanc; ***Measurement and Prediction of Parallel Program Performance*****; (CCR-9510173); $77,095; 12 months.**

The goal of this research is to develop a set of tools and techniques for performance analysis and tuning of parallel programs that bridges this gap between measurement and modeling, and theory and practice. The approach to performance measurement and modeling is called "lost cycles analysis", and is based on the observation that the distinction between productive computation and parallel overhead is useful both for performance diagnosis and for performance prediction. This approach decomposes all parallel overhead into mutually-exclusive, yet comprehensive categories (such as load imbalance and communication overhead), each of which can be measured and modeled separately. These categories express overheads in common units, called "lost cycles", and together represent the portion of the total execution time not devoted to useful computation. This research effort's goal is to build a lost cycles toolkit, integrating empirical model-building techniques from statistics with the measurement and modeling techniques of lost cycles analysis to generate accurate models of application performance. The toolkit will use a priori knowledge of the sources and characteristics of overhead in parallel systems to guide and constrain the modeling process. It will incorporate techniques for optimum experimental design into an experiment generation tool that will generate a "script" of representative program executions. During these executions it will use a measurement tool to record the lost cycles for each overhead category as a function of the variables of interest (including, but not limited to, the number of processors and the problem size). It will use the measurement results obtained from the executions to select from a database of standard models for overhead categories. In addition, it will use a tool for fitting models of overhead categories to experimental data to establish the constants associated with a particular hardware and software system. By automating performance modeling as much as possible, this effort intends to make lost cycles analysis practical for everyday use by parallel programmers.

University of Washington; Henry Levy and Anna R Karlin; *Distributed Resource Management in Workstation and PC Clusters*; (CCR-9632769); $202,408; 24 months.

Advances in network and processor technology are greatly changing the communication and computational power of local-area clusters of workstations and personal computers. In particular, the replacement of Ethernets by newer switched network technologies, which provide both higher throughput and lower latency, presents an opportunity for operating systems to better use network resources. This research seeks to improve the performance of applications through distributed (cluster-wide) resource management. The project will develop algorithms that permit the effective sharing of all network resources: processors, network, memory, and disk storage. For example, programs on heavily- loaded nodes can use available memory on idle nodes for both file system and virtual memory backing store, rather than disk, thereby reducing the data latency on access to that data. As another example, applications may migrate to nodes that provide either a better processing environment or a better memory environment for their execution. A key issue in developing algorithms for such closely-coupled system is finding the right balance in order to optimize system-wide performance while minimizing any degradation of performance on individual systems as a result of network-wide resource sharing.

California Institute of Technology; Paul Messina; *Scalable I/O Initiative*; (CDA-9415797 A001); $300,000; 12 months. (Joint support with the Experimental Systems Program Program and the CISE Institutional Infrastructure Program)

This award supports the infrastructure for the Scalable I/O national project. This national project will investigate many of the issues surrounding the scalability of Massively Parallel Processing (MPP) System input and output systems. The national project consists of 18 application groups involving more than 30 researchers. The equipment provided by this award consists of I/O nodes for an MPP and a small MPP for crashable experimentation. The infrastructure will be available to all of the application groups for their experimental research needs.

Rutgers University; Naftaly H Minsky; *Law-Governed Coordination and Access Control in Distributed Systems*; (CCR-9626577); $132,578; 24 months.

Computing increasingly involves interactions between autonomous and heterogeneous agents distributed throughout a network, agents that have been designed separately, by different people, with little, if any, knowledge of each other. For such disparate agents to work safely and harmoniously together, they must be able to trust each other to observe certain unifying principles and protocols. Unfortunately, in such an open system one cannot expect universal voluntary compliance with any given principle. If required, such a principle must be imposed on all components of a given system. A flexible, scalable, and fault tolerant mechanism for imposing such principles and protocols over open groups of distributed agents is being developed under this project. This mechanism, implemented as a toolkit called Moses, is based on the concept of law-governed interaction (LGI), under which all messages exchanged between the members of a given group of agents are governed by an explicit, declarative, and locally enforced set of rules, called the law of this group. Some of the the potential applications of LGI are being explored.

Iowa State University; Prasant Mohapatra; *An Integrated Processor Management Scheme for Parallel Computers*; (CCR-9634547); $92,447; 36 months.

The performance of massively parallel multicomputers depends largely on their processor management scheme. Unfortunately, the schemes employed in the current generation of commercial multicomputers do not efficiently utilize the system resources. Processor management is concerned with processor allocation and job scheduling. Processor allocation also affects the interprocessor communication latency. Researchers have addressed these issues independently of each other. However, it is essential to evaluate the combined effect of these issues on system performance. Thus, this project focuses on an integrated study of processor allocation, job scheduling, and interprocessor communication. The research objectives will be accomplished in three phases. The first phase encompasses the design and development of efficient processor allocation and job scheduling schemes while employing a suitable message routing algorithm for interprocessor communication. In the second phase, the integrated effect of allocation, scheduling, and interprocess communication on system performance is to be studied. The third phase of the proposed research is concerned with the experimental validation of these schemes.

University of California – Davis; Ronald A Olsson; *Concurrent Programming Language Support for Invocation Handling: Design and Implementation*; (CCR-9527295); $105,855;

24 months. (Joint support with the Programming Languages and Compilers Program)

This project examines new language mechanisms for accessing message invocations in message passing based concurrent programming languages. Current mechanisms for invocation selection control, including those that allow invocation parameters to be examined, are not expressive enough. In particular, they do not provide a simple and efficient way to examine pending invocations of an operation or allow the selection decision to be based on more than just a single invocations parameters. Examining invocations and performing cross-invocation or cross-operation selection are important in a number of real applications, including debugging, visualization, and scheduling algorithms. Using current mechanisms to solve these problems results in cumbersome and inefficient code. This project involves a mix of conceptual, design, implementation, and evaluation work. To make the work concrete, the new language mechanisms are being defined and implemented as extensions to the SR concurrent programming language. The results of the qualitative and quantitative evaluations are being used to refine the semantics and implementations of the new language mechanisms.

University of Illinois – Urbana; Daniel A Reed; ***Workshop on Software Tools for High-Performace Computing Systems, Chatham, Massachusetts, October 1996*****; (ASC-9612753); $10,000; 12 months. (Joint support with the New Technologies Program)**

The critical importance of robust, flexible, and efficient software tools has long been recognized, and has been a major discussion topic at several national meetings, including those organized by the PI. These meetings focused on the difficulties associated with building good tools, and the lack of infrastructure for tool development and test. This proposal is for a workshop to discuss design and implementation solutions to the lack of these software tools for high performance computing systems. The discussions will focus will focus primarily on issues of debugging, performance measurement, analysis, and presentation as they relate to these needs. By including academic researchers, vendors, and government agency representatives, a wide variety of opinions, viewpoints, and needs can be presented and explored to the mutual benefit of all the groups.

University of Arizona; Richard D Schlichting; ***Designing and Implementing Adaptive Distributed Systems*****; (CCR-9633336); $100,000; 24 months. (Joint support with the Software Engineering Program)**

An adaptive computing system is one that evolves its behavior based on changes in the environment. For an adaptable system built on a distributed architecture where multiple machines are connected by a network, such changes can range from site failures to updated application requirements. The evolutionary action taken when a change occurs can vary as well, ranging from dynamically reconfiguring the system software to maintain the same service guarantees under changed conditions, to negotiating new service guarantees with the application. The goal of this project is to develop fundamental new techniques for designing and implementing adaptive distributed computing systems. The specific focus is on providing a distribution support layer, a software layer implementing strong execution guarantees regarding the distributed state of the application, that can adapt to changes. The research concentrates on two lines of investigation: (1) the development of a conceptual model for systems of this type and (2) a new approach to implementing adaptive distributed systems based on fine-grain code modules called micro-protocols. The framework is being extended to support dynamic evolution of software, including reconfiguration, dynamic code linking, and runtime code generation.

Emory University; V. S Sunderam; ***Interfaces for Parallel and Distributed I/O*****; (CCR-9523544); $95,431; 24 months.**

The PIOUS system is an experimental software architecture for efficient data management and I/O in heterogeneous parallel and distributed computing environments. PIOUS offers scalable concurrency control with standard consistency semantics, presents global or segmented views of parallel files to applications, and declusters physical data for improved performance. Concurrency control in PIOUS is based on transactions, and permits either volatile or stable file access semantics for increased throughput or resilience to system failures. The goal of this project is to complete and implement the PIOUS prototype system on the PVM message passing substrate. The system will consist of an API library that is designed to present familiar file system syntax and semantics, while permitting multiple concurrent accesses and declustering across multiple disk devices. The PIOUS infrastructure will then be further enhanced by designing appropriate file system primitives that will enable the efficient implementation of high level file access semantics and permit large structured and heterogeneous data sets to be efficiently stored and accessed in distributed networked environments. An appropriate file model will be defined, based on an extensible and flexible, multilayered architecture, and will specify a set of access interfaces; these methods will be realized as a suite of portable and efficient runtime libraries and support software.

University of Texas – Austin; Harrick M Vin; ***CAREER: Designing an End-To-End Architecture for Distributed Multimedia Systems*****; (CCR-9624757); $99,034; 24 months.**

The main objectives of the proposed research are: (1) to develop an integrated multimedia computing environment that supports efficient mechanisms for storage, transmission, and presentation of information in multiple forms (e.g., video, audio, images, text, and numeric data); and (2) to utilize it to develop innovative methods for delivering education. To achieve these objectives, the proposed research promises to develop an end-to-end architecture consisting of: (i) a multimedia file system that supports efficient means for storage and retrieval of multi-resolution multimedia objects, (ii) network algorithms and protocols for efficient transmission of multimedia objects over networks, and (iii) processor scheduling mechanisms as well as transport and higher layer protocols that support efficient processing and streaming of data at the end-station. The end-to-end system architecture thus created will form the basis for: (a) creating a digital repository of educational material containing video-lectures, notes, presentation slides, simulation packages, animation, images, etc., on several inter-disciplinary topics, and (b) providing efficient methods for organizing and browsing through the information space. In addition to being a realistic testbed for evaluating the end-to-end system architecture, this digital library prototype will provide an excellent environment for evaluating the effectiveness of various methods of distance learning and self-paced education. The findings of this research will significantly impact the design of large-scale digital libraries and distance learning tools of the future.

Rice University; Willy Zwaenepoel; *An Integrated Compiler-Runtime Approach to Software Distributed Shared Memory*; (CCR-9521735); $175,000; 24 months.

This research effort focuses on an integrated compiler-runtime approach to support software distributed shared memory (DSM). The goal is to achieve performance close to hand-tuned message passing programs while still presenting the user with a shared memory programming model. The work is primarily directed towards "networks of workstations", but it is also applicable to dedicated distributed memory machines like the IBM SP-2. Notwithstanding advances in the design and implementation of software DSM systems over the last few years, message passing has remained the most commonly used method for programming distributed memory machines, because message passing systems, in general, perform better than DSM systems. This project addresses this issue by building an integrated compiler-runtime DSM system, including the optimization of explicitly parallel shared memory programs through compiler analysis. Specifically, the ability of the compiler to predict access patterns, rather than merely react to them as the runtime system does, will allow prefetching and aggregation of messages, and will avoid unnecessary data movement and consistency actions. The work will be based on the TreadMarks DSM system developed at Rice. The proposed work will contribute to the better utilization of networks of workstations and, more generally, distributed memory machines for parallel computing. Earlier work on TreadMarks has already contributed to that goal, in that it has produced one of the most advanced runtime systems for DSM, that has been used by other researchers for work in computational genetics, mixed integer programming, and seismic modeling.

Security, Reliability and Fault Tolerance

University of Texas – El Paso; Glenn A Gibson and Sergio D Cabrera; *Porting and Enhancement of an Architecture Simulation Package to a Workstation Environment*; (CCR-9522810); $50,000; 24 months.

This is a NSF-CONACyT collaborative award. The primary objective of the project is to enhance and port a software simulation package, SIMARC, to a Sun workstation network. The purpose of the SIMARC package is to enable the user to quickly develop Modularly Configured Attached Processor (MCAP) architectures and simulate the execution of various algorithms on these architectures. The SIMARC package has already been developed at the University of Texas at El Paso (UTEP) for IBM PC compatible computers. The simulator and assembler portions of SIMARC would be ported by a group at UTEP and the architecture editor portion would be ported by a group at the Instituto Technologio de Estudios Superiores Monterrey Campus Ciudad de Mexico. Also both groups would use the package to develop high-performance architectures for executing signal processing and communications algorithms. The principal result of the project would be a software package that could be used for the fast prototyping of customized attached processors designed to efficiently execute specified sets of algorithms.

University of Michigan – Ann Arbor; Farnam Jahanian; *CAREER: Probing and Fault Injection of Distributed Systems*; (CCR-9502341 A001); $5,000; 0 months.

This research will investigate fault injection techniques for the evaluation and validation of dependable protocols. The project will focus on techniques that can be used to demonstrate three aspects of a target protocol: (i) detection of design or implementation errors, (ii) identification of violations of protocol specifications, and (iii) insight into design decisions made by the implementors. The end goal of the project is the investigation of techniques for testing the fault-tolerance capabilities of dependable protocols and the development of a tool for the experimental evaluation of commercial and prototype protocols. The approach will center around a technique called 'script-driven probing and fault injection.' A distributed protocol can be viewed as an abstraction through which a collection of participants communicate by exchanging a set of messages. Hence, no distinction is made between application-level protocols, communication protocols, or device layer protocols. A probe/fault injection (PFI) layer is inserted below a target protocol to filter and to manipulate the messages that are exchanged between participants. The fault injection layer supports the execution of deterministic or randomly-generated test scripts to probe the participants and inject various faults into the system. In particular, by intercepting messages between two layers in a protocol stack, the fault injection layer can delay, drop, reorder, duplicate, and modify messages. Furthermore, it can introduce spontaneous messages into the system to probe the participants and to orchestrate the system execution into a particular path. Since timing constraints are often an important component of the dependability requirements on a system, the intrusiveness of a fault injection mechanism on a target system is a key issue which will be addressed by this research. The educational component of this CAREER project will include 'undergraduate mentoring' and 'learning via experimentation.' Mentoring exposes students to one-on-one technical interactions with instructors and more senior students. Learning via experimentation has a two pronged effect. It allows students to 'practice' how to apply knowledge and to'learn' how to obtain new information. These two aspects of the plan will be formalized into the curriculum with the following educational activities: (1) Refocus the Undergraduate Directed Study to involve junior and senior-level students in some of the ongoing prototyping activities in research projects and in various laboratories on campus; (2) Develop a curriculum track that emphasizes experimentation to enhance learning of software systems concepts.

University of Pittsburgh; Daniel Mosse'; *Research Initiation Award (RIA): An Integrated Framework for Fault-Tolerant Applications in Real-Time and Non-Real-Time Systems*; (CCR-9308886 A003); $4,950; 0 months.

This research studies a new model of computation, which extends the object-oriented model with timing constraints and resource requirements specifications. A set of scenarios (a set of applications executing at a time) and a scenario change mechanism that allows for different sets of applications to be executed according to the situation are defined. These scenario changes are triggered by an error handler, by an operator or by a user. Scenario changes also allow uniformity in multiple types of fault-tolerant behavior such as graceful degradation, load-shedding, etc.. The latency of these scenario changes and quantify their applicability to real-time and non-real-time systems are studied. Methods are developed for the detection of situations in which these scenario changes can be applied, and develop a theory for the adaptation of these scenario changes to non-time-constrained applications and systems. More accurate models are developed based on real-life workloads and situations, to model realistic fault models, encompassing transient as well as permanent faults. These are of temporal or physical nature (e.g., transient overloads, site crashes or link failures). The implementation of a prototype of a fault-tolerant real-time system that supports this model of computation follows.

SRI International; Xiaolei Qian; *Trusted Interoperation of Health-Care Information Systems*; (CCR-9422688 A001); $435,461; 12 months. (Joint support with the CISE Institutional Infrastructure Program, the Software Engineering Program, and the Computational Engineering Program)

This award is made in support of the NSF initiative ENG-CISE-IITA94, National Challenges. SRI International, in collaboration with Stanford University Department of computer Science and Medicine, and Inova Health System, will develop and demonstrate techniques for the trusted interoperation of autonomous heterogeneous health care databases containing sensitive data that mismatch in semantics, representations, and security/privacy policies. Semantic connections between component databases are encoded in a knowledge base, which are then used by a mediator to automatically resolve mismatches and transform queries between component databases using the local language and schema, making a local database access data in multiple databases using the local language and schema, making interoperable both the data and the applications accessing the data in legacy databases. Technique will be developed and demonstrated for solving three critical problems: (1) A security/Privacy Policy Framework. What is the relationship and difference between security and privacy? How can the security/privacy policy of a health care database be characterized and formalized? How can we compare the security/privacy policies of autonomous heterogeneous health care databases? (2) Trusted Query Mediation. When autonomous heterogeneous health care databases are interoperating through query mediation, how can we guarantee the autonomy and security/privacy of component databases. (3) Schema Mediation. In constructing a physician's workstation as a uniform interface to multiple autonomous heterogeneous health care databases, how can a consistent view (schema) be constructed that (a) captures exactly the data that are relevant to users of the workstation, and (b) resolves all the mismatches in the relevant data? These techniques will be implemented in a demonstration system that leverages off an existing one developed for the query mediation framework.

Cornell University; Sam Toueg; *Unreliable Failure Detectors for Reliable Distributed Systems*; (CCR-9402896 A001); $70,000; 12 months. (Joint support with the Computer Systems Architecture Program and the Theory of Computing Program)

The starting point for this research is a fundamental problem in fault-tolerant distributed computing: reaching agreement among processes in a system that is subject to failures. It is well-known that this problem, called Consensus, has no deterministic solution in asynchronous systems, even if it is assumed that communication is reliable and no more than one process may fail. The impossibility of solving Consensus (and other related problems such as Atomic Broadcast) is one of the most severe obstacles to implementing fault-tolerant applications in asynchronous systems. In recent work the PI has introduced a novel approach to circumvent such impossibility results: he showed that unreliable failure detectors can be used to solve Consensus (and Atomic Broadcast), even if the information that they provide about failures is highly inaccurate, e.g., even if they make an infinite number of mistakes. Since such failure detectors can be implemented in realistic distributed systems, and since various considerations make the asynchronous models especially attractive, this work suggests an approach to fault-tolerance that is viable in practice. The objectives of this research are to broaden the applicability of this approach by removing the limitations of the earlier work, and to explore in more concrete terms its practicability. Specific goals include: (1) tolerating communication failures, including network partitions (the earlier work assumed reliable links); (2) tolerating process failures of various types (the earlier work dealt with crash failures only); (3) considering shared-memory systems (the earlier work dealt with message-passing systems); and (4) solving other problems that are central to fault-tolerant distributed computing, including Group Membership and Group Multicasts (the earlier work solved Consensus and Atomic Broadcast). In order to assess the cost and benefit of using unreliable failure detectors complexity questions are also explored. Finally, the practicality of this approach is validated by implementation on an experimental platform.

Michigan State University; I-Ling Yen; *Processor Specialization in Fault-Tolerant Distributed Systems*; (CCR-9634249); $115,343; 24 months.

The goal of this project is to develop a scaleable fault tolerance framework that can be customized to economically cover a wide range of fault models, applications, and computational environments. The system is based on the N-modular redundancy (NMR) concept to uniformly handle all types of failures. A specialized NMR (SNMR) protocol that can tolerate malicious failures has been developed to improve the system performance and to enhance its scalability. In this research, the SNMR protocol will be enhanced in several ways, including adaptations for wide- area network applications and for efficiently tolerating general and fail-stop failures. By adjusting the parameters, such as the replication level and error-checking algorithms, the system can be customized to satisfy various application requirements and cost constraints. A prototype of the SNMR core will be implemented in the two-year period. The customizability, applicability, performance, fault coverage, and scalability of the SNMR system will be thoroughly studied and compared with existing approaches.

Performance Evaluation, Scheduling, and Resource Management

University of North Carolina – Chapel Hill; James H Anderson and Kevin Jeffay; *Object-Sharing Technology for Real-Time Systems*; (CCR-9510156 , A001); $110,329; 18 months. (Joint support with the Theory of Computing Program)

The objective of this project is to develop an object-based framework for synchronization and interprocess communication in real-time computer systems. This framework will integrate work on scheduling theory with work on delay-resilient shared objects. The term “scheduling” refers to the problem of sequencing tasks on a shared execution unit such that it is possible to determine a priori if the timing requirements of tasks are respected. A variety of scheduling policies will be considered in this project, spanning the spectrum from hard to soft real-time systems. By focusing on delay-resilient shared objects, this project represents a significant point of departure from past work on real-time synchronization. Delay-resilient objects are distinguished by the fact that they are not locked. The notion of a delay-resilient shared object builds upon recent promising work on wait-free and lock-free synchronization in shared-memory multiprocessors. Although such work has been ongoing for several years, this project represents the first comprehensive effort to integrate such objects within an object-sharing framework for real-time computing. From a real-time perspective, the key advantage of delay-resilient objects is that they do not give rise to priority inversions. Traditional object-sharing approaches for real-time systems rely on significant operating system support to deal with this problem. The proposed approach does not require such support, and thus entails lower synchronization overhead than traditional approaches. Another advantage of the proposed approach is that it can be applied across a variety of diverse platforms. Target platforms for this project include uniprocessors, shared-memory and message-passing multiprocessors, and workstation clusters.

Columbia University; Leana Golubchic; *CAREER: Towards a New Generation of Multimedia Storage Systems*; (CCR-9625013); $100,000; 24 months.

Efficient storage organization and access is of central importance to the advent of large scale multimedia systems. Multimedia applications, ranging from video-on-demand (VOD) to medical tele-presense or large scale scientific computations, involve storage and retrieval of large volumes of information subject to a spectrum of real-time constraints. Current storage systems are inadequate to address the needs of multimedia systems. Indeed, various recent experiments in delivering VOD to populations of several thousand customers have established storage access as a key barrier to success. There is thus a growing need for a new generation of high-performance storage technologies that can efficiently support large scale multimedia systems. The main goal of the proposed research is to develop novel architectures and resource management algorithms for multimedia storage systems that will result in orders of magnitude performance improvements. The main thesis pursued here is that a storage system can exploit intrinsic, non-random, access characteristics of multimedia applications to accomplish dramatic performance improvements. This research effort will have several important contributions to the study of multimedia systems in general and to their storage technologies in particular. It will seek to characterize storage access patterns of the broad spectrum of multimedia applications and study techniques to take advantage of these characteristics in improving storage and access efficiency. The proposed study will consider the entire storage hierarchy, study the entire pool of resources and how these resources interact, as well as devise methods for incorporating application and workload constraints and characteristics into resource management techniques. Such media-aware, workload-aware, and application-aware storage organization and access mechanisms can significantly improve performance of a storage server.

University of California – Irvine; Sandra Irani; *Competitive Analysis of Problems in Resource Allocation*; (CCR-9625844); $166,809; 36 months. (Joint support with the Theory of Computing Program)

This project focuses on algorithms that work with limited information. This research investigates on-line algorithms in the context of various resource allocation problems in operating systems. This project is taking a twofold approach to each problem. The first approach is to perform a theoretical study using competitive analysis. The competitive measure is used to evaluate on-line algorithms by measuring their performance in relation to the optimal off-line algorithm which has available to it complete information about the problem instance at hand. This effort plans to perform empirical tests to evaluate the performance of various on-line algorithms in a realistic setting. The first two problems addressed in this project are related to memory management. The first of these addresses ways in which information about the access pattern of a program available at compile time can be used to guide page replacement and prefetching policies. The investigator will develop methods which can take a program and automatically generate appropriate memory management instructions to be inserted in the code. The second problem addresses allocating memory resources among users in a multiprocessing environment. This work addresses this problem for general multiprocessing environments as well as methods specifically designed for database systems. The third problem examines preemptive load balancing strategies over local area networks under a probabilistic model for process duration. The last problem is an investigation of disk scheduling strategies. This work is motivated by the fact that the availability of larger cheap memory has made it possible to have long queues of disk updates which means that more sophisticated scheduling policies can have a greater impact on disk utilization.

University of Illinois – Urbana; Klara Nahrstedt; *CAREER: Time-variant QoS Management*; (CCR-9623867); $100,000; 24 months.

The proposed research is directed towards the advancement of the exploration of complex multi-dimensional timing control for multimedia streams in distributed multimedia systems. To achieve this goal, this research effort focuses on the exploration and development of a true generic system for the advanced timing control of resources used by multimedia streams. Timing control is particularly important for interactive multimedia distributed applications such as telerobotics, teleoperation in manufacturing, and virtual reality applications with force feedback including telesurgical training and audio-video conferencing. The focus of this research is on the study of algorithms and techniques to provide distributed control of multi-dimensional timing behaviors over different resources in an integrated fashion. While there have been advances in the control of the one dimensional case (timely delivery), and of the two- dimensional case (media synchronization), there has been little improvement in providing a third dimension in which the systems can control the adaptation of timely delivery and/or media synchronization specified by quality of service (QoS) parameters such as end-to-end delay and jitter. This research addresses integrated time-variant QoS management for the above defined three dimensions. To achieve this integration, the research effort calls for: development of concepts for time-variant QoS management; development of a theoretical framework for formal specifications for time-variant QoS and the modeling of distributed control to accommodate QoS requirements; provision of an experimental testbed for multimedia transmission according to QoS requirements; and, evaluation of the framework through an experimental verification.

Stanford University; Mendel Rosenblum; *NSF Young Investigator (NYI) Award: Storage Management for High Performance Computer Architecture*; (CCR-9257104 A004); $62,500; 12 months.

This project addresses research in storage management for high performance computer architectures. There is need for technology for comparing storage managers relative to a given workload characterization to determine which will perform best. Research techniques to accomplish this are based on a combination of analysis, simulation, and prototype implementations. Applications such as audio and video are imposing new requirements for performance and guaranteed response times. The introduction of object oriented systems introduces new and different request patterns. This illustrates the direction of new hardware technology and new application demands that will provide a focus for the research.

State University of New York – Buffalo; Min-You Wu; *Parallelization of Scheduling Algorithms*; (CCR-9625784); $101,438; 24 months.

Static scheduling utilizes the knowledge of problem characteristics to reach a global optimal, or near optimal, solution. Although these scheduling algorithms apply to parallel programs, the algorithms themselves are sequential, and are executed on a single processor system. The memory and speed of a single processor limits the size of the graph that can be handled. Using multiprocessors to schedule tasks to multiprocessors is a natural solution to solve these problems. In fact, a scalable scheduler is not feasible without parallelizing the scheduling algorithm and running it on a parallel computer. There are two major data domains in a scheduling algorithm, the source domain and the target domain. The source domain is the macro-dataflow graph; the target domain is the schedule for target processors. Two approaches are proposed, the vertical scheme and the horizontal scheme. The advantages and disadvantages of the two approaches will be studied. This research has already accomplished the primary stage. New parallel scheduling algorithms will be developed. An incremental parallel scheduling is also proposed for applications with dynamic and unpredictable features. The goal of this project is to investigate this approach and evaluate its performance for some important applications, and to integrate these results into a toolset that will provide support for scheduling of dynamic and irregular problems.

Application of Formal Techniques

Auburn University; Jennifer McManis; *Research Planning Grant (RPG): Software Tools for Verification of Hybrid Systems*; (CCR-9630022); $18,000; 18 months.

The hybrid automaton is a promising model for studying the behavior of hybrid systems. One drawback of this model is the difficulty in deriving decidable or semidecidable procedures. The suspension automaton is a subclass of linear hybrid automata for which decidable language containment procedures exist. The objective of this project is to investigate the use of suspension automaton to accurately model and verify real-time systems. Due to the complexity of these problems, a methodical way to study them is called for, such as would be provided by a software tool. This project is designed to explore the issues involved in building an implementation of verification techniques based on the suspension automaton. Two specific issues must be addressed. First, the relation of this work to existing work, particularly in the area of real-time systems, must be established. Second, a specific approach to building such a tool must be formulated. Both of these goals may be aided by establishing contact with researchers in the real-time area, particularly those who are have produced software tools to study real-time systems.

University of Texas – Austin; Jayadev Misra; *Modular Multiprogramming*; (CCR-9504190 A001); $10,000; 0 months.

Two major goals in multiprogramming research are: (1) to design and understand the modules (e.g., processes or data objects) of a program in isolation, without considerations of interference by the other modules, and (2) to implement the modules on separate processors with a fine grain of interleaving so that no processor is ever locked out of accessing common data for long periods of time. Further, it is essential to devise a model in which the distinction between computation and communication is removed; in particular, the methods for designing and reasoning about the interfaces should be no different from those employed for the computations at the nodes of the network. The project is based on the hypothesis that many of the programming concepts developed for databases, object-oriented programming and designs of reactive systems can be unified into a compact model of distributed programs that meets the desired goals. In the past, much effort has gone into limiting interference among the modules by employing a variety of synchronization mechanisms: locks or semaphores, critical regions, monitors and communication. In this project, no specific communication or synchronization mechanism, except procedure call, is built into the model. Yet, the traditional schemes for communication using bounded or unbounded channels, semaphores, and accesses to shared memory can be encoded as modules. This project uses two distinct kinds of procedures, to model sequential and concurrent aspects of programming. The programming model allows for disciplined interactions between these two types of procedures. The research seeks to: (1) develop the model, in particular for dynamic creations of modules (2) implement the model on top of an existing language, and (3) develop the appropriate formalism to specify and verify the properties of programs.

New York University; Robert A Paige; ***Improving Productivity of Algorithm and System Implementation in Scaled up Applications*****; (CCR-9616993); $50,000; 24 months. (Joint support with the Software Engineering Program and the Programming Languages and Compilers Program)**

Improving productivity and reliability of labor-intensive high-performance software is an important scientific problem with obvious economic ramifications. This research is testing the feasibility of a radically new compilation and meta-compilation methodology to improve productivity and reliability for high-performance algorithmic software in the C programming language. The research combines partial evaluation for deeply-nested SETL data structures, data structure and algorithm selection techniques, and transformational techniques from relational query optimization. This is leading to a statically typed variant of SETL from which it is expected large-scale C programs can be generated at a high translation rate.

University of Illinois – Urbana; Uday S Reddy; ***Formal Methods for Object-Based Computing*****; (CCR-9633737); $160,000; 24 months.**

Objects are programming abstractions that encapsulate mutable internal state and provide operations to manipulate the state. Object-based computing is the programming paradigm that uses objects to structure the state-manipulation carried out in programs. The goal of this project is to develop formal methods for specifying, conceptualizing, and reasoning about object-based programs. These formal methods include: (1) Type systems: what kinds of semantic entities are involved in object-based programs, and how does one capture them in type systems for programming languages? (2) Denotational semantics: what mathematical abstractions are involved in capturing the bahaviors of object-based programs? (3) Specifications: How does one specify the behaviors of individual objects? (4) Reasoning methods: what methods are appropriate and useful for proving properties of object-based programs? The Direction being pursued is to build on the several decades of research in functional programming so as to achieve a "tight integration" of programming concepts and mathematical tools in working with object-based programs. Of primary focus are algebraic specification and equational reasoning techniques using lambda-calculus-based higher-order language frameworks.

SRI International; John Rushby, Patrick D Lincoln, and Natarajan Shankar; ***An Integrated Framework for Mechanical Verification of Fault-Tolerant Systems*****; (CCR-9509931); $345,147; 36 months. (Joint support with the Software Engineering Program, the Computer Systems Architecture Program, the Numeric, Symbolic & Geometric Computation Program, and the Theory of Computing Program)**

Formal verification is a technique for reducing certain questions about the requirements, designs, or implementations of computer software and hardware to a mathematical form that can examined by methods having the character of calculation. Similar to the way in which questions about the properties of an aerofoil can be reduced to calculations that can be mechanized by computational fluid dynamics, mechanized formal verification uses automated theorem proving and related methods (e.g., model checking) to perform logical calculations about the properties of computer systems. Whereas testing can examine only some of the behaviors of a complex system, mechanized formal verification can examine all the behaviors of a model of the system (again, the situation is analogous to flight test vs. computational fluid dynamics). Using mechanized formal verification, suitably realistic models can provide strong evidence for the correctness of computer systems, and simplified models are often useful for detecting bugs that testing does not encounter. Mechanized formal verification is most effective when applied to the hardest and most critical elements of design. An area where it holds great promise is in the analysis of fault-tolerant systems, where the presence of faults can produce a complexity of behavior that defies other forms of analysis. The increased assurance provided by mechanized formal verification can be of particular value for fault- tolerant systems because these are generally found in applications that are critical to safety (such as aircraft and other vehicles) or national infrastructure (such as the telephone, power-distribution, and banking systems). The objectives of this research are to reduce mechanized formal verification of fault-tolerant distributed algorithms from tours-de- force to routine practice, and to develop a framework for mechanical verification of the architectures of complete fault-tolerant systems composed from such algorithms. The research focuses on developing systematic methods for the formal specification of classes of fault- tolerant algorithms, and mechanized strategies for their exploration and verification. These strategies will build on techniques that combine decision procedures and model checking within an interactive framework. The algorithms considered will include those for distributed consensus, synchronization, diagnosis, and group membership. The work is expected to yield improved understanding of the assumptions, constraints, and fault models underlying the algorithms studied, more refined arguments for their correctness, and coherent arguments for their synthesis into an overall architecture. In addition, the research is expected to yield new results in the construction and application of formal verification systems, and improved methods for the formal representation and analysis of fault- tolerant algorithms.

Rice University; Moshe Y Vardi; ***Development of Reliable Distributed Protocols*****; (CCR-9628400); $130,654; 24 months.**

This project focuses on the verification aspect in the development of reliable, fault-tolerant, distributed protocols. High-performance computing technology, such as distributed real-time systems, high-performance memory systems, wide-band I/O systems, modular operating systems, high-speed network interfaces, commodity workstation clusters, and the like, is reliant on an underlying layer of distributed software (often implemented in hardware), which consists of suites of protocols for check-pointing, communication, coordination, data distribution, recovery, and synchronization of distributed components. These protocols provide a critical software layer for high-performance computing systems. Having such a standardized layer is crucial to interoperability, portability, scalability, and reusability of existing software. Unfortunately, developing reliable software for distributed environments is a notoriously difficult task. Here reliability refers to pre-committed real-time performance, dynamic reconfigurability, security, robustness, and high assurance. The difficulty in producing reliable distributed software is twofold. First, designing reliable distributed software is extremely challenging. Not only does the software have to be free from programming errors, but it also has to be resilient against hardware and software faults. Second, because of the nondeterministic behavior of distributed environment, it is very difficult to verify that the software meets its desired specification. The focus of this project is on the study of algorithms and data structures applicable to the verification of programs. It draws upon ideas and results from logic, graph theory, and automata theory, combining both theoretical and experimental aspects.

PROGRAMMING LANGUAGES AND COMPILERS

Programming Languages

University of California – Berkeley; Alexander Aiken; ***NSF Young Investigator (NYI) Award: Constraint-Based Program Analysis*****; (CCR-9457812 A002); $62,106; 12 months.**

In this project the work is focused on the design and implementation of advanced set constraint based program analysis algorithms, and their application to program optimization for a wide variety of programming languages. The basic theory, key algorithms, implementation techniques, and applications of the areas are simultaneously explored. Within the teaching plan, the constraint-based program analysis system is further developed and disseminated in support of the graduate programs. Work at the undergraduate level focuses on preparing students to work in the software industry.

University of Oregon – Eugene; Zena M Ariola; ***Research Initiation Award (RIA): Intermediate Languages Based on Graph Rewriting*****; (CCR-9410237 A002); $30,000; 12 months.**

This research pursues the development of a formal semantics for real implementations of functional languages, modeling sharing and cycles, and for parallel functional languages with state. Traditionally both the operational semantics and the compilation process of functional languages are given by translating them to either the lambda-calculus, combinatory logic or abstract reduction systems. However, these computational models do not capture sharing and direct recursion, and thus are inadequate for reasoning about compiler issues and implementations. This research develops a sound theory for cyclic structures involving lambda abstraction. The need for new mathematical foundations arises when state is introduced in functional languages. Previous theories apply only to languages with sequential semantics. This research seeks to provide a framework for reasoning about state in a parallel setting by extending the above graph calculus.

Carnegie Mellon University; Guy E Blelloch; ***NSF Young Investigator (NYI) Award: A Functional Data-Parallel Language for High Performance Computers*****; (CCR-9258525 A004); $25,000; 12 months.**

Current research on a first-order functional language (NESL) that supplies parallelism through a set of data-parallel constructs based on vectors is extended in several directions. One is the testing of the language via scientific applications algorithms. Another is the addition of polymorphic types to the language. Yet another is the examination of lazy load-balancing based upon dynamic testing. In addition, implementations on several newer high performance machines is carried out.

University of Minnesota – Twin Cities; Daniel Boley; ***Robust Fault Tolerance for Computations in Linear Algebra and Signal Processing*****; (CCR-9628786); $59,862; 24 months. (Joint support with the Computer Systems Architecture Program)**

The objective of this project is to develop and test fault tolerant algorithms for numerical computation in various domains, including sparse matrices, recursive filtering and signal processing. The methods used are a combination of checks made during the computation based on preconditioned checksums and thresholds, and a-posteriori checks made on the final computed results based on backward error assertions (BEAs). In spite of the presence of rounding errors, BEAs allow very tight, cheap, validation of the final results independent of the specifics of the implementation or sources of error, and checksums and thresholds supplement the BEAs by providing checks during the computation. These methods will be extended beyond prototype examples to more realistic large computations, demonstrating their scalability and applicability to large domains.

University of Washington; Alan H Borning; ***Constraint Imperative Programming*****; (CCR-9402551 A002); $62,000; 12 months.**

This project continues work on Constraint Imperative Programming (CIP) languages. A constraint represents a relation that should be maintained by a computer program. Constraints have a number of desirable properties that make them useful in building systems such as interactive graphical applications. They are declarative statements, rather than imperative statement; a programmer need only to state "what" rather than "how". Additional advantages result from embedding support for constraints in the programming language itself. First, it becomes convenient for programmers to use constraints. Second, the language implementation can ensure that the constraints are checked automatically. Third, a more uniform and elegant language can result from expressing standard constructs as constraints. Finally, more opportunities for optimization are opened to the compiler. In this project, the activities include investigating techniques for compiling object-oriented constraint imperative languages, such as Kaleidoscope, to produce efficient code. To accomplish this, the system does as much compile time analysis of the constraints as possible. Other activities include developing tools for writing, browsing, and debugging constraint imperative programs, investigating the compilation of Kaleidoscope programs to run on shared memory MIMD machines, and testing the utility of this class of languages by programming a set of representative applications. Another important focus is the development of a more efficient and powerful constraint satisfaction algorithm. A final activity is work on debugging support for constraint systems. The work on constraint satisfaction algorithms and debugging is useful both to support constraint imperative languages, as well as other kinds of constraint-based systems.

Wake Forest University; Jennifer J Burg; ***Research at Undergraduate Institution (RUI) : Strategies for Fast Execution of Constraint Logic Programs Over a Real-Number Domain*****; (CCR-9408298 A001); $5,000; 0 months.**

Constraint logic programming is a declarative programming paradigm which has proven to be expressive for computationally difficult problems such as scheduling, air traffic flow management, simulation of dynamic physical systems, solution of differential equations, and options trading analysis. However, while CLP languages are expressive, their execution is not always efficient enough for the problems to which they apply. The goal of this research is to develop strategies for fast execution of constraint logic programs, particularly in the domain of real or rational arithmetic constraints. To this end, incremental Gaussian elimination-based and simplex-based constraint satisfaction algorithms are compared for their efficiency in the context of CLP; the feasibility of intelligent backtracking and independent and-parallelism is investigated; work on asynchronous parallel execution of the inference engine and solver is continued; and parallel incremental algorithms for constraint solving are being developed. The research program has strong undergraduate involvement and research components. Students are involved in the design, implementation, and testing of algorithms, and are exposed to concepts from the theory of computation, programming language design, the linear programming problem, AI search and backtracking strategies, and parallel computation.

Rice University; Robert S Cartwright and Matthias Felleisen; ***Smart Programming Environments*****; (CCR-9633746); $159,980; 24 months. (Joint support with the Software Engineering Program)**

Type safety is an essential feature of advanced programming environments for the rapid development of reliable software, but it is not sufficient. Programmers try to eliminate as many errors as possible by reasoning about the program's behavior. To support this activity, this research seeks to design and construct smart programming environments. A smart programming environment is a system for writing, editing, executing, and analyzing programs. For the latter, the environment includes "soft type checkers", tools for verifying simple program invariants. Most notably, these are the preconditions for each program operation. Two prototypes of soft type checkers for Scheme have been developed, based on different analysis methods. One uses an adaptation of Hindley-Milner type inference (for a non-standard type algebra); the other uses a form of set-based analysis. This research would extend the tools, and seeks (1) to develop a new class of soft types that combine the advantages of Hindley-Milner and set-based analysis, (2) to incorporate additional verification technology, (3) to develop alternative interfaces for exploring the verifier's reasoning, and (4) to extend the tools for modular programming. In addition, ways are being explored to transfer the technology to lower-level languages, such as a safe sublanguage of C++ (similar to Java) in the programming environment.

Brandeis University; Jacques Cohen and Timothy Hickey; ***Parallel Processors for Constraint Languages*****; (CCR-9403427 A002); $68,936; 12 months.**

The focus of this project is the development of compilers and interpreters for Logic and Constraint Programming (CLP) languages executing on sequential and parallel machines. Prolog is an example of such a language, and it offers the users capabilities of solving problems in domains including reals, booleans and strings. The four specific topics of the project are: (1) Constraint Parallelism. Constraint solvers for several domains (including integers, reals and booleans) are designed, implemented and analyzed on parallel machines. These solvers are used to prototype constraint parallel Prolog systems; (2) Program Analysis and Transformations. Tools for the analysis of parallel constraint and logic programs are developed. Some of the techniques that are considered are partial evaluation and abstract interpretation based global analysis; (3) Extension of CLP. High-level, C-like languages based on constraints and interval arithmetic are designed, implemented and analyzed. The possibility of integrating, functional, set-theoretic and CLP languages in a single higher order logic programming framework is also studied; and (4) Applications of Parallel CLP. Several classes of parallel CLP applications are considered, including genome parsing, combinatorial search problems (such as scheduling), and optimization of non-linear constraint systems.

University of Arizona; Saumya K Debray; ***Scalable Parallel Implementations of Declarative Programming Languages*****; (CCR-9502826 A001); $60,177; 12 months.**

This project will develop compilation and implementation techniques for declarative programming languages on distributed memory architectures. The project consists of three parts. First, it will develop compile-time techniques for cost analysis of message-passing implementations of declarative languages. Second, it will develop and implement algorithms for granularity control and task partitioning of languages with fine-grained implicit concurrency on distributed memory systems. Finally, the project will develop compile-time and runtime techniques for efficient execution of declarative languages in the context of multi-level memory hierarchies.

New Mexico State University; Gopal Gupta; ***Implementation Techniques for Parallel Logic Programming: Systematic Development of Parallel Prolog Engines*****; (CCR-9625358); $99,992; 24 months.**

An important property of logic programming languages, such as Prolog, is that their declarative semantics is largely independent of the order of execution of different operations in the program. The different operations can also be performed in parallel. As a result, it is possible to exploit substantial parallelism from Prolog programs, written for sequential machines, without making any modifications to them. There are various forms of implicit parallelism found in Prolog programs: or-parallelism, independent and-parallelism, dependent and-parallelism, stream and-parallelism, data or-parallelism, data and-parallelism. Research so far has only focused on exploiting a single form of parallelism. This project's objective is to develop parallel implementation techniques that will allow one to extend a high performance sequential Prolog system to exploit the various forms of parallelism simultaneously. The focus of this project is upon self-contained parallel implementation techniques that are simple and that will allow addition of the various forms of parallelism to an existing sequential Prolog system in a systematic and incremental way. The techniques to be developed here will be tested and demonstrated for efficacy by incorporating them in the SICStus sequential Prolog engine. Relevant tools for compile-time analysis will also be developed. General runtime optimization techniques for reducing the parallel implementation overhead will also be researched.

Association Computing Machinery; Susan Horwitz; ***Special Project: Group Travel Grant for Faculty at Minority Institutions and Women's Colleges to attend PLDI 96*****; (CDA-9625408); $20,000; 4 months. (Joint support with the Program)**

As a step toward improving the representation of women and minorities in computer science, the ACM Special Interest Group on Programming Languages (SIGPLAN) is working to improve the computer science background of faculty members from colleges with a large minority and/or female student enrollment. This travel grant provided partial support for selected individuals from such institutions to attend the annual conference on Programming Languages Design and Implementation (PLDI). In addition to the conference, the participants attended tutorials targeted specifically to teachers from non-research oriented institutions. For the selected participants, SIGPLAN provided for the tutorial and conferences fees and NSF funds provided for travel and subsistence costs.

Association of Computing Machinery; Mary J Irwin; ***The 1996 Federated Computing Research Conference Philadelphia, PA*****; (CCR-9628276); $28,240; 9 months.**

The Federated Computing Research Conference (FCRC) is an ensemble of 16 small computer science research conferences, symposia and workshops, sponsored by four different professional societies and collocated in Philadelphia, PA from May 20 through 28, 1996. Many of the constituent meetings are held on an annual basis. The purpose of the Federated Conference is to stimulate interaction among scientists in sub-disciplinary areas who might otherwise not communicate with one another. Registrants at FCRC are allowed and encouraged to attend any and all sessions of the constituent meetings. The first FCRC was held in 1993 in San Diego, CA and proved to be successful. The content of FCRC was enriched by five plenary sessions featuring internationally distinguished, invited speakers.

Loyola University of Chicago; Radhakrishnan Jagadeesan; ***CAREER: ATDCC – A Language for Communicating Reactive Process*****; (CCR-9501943 A001); $34,430; 12 months.**

The focus of this CAREER research project is the design, semantics, and implementation of ATDCC, a language for a network of asynchronously communicating reactive programs. Reactive systems are tightly coupled, event-driven systems that react continuously with their environment at a rate controlled by the environment. Asynchronous systems are those in which processes are loosely coupled and communication can take arbitrary amounts of time. Complex applications, arising commonly in practice, typically require the combination of both these ideas, thus motivating ATDCC. ATDCC is expected to have the following advantages: (1) support for hierarchical construction of programs/specifications; (2) good foundational properties; (3) efficient implementation as a network of asynchronously communicating automata; (4) verification of safety properties, including real time constraints. This research is aimed towards the creation of a complete programming environment for modeling and analyzing complex electromechanical systems. This CAREER project includes active involvement in education, in conjunction with the research program. The research program itself involves the participation of students, who will benefit from the exposure to industrial experience. The education plan includes course and curriculum development, for both graduate and undergraduate students, incorporating current research in the general area of interest in the project. An undergraduate parallel programming laboratory is planned to collaborate, disseminate and share resources with institutions in the Chicago area. The laboratory and curriculum development are intended to lead to undergraduate participation in research.

Kestrel Institute; Richard Jullig; ***International Workshop on Formal Foundations of Software Systems: Rio de Janeiro, Brazil, May 5-9, 1997*****; (CCR-9617781); $37,200; 6 months. (Joint support with the Software Engineering Program, the Numeric, Symbolic & Geometric Computation Program, and the Theory of Computing Program)**

This grant supports an international workshop on Formal Foundations of Software Systems in Rio de Janeiro, Brazil, May 5-9, 1997, to explore future directions in formal methods, programming logics, and programming languages. The workshop combines 3 days of technical exchange with 2 days of discussions to produce recommendations to the science funding agencies of the U.S. and Brazil: NSF, CNPq, and FINEP. This workshop's goal is to promote collaboration among Brazilian and U.S. researchers and provide information and recommendations for future research efforts and funding initiatives.

Stanford University; John C. Mitchell; ***Object Systems: Programming Languages and Software Security*****; (CCR-9629754); $239,735; 36 months.**

Both the science of software production and software security are considered important areas, with increasing concern for security likely to lead to increasing interest in relevant software analysis methods. This project addresses problems in the areas of object-oriented programming languages and software security. The goals are to simplify the structure and process of producing software systems and increase their resistance to certain forms of insecurity. The techniques are both analytical, drawing on ideas from type theory and mathematical logic, and experimental, involving the implementation of language designs and experimental development of prototype systems. One main focus of the work is the design and implementation of an extension of Standard ML with object-oriented features. Similar language ideas are also considered for C++ or Java variants, or design or debugging tools. An experimental evaluation of either language may involve design or redesign of existing software in order to improve its security properties. This might involve network servers such as ftp or http, or an interpreter for a network scripting language.

University of California – Davis; Ronald A Olsson; ***Concurrent Programming Language Support for Invocation Handling: Design and Implementation*****; (CCR-9527295); $105,855; 24 months. (Joint support with the Operating Systems and Systems Software Program)**

This project examines new language mechanisms for accessing message invocations in message passing based concurrent programming languages. Current mechanisms for invocation selection control, including those that allow invocation parameters to be examined, are not expressive enough. In particular, they do not provide a simple and efficient way to examine pending invocations of an operation or allow the selection decision to be based on more than just a single invocations parameters. Examining invocations and performing cross-invocation or cross-operation selection are important in a number of real applications, including debugging, visualization, and scheduling algorithms. Using current mechanisms to solve these problems results in cumbersome and inefficient code. This project involves a mix of conceptual, design, implementation, and evaluation work. To make the work concrete, the new language mechanisms are being defined and implemented as extensions to the SR concurrent programming language. The results of the qualitative and quantitative evaluations are being used to refine the semantics and implementations of the new language mechanisms.

New York University; Robert A Paige; ***Improving Productivity of Algorithm and System Implementation in Scaled up Applications*****; (CCR-9616993); $50,000; 24 months. (Joint support with the Software Engineering Program and the Operating Systems and Systems Software Program)**

Improving productivity and reliability of labor-intensive high-performance software is an important scientific problem with obvious economic ramifications. This research is testing the feasibility of a radically new compilation and metacompilation methodology to improve productivity and reliability for high-performance algorithmic software in the C programming language. The research combines partial evaluation for deeply-nested SETL data structures, data structure and algorithm selection techniques, and transformational techniques from relational query optimization. This is leading to a statically typed variant of SETL from which it is expected large-scale C programs can be generated at a high translation rate.

Winston – Salem State University; Atanas A Radenski; ***Research at Undergraduate Institution (RUI) : Parallel Object-Oriented Programming with Multiple Inheritance*****; (CCR-9509223 A001); $32,108; 12 months.**

The transition from sequential object-oriented programming to parallelism has been the focus of research over the past decade. Experimental languages that try to integrate objects and parallelism often compromise the inheritance capability for parallel objects. Even the languages that permit some parallelism and inheritance support only single-class inheritance. In this project support for multiple-class inheritance for parallel objects will be investigated. The technique employed is based on 'separate' methods which generate processes and provide rendezvous type coordination. The project will apply this technique to Oberon language. The research will be performed under the Research for Undergraduate Institutions (RUI) program.

University of Illinois – Urbana; Edward M Reingold and Nachum Dershowitz; *U.S.-France Cooperative Research: Specifying and Finding Solutions to Constraint and Search Problems (with INRIA)*; (INT-9507248); $15,000; 36 months. (Joint support with the Western Europe Program Program and the Theory of Computing Program)

This three-year award will support U.S.-France cooperative research in theoretical computing and programming languages between research groups at the University of Illinois and at INRIA (the French National Institute for Research in Computer Science and Applied Mathematics). The U.S. investigators, Edward Reingold and Nachum Dershowitz, propose to collaborate with Claude Kirchner and Michael Rusinowitz at INRIA/Lorraine and with other researchers at the University of Nancy and the French Center for Research in Informatics at Nancy. Their research deals with problems in term rewriting and search techniques and in solutions of constraint and search algorithms. The U.S. investigators bring to this collaboration expertise in algorithms and equational reasoning. This is complemented by the French investigators' expertise in programming languages, constraint-based reasoning, and term rewriting techniques. The problems under study have important applications in image processing, programming languages, and automated deduction.

Carnegie Mellon University; John C Reynolds; *The Design, Definition, and Implementation of Programming Languages*; (CCR-9409997 A001); $104,388; 12 months.

This project conducts research on the design, definition, and implementation of programming languages, with emphasis on type structure and on bridging the gap between functional and imperative programming. In studying how to implement the programming language Forsythe, it investigates better methods of type checking for intersection types. Considerable improvement is expected from using a top-down rather than the present bottom-up algorithm. A more flexible approach to the implementation of procedures is also investigated, where a portion of the procedure body is compiled into closed code while the rest of the body is compiled into inline code at each call site. This is expected to substantially alleviate the overhead associated with call by name. The research seeks to unify two of the intractable dichotomies that complicate and divide the study of programming languages: (1) Semantics based on domains, which permits partial and infinite computations, with (2) semantics based on partial equivalence relations and intuitionistic type theory, which is limited to always-terminating computations. Also investigated is the unification of earlier work on the syntactic control of interference in Algol-like languages with the more recent development of low-level functional languages based on linear logic. In particular, this research would connect the concept of passive types with the modal operators of linear logic.

Brown University; Pascal Van Hentenryck; *NSF Young Investigator (NYI) Award: Constraint Programming Languages*; (CCR-9357704 A003); $62,500; 12 months.

Constraints are ubiquitous in many application areas including operations research (packing and scheduling), hardware design (circuit verification), computational linguistics, computer graphics, and software engineering. Constraint programming languages hold the promise of reducing the development time of these applications significantly, while preserving a reasonable efficiency. The purpose of this project to study the design, implementation, and applications of constraint programming languages. Both applicative and object-oriented languages are considered. Design issues focus on how best to support established paradigms for the applications considered. Implementation research considers the design of algorithms for constraint solving, entailment, and generalization for a variety of constraint systems. It also investigates the static analysis of constraint programming languages to reduce the performance gap between procedural and constraint-based languages, leading to the development of optimizing compilers for a variety of languages. Finally, applications validate the technology and implementation developed in the project and demonstrate the potential impact of the technology.

Semantics Theory

University of Oregon – Eugene; Zena M Ariola; *CAREER: A Comprehensive Rewriting Framework for Compilation, Optimization, and Run-Time Environments*; (CCR-9624711); $100,000; 24 months.

This research plan pursues the development of a unifying framework for reasoning about execution, compilation, and optimization of programs. The goal is to provide the programmer, compiler writer, and implementor with an abstract laboratory in which to investigate the interplay among various techniques and to reason about their correctness and efficiency with respect to both time and space. Specifically, the aim is to formalize an evaluator, a compiler, an optimizer, and a parser as a rewriting process. This requires enhancing current rewriting technology with the ability to deal with graph structures, including cyclic and disconnected graphs. This is important since graphs are ubiquitous in a program development system. They occur at the source level, after parsing, in the intermediate program representation and during program execution. The rewriting technology will be enhanced further by associating priorities and conditions to the rules. This will allow direct analysis of the impact of different application schemes on efficiency. Both strict and nonstrict languages will be considered, and also their extensions to handle state. This study constitutes a first step towards providing automatically generated tools, such as optimizing compilers. The development of the abstract framework addressed in this CAREER research project is also important for educational purposes, since it provides a simple yet rigorous, basis for introducing students to concepts, practical issues and, at the same time, to formal reasoning about properties of programs without having to introduce complex mathematical machinery. The proposed research will allow an integration across the undergraduate language-compilation implementation sequence, establishing a connection between abstraction and application. Exposure of students to research is planned through collaborations with other institutions and student exchanges, which can serve to encourage students to consider advanced education and research careers.

Carnegie Mellon University; Stephen D Brookes; ***Semantics of Parallel Programs*****; (CCR-9412980 A001); $65,000; 12 months.**

This project seeks to develop denotational semantics for a variety of programming languages, tailored to support compositional reasoning about various important forms of program behavior. It studies a typical functional programming language, and two typical parallel imperative languages: a language for shared-variable parallelism, and a language of communicating processes related to Hoare's CSP and Milner's CCS. Notions of behavior considered for the parallel languages include safety and liveness properties, and partial and total correctness, assuming various forms of fairness properties that correspond to natural assumptions about the runtime execution of concurrent processes. This should permit reasoning about parallel programs to be carried out without knowledge of or reference to details concerning scheduling. There also is a research focus on intensional properties concerning runtime and efficient use of resources. Technically, the project aims to develop fully abstract semantics, with respect to a variety of important notions of program behavior. A semantics is fully abstract if it gives the same meaning to two program phrases precisely when they induce identical behavior in all program contexts. This provides a rigorous criterion for judging the utility of a semantics in supporting compositional or modular reasoning about program behavior. The long term goal is the development of a unifying framework suitable for establishing and exploiting relationships between languages, models, and proof methods for reasoning about programs. In particular, the research seeks to establish a mathematically tractable theory and use it to develop practical techniques for modular design and analysis of parallel programs.

Williams College; Kim B Bruce; ***Research at Undergraduate Institution (RUI) : Semantics and Design of Object-Oriented Languages*****; (CCR-9424123 A001); $50,499; 12 months.**

This project researches the design and semantics of statically-typed object-oriented programming languages. The project has several goals. (1) It builds on and extends the design of the polymorphic, statically-typed object-oriented programming language, PolyTOIL, to provide more power and expressibility while retaining type-safety. (2) It investigates the design of an object-oriented language whose primary relation between types is matching rather than subtyping. This will likely require the development of a partial type-inference system to allow the programmer to avoid the proliferation of bounded type parameters. (3) It seeks to create better mathematical (denotational) models for foundational calculi for object-oriented languages. (4) It investigates the development of verification techniques for object-oriented programming languages. (5) It applies current knowledge of the semantics of object-oriented languages to the understanding of concurrent object-oriented languages. These problems are being attacked using formal specifications of type-checking rules and operational semantics. Subject-reduction-style theorems are being used to prove the safety of type systems. Interpreters are being written in order to provide a better understanding of the strengths and weaknesses of the systems and in order to investigate scaling up of theoretical designs to full-strength programming languages. The project seeks to validate verification techniques by proving the consistency of an operational and axiomatic semantics.

Cornell University ; Robert L Constable; ***Exploring New Constructs in Computational Type Theory*****; (CCR-9423687 A001); $77,000; 12 months.**

Computer science has become increasingly concerned with the concept of type. From the 60's onward this notion has demonstrated its value as an effective organizing concept_first in programming and programming languages, then in knowledge representation, in automated reasoning, in formal methods, and now perhaps in computer algebra systems. Previous NSF-supported research has led to a coherent family of constructive type theories and to their implementation in the Nuprl and related systems. The plan of the investigation is to find the right concepts to explain the use of types in computer algebra systems and in formalizations of abstract computational mathematics (especially algebra). It is conjectured that in some cases the right concepts are generalizations of algebraic data types to higher-order ADT's. In other cases, new type constructors may be necessary, such as the very dependent types. These concepts are expected to clarify both pragmatic issues and foundational questions. Moreover, the method of investigation seeks to formalize the object oriented features of algebra systems and generalize them to other programming environments. The result of this research enables the design of a new generation of computer system that integrates subsystems (such as compilers, concrete and symbolic evaluators, transformation systems, provers, and data base managers) around a rich notion of type used to specify and verify the linking protocols. Investigation of computational type theory also deepens understanding of the foundations of computer science and its links to mathematics and computational science.

University of Pennsylvania; Carl Gunter; ***Relating Static and Dynamic Semantics of Programs*****; (CCR-9415443 A001); $63,740; 12 months.**

A key problem in programming languages is the design and implementation of a balance between static (compile-time) and dynamic (run-time) phases of computation. A good balance allows the generation of good code without an undue burden on programmers or compilers. Finding this balance relies critically on intuition and experimentation, but these methods are not enough by themselves. As type systems and optimizing transformations have become more sophisticated, it has become essential to provide rigorous means to express static/dynamic relationships at suitable levels of abstraction. Programmers and compiler writers rely on these relationships, so a precise but approachable description of them is important. Using intuition and experimentation alone may allow errors or valuable opportunities for generalization to go unnoticed. The goal of this project is to develop and apply techniques for rigorously expressing and proving the relationships between static and dynamic semantics of programming languages. Primary objectives include (i) finding ways to specify meanings at levels of abstraction suitable for reasoning about sharing, copying, and liveness; (ii) developing ways to represent and manipulate semantic specifications in a formal manner; (iii) aiding the development of new interface definition languages by formulating techniques for describing the properties that compatibility of modules guarantees for systems that integrate them.

Carnegie Mellon University; Robert W Harper; ***CAREER: Type Theory and Operational Semantics for Programming Languages*****; (CCR-9502674 A001); $35,000; 12 months.**

This CAREER project investigates the use of type theory and operational semantics in the design and implementation of programming languages. Type theory has proved to be an important organizing principle in programming language design. This is well exemplified by the use of type theory in the design of modularity and abstraction constructs. This investigation seeks to consolidate advances in type theory into the design of the ML2000 programming language. Type theory has proved important for the implementation of programming languages. Conventional syntax-directed compilation methods are generalized to type-directed compilation, allowing type information to be exploited during compilation. Sophisticated type systems such as those derived from the Girard-Reynolds polymorphic lambda-calculus suggest new implementation strategies based on passing type information at link-and run-time. This research investigates the use of type-based compilation techniques. Operational semantics provides a suitable framework for addressing issues of compiler correctness and proving properties of programs. Typed programming languages such as Standard ML provide a rich setting in which to discuss high-level programming techniques such as data abstraction, modularity, and separate compilation. Types are essential for reasoning about programs. A procedure can in general be deemed to have certain input/output properties only under the assumption that its arguments have suitable types. Operational semantics is a useful pedagogical tool for teaching undergraduate programming, both as an explanatory device and as the basis for proving properties of programs and languages. Important advantages are expected through the interplay between research and education.

Boston University; Assaf J Kfoury, Pawel Urzyczyn, and Jerzy Tiuryn; ***Combinatorial Problems in Typed Lambda-Calculi*****; (CCR-9417382 A001); $100,000; 12 months. (Joint support with the Theory of Computing Program)**

This award supports research in the area of typed lambda-calculi, with special emphasis on the combinatorics, complexity, and logic of type inference. Specific problems concern type-reconstruction, type-checking, and related aspects, and are motivated by design and implementation issues in programming languages. Particular attention is given to algebraic characterizations of these problems. Such characterizations give rise to new and more general forms of unification, in particular semi unification, often with applications in other parts of computer science (term-rewriting systems, database theory, computational linguistics). Based on some of the same techniques, the project explores problems in two research areas, both with important applications: (1) higher-order matching (for automatic theorem proving) and (2) extensions of ML-style polymorphism (for polymorphically typed programming languages).

Rutgers University; Paul J Krasucki; *Research at Undergraduate Institution (RUI) : Knowledge and Concurrency*; (CCR-9528919); $93,938; 24 months. (Joint support with the Software Engineering Program and the Theory of Computing Program)

This Research at Undergraduate Institutions (RUI) and Cooperative US-India Research Project investigates distributed systems, using reasoning about knowledge: During the run of the distributed system, processes acquire knowledge about the state of the system through messages. The goals in this area are to (a) characterize changes in the state of knowledge of processes due to communications and (b) relate knowledge structure of the system to the temporal structure on event occurrences in the system. The third research goal of this project is to modify a model, created by economists to reason about rational agents communicating values of certain agreed-upon functions, so that it can be used to design and analyze protocols leading to distributed consensus.

Colgate University; Philip S Mulry; *Research at Undergraduate Institution (RUI) : Categorical Methods in Computation and Semantics*; (CCR-9506383 A001); $32,870; 12 months.

This project exploits recent advances in the use of sophisticated mathematical methodologies to address related fundamental questions in the foundations of computation theory and programming language semantics. Many important questions arise out of a desire and need to capture in a semantic setting both minor and major variations in computational strategies. The significance in investigating such variations lies in the fact that they can both provide a means of interpreting increasingly sophisticated programming language constructs, and also isolate fundamental mathematical processes at work that help define and distinguish between different notions of computation. The main focus of the research entails the utilization of new approaches in the use of algebras, monads and comonads, categories, logic and sheaves to provide a systematic and conceptual approach to the interpretation and solution of these questions. Questions to be addressed include: finding an algebraically robust interpretation for partial data types and partial maps; generating an intrinsic characterization of fixed point constructions, including the generation of recursive data types, and isolating the special role of invariance in this process; defining a categorical foundation for semantics in general and describing a synthetic domain theory in particular; and identifying and utilizing in a semantic setting logical properties resident in sheaf-theoretic structures. General areas of investigation include partial data types and computation, fixed point semantics, monadic and comonadic computation, invariance and algebras, and partial sheaf semantics. Technical tools include functorial liftings on categories of algebras, refined partial map classifiers, algebraic uniformity, induction and coinduction principles, natural and dinatural transformations, monadic and comonadic strength, domain theory and extensional Partial Equivalence Relations.

University of Central Florida; Rebecca J Parsons; *CAREER: A Semantic Framework for Static and Dynamic Optimization*; (CCR-9624932); $100,000; 24 months.

Acceptance of an object-oriented approach to scientific computing is being hindered by the performance differential between programs written in traditional languages and those written in object-oriented languages. These performance issues stem from the limitations on static dependence analysis and compiler optimization posed by the presence of objects. An approach to increasing the degree of optimization in object-oriented environments is incorporating information about the object semantics and performing runtime optimizations. As seen in compiler optimization using static dependence information, informal semantics and intuition are inadequate. A semantic foundation for optimization which includes support for dynamic dependence and for the richer data types exploited in object-oriented languages provides an important foundation for advancing compiler research. The semantic model studied in this research is intended to unify reasoning both in the compiler (static) environment and the runtime (dynamic) environment. Support for dynamic object creation and even dynamic class definition can be considered in a setting exploiting both static and dynamic dependence information. The resulting model, utilizing tools such as denotational semantics and equational calculi, will allow the creation of automated tools to assist in the development of advanced compiler optimizations with the additional formal assurances that these optimizations do not alter the meaning of the program. The educational goals of this research are to alter the programming languages and compiler curricula to address, at all levels, issues of semantics, optimization, and the degree to which each affect the other. The computational biology work provides a promising framework within which to study these issues. Thus, the computational biology curricula is being developed in parallel with the languages work. This research is expected to contribute to an understanding of the nature of program dependences; how different language constructs and program usage contribute to different patterns of dependence; and how these patterns can be exploited in language design or in language translation. A goal of this research is to support the mixing of languages without sacrificing the performance of the program, allowing the programmer to select the most appropriate language for the individual parts of a problem.

Yale University; Zhong Shao; *CAREER: Type-Directed Compilation*; (CCR-9501624 A001); $35,729; 12 months.

Compile-time type information is believed to be very valuable in efficient compilation of functional languages such as Standard ML, but little is known on how to support type-directed compilation in real compilers and how much performance gain one can get from various type-based optimizations. The goal of this research is to design and identify important type-directed compilation techniques that can significantly improve the performance of ML programs. These new compilation techniques are expected to lead to the following results: (1) data objects (including recursive data structures) in ML programs will be able to use unboxed data representations that are as efficient as those in C programs; (2) polymorphic functions and parametrized modules will be implemented efficiently through cross-module type specialization and inlining expansion; (3) type information propagated into the compiler back end will be used to assist program analysis such as closure analysis and instruction scheduling. The educational component of this CAREER award seeks two objectives: the construction of a new compiler lab kit (based on the above type-based ML compiler) and the incorporation of state-of-the-art compiler technology into the undergraduate and graduate curriculum.

Florida International University; Geoffrey S Smith; *Type Systems for Secure Remote Evaluation*; (CCR-9612176); $118,072; 36 months. (Joint support with the Software Engineering Program)

A remote evaluation system is characterized by a client that sends a program to a compute server for execution on its behalf. The server returns to the client the results of executing the program. This is in contrast to RPC and CGI scripts which are examples of server-side programs, not remote evaluation. Remote evaluation appears to be gaining currency as evidenced by new programming languages emerging to support it. Among them are Telescript, Tcl, Oblique and Java. The download-and-execute paradigm supported by Java and Java-compatible Web browsers is an example of remote evaluation; in this case, however, a Web server sends a Java-enhanced HTML file to a browser which the browser then executes. Server security is a major problem with remote evaluation. A server's availability, privacy and integrity must not be compromised. Current techniques for ensuring server security is ad hoc. There is no precise characterization of what kind of security properties these techniques can guarantee. Consequently, the typical reaction is to limit, or in most cases prohibit altogether, an application's access to server resources. This greatly weakens the remote evaluation paradigm. This research investigates the role of programming language design and type systems in ensuring the security of servers in remote evaluation systems. Specifically, the long-term objective is to develop static analyses, within the context of type systems, to ensure that remotely-evaluated programs do not violate a server's privacy or integrity. Type soundness theorems are being developed and proved with respect to a formal semantics for a core imperative language. They precisely characterize the security properties of well-typed programs written in the language. The impact of features such as concurrency and exceptions in a programming language on security properties is also investigated.

Johns Hopkins University; Scott F Smith; *Semantic Tools for Realistic Programming Languages*; (CCR-9301340 A002); $58,188; 12 months.

This project addresses the problem of giving full and formal semantics to programming languages that combine a number of complex features, including memories, object-orientedness, rich notions of type, exceptions, and concurrency. An operational approach is taken to semantics since it is the only method with immediate potential for applicability to such rich languages. The resulting operational theory also includes a rich theory of equivalence on programs to aid compiler writers and language transformers. This research focuses on: (1) lifting concepts from domain theory, such as finiteness and omega-algebraicity, into full-featured languages, (2) giving semantic type descriptions of full-featured languages, in particular the difficult case of object-oriented languages, and (3) studying full-featured languages as one piece to discover interactions and conflicts, rather than addressing each feature in a separate study.

Carnegie Mellon University; Richard Statman; *Lambda Calculus*; (CCR-9624681); $110,470; 36 months.

Lambda Calculus is both a mathematical theory of functions and a model of computation. As such it is part of the interface of applied logic and theoretical computer science. As a theory, simply typed Lambda Calculus is part of Church's theory of functions of higher type. Higher type theorem proving reduces questions of validity of sentences to unification and conversion, which are purely equational questions of Lambda Calculus. Lambda Calculus also enters implicitly into the proof theory of type theory via the well known Curry-Howard isomorphism. As a formal model of computation Lambda Calculus contributes to understanding of higher-type and type-free programming constructs. Denotational semantics reduces questions of program synthesis and correctness to finding and verifying solutions to certain combinatory functional equations. Lambda Calculus also enters explicitly into the syntactic foundations of applicative programming languages as a paradigm of sequential computation. In short, Lambda Calculus enters both explicitly and implicitly into the theory of highertype theorem proving and the theory of programming languages. Lambda Calculus is the study of certain computation rules, programs, or algorithms. This research singles out those rules whose execution depends only on the fact that some of the data are themselves computation rules of the same sort. It is not immediately obvious that there are any non-trivial examples of such rules. The rich deep structure of the Lambda Calculus had to be discovered by Church, Bernays, Curry, Kleene, and those who followed them. The research focuses on the deep structure of pure Lambda Calculus with at most algebraic types, to address a number of open questions in Lambda Calculus. These include: (1) Is it decidable whether a given finite set of proper combinators forms a basis? (2) Is there a recursive one-step Church Rosser strategy? (3) Is there a uniform universal generator? (4) Is the word problem for all proper combinators of order < 3 decidable? (5) Is the matching problem for the pure simply typed Lambda Calculus decidable?

Wesleyan University; Ramesh Subrahmanyam; *Research Initiation Award (RIA): Types, Abstraction and Programming Language Expressiveness*; (CCR-9406202 A002); $34,100; 12 months.

The notion of types is an important concept in both the design of programming languages and in the theoretical investigations of programming languages. While a significant understanding of type theoretic issues in a programming language context exists, fairly basic questions remain open. Similarly, while it has been suggested that sophisticated type systems enhance the expressive power of languages, very little has been done to set-up a formal work to investigate this idea. Two related issues form the core of this work: language expressiveness, and types and operational semantics. A notion of expressiveness for typed programming languages is described. The expressive power of a language relative to another is measured by comparing the sets of abstract data types implementable in the two languages. Preliminary results suggest that the resulting expressiveness hierarchy of programming languages in non-trivial. This research conducts a detailed investigation into the relationship between this notion of expressiveness, and various type disciplines and language mechanisms. Answers to these questions have a definite impact on the design of programming languages. The research investigates some questions about types in an operational semantics context; these are related to questions studied in the context of lambda calculus and type theory. Two specific classes of questions that are addressed follow: (1) orthogonal extension questions: does type system A, when added to a languages based on type system B, invalidate any existing observational equivalences; and (2) type isomorphism questions: given a pair of types in a programming language is there a pair of definable functions (from one to the other) which are observationally inverse. These lead to further questions about axiomatization of observational equivalence in fragments of typed programming languages.

University of Pennsylvania; Val Tannen and O. Peter Buneman; *Collection Types in Programming Languages and Databases*; (CCR-9216122 A002); $70,000; 12 months. (Joint support with the Database & Expert Systems Program)

Structural recursion provides a unifying paradigm for the forms of manipulation of collection types that are common in programming languages and databases. As evidence for this claim a general characterization is developed for the semantics of programs that exploit structural recursion on lists, bags, and sets. An intimate connection is identified between structural recursion and a range of existing languages that manipulate complex objects. This project conducts an investigation into the general principles of collection types, including heterogeneous collections, the semantics of collections in the presence of divergence, the parallel semantics of collection languages, the equational theories of collection languages as a basis for discovering and validating optimizations, and extended pattern-matching with collection types. It also involves the design and prototype implementation of a language demonstrating most of these ideas.

Naval Postgraduate School; Dennis M Volpano; *Type Systems for Secure Remote Evaluation*; (CCR-9612345); $115,093; 36 months. (Joint support with the Software Engineering Program)

A remote evaluation system is characterized by a client that sends a program to a compute server for execution on its behalf. The server returns to the client the results of executing the program. This is in contrast to RPC and CGI scripts which are examples of server-side programs, not remote evaluation. Remote evaluation appears to be gaining currency as evidenced by new programming languages emerging to support it. Among them are Telescript, Tcl, Oblique and Java. The download-and-execute paradigm supported by Java and Java-compatible Web browsers is an example of remote evaluation; in this case, however, a Web server sends a Java-enhanced HTML file to a browser which the browser then executes. Server security is a major problem with remote evaluation. A server's availability, privacy and integrity must not be compromised. Current techniques for ensuring server security is ad hoc. There is no precise characterization of what kind of security properties these techniques can guarantee. Consequently, the typical reaction is to limit, or in most cases prohibit altogether, an application's access to server resources. This greatly weakens the remote evaluation paradigm. This research investigates the role of programming language design and type systems in ensuring the security of servers in remote evaluation systems. Specifically, the long-term objective is to develop static analyses, within the context of type systems, to ensure that remotely-evaluated programs do not violate a server's privacy or integrity. Type soundness theorems are being developed and proved with respect to a formal semantics for a core imperative language. They precisely characterize the security properties of well-typed programs written in the language. The impact of features such as concurrency and exceptions in a programming language on security properties is also investigated.

Compilation and Optimization

Princeton University; Andrew W Appel; ***Framework, Algorithms, and Applications for Cross-Module Inlining*****; (CCR-9625413); $180,331; 24 months.**

Modern programming languages encourage programmers to use abstraction and encapsulation to hide one module's data representation details from other modules. But this leads to slow execution without in-line expansion (substitution of procedure bodies for procedure calls). The problem is particularly acute in object-oriented and functional programming languages that support method and function abstraction. Most previous work on in-line expansion (in imperative languages) fails to inline higher-order and recursive functions, but in functional languages (and object-oriented languages, with dynamic methods) these can be crucial. Since modules can be compiled separately from each other, cross-module in-line expansion complicates program linkage. This work uses a lambda-calculus framework for cross-module in-line expansion, which cleanly separates the compilation-management issues from the inlining algorithms. The objective of this project is to implement this framework, and then design inlining algorithms within the framework that can be tested in the field.

Northwestern University; Prithviraj Banerjee; ***Efficient Compilation Issues for Scalable Distributed-Memory Multicomputers*****; (CCR-9526325); $91,408; 24 months.**

Distributed-memory, massively-parallel multicomputers can provide the high levels of performance required to solve the Grand Challenge computational science problems. Multicomputers offer significant advantages over shared-memory multiprocessors in terms of cost and scalability. Unfortunately, extracting all the computational power from these machines requires users to write efficient software for them, which is an extremely laborious and error-prone process. The distribution of data across processors is of critical importance to the efficiency of the parallel program in a distributed memory system. In this project the problem of automated data distribution in such machines is being investigated. The approach is unique in that it is based on sophisticated cost models of communication and computation that are parameterized by architectural metrics empirically measured for different target machines. Using the cost estimates, data distributions are selected to minimize the overall execution time of the program by maximizing parallelism (while maintaining load balance) and minimizing the amount of communication overhead (by maximizing data locality). Both static and dynamic distribution will be supported in a unified framework to automatically select data distributions which can dynamically change over the course of a program's execution in order to provide scalable parallel performance for large, complex applications. This approach not only has a solid theoretical basis, but is being integrated into a sophisticated compiler which can actually generate code for a variable number of processors supporting all block, cyclic, and block-cyclic data distributions. The compilation techniques investigated in this project will also be integrated with simultaneous support for regular and irregular accesses in parallel applications through a novel interval-based representation.

Michigan Technological University; Steven M Carr; ***Research Initiation Award (RIA): Improving Cache Performance in Scientific Applications*****; (CCR-9409341 A003); $35,299; 12 months.**

This project explores approaches for automatically restructuring code to improve the performance of cache memories for scientific applications. Specifically, the project examines experimentally the following techniques: (1) loop optimization guided by a new performance metric that includes the effects of cache performance and floating-point computation on execution speed, (2) the enhancement of cache line replacement algorithms by using compiler generated knowledge of future reference patterns, and (3) the use of compiler generated cache hit and miss information by the instruction scheduler to hide latency on architectures having non-blocking caches. Each of these methods is implemented in an experimental compiler and evaluated on advanced microprocessors. The success of the project represents a significant step towards achieving fully automatic memory hierarchy management.

University of California – San Diego; Larry Carter and Jeanne Ferrante; ***Hierarchical Tiling: A Framework for Multi-level Parallelism and Locality*****; (CCR-9504150 A001); $5,000; 0 months.**

To achieve maximum benefit from tiling, compiler optimization must exercise more control over data movement and storage assignments than is commonly done. Hierarchical tiling takes responsibility for several phases of compilation and code improvements that are traditionally done separately, such as scalar replacement, register allocation, generating message passing calls, and storage mapping. It uses the mechanisms of explicitly naming and copying data to control the movement of data up and down the memory hierarchy and to exploit all levels of parallelism. Its effectiveness as a systematic framework for hand-crafting highly optimized code has been tested on scientific applications for IBM SP1 system. This project will extend the research in compiler optimizations by investigating the following: (1) Develop a parameterized machine model that captures the architectural information needed to guide hierarchical tiling. (2) Study interactions between tiling at various levels of granularity. (3) Incorporate hierarchical tiling into the SUIF toolset. (4) Extend the work to disk storage, explicit I/O and implicit paging virtual memories. (5) Validate the approach using various programs on different parallel machines.

University of Washington; Craig Chambers; *Analysis and Optimization of Object-Oriented Languages*; (CCR-9503741 A001); $49,499; 12 months.

Object-oriented languages offer one promising approach to making programs easier to write and easier to modify and extend. Unfortunately, object-oriented language constructs often incur a substantial run-time cost. To improve the run-time performance of object-oriented languages, previous research applied intraprocedural class analysis and compile-time class prediction to determine the actual classes of arguments to a dynamically bound call at compile-time, thereby statically binding the call to the target procedure and enabling inline expansion to eliminate the overhead of the call. This work was performed primarily in the context of purely object-oriented languages. The goal of this project is to extend previous work in three directions. First, the scope of analysis will be expanded to the whole application program. This necessitates developing and evaluating techniques for constructing call graphs for object-oriented languages and performing interprocedural class analysis. Second, profile information will be incorporated into the analysis and optimization process. In particular, profile information can provide much more accurate class prediction information than can static estimation techniques. Third, procedure specialization will be selectively applied to the optimization of object-oriented languages. The goal will be to investigate the relative trade-offs in cost and benefit among inline expansion, procedure specialization, and interprocedural analysis for object-oriented languages. The suitability of these techniques across a range of object-oriented programming languages, both pure and hybrid, will be evaluated in this effort.

Syracuse University; Alok Choudhary; *NSF Young Investigator (NYI) Award: Compiler and Runtime Optimization Techniques for Parallel Programming on Distributed Memory Machines*; (CCR-9357840 A003); $0; 12 months. (Joint support with the Operating Systems and Systems Software Program)

The objectives of this research are to develop compiler and runtime optimization techniques for parallel programming on homogeneous and heterogeneous high-performance parallel computers. Specifically, this research addresses the following issues: (1) Scalable Parallel I/O-Scalable massively parallel computers must include scalable I/O systems to provide a balance between computing power and I/O. The thrust of this research is to investigate compiler transformation for parallelizing I/O operations, strip-mining and prefetching data, overlapping computation and I/O, generate efficient I/O access schedules, develop runtime support that can recognize and use access patterns to perform efficient parallel I/O, and develop language primitives that allow users to specify parallel I/O operations; (2) Automatic Data Distribution for Homogeneous and Heterogeneous Systems-The goals are to develop algorithms to automatically generate data distributions that incorporate global analysis, use realistic cost models, and that allow dynamic alignments, distributions and data replication; and (3) Runtime Support for Dynamic Data Redistribution-Dynamic data redistribution may be necessary due to explicit program constructs, compiler generated distributions, across procedure boundaries, or adaptive nature of a problem. The goals are to develop parallel algorithms (including communication scheduling, index conversion, and space optimization) for runtime support for dynamic data redistribution.

Dartmouth College; Thomas H Cormen; *ViC*: A Compiler and Runtime System for Out-of-Core C; (CCR-9625894); $109,185; 24 months.**

The proposed project completes the implementation, already underway, of the ViC* compiler and run-time system. The compiler translates the data-parallel language C* with some data explicitly declared to be "out of core." Out-of-core data reside on a parallel disk system. The compiler produces C for a distributed-memory multiprocessor. It adds I/O and library calls to access data on disk, and it also produces virtual processing loops to allow each physical processor to be shared by multiple parallel-variable elements. A ViC* program will run significantly faster than one run with traditional sequential virtual memory on the individual nodes of a MIMD machine. Unlike demand paging, ViC* will take advantage of aggregate data-parallel operations by optimizing away many I/O calls. Moreover, the ViC* runtime library will execute optimal parallel-disk algorithms that have appeared in the literature recently. There are two major components to ViC*: a compiler and a run time library. Both are designed to minimize the number of parallel disk accesses made by the executing program. The runtime library will implement the entire C* library for both incore and out-of-core parallel variables. It will run on any platform that supports PVM and any of the following file system interfaces: Scotch, Galley, Scalable I/O Initiative API, and UFS for multiple disks.

Washington University; Ron K Cytron; *SPAR: Synergistic Program Analysis and Restructuring*; (CCR-9402883 A003);

$71,873; 12 months.

This project investigates the introduction of ambitious program analysis and program restructuring techniques into automatic program optimization systems, with the goal of significantly improving the quality of code produced by compilers for modern computer architectures. Although a variety of compiler optimization techniques exist, they are not used to their full extent in production compilers because, the analysis and transformations are computationally intensive. In addition, the information obtained by analyses must be maintained and updated across the various compiler passes and program transformations. The project relies on the previous work on "sparse" data flow frameworks and explores three methods for overcoming the current limitations to using ambitious program analyses in production compilers: (1) decomposing a dense data flow framework into an interconnected set of sparse data flow frameworks, (2) selectively computing and using costly information on behalf of a particular program optimization. and (3) integrating program analysis and transformation.

Massachusetts Institute of Technology; Jack B Dennis; *Implementation of Stream-Based Functional Programming for Real-Time Applications*; (CCR-9527253); $186,262; 24 months.

The project's goal is to investigate applications of stream-based functional programming methodology to the implementation of multiprocessor based real-time systems, and specifically to signal processing applications. The work is based on the Sisal programming language. It will use Sisal's stream data types, static analysis methods, and automatic mapping to map parallel computations on processors in distributed computer systems. The Paradigm compiler will be redesigned and extended to perform static analysis and mapping of continuous stream processing computations. A related objective of this effort is to develop analytical methods for performance modeling of stream-based computations on multiprocessor systems.

University of Pittsburgh; Rajiv Gupta; *Presidential Young Investigator (PYI) Award: Loop Transformations and Scheduling Strategies for Parallelization Software*; (CCR-9157371 A006, A007); $16,316; 12 months.

A framework for studying the interactions among loop transformation and scheduling strategies and a parallelization tool that allows experimentation with alternative parallelization strategies will be developed. The tool will accept architectural information to assist in the application of transformations and selection of schedules. The results obtained from studying these interactions will be applied in the development of an integrated approach to loop transformation and scheduling. In this approach the loop parallelization process will be composed of three major phases. The first phase, the parallelism detection phase, will perform dependency analysis to identify loops that can be executed in parallel. In addition, it will perform transformations, such as renaming, to convert sequential loops into parallel loops. The second phase, the parallelism scheduling phase, will select an appropriate scheduling strategy. The goal here will be to choose a strategy that involves minimal run-time scheduling overhead and results in schedules that provide good load balancing and high processor utilization. The final phase, the parallelism enhancement phase, will transform the loops to increase the efficiency with which parallelism is exploited. The transformations that exploit architectural characteristics will be applied to reduce synchronization and communication overhead. Any decisions to utilize special hardware features will be made in this phase. Decisions of activities performed in the later phases will be based upon the decisions made in the earlier phases. An integrated strategy will allow convenient application of transformations after schedules have been generated. It will also ensure that the scheduling strategy does not undo the effects of transformations that had been applied.

State University of New York – College New Paltz; Michael Hind; *New Paltz Interprocedural Compiler*; (CCR-9633010); $88,662; 24 months.

The focus of this research is on interprocedural program analysis; the specific goal of this project is to quantify the tradeoffs between precision and compilation speed for a number of interprocedural analyses. The initial work on this project has focused on implementing a flow-sensitive interprocedural pointer alias analysis. The current activities on this project are aimed in the following three directions. (1) A flow-insensitive interprocedural pointer alias algorithm will be implemented. The tradeoffs in precision and efficiency between this algorithm and the flow-sensitive variant will be quantified. (2) A sparse representation of alias information will be developed that avoids storing redundant information at the cost of retrieving it during an alias query. (3) Experiments will be conducted to demonstrate the effectiveness of this representation. A natural extension of a flow-sensitive pointer alias analysis for C++ type analysis will be implemented. Experiments will be conducted to demonstrate its effectiveness and compare it to previous work.

University of California – Santa Barbara; Urs Hoelzle; *CAREER: Efficient Implementation of Object Oriented Languages on Superscalar Processors*; (CCR-9624458); $100,000; 24 months.

Programmers have long been faced with the dilemma of choosing between abstraction and efficiency: if they use abstraction pervasively, their programs become more maintainable but may execute too slowly. Ideally, hiding a design decision behind an abstract interface should have no impact on performance, so that programmers can reap the full benefits of encapsulation without fear of performance problems. Thus, abstraction-oriented languages must employ aggressive compiler optimizations to reduce the run-time overhead of abstraction. Since the ultimate goal is to increase programmer productivity, these optimizations must be compatible with interactive, exploratory programming environments. The focus of this work is on reducing the cost of abstraction in object-oriented languages, as well as, complementary educational goals. The objective of this project is to develop adaptive customization, a compiler technique that aims to make aggressive optimization practical by improving performance without unduly increasing code size. Being suitable for interactive programming environments, adaptive customization could dramatically improve the performance of real-world applications written in dynamic languages like Smalltalk, Self, Java, Dylan, or CLOS. Not all dispatches can be optimized away, and the remaining dispatches may become more expensive on superscalar processors. Thus, the goal is to develop new dispatch techniques specifically optimized for such processors. A comprehensive study of dispatch performance on superscalar processors will be undertaken in this project. Benchmark suites and tools will be developed to support this research.

University of Illinois – Urbana; Wen-mei Hwu; *Data Speculation for Instruction-Level Parallel Processing Systems*; (CCR-9629948); $109,328; 24 months. (Joint support with the Operating Systems and Systems Software Program)

Data speculation refers to the execution of instructions before it has been determined that the data being used is valid. Data speculation depends on dynamic memory disambiguation to determine if the data was valid, and forces recovery code to execute if the data was invalid. A data speculation compiler generates recovery code to handle the situation where such execution violates a memory dependence. This project examines data speculation techniques, and attempts to understand how to utilize them more effectively. The issues addressed are (1) using data speculation with predication, (2) improving the scheduling algorithms used with it, and (3) using static analysis to help determine when it should be applied. This work will be done using the Memory Conflict Buffer (MCB) technique for dynamic disambiguation. The concepts will be developed so they are applicable to other forms of dynamic memory disambiguation.

Stanford University; Monica S Lam; *NSF Young Investigator (NYI) Award: Language and Compiler Technology for Parallel Machine Organizations*; (CCR-9257105 A004); $62,500; 12 months.

Topics pertinent to parallel system development are examined. Strategies for optimizing interprocessor communication (such as grouping computations which use common data to improve data locality or reduce remote access by the allocation of data to local processor memory) are investigated. Also, tools for simplifying the programming of applications for distributed memory machines are developed.

University of Rochester; Wei Li; *Research Initiation Award (RIA): A Parallelizing Compiler for Distributed Shared Memory Machines*; (CCR-9409120 A002); $29,413; 12 months.

This project develops a parallelizing compiler for distributed shared memory systems, and incorporates a vertically integrated system to manage data and thread locality in parallel programs. The compiler locates threads and data statically whenever sufficient compile time information is available to make near-optimal decisions. When such information is not available, the compiler provides the underlying system with whatever information is available, and lets the underlying system utilize the information. The project investigates the following compile time issues as related to distributed memory systems: (1) automatic parallelism for relaxed models of memory consistency, (2) elimination of false sharing, (3) adaptive coherence protocols, and (4) communication optimizations. Solutions to the issues are implemented within a compiler known as Rochester Optimizing Compiler (ROC). These solutions will be evaluated using standard benchmark programs such as LINPACK, SPEC, Perfect Club, SPLASH and the HPF benchmark suite of Syracuse.

University of Massachusetts – Boston; Joan D Lukas; *Research Initiation Award (RIA): Data Locality Optimization for Data Parallel Extensions to C*; (CCR-9409360 A002); $31,564; 12 months.

In this project, compiler algorithms for the enhancement of data locality on parallel computer systems are developed. These algorithms are implemented in a compiler for a data parallel extension to C targeted to MIMD systems. Enhancements to data locality improve the run time performance of programs by reducing the need for communication and/or synchronization. This work extends compiler design and implementation work on data parallel extensions to FORTRAN (e.g., FORTRAN 90 with extensions) targeted to SIMD systems. Both language extensions and compiler optimizations are investigated. The language extensions are aimed at improving the expressiveness of data parallelism. Generalizations of data locality representations and algorithms to a wider class of architectures and languages are designed and implemented. Automatic techniques for localizing data are investigated with programmer directives.

University of Massachusetts – Amherst; Kathryn S McKinley; *CAREER: Compiling for Heterogeneous Systems*; (CCR-9624209); $100,000; 24 months.

Heterogeneous parallel systems incorporate diverse models of parallelism within a single machine or across machines. These systems provide high performance for large, diverse applications. Heterogeneity unfortunately complicates software development. Users currently hand parallelize each task for each architecture. These programs are not portable and must be recoded if changes occur (e.g., machines are added or removed). The user is also burdened with performing the appropriate optimizations to achieve high performance. A compiler can solve these problems. The most important advantages of a compiler based approach are: (1) Portability and Ease of Use: Users program the appropriate model of parallelism for their algorithms. The compiler automatically maps them on to the specifics of the available architectures. (2) High Performance: The compiler optimizes with aggressive, target specific optimizations and coordinates diverse hardware to achieve excellent performance. To manage heterogeneity the compiler must: target diverse architectures, adjust optimization strategies to suit targets, exploit run-time changes in machine availability, and allow languages to evolve. The objective of this research is to investigate a compilation system to support these requirements. The system will use an innovative compiler organization and representations in order to provide the flexibility that heterogeneity requires. This project concentrates on the design and implementation of this compiler infrastructure. The system will be used to investigate automatic parallelization and partitioning strategies for a heterogeneous target. It will address issues such as how to map programmer models of parallelism on to the corresponding machine and how to further optimize it for total system performance. This compiler-based approach will provide an infrastructure capable of performing the advanced dynamic and static optimizations that heterogeneous systems demand for high performance.

Appalachian State University; Cindy Norris; *CAREER: Cooperative Compiler Optimizations for Fine-Grained Parallel Architectures*; (CCR-9625219); $126,898; 24 months.

In order for compilers of fine-grain parallel architectures to be as effective as possible, compiler writers must understand the interactions between the phases of the compiler. These phases include instruction scheduling which rearranges code in order to increase available fine-grain parallelism. In addition, code transformations to increase memory performance are applied, such as tiling and loop interchange to increase cache performance, prefetching to prefetch values from memory in order to eliminate the memory latency, and scalar replacement which changes array references to scalar references to decrease memory accesses. The compiler also performs register allocation, which allocates registers to values so as to minimize the number of memory accesses. The goal of this proposed work is to investigate the interactions of these phases. This effort is being undertaken along two directions. The first direction is to continue previous work in cooperative register allocation and instruction scheduling by (1) incorporating cooperative software pipelining and performing the cooperative register allocation and scheduling optimizations over a single framework, the program dependence graph (2) performing a thorough analysis and experimental study of the effects of register assignment strategies on scheduling to increase fine-grain parallelism. The second direction is to investigate the interactions between code transformations improving memory performance and other compiler phases. In particular, the proposed work will investigate (1) the impact of scalar replacement and software prefetching on instruction scheduling to increase fine-grain parallelism and on register allocation (2) strategies for enabling communication between code transformations to increase memory performance, register allocation, and instruction scheduling.

Louisiana State University & Agricultural and Mechanical College; Jagannathan Ramanujam; *NSF Young Investigator (NYI) Award: Languages, Compilers, and Runtime Systems for Parallel Architectures*; (CCR-9457768 A003); $49,400; 12 months.

The goal of this research is to ease the burden of programming high-performance parallel computers. The project investigates several key issues in this regard: (a) the design of language extensions or new languages for machine-independent parallel programming; (b) algorithms for automatic code and data mapping (both static and dynamic) for scientific programs; (c) code generation, and optimization techniques for locality, inter-processor communication etc.; (d) analysis and optimization of unstructured computations such as those arising in fluid dynamics codes; (e) runtime compilation of irregular sparse matrix codes that exploits both parallelism and data locality; (f) interprocedural analysis and optimization; (g) integration of task and data parallelism; and (h) high-performance input and output.

Oregon Graduate Institute of Science & Technology; Tim Sheard and John Launchbury; *1996 Summer School on Advanced Functional Programming; Pacific Software Research Center, Portland, Oregon*; (CCR-9614784); $10,000; 12 months.

The Pacific Software Research Institute held a summer school on advanced functional programming techniques. The school was designed to be of use to university faculty, industry and government workers interested in software technology. It also brought together a group of leading researchers in advanced functional programming to lecture and interact with the other participants. This allowed in-depth discussion by both researchers and practitioners in the field, providing an assessment of the state of the art and new directions of research needed.

State University of New York – Stony Brook; Scott A Smolka and Philip M Lewis; ***Practical Techniques for the Design, Specification Verification, and Implementation of Concurrent Systems*****; (CCR-9505562); $308,036; 36 months. (Joint support with the Software Engineering Program)**

Over the past decade, researchers have developed tools that offer automated support for applying formal methods to the design and analysis of concurrent software systems. In principle, these tools allow system designers to utilize formal system verification techniques without requiring them to know all the details of the underlying mathematical theories. In practice, however, the impact of the tools has been limited because of the relatively small scale of the problems they can handle and because of their lack of support for more traditional development methods for software, such as structured design techniques, simulation, and debugging. This research project aims at extending the applicability of existing automated formal techniques for concurrent systems to problems of real-life scale and complexity. The focus is on two broad issues: (1) increasing the efficiency and scalability of existing approaches; and (2) enhancing the usability of existing techniques by integrating them with more traditional software development methods. More specifically, it seeks to develop advanced state-space management techniques for checking system specifications against logical properties expressed in the modal mu-calculus, and techniques for graphically conveying diagnostic information for systems that fail to meet their logical specifications. The impact of the research is being assessed by analyzing several real-life applications, such as sliding-window communications protocols and cache coherency protocols for distributed shared memory. This research builds on the investigators' prior work on the Concurrency Factory, an integrated toolset for specification, simulation, verification, and implementation of concurrent systems.

Michigan Technological University; Philip H Sweany; ***Research Initiation Award (RIA): Global Instruction Scheduling Without Copies*****; (CCR-9308348 A003); $5,000; 0 months.**

Adding instruction-level parallelism to an architecture is a technique becoming popular for increasing power in modern computers. One consequence of this trend is the need to include instruction scheduling in a compiler. "Global Instruction Scheduling Without Copies" investigates a method of global scheduling which, unlike other global scheduling techniques does not suffer from the need for compensation copies. The work's goal is refining a technique for scheduling which will work better than existing methods for ILP architectures with limited parallelism, a characteristic of ILP architectures found today, and expected for superscalar architectures in the future. The study investigates dominator-path scheduling (DPS) in some detail and compare DPS to other global scheduling methods. Since DPS is better suited than other scheduling techniques for the levels of ILP found in current architectures, one outcome of the research is the ability to generate better code for currently popular computers. Additionally, the work quantifies the parallelism available in imperative programs when compensation copies are not allowed. This measurement of such parallelism will influence the design of ILP architectures in the future.

University of Tennessee – Knoxville; Bradley T Vander Zanden; ***Adaptive, Profile-Based Compilation of Highly Interactive Graphical Applications by Demonstration*****; (CCR-9633624); $168,729; 36 months.**

Rapid prototyping environments provide high-level operations that allow users rapidly to create and modify interactive applications, such as graphical interfaces, scientific and program visualizations, and simulations. These environments often sacrifice speed and storage for increased flexibility. Researchers have proposed optimizations that can reduce storage and improve performance. However, the programmer must manually apply these optimizations, which impedes rapid, iterative development of applications. The objective of this project is to develop adaptive, profile-based compilation techniques that automatically apply these optimizations so that the programmer can make use of rapid prototyping techniques while still retaining excellent time and storage performance. Constraints and structured graphics are targeted by this research since studies show that these two techniques facilitate the rapid, iterative development of graphical user interfaces. This project is developing techniques whereby: (1) a profiler monitors operations as they execute and extracts information about constraint evaluation and variable access, and (2) an adaptive compiler uses this information to dynamically implement various performance and storage optimizations.

State University of New York – Stony Brook; David S Warren and I. V Ramakrishnan; ***Enhancing the Power and Performance of Tabled Evaluation of Logic Programs*****; (CCR-9404921 A003); $106,000; 12 months. (Joint support with the Knowledge Models and Cognitive Systems Program and the Numeric, Symbolic & Geometric Computation Program)**

This project continues work on the development of XSB system. XSB is a system that integrates logic programming, deductive databases and non-monotonic reasoning. It uses SLG resolution, a tabled evaluation strategy, to combine Prolog's SLD and deductive database's bottom up evaluation in a computationally effective way. Current benchmarks show that the XSB system is significantly faster than more traditional approaches to deductive database query evaluation, and shows promise of being very efficient for certain types of non-monotonic reasoning. In the current project, the XSB implementation will be improved by; (1) improving the underlying WAM-based engine to make it more efficient, and (2) developing program transformation techniques to permit XSB programs to be optimized at the source level.

California State University – Los Angeles; Nancy J Warter-Perez; *NSF Young Investigator (NYI) Award: Aggressive Scheduling for Instructor-Level Processors*; (CCR-9457395 A002); $85,409; 12 months.

In this project, aggressive instruction scheduling techniques for instruction-level parallel (ILP) processors are investigated. ILP processors, such as superscalar and VLIW processors, are used effectively in both general purpose and scientific computer systems. For general purpose systems, the high frequency of branch instructions creates the need for aggressive global scheduling techniques that support speculatively and conditionally executed instructions. For scientific systems, cyclic scheduling techniques such as software pipelining have been developed to aggressively schedule inner loops. The four areas that are explored in this project are: (1) Investigate issues associated with the software pipelining technique and Modulo scheduling. These include scheduling while loops, handling loops with cross-iteration dependencies and removing the uniform initiation interval restriction; (2) Formalize an operation-level intermediate program representation that allows the application of well-known local scheduling techniques to achieve globally scheduled code. This approach has been successfully applied to Modulo Scheduling and is extended to acyclic scheduling; (3) Develop a task granularity adjustment methodology, similar in concept to vectorization, for software pipelining in a superscalar/VLIW based multiple processor system; and (4) Analyze the control architectures for superscalar implementations of existing processors by studying the hardware and compiler design complexities associated with scheduling multiway branching and partial predication architectures.

Slippery Rock University of Pennsylvania; Deborah L Whitfield; *Research at Undergraduate Institution (RUI) : Acquisition of Combining Optimizations to Create Powerful Parallelization Techniques*; (CCR-9628760); $94,319; 24 months. (Joint support with the Science and Technology Centers Program and the NSFNET Program)

This is a Research at Undergraduate Institution (RUI) project. This project will investigate both the techniques for combining optimizations in compilation and the potential benefit of applying combined optimizations. A determination of when optimizations can be combined has not been discovered. Likewise, a tool to automatically generate an optimizer for these more powerful combined optimizations has not been developed. The objectives of this ongoing project are: investigate the properties of optimizations that permit their combination, describe when two optimizations may be combined, and propose a tool which generates an optimizer for the newly combined optimizations. A generalization of the description to combine chains of optimizations will also be investigated. This project will be used as a vehicle to introduce undergraduate students to the research process and actively involve students in the theoretical investigation, the implementation, the experimentation, and the dissemination of the results. Students will be encouraged to involve themselves in all aspects of this project.

University of Texas – Austin; Paul R Wilson; *Research Initiation Award (RIA): Automatic Storage Management for Real-Time, Near-real-time, and Distributed Systems*; (CCR-9410026 A002); $23,982; 12 months.

This project explores techniques for automatic storage reclamation, especially for real-time and distributed systems. It compares techniques for incremental garbage collection in terms of efficiency, real-time guarantees, and performance for distributed data. The focus is on techniques that support high-performance on standard hardware, and standard compiler optimizations. The three major phases of the project are; (1) implementation and evaluation of a new real-time uniprocessor garbage collection algorithm which provides better real-time guarantees than incremental copying due to weaker coupling between the application and the garbage collector, (2) extensions of the algorithm for near real-time systems where some real-time guarantees can be sacrificed for higher performance, and (3) application of these incremental techniques as local garbage collectors in a distributed garbage collection system, where they will enable higher performance garbage collection of poorly distributed data.

Oregon Graduate Institute of Science & Technology; Michael Wolfe; *Travel Support for US Participants at an International Workshop - April 15-19, 1996*; (IRI-9527926); $22,000; 12 months.

The International Workshop on Loop Parallelization was held in April 1966 in Wadern, Germany. Its goal was to bring together two research communities which are both working on Loop Parallelization, but applying differing models and criteria. One of these communities has concentrated on parallelizing compilers for sequential loops, pragmatically developing heuristic methods that may work at high speed for a variety of loops, but are not provably optimal, while the other community works on regular array design, identifying optimal parallelizations for highly regular loop nests. The purpose of the Workshop was to get the groups working together on a variety of common problems. This grant supported participation of U.S. scientists from the two relevant communities working on this important aspect of parallel computation.

University of Colorado – Boulder; Benjamin G Zorn; *Improving the Performance of Dynamic Storage Allocation Using Behavior Prediction*; (CCR-9404669 A002); $62,819; 12 months.

This project develops optimizations based on object behavior prediction to improve the performance of manual and automatic storage reclamation algorithms. Programs written in languages such as C, C++, or Fortran that use manual garbage collection, and programs written in languages such as Common Lisp, ML, Smalltalk, or Modula-3 which uses automatic storage reclamation benefit from the techniques developed in this project. The project uses profile-based optimizations to predict properties of allocated objects at the time they are allocated. The properties include object lifetime and object reference behavior. By correctly predicting these properties, the space usage, CPU overhead, and reference locality of programs using either manual or automatic storage reclamation is improved. The project first investigates empirical properties of programming languages using manual and automatic storage reclamation and determines if behavior prediction is feasible. Then the project develops and evaluates alternative algorithms that exploit observed predictable behaviors. Finally, the project prototypes the most promising algorithms and measures their performance.

Automated Program Generation and Verification

University of Kansas Center for Research Inc; Michael J Ashley; *CAREER: Employing Flow Analyses in Practical Program Transformation Environments for Mostly-Functional Languages*; (CCR-9623753); $100,000; 24 months.

A flow analysis is an automatic tool that collects information about programs. Flow information is crucial, for without it the correctness of sophisticated program transformations cannot be verified. Such transformations are important, because they can transform a program written cleanly to solve a problem into a new program that, for example, is more efficient and fault tolerant. Ultimately, flow-analysis-justified transformations can boost software reliability and programmer productivity. Despite their utility, however, existing flow analyses for mostly functional languages are impractical for use in production-quality environments. They have high algorithmic complexity, they can only process language subsets, and they are accurate only when programs are written in a certain style. This CAREER project has three goals to address these problems and incorporate advanced analysis techniques into the undergraduate curriculum. The first goal is to use a practical flow analysis in a production Scheme compiler to justify optimizations. The second goal is to use the flow analysis and the experience gained from its application to develop a program transformation toolkit for building automatic, flow analysis-based program transformers. The third goal is to incorporate the results into the undergraduate curriculum so that students understand what a flow analysis is, how it works, and how it can be profitably applied. Achieving these goals advances both research and education. Besides solving the above problems, the research aspect of this project pursues contributions in the areas of language implementation and semantics-based program manipulation. In particular, optimizations are being implemented that can improve the performance of dynamic languages, and program transformation techniques are being developed that allow more realistic programs to be manipulated. The transformation toolkit developed is expected to provide the research community with needed infrastructure for experimenting with transformations. The educational aspect of this CAREER investigation seeks to improve the undergraduate programming languages curriculum by addressing the distinction between static and dynamic properties of programs and how static information can be used to optimize programs. Such an understanding is essential to undergraduate education as programming languages are playing an increasingly prominent role in customizable software.

Pennsylvania State University – University Park; John J Hannan; *CAREER: Specification and Verification of Compiler Algorithms*; (CCR-9502356 A002, A003); $45,000; 12 months.

The research goal of this CAREER project is the formal specification and verification of algorithms that support the implementation of programming languages. Such formal systems can aid the development of new languages and efficient implementation of these languages. This support arises from both the verification of algorithms, which increases confidence in the correctness of a compiler, and from the improved understanding of compilation techniques allowed by formal specification. Recent work has focused on using logical frameworks and mechanized proof assistants to specify deductive systems representing aspects of compilation and to verify the correctness of these specifications. This research project extends these results by considering advanced, state-of-the-art algorithms found in current compilers, formalizing theories for these algorithms, and verifying the correctness of these algorithms. One specific problem area to be addressed involves closure conversion algorithms and the concept of 'safe-for-space complexity.' An important aspect of this project is extending the use of mechanized proof assistants to support the verification of complex flow analyses and optimizations of programs. The goal of the education plan is to provide both a general, but rigorous, introduction to the study of programming languages at the undergraduate level and would introduce, at both the undergraduate and graduate level, the recently emerging field of logic and computation which studies the relationship between constructive logics and typed lambda calculi. Although this is typically considered advanced material and not suitable for undergraduates, it can be made accessible to students.

SRI International; Natarajan Shankar; ***An International Workshop on Applied Formal Methods: Hyderabad, India, December 14-17, 1996*****; (CCR-9633644); $10,349; 6 months.**

This grant supported the American participants in a workshop on Applied Formal Methods held in Hyderabad, India, in December 1996. The goal of the workshop is to bring together several international experts from the U.S., India, and elsewhere with the purpose of capturing the state of the art in the technology and the state of practice in the applications of formal methods. The workshop scope includes the study of specification languages and logics and verification methods, such as theorem proving, model checking, and safety analysis. Formal methods are increasingly applied to the core algorithms in safety-critical systems and in the verification of hardware circuit designs. A workshop of this nature provides a timely means for developing a common understanding of the capabilities and the potential applications of these technologies. The organizers will edit a technical report providing an up-to-date survey of applied formal methods.

SOFTWARE ENGINEERING

Specification and Design

Florida Atlantic University; Robert B France; *Research Initiation Award (RIA): Incorporating Formal Specification Techniques into Structure and Object-Oriented Analysis and Design Processes*; (CCR-9410396 A001); $19,200; 0 months.

A significant reason for the rare use of formal specification techniques (FSTs) in the software development industry is a lack of understanding of how FSTs can be incorporated into current industrial-strength software development processes. Current industrial software development environments tend to favor the use of informal Structured Specification Methods(SSMs) and Object-Oriented Methods(OOMs) because of their use of relatively simple, intuitive notations and concepts. Lack of understanding of how FSTs can be effectively applied in the context of practical development processes based on SSMs and OOMs often leads to the perception that the use of one precludes the other. This research is concerned with bridging the gap between FSTs and the less formal SSMs and OOMs. The research program has three components: the technical component is concerned with how integration of formal and informal techniques is accomplished, the organizational component is concerned with when, how, and under what circumstances integrated specification techniques can be applied, and the experimental component is concerned with evaluating the techniques, supporting tools and processes through experimental application.

University of Minnesota – Twin Cities; Mats Per Erik Heimdahl; *CAREER: Requirements Specification for Embedded Systems: An Integrated Framework*; (CCR-9624324); $128,583; 24 months.

The vast majority of serious software design flaws are due to incomplete, inconsistent, or obscure requirements documents. It is in the very first stages of software development, the requirements specification phase, that the fewest formal procedures exist to aid the analyst. Techniques to capture, verify, and validate requirements specifications for embedded systems are of demonstrable importance. This project focuses on validation of requirements specifications, specifically on executable specifications. An integrated framework is being constructed for the development, analysis, and execution of high-level requirements specifications for embedded systems. The framework is intended to support (1) execution and simulation of embedded system requirements (hardware-in-the loop simulation), and (2) requirements reuse and requirements synthesis based on the structure and design of the physical system. This environment will enable an executable high-level requirements specification to interact with: (1) models of the assumed behavior of the environment, (2) software simulations of the environment, or (3) the actual physical components in the environment. The approach taken is to extend an existing formal specification language, RSML, with the capability to define system-level intercomponent communication in an embedded system. The educational component of this CAREER project seeks to increase industrial involvement in undergraduate and graduate education through case studies and industrial projects. A remotely accessible embedded systems laboratory is being established, where students can gain realistic hands-on experience with software development for real-time systems.

Kestrel Institute; Richard Jullig; *International Workshop on Formal Foundations of Software Systems: Rio de Janeiro, Brazil, May 5-9, 1997*; (CCR-9617781); $37,200; 6 months. (Joint support with the Programming Languages and Compilers Program, the Numeric, Symbolic & Geometric Computation Program, and the Theory of Computing Program)

This grant supports an international workshop on Formal Foundations of Software Systems in Rio de Janeiro, Brazil, May 5-9, 1997, to explore future directions in formal methods, programming logics, and programming languages. The workshop combines 3 days of technical exchange with 2 days of discussions to produce recommendations to the science funding agencies of the U.S. and Brazil: NSF, CNPq, and FINEP. This workshop's goal is to promote collaboration among Brazilian and U.S. researchers and provide information and recommendations for future research efforts and funding initiatives.

University of Washington; Nancy G Leveson; *Software Requirements and the Human-Computer Interface*; (CCR-9520813); $94,999; 12 months.

Previous research led to software requirements completeness criteria for process control systems. These criteria, however, did not include the human-computer interface. This research involves extending these criteria to specify appropriate abstractions and completeness requirements for man-in-the-loop systems and devising hazard analysis techniques for the human-computer interface. Five questions are examined: (1) What information needs to be specified in a software requirements specification to ensure safety for software that includes a human-computer interface; (2) How can traditional hazard analyses be extended to include human-computer interface issues; (3) How can the information from these hazard analyses be used to improve the human-computer interface design; (4) How can a state-machine-based requirements specification be analyzed for compliance with some of these human-interface criteria; (5) How can specification languages be designed to enhance the ability to perform this analysis, and to enhance the ability to generate safe human-computer interface designs.

University of Washington; Nancy G Leveson; *Specification and Design Support for Software Evolution*; (CCR-9633428); $70,000; 24 months.

This research is investigating specification and design support for software evolution by applying hierarchical abstractions based on intent rather than the aggregation and information-hiding abstractions commonly used in software engineering. Cognitive psychologists have shown that intent abstraction is consistent with human problem-solving processes. The hypothesis being evaluated in this research is that new types of specifications and designs based on this concept can assist in understanding existing software, capture the most important design rationale information in an efficient and economical way, and support the process of identifying and analyzing required changes to minimize the introduction of errors during maintenance. To evaluate these hypotheses, protocol analysis techniques are being used to explore the processes involved in software evolution tasks, and other empirical studies are being used to evaluate the impact of this new form of specification and design on these processes.

University of Illinois – Chicago; Tadao Murata and Sol M Shatz; *Adapting Petri Net Theory and Methods to Specification and Analysis of Concurrent Software*; (CCR-9321743 A002); $93,644; 12 months.

One of the significant problems in concurrent-software analysis research is achieving a balance between general theoretical concepts and specific practical techniques. This is the motivation for this research. The focus of the project is on applying and adapting Petri net theory for solutions to pragmatic problems in concurrent-software analysis. The specific research topics addressed represent natural points of extension from previous efforts. The research explores Ada tasking analysis issues within the framework of an existing experimental toolkit and also expands on research in analysis of Petri nets using some new reachability concepts and techniques. One area of emphasis is in empirical studies of various deadlock detection techniques and the applicability of these techniques to generic language structures beyond the example of Ada tasking. A second emphasis is the investigation of a generalized form of reachability graph and its construction. This area of work exploits some identified connections to recent research in use of Binary-Decision Diagrams (BDDs) for representation of symbolic state spaces.

University of Washington; David S Notkin; *Reflexion Models: Bridging the Gap Between Source and High-Level Models*; (CCR-9506779 A001); $81,000; 12 months.

Given an existing software system and a task to accomplish, software engineers decide whether and how to proceed based on experience, judgment, and system-specific knowledge. For even modest-sized software systems, there is a substantial risk of making poor decisions because it is difficult (if not impossible) for software engineers to maintain adequate knowledge of the source. To manage this risk, engineers often use high-level models to focus on particular aspects of the system relevant to the task. These models help the engineer reduce risk only if they are sufficiently accurate representations of the actual system. This research investigates an approach that may help engineers determine whether their task-specific models are good enough to make informed decisions. An engineer defines a high-level structural model appropriate for the task at hand and also specifies how the model maps to the source. A tool then computes a reflexion model that shows the engineer where the high-level model is consistent with and differs from the source. Analyzing the reflexion model and applying personal experience and judgment, an engineer can then assess the risk of proceeding with the task. In some cases, especially when there are unexpected inconsistencies between the model and the source, computation of additional reflexion models may be needed before an informed decision can be made. This project concentrates on two activities intended to determine whether reflexion models genuinely help improve our ability to effectively engineer software. The first activity focuses on specific enabling technical problems related to reflexion models. The second activity consists of case studies with industrial partners. Kevin J. Sullivan, University of Virginia participates through a research subcontract.

Georgia Technology Research Corporation – Georgia Institute of Technology; Colin Potts; *Improving Requirements Completeness through Inquiry-Based Goal Refinement*; (CCR-9503863); $49,966; 24 months.

Requirements completeness is the extent to which an application meets user needs. This project is to develop and evaluate novel methods for improving and predicting the completeness of requirements with an emphasis on interactive applications for individual users and small groups. It is assumed that requirements can be derived from user and enterprise goals, and so completeness is defined as the fit between a feature set and the goals and goal-thwarting obstacles that arise in the use context. There are two research challenges: (i) development of methods for analyzing the use of context to uncover goals and obstacles, and (ii) development of methods for deriving requirements from goals and obstacles. Approaches are to be developed for performing these transformations: identifying goals, decomposing goals into subgoals, clarifying goal dependencies, transforming goals into operations, establishing the boundary between application and environment, allocating operations to application and users, identifying obstacles, inventing scenarios for analysis of obstacles, and adding operations or goals to deal with obstacles. The project also seeks to extend and specialize for goal refinement the Inquiry-Based Requirements Analysis process (in which requirements derivation intertwines the expression of models, discussion of their content, and transformation). Development and evaluation of the goal refinement strategy are proceeding simultaneously. The number and quality of additional requirements that are identified as a result of applying the strategy to realistic case studies will be measured. Industrial case studies are carried out in parallel with industrial funding.

SRI International; Xiaolei Qian; ***Trusted Interoperation of Health-Care Information Systems*****; (CCR-9422688 A001); $435,461; 12 months. (Joint support with the CISE Institutional Infrastructure Program, the Operating Systems and Systems Software Program, and the Computational Engineering Program)**

This award is made in support of the NSF initiative ENG-CISE-IITA94, National Challenges. It is a joint initiative between the Engineering Directorate and the Computer Information and Engineering Directorate. SRI International, in collaboration with Stanford University Department of computer Science and Medicine, and Inova Health System, will develop and demonstrate techniques for the trusted interoperation of autonomous heterogeneous health care databases containing sensitive data that mismatch in semantics, representations, and security/privacy policies. This effort will be built on the query mediation framework developed under ARPA and Rome Laboratory sponsorship. The framework supports the automated and efficient mediation of queries between autonomous heterogeneous databases. Semantic connections between component databases are encoded in a knowledge base, which are then used by a mediator to automatically resolve mismatches and transform queries between component databases using the local language and schema, making a local database access data in multiple databases using the local language and schema, making interoperable both the data and the applications accessing the data in legacy databases. Technique will be developed and demonstrated for solving three critical problems that hinder trusted query mediation. In particular, the following questions will be answered: 1. A security/Privacy Policy Framework. What is the relationship and difference between security and privacy? How can the security/privacy policy of a health care database be characterized and formalized? How can we compare the security/privacy policies of autonomous heterogeneous health care databases? 2. Trusted Query Mediation. When autonomous heterogeneous health care databases are interoperating through query mediation, how can we guarantee the autonomy and security/privacy of component databases. In other words, how can queries be mediated such that neither the rights nor the security/privacy of components databases are compromised? 3. Schema Mediation. In constructing a physician's workstation as a uniform interface to multiple autonomous heterogeneous health care databases, how can a consistent view (schema) be constructed that (1) captures exactly the data that are relevant to users of the workstation, and (2) resolves all the mismatches in the relevant data? These techniques will be implemented in a demonstration system that leverages off an existing one developed for the query mediation framework.

Washington University; Gruia C Roman; ***Reasoning About Program Interactions in the Presence of Mobility*****; (CCR-9624815); $132,270; 24 months. (Joint support with the Networking Research Program)**

Mobile computing is emerging as an important new paradigm which has the potential to reshape thinking about distributed computing. Mobility has far-reaching implications for what designers and users can assume about communication patterns, resource availability, and application behaviors. Components move from one location to another, joining or leaving groups of other components in their vicinity. New distributed algorithms are expected to be required as the nature of applications shifts with the emergence of this new kind of computing environment. Ad-hoc wireless networks (AWNs) supporting context-aware applications are studied in this research. Such networks have practical significance for crisis management, communication for small enterprises, and organization of transient collaborations. The somewhat extreme characteristics of AWNs provide a challenging testbed for studying mobile computing. This research examines whether existing models of concurrency can be extended to mobile computing and thus provide appropriate means for the specification, design and verification of mobile systems. It also considers related support system software, such as communication protocols. To accomplish this, the notion of location must play a more central role in the treatment of concurrency, even in cases where the ultimate goal is location-transparency. New forms of modular interaction are investigated in this research. Of particular interest are techniques to gain the advantages of formal methods without the costs of developing formal proofs.

Design schemas and patterns with desirable properties are verified once and reused in many applications.

University of Arizona; Richard D Schlichting; ***Designing and Implementing Adaptive Distributed Systems*****; (CCR-9633336); $100,000; 24 months. (Joint support with the Operating Systems and Systems Software Program)**

An adaptive computing system is one that evolves its behavior based on changes in the environment. For an adaptable system built on a distributed architecture where multiple machines are connected by a network, such changes can range from site failures to updated application requirements. The evolutionary action taken when a change occurs can vary as well, ranging from dynamically reconfiguring the system software to maintain the same service guarantees under changed conditions, to negotiating new service guarantees with the application. The goal of this project is to develop fundamental new techniques for designing and implementing adaptive distributed computing systems. The specific focus is on providing a distribution support layer, a software layer implementing strong execution guarantees regarding the distributed state of the application, that can adapt to changes. The research concentrates on two lines of investigation: (1) the development of a conceptual model for systems of this type and (2) a new approach to implementing adaptive distributed systems based on fine-grain code modules called micro-protocols. The framework is being extended to support dynamic evolution of software, including reconfiguration, dynamic code linking, and runtime code generation.

University of Illinois – Chicago; Jeffrey P Tsai and Aravinda P Sistla; ***A Formal Approach for a Specification-Based Software Architecture of Real-Time Systems*****; (CCR-9633536); $170,001; 24 months. (Joint support with the ARPA-EVOLUTIONARY DESIGN OF COMPLEX SYSTEMS Program)**

Real-time systems are becoming increasingly important to everyday life. With increasing computing power and high demand for multimedia applications, larger and more complex real-time systems are being continuously developed. However, the development of real-time systems is difficult and takes more time than the development of a system without real-time constraints, and a significant portion of the entire effort remains devoted to the maintenance activities of the software life cycle. The concept of software architecture has recently emerged as a new way to improve our ability to effectively construct large-scale software systems. In spite of many attempts to develop architecture specification languages, this area still lacks a solid formal foundation. Without a solid foundation, it is difficult to develop and apply rigorous analysis methods to assure the correctness of real-time software systems as well as to analyze the impacts of changes on various system properties such as safety, consistency, and reliability. This research seeks to develop a formal foundation for an architecture specification language using the integration of non-monotonic logic and temporal logic. The syntax of the language is based on the concepts of object, frames and production rules. This allows developers to easily describe the attributes and constraints of each software component, and also the hierarchical relations among the software components in the software architecture. The goal is to localize changes to software components in the low-level architecture specification and avoid modifying the high-level architecture specification. The semantics of multiple inheritance and exception handling will be established using a non-monotonic logic. Based on the formal foundation of the architecture specification language, it is expected that new possibilities will exist for analyzing software properties such as liveness, safety, consistency, and temporal constraints. This foundation is also expected to provide a basis for analyzing the impact of changes to a real-time software system.

Program Testing, Analysis, and Verification

University of California – Berkeley; Alexander Aiken; ***Program Analysis: Logics, Algorithms, and Application*****; (CCR-9416973 A001); $56,412; 12 months.**

This project undertakes a systematic investigation of a general class of structural program analysis problems. The term structural program analysis refers to compile-time program analyses that deal with the types and structure of values computed by programs. This class of program analyses is broad, including within it the classical dataflow analyses of C and FORTRAN compilers, type inference for languages such as ML, and analysis of class membership of values in object-oriented languages. The primary goal of this work is the synthesis of set constraint techniques for structural program analysis in a common framework; in particular, an implementation of set constraints suitable for use in a general class of program analysis problems will be developed and distributed for use in research and teaching. There has been considerable activity directed towards using set constraints for structural analyses in recent years. Most of the work has been directed either towards studying the theory of set constraints itself, or at applications of set constraints to particular problems. Set-based analysis is the exception; set-based analysis provides a high-level characterization of a class of program analyses in terms of set constraints. Currently, however, there is no general-purpose implementation of set constraints available for experimentation with set-based (or any other class of) analyses. This system is expected to be expressive enough to encompass all set-based analyses and many analyses that are not set-based. Aside from the primary goal of developing an 'off-the-shelf' implementation of set constraints that can be used in many settings, the project also pursues three secondary goals. The first goal is to develop a common technical language for structural analyses. The second goal is to describe the structure of the design space for structural program analyses using this common technical framework. The tradeoffs between accuracy and speed in structural analyses are poorly understood. By studying these problems in an abstract setting, it should be possible to make general statements about the relative efficiency and precision of various program analysis techniques that apply across a wide spectrum of applications. The third goal is to design algorithms using the above results. General algorithms are sought that provide the best precision-speed tradeoffs for various classes of analysis problems.

University of Massachusetts – Amherst; George Avrunin; *Automated Analysis of Concurrent Systems*; (CCR-9407182 A002); $66,848; 12 months.

The major obstacle to developing automated tools for the detection of faults in concurrent software systems is the state explosion problem: the number of possible states of a concurrent system is, in general, exponential in the number of processes making up the system. A number of techniques have been proposed to address this problem, but most real concurrent software systems remain beyond the reach of existing approaches to analysis. This project continues the development of an approach to analysis based on generating necessary conditions for the existence of certain executions of a concurrent system and using integer programming to check the consistency of those conditions. This approach does not require enumeration or explicit representation of the states of the system being analyzed, and can be used to check such properties as deadlock, starvation, mutual exclusion, and minimum and maximum elapsed time between events. This project includes research to take advantage of the compositional structure and symmetries of a concurrent system and investigates connections with other automated analysis methods.

Carnegie Mellon University; Edmund M Clarke; *Automatic Verification of Finite-State Concurrent Systems in Hardware and Software*; (CCR-9217549 A003); $150,000; 12 months. (Joint support with the Design Automation Program Program)

Logical errors in sequential circuit designs and communication protocols have always been an important problem. They can delay getting a new product on the market or cause the failure of some critical device that is already in use. The most widely used method for verifying such systems is based on extensive simulation and can easily miss significant errors when the number of possible states is very large. This research deals with developing an alternative approach based on a technique called temporal logic model checking. In this approach specifications are expressed in a propositional temporal logic, and sequential circuits and communication protocols are modeled as state transition systems. An efficient search procedure is used to determine automatically if the specifications are satisfied by the transition system. The technique has been used in the past to find subtle errors in a number of non-trivial designs.

North Carolina State University – Raleigh; W. Rance Cleaveland; *NSF Young Investigator (NYI) Award: Theoretical Underpinnings of Formal Analysis of Concurrent Systems*; (CCR-9257963 A004); $25,000; 12 months.

This research addresses automated tools for design and verification of finite-state concurrent systems and the theoretical underpinnings of formal analysis of concurrent systems. Application of the tool to verification of communications protocols and to other real-world concurrent systems are examined. The theoretical investigations provide insight into reasoning about real-time and probabilistic behavior of systems.

University of California – Santa Barbara; Laura K Dillon; *Automated Support for Testing and Debugging of Real-Time Programs Using Oracles*; (CCR-9505392 A001); $74,961; 12 months.

One of the great challenges facing today's software engineers is the development of reliable systems which are known to perform correctly in all circumstances. The most critical applications often involve concurrency and real-time, which increase the difficulty of system development and validation. Testing and debugging cannot prove a system is correct; however, they are currently the most effective methods available for validating real applications. This project investigates the use of oracles created from Graphical Interval Logic (GIL) specifications in testing and debugging of real-time programs. GIL is a visual logic for specifying and reasoning about properties of real-time systems. The existing GIL tools allow mechanical checking that system specifications guarantee critical correctness requirements. This project seeks to produce a method for relating GIL specifications to real-time programs that permits deterministic finite-state automata constructed from GIL specifications to be used in verifying that executions of a program are correct. When an execution violates a specification, the associated oracle would construct a GIL formula describing a fault in the execution. This formula would be displayed graphically, appropriately aligned with an execution trace, to help the user see where in the trace the fault occurred and the nature of the fault. Formal methods must be automated if software engineers are to use them in the development of real systems. The implementation and experimental evaluation of prototypes is therefore an important facet of this project, which combines analytical research on formal methods with experimental research aimed at assessing the practical utility of formal methods.

University of Texas – Austin; E. Allen Emerson; *Formal Reasoning about Reactive Systems*; (CCR-9415496 A001); $79,944; 12 months.

There is a growing need for effective formal methods of constructing correct and reliable computer programs. This need becomes especially crucial in the case of concurrent, distributed, real-time, or, more generally, reactive systems. Such systems typically involve a number of delicate and subtle aspects such as parallel operation of distinct, often geographically dispersed components, and ongoing, ideally nonterminating computation where each component repeatedly interacts with neighboring components in order to achieve a global goal or maintain a global invariant. Many safety critical real-world applications programs can be counted among the reactive systems. Examples include: computer operating systems, computer network communication protocols, air traffic control systems, on-board avionics control systems, as well as automated banking networks. The approach taken in this research to reasoning about such reactive systems centers around the use of Temporal Logic, a formalism that can be used for specification, verification, design, and synthesis of reactive systems. This research investigates to what extent efficient automation of temporal reasoning about reactive systems is possible. Considerable attention is devoted to the following particular aspects: attacking the state explosion problem, reasoning about many homogenous processes, real-time reasoning, and synthesis of reactive systems.

Carnegie Mellon University; David Garlan; *Foundations for Reasoning About (Practical) Implicit Invocation Systems*; (CCR-9633532); $75,000; 36 months.

A critical issue for large-scale systems design and evolution is the choice of an architectural style that permits the integration of separately developed components into larger systems. Familiar styles include those based on remote procedure call, shared variables, asynchronous message passing, etc. One increasingly important architectural style for system composition is implicit invocation (II). However, there is no established basis for reasoning about II systems. In particular it is difficult to answer questions like: What will be the effect of announcing a given event? Have enough event bindings been declared? Does a given component announce sufficient events to permit effective integration? If I add a new component to an existing system, will it break the existing system? The goal of this project is to provide such a basis for reasoning about systems designed using the II architectural style. Two central benefits are expected. The first benefit is a step towards being able to reason fully about systems constructed in this style. The second and probably more important benefit is to replace current ad hoc reasoning approaches used by practitioners who use the II style with a collection of sound ideas that allow better informal reasoning about such systems. The research is carried out in collaboration with David Notkin of the University of Washington.

Massachusetts Institute of Technology; Stephen J Garland; *Automated Reasoning in Software Engineering*; (CCR-9504248 A001); $75,000; 12 months.

The aim of this research project is to increase the benefits and to reduce the costs of automated reasoning across a wide variety of applications. The project both supports and derives motivation from the ongoing development of the Larch Prover (LP), an interactive proof assistant for multisorted first order logic. Experience with previous versions of LP suggests four important ways to improve the cost/benefit ratio: reducing the amount of detail required in proofs, supporting modular descriptions of theories and proofs, closing the gap between the logical and operational meaning of theories, and integrating LP into applications. To reduce the amount of detail required in proofs, the project seeks to combine different reasoning techniques in seamless and efficient ways. The project explores fruitful interactions between equational term-rewriting and decision procedures. The project addresses both logical and operational issues for using independently developed logical modules in larger proofs and for ensuring that proof scripts are replayable after changes to formalizations. For almost all applications of automated reasoning, there is an unavoidable gap between logical content (e.g., the consequences of a specification) and operational content (e.g., the proof techniques and/or sequence of lemmas required to prove a particular consequence). The project studies how to narrow this gap, where possible, and how to provide useful information about the nature of the gap, particularly when proofs fail. Finally, the project seeks to improve the cost/benefit ratio for formal reasoning by integrating LP into applications. The goals of tighter integration are to enable users to think in application-specific terms, to make LP work better in the context of particular applications, and to understand the interface reasoning tools should present to applications. The project is seeking to achieve tighter integration by eliminating the need for transcribing designs and by developing application specific improvements in LP's expressive and deductive power (e.g., for reasoning about probability).

University of Wisconsin – Madison; Susan Horwitz; *Program Dependence Graphs and Dataflow Analysis*; (CCR-9625656); $160,787; 24 months.

The goals of this project are to advance the state-of-the-art in language-based tools for software engineering. There are three related aspects to the research: (1) extending dependence-graph-based tools, (2) carrying out theoretical and practical studies of pointer analysis, and (3) quantifying trade-offs in dataflow analysis. The work builds on the prior research in two areas: the design and implementation of software engineering tools, and the design, implementation and evaluation of algorithms for dataflow analysis. The first part of the work includes the continuation of previous NSF-supported research extending underlying theory to permit graph-based tools to be used for a wider range of languages than is currently possible, and improving existing graph based tools. It also includes new directions: the design and implementation of new tools using a hybrid approach that combines text or syntax-based tools with semantic checks. The research on pointer analysis and dataflow analysis tradeoffs is complementary to the work on dependence graphs, since the results are expected to improve the graph-based tools. They are also expected to be useful to other researchers and practitioners concerned with applications that rely on the results of static analysis.

Carnegie Mellon University; Daniel Jackson and Jeannette M Wing; *Model Checking of Software Systems*; (CCR-9523972); $102,735; 12 months.

Software systems keep growing in size and complexity. Our reliance on such systems is likewise increasing, making it critical that they obey the properties their designers intended. Because these properties are global, their consideration cannot be left to the implementation of modules; rather, they must be ensured by design early on. Achieving critical properties thus depends on being able to analyze designs, so that the consequences of design choices can be explored without the cost of implementation, and potentially disastrous flaws can be caught while their repair is still relatively cheap. Analysis methods based on formal, rather than informal, techniques are amenable to mechanical reasoning. Formal methods have had a dramatic impact in hardware verification – principally due to the success of model checking – and are being adopted by industry. In software development, on the other hand, formal methods have not yet had the same impact. This project has two aspects. First, it investigates the application of model checking techniques to software. This involves both the identification of appropriate abstract representations of large systems, along with the properties they should obey, and the invention of new abstraction techniques to relate such representations to software designs. The abstraction techniques pursued in this research arise in three ways: (1) from the form of the property to be checked (for example, merging states to preserve a universally quantified property); (2) from the application area (for example, not distinguishing the varieties of failure in a distributed system); and (3) from the application itself (for example, abstracting the data held in cache lines in the evaluation of a cache protocol). Second, new checking techniques are being developed for specifications that are not currently well served by model checking. Software designs often involve operations on complex data structures that are defined implicitly. In this case, there is no apparent model to be checked; rather, it is necessary to generate models from the specification and determine whether they satisfy intended properties. The project is developing a language, NP, that is roughly a subset of the Z specification language, and a checker, Nitpick, that can both simulate the execution of operations specified implicitly and check inductive properties of operations, by generating models of a relational formula. While emphasizing principles that are expected to find broader application, the project is driven by realistic case studies: railway switching, telephone features, and security protocols.

University of Texas – San Antonio; Clinton L Jeffery; *Exploratory Programming Techniques for Program Execution Monitoring*; (CCR-9409082 A001); $10,000; 0 months.

This project seeks to develop techniques that reduce the cost of developing program execution monitors for systems programming languages. These techniques include automated instrumentation of the application, a shared-memory coroutine relationship between the monitor and target program, and dynamic loading and binding of monitors to target programs. The resulting monitoring framework is expected to be used to explore a particular problem in execution monitoring: the control of execution by graphic specification. Execution monitors that use visualization techniques would provide execution controls using the same visual metaphors for graphic input as are employed for output.

Rutgers University; Paul J Krasucki; *Research at Undergraduate Institution (RUI): Knowledge and Concurrency*; (CCR-9528919); $93,938; 24 months. (Joint support with the Programming Languages and Compilers Program and the Theory of Computing Program)

This Research at Undergraduate Institutions (RUI) and Cooperative US-India Research Project investigates distributed systems, using reasoning about knowledge: During the run of the distributed system, processes acquire knowledge about the state of the system through messages. The goals in this area are to (a) characterize changes in the state of knowledge of processes due to communications and (b) relate knowledge structure of the system to the temporal structure on event occurrences in the system. The third research goal of this project is to modify a model, created by economists to reason about rational agents communicating values of certain agreed-upon functions, so that it can be used to design and analyze protocols leading to distributed consensus.

Iowa State University; Gary T Leavens and Don Pigozzi; ***A Theoretical and Practical Basis for Applying Formal Methods to Object-Oriented Programming and C++;*** **(CCR-9503168 A001); $79,999; 12 months.**

This research seeks a theoretical foundation for specifying and verifying object-oriented programs, and a practical foundation for the use of formal methods with the programming language, C++. The project is extending the model theory and proof theory of abstract data types to help characterize when one abstract data type is a behavioral subtype of another. Previous work has given sufficient conditions for behavioral subtyping among abstract data types with immutable objects. These conditions are based on type specifications, as reflected in algebraic models of these specifications. Behavioral subtyping allows modular specification and verification of programs, using static type information without using case analysis for each subtype. Separately, one proves that the subtype relationships satisfy the semantic conditions of behavioral subtyping. An important problem is to find necessary and sufficient conditions for behavioral subtyping for abstract types whose objects have time-varying state (i.e., that are mutable), since these types occur frequently in practice. Needed are ways to prove behavioral subtype relationships from type specifications. This research extends the work on modular specification and verification to languages with mutation and non-determinism. The practical work is aimed at providing the fast-growing community of C++ programmers with a foundation for the use of formal methods with this language. Essential to systematic development of code, program verification, code reuse, and other formal development activities, is a formal specification language. This project advances Larch/C++, and interface specification language tailored to specify C++ modules. Tools, including a type-checker for Larch/C++, are being developed, as are a suite of teaching materials, and tutorial and larger worked examples.

University of Pennsylvania; Insup Lee; ***Hierarchical Specification, Analysis, and Testing of Real-Time Systems;*** **(CCR-9415346 A001); $63,110; 12 months.**

As computers become ubiquitous, they are increasingly used in safety critical environments. Typical safety applications are control systems, monitoring systems and communication systems. Any failure of such computer systems may cause a great financial loss, environmental disaster or even the loss of lives. The potential high cost associated with an incorrect operation of these systems has created a demand for a rigorous framework in which various design alternatives can be formally specified and rigorously analyzed and tested before implementation. One common aspect of all safety critical systems is that they must respond under stringent real-time constraints. That is, their correctness depends not only on how concurrent components interact, but also on the time at which these interactions occur. Thus, it is important to develop a formal framework that supports automatic and computer-aided analysis and testing to effectively cope with increased complexity. The major thrust of this research is to develop analysis and testing techniques for timing properties, and to implement supporting software tools based on real-time temporal logic and state machine models. The proposed tools will provide an environment in which software engineers can generate tests from behavioral specifications and performance requirement constraints and simulate and test to validate the specification.

Tufts University; Karen P Lentz; ***Compression and Interaction Algorithms for Modeling and Simulation Environments;*** **(MIP-9528194); $17,996; 18 months. (Joint support with the Design Automation Program Program)**

This is a research planning grant developing modeling and simulation algorithms to assess system performance and to recreate events so as to interpret cause- effect relationships. The research planning activities focus on investigating the potential for developing accurate, general purpose modeling techniques that include compressing multiple experiments (variations of effects on the model) while allowing internal observation to expose the perturbations that caused the outcome. The grant resources are being used to contact other researchers involved in related efforts; and to investigate issues involved in expanding existing concurrent algorithms to areas beyond digital logic modeling and experimentation.

Massachusetts Institute of Technology; Nancy Lynch and Nir Shavit; ***The Topological Approach to Asynchronous Computability;*** **(CCR-9520298); $76,034; 12 months. (Joint support with the Theory of Computing Program)**

The choice of primitive coordination operations, whether reading and writing a shared memory, message-passing, or more complex operations, has become a crucial design issue. To make such choices, one must analyze the behavior of multiprocess executions, reasoning about the multitude of possible combinations of concurrent operations. This is an unusually delicate and laborious task. The goal of this research is to replace this need to argue about concurrent executions with combinatorial conditions on the geometric representation of problems. This should allow researchers and designers, instead of analyzing behaviors, to apply powerful mathematical tools to recognize when certain protocols are impossible, to evaluate the power of alternative synchronization primitives, or to make explicit the assumptions needed to make a given problem solvable. This project pursues the problem of characterizing and classifying the power of synchronization primitives in different models. Prior research has developed powerful new tools based on classical topology for analyzing fault-tolerant concurrent algorithms and data structures in a variety of models and architectures. This research has shown that a large class of synchronization problems can be associated with a high-dimensional geometric structure called a simplicial complex, and that the computational complexity of solving the problem using a given underlying architecture is captured by certain topological properties of the complex. These properties have been used to give a complete characterization of the synchronization problems solvable in asynchronous read/write memory, and to provide impossibility results and lower bounds for several known open problems in both the asynchronous model and in the synchronous message-passing model. These techniques are being applied toward a comprehensive theory of multiprocessor synchronization.

Stanford University; Zohar Manna; *Tools for the Modular Verification and Refinement of Reactive Systems*; (CCR-9527927); $200,375; 24 months.

This research is developing formal frameworks and methodologies for the verification and development of reactive systems: systems that maintain an ongoing interaction with their environment. The research concentrates on two tasks: modular verification, so that the correctness of an entire system can be derived from the correctness of its smaller components, and refinement, so that a high-level functional specification can be refined repeatedly, where each refinement step is formally proven to be a correct implementation of the former one. The techniques are being integrated into the STeP (Stanford Temporal Prover) system.

Johns Hopkins University; Gerald M Masson and Gregory Sullivan; *The Certification Trail Approach to Software Fault Tolerance and Testability*; (CCR-9319945 A002); $96,801; 12 months.

This research investigates a technique which allows computer programs to be constructed which are more easily and efficiently tested and which can be incorporated into a design methodology for producing efficiently testable and fault tolerant software systems. The technique is to modify a primary applications program so that, with essentially negligible overhead, the primary program not only produces its required output but also generates additional data, referred to as a "certification trail." The certification trail allows the correctness of the primary program's output to be completely and efficiently certified by means of a second, simpler and more efficiently executed program, referred to as a "certifier" which uses the certification trail to also produce the output so that a subsequent comparison operation can take place. The relative simplicity of the certifier has significant advantages beyond that of speed of execution. For example, the certifier program can be formally verified for correctness in cases where doing so for the primary program is much more complicated — perhaps prohibitively so. In general, this project develops a formal, theoretical framework for the certification trail technique, as well as for building upon proof-of-concept performance experimentation by means of implementing the technique and generating performance data for primary programs implementing a variety of well-known algorithms.

University of Washington; David S Notkin; *Foundations for Reasoning About Implicit Invocation Systems*; (CCR-9633462); $75,000; 36 months.

A critical issue for large-scale systems design and evolution is the choice of an architectural style that permits the integration of separately developed components into larger systems. Familiar styles include those based on remote procedure call, shared variables, asynchronous message passing, etc. One increasingly important architectural style for system composition is implicit invocation (II). However, there is no established basis for reasoning about II systems. In particular it is difficult to answer questions like: What will be the effect of announcing a given event? Have enough event bindings been declared? Does a given component announce the sufficient events to permit effective integration? If I add a new component to an existing system, will it break the existing system? The goal of this project is to provide such a basis for reasoning about systems designed using the II architectural style. Two central benefits are expected. The first benefit is a step towards being able to reason fully about systems constructed in this style. The second and probably more important benefit is to replace current ad hoc reasoning approaches used by practitioners who use the II style with a collection of sound ideas that allow better informal reasoning about such systems. This work is carried out in collaboration with David Garlan of Carnegie Mellon University.

North Carolina State University – Raleigh; S Purushothaman-Iyer; *Analysis of Infinite State Systems*; (CCR-9404619 A001); $46,029; 12 months.

The goal of this research is the automated analysis of infinite state concurrent systems for properties like local and global deadlocks, and conformance against specifications in temporal logic. These analyses will support the design of protocol validation systems and the semantic analysis phase of compilers for distributed programming languages. In general, the problems of interest are undecidable in the context of arbitrary infinite state systems. Thus, this project investigates subclasses of infinite state systems where the problems of interest are decidable, and examines the use of flow analysis as a tool for approximative reasoning.

Rutgers University; Fred S Roberts; ***Science and Technology Center for Discrete Mathematics and Theoretical Computer Science*****; (CCR-9119999 A010, A011); $2,139,250; 0 months. (Joint support with the Science and Technology Centers Program)**

Rutgers University, in cooperation with Princeton University, AT&T Bell Laboratories (Bell Labs), and Bell Communications Research (Bellcore) has established a Science and Technology Center for Discrete Mathematics and Theoretical Computer Science. The Director of the Center is Professor Daniel Gorenstein of Rutgers University. The Center in Discrete Mathematics and Theoretical Computer Science strengthens the development of these two overlapping areas by focusing research activity in a continuing series of year-long research programs, workshops and conferences. Approximately one hundred researchers from the four collaborating institutions provide the background coherence and direction for this activity. The core of the research activity is carried out by a vigorous program of visitors (both senior and junior) from other institutions (academic, industrial, government laboratory). Postdoctoral fellows and graduate students are affiliated with the Center for varying lengths of time. Each year, the Center designates specific areas for special attention, inviting leading experts in the designated fields to organize a research program in those areas of concentration. These areas are chosen to take advantage of the expertise of the participating personnel from the collaborating institutions while addressing questions of demonstrable interest to researchers in both the discrete mathematics and theoretical computer science communities. Within these general guidelines, there are several avenues of potential emphasis that involve both serious theoretical inquiry and relevant application. For example, our modern "information society" depends on clear, efficiently designed and reliable telecommunications and communications systems. These are easily described by network theory and can be modeled using linear programming techniques. Transportation problems require similar techniques; and optimization and linear programming, possibly enhanced by parallel computations, are critical in VLSI design and manufacture. Discrete and computational geometry are foundational disciplines for computer graphics and robotics. As the geometries required for the applications become increasingly more complex, research in computational geometry becomes ever more important. Yet there are fundamental questions in discrete and computational geometry that must be addressed before the applications can be realized.

SRI International; John Rushby, Patrick D Lincoln, and Natarajan Shankar; ***An Integrated Framework for Mechanical Verification of Fault-Tolerant Systems*****; (CCR-9509931); $345,147; 36 months. (Joint support with the Operating Systems and Systems Software Program, the Computer Systems Architecture Program, the Numeric, Symbolic & Geometric Computation Program, and the Theory of Computing Program)**

Formal verification is a technique for reducing certain questions about the requirements, designs, or implementations of computer software and hardware to a mathematical form that can examined by methods having the character of calculation. Mechanized formal verification is most effective when applied to the hardest and most critical elements of design. An area where it holds great promise is in the analysis of fault-tolerant systems, where the presence of faults can produce a complexity of behavior that defies other forms of analysis. The objectives of this research are to reduce mechanized formal verification of fault-tolerant distributed algorithms from tours-de-force to routine practice, and to develop a framework for mechanical verification of the architectures of complete fault-tolerant systems composed from such algorithms. The research focuses on developing systematic methods for the formal specification of classes of fault-tolerant algorithms, and mechanized strategies for their exploration and verification. These strategies will build on techniques that combine decision procedures and model checking within an interactive framework. The algorithms considered will include those for distributed consensus, synchronization, diagnosis, and group membership. The work is expected to yield improved understanding of the assumptions, constraints, and fault models underlying the algorithms studied, more refined arguments for their correctness, and coherent arguments for their synthesis into an overall architecture. In addition, the research is expected to yield new results in the construction and application of formal verification systems, and improved methods for the formal representation and analysis of fault-tolerant algorithms.

Rutgers University; Barbara G Ryder and William A Landi; ***Scalable Static Techniques for Exhaustive and Incremental Analyses of C Systems*****; (CCR-9501761 A001); $125,011; 12 months.**

Incremental update algorithms, which only calculate data flow information affected by the program changes, avoiding total recalculation, efficiently provide consistent documentation for a large evolving software system. Previously, incremental techniques for Fortran-like languages and exhaustive analyses for C programs have been developed. When these analyses of C systems produce solutions, they are inexpensive and high quality; however, often the analyses exhaust available resources without yielding a complete solution, especially for larger programs (over 5,000-10,000 lines of code). This research aims at expanding the applicability of exhaustive techniques to handle a wider class of programs and developing incremental techniques for C systems. The work is performed by an academic and industrial team, focusing on the development of new analysis techniques and their testing on research industrial applications. Goals are to: (i) scale the existing prototype to handle programs with 10,000 to 100,000 lines of code, (ii) develop techniques for analyzing modules without analyzing the entire program (i.e., a separate compilation mode of analysis),(iii) develop incremental techniques for interesting interprocedural static analyses for languages with general purpose pointers, and (iv) enhance the theoretical infrastructure for semantic analysis of modern programming languages (i.e., C, C++, and Fortran 90). The research explores techniques for varying the precision of alias information by designing new, flow-insensitive methods, to augment our current flow-sensitive ones. Partitioning the program variables allows for an aliasing solution with varying degrees of precision for different sets of variables. Experiments are conducted varying the value of k for these different sets, using k-limiting to handle dereferencing in recursive structures. A hybrid incremental analysis algorithm is used as a basis for incrementalizing side effect analysis of C programs.

SRI International; Natarajan Shankar; ***An International Workshop on Applied Formal Methods: Hyderabad, India, December 14-17, 1996*****; (CCR-9633644); $10,349; 6 months. (Joint support with the Programming Languages and Compilers Program)**

This grant supported the American participants in a workshop on Applied Formal Methods to be held in Hyderabad, India, in December 1996. The goal of the workshop is to bring together several international experts from the U.S., India, and elsewhere with the purpose of capturing the state of the art in the technology and the state of practice in the applications of formal methods. The workshop scope includes the study of specification languages and logics and verification methods, such as theorem proving, model checking, and safety analysis. Formal methods are increasingly applied to the core algorithms in safety-critical systems and in the verification of hardware circuit designs. A workshop of this nature provides a timely means for developing a common understanding of the capabilities and the potential applications of these technologies. The organizers will edit a technical report providing an up-to-date survey of applied formal methods.

University of Illinois – Chicago; Aravinda P Sistla; ***Formal Methods in Concurrent and Distributed Systems*****; (CCR-9623229); $107,127; 36 months.**

With the widespread use of distributed and concurrent systems and with the increase in the complexity of software for such systems, it becomes important to develop various methods for ensuring the quality of concurrent software systems. This project explores various methods for automated and semi-automated analysis and verification of concurrent and distributed systems. In particular, the research seeks methods based on temporal logic model checking for verification of concurrent systems. Although model checking has been applied fairly successfully for verification of quite a few real-life systems, its applicability to a wider class of practical systems has been hampered by the state explosion problem (i.e. the enormous increase in the size of the state space). In this research, symmetry based techniques are used to overcome the state explosion problem. In particular, symmetry is exploited for verification of liveness properties of real-life concurrent and distributed software systems. The effectiveness of symmetry-based methods for model checking requires efficient solutions to the orbit problem. A solution to the orbit requires checking if two given global states are equivalent under the symmetry. An efficient solution to this problem facilitates the construction of the quotient structure by collapsing each set of equivalent states into a single state. The project investigates various efficient algorithms for the orbit problem for many of the symmetries that occur in practical systems. It is implementing an on-line model checking system that exploits symmetry by simultaneously constructing reduced global state graph (i.e. the quotient structure) and exploring the partially generated quotient structure for existence of certain fair strongly connected components containing an incorrect computation.

State University of New York – Stony Brook; Scott A Smolka and Philip M Lewis; ***Practical Techniques for the Design, Specification Verification, and Implementation of Concurrent Systems*****; (CCR-9505562); $308,036; 36 months. (Joint support with the Programming Languages and Compilers Program)**

Over the past decade, researchers have developed tools that offer automated support for applying formal methods to the design and analysis of concurrent software systems. In principle, these tools allow system designers to utilize formal system verification techniques without requiring them to know all the details of the underlying mathematical theories. In practice, however, the impact of the tools has been limited because of the relatively small scale of the problems they can handle and because of their lack of support for more traditional development methods for software, such as structured design techniques, simulation, and debugging. This research project aims at extending the applicability of existing automated formal techniques for concurrent systems to problems of real-life scale and complexity. The focus is on two broad issues: (1) increasing the efficiency and scalability of existing approaches; and (2) enhancing the usability of existing techniques by integrating them with more traditional software development methods. More specifically, it seeks to develop advanced state-space management techniques for checking system specifications against logical properties expressed in the modal mu-calculus, and techniques for graphically conveying diagnostic information for systems that fail to meet their logical specifications. The impact of the research is being assessed by analyzing several real-life applications, such as sliding-window communications protocols and cache coherency protocols for distributed shared memory. This research builds on the investigators' prior work on the Concurrency Factory, an integrated toolset for specification, simulation, verification, and implementation of concurrent systems.

University of Pittsburgh; Mary L Soffa and Rajiv Gupta; ***Demand Driven Computation of Partial Data Flow and its Application in Software Engineering*****; (CCR-9402226 A001); $80,000; 12 months.**

Although data flow analysis was first developed for use in compilers, its usefulness is now recognized in many software engineering tools. Because of the origin of global data flow analysis, that is, compilers, the computation of data flow for software tools is based on the traditional data flow framework. However, this framework for computing data flow is not producing the efficiency, the preciseness, or the type of data flow problem needed for software tools. These problems are due to the requirements of exhaustive data flow that information is computed for all possible executions, at all points, and all information is computed at each program point. This work investigates the definition of data flow problems when the properties of all executions, all program points and all information at each program point of exhaustive data flow are relaxed, producing partial data flow information computed on demand. The goal is to define a formal framework, in the vein of the traditional framework, where partial data flow problems can be defined and computed. Because the number of problems will grow with partial data flow, and the data flow problems will not be known in many cases until run time, this project investigates the implementation of partial data flow analysis by defining a specification technique to express the demand driven problem and automatically generating the algorithm that computes the information. In order to define partial data flow, information has to be provided that reduces the scope of data flow analysis. This information can either be static information, provided before execution, or dynamic information that is used to either further refine static data flow or to drive its computation. This work first focuses on debugging and testing tools and then explores other applications in software engineering.

Georgia Technology Research Corporation – Georgia Institute of Technology; John T Stasko; ***Understanding Object-Oriented Systems Through Visualization*****; (CCR-9628149); $116,579; 24 months.**

The object-oriented approach to program design has made a significant impact on the way software is developed and maintained. Programmers have adopted the object-oriented approach in order to take advantage of better data abstraction, re-use through inheritance, and polymorphism. However, the very features which make the object-oriented paradigm attractive also make object-oriented programs more difficult to understand. The majority of current object-oriented software development environments and aids provide textual interfaces through which programmers can answer queries about the program state and assess program performance. The hypothesis of this project is that textual tools for understanding object-oriented programs, although absolutely essential, either fail to provide all the information required by programmers about the dynamics and global behavior of their code, or they provide information in a manner that is cumbersome to acquire. Program visualization, through the use of graphics and animation, provides users with a more complete understanding of their code, both statically and dynamically. The goal of this project is to develop a visualization framework with innovative graphical views that assist programmers to understand, analyze, test, debug, and optimize their code. The views illustrate both important details about a program and the global context of how particular pieces of code relate to the entire program. An initial technique being used is called the Information Mural, which presents very large program event traces using a focus-plus-context display. One important aspect of the work with the Information Mural is to identify patterns in the execution of a program.

Measurement and Process

University of Hawaii – Manoa; Philip Johnson; ***Improving Software Quality through Instrumented Formal Technical Review*****; (CCR-9403475 A001); $51,843; 12 months.**

Formal technical review (FTR) encompasses a range of structured, collaborative methods employing technical personnel to improve the quality of a software development artifact. Research shows that FTR, when appropriately practiced, provides significant benefits. However, research also shows that FTR introduces substantial new overhead, and that its cost-effectiveness is easily compromised by a variety of group process obstacles. In addition, current FTR practice is difficult and costly to investigate effectively, and thus many important questions about its nature and comparative effectiveness remain unanswered. This project uses CSRS, a client-server, instrumented, collaborative software review system to investigate issues in research and practice of FTR. Using CSRS, controlled empirical studies provide insight into factors (such as the presence or absence of meetings) influencing the cost-effectiveness of review. The project is also enhancing CSRS to support its public release and use in case-study research in industrial settings. These efforts provide insight into technology transfer issues arising in the shift from manual to computer-supported software review. They also support evaluation of novel forms of FTR made possible by computer support, such as software review involving 20 or more participants, and software review where participants are geographically distributed.

University of Virginia; John C Knight; *The Somerset Laboratory: An Open Facility for Experimentation in Safety-Critical Systems*; (CCR-9525843); $40,000; 12 months.

The Somerset Laboratory is a facility being developed at the University of Virginia that provides a focus for experimental research in safety-critical software engineering. Over the long term, the intent of the laboratory is to establish empirical research with safety-critical systems as a sub-field of software engineering. The laboratory provides the following to researchers world wide, via the Internet: (1) A comprehensive set of artifacts associated with a small number of safety-critical applications that, to the extent possible, describe the applications accurately and comprehensively. (2) A repository of the results of research projects undertaken using the available applications. (3) A vehicle for accurate comparison of software engineering techniques for safety-critical systems. Using the artifacts associated with the safety-critical applications in the laboratory, researchers are able to evaluate, compare, and contrast tools and techniques. The first safety-critical application that the laboratory has chosen to support is a nuclear reactor. The reactor is a low-power, research reactor that employs a conventional electromechanical control system. Though a small reactor, it presents significant challenges to researchers interested in software-based control and safety systems. The artifacts being developed for the laboratory include details of the reactor's physical structure, the safety rules that must be enforced, digitized images of key elements of the physical equipment, sample traces of all the reactor's sensors, a reactor simulator, and complete wiring diagrams. The goal of this planning grant is to obtain an assessment of the likely level of interest in the laboratory by the research community and to develop a detailed plan for the implementation of the laboratory.

University of California – Irvine; David Redmiles; *CAREER: Improving the Design of Interactive Software*; (CCR-9624846); $100,000; 24 months.

Poor design of interactive systems is often attributable to insufficient communication between developers and end users. Usability practices from the field of human-computer interaction are not currently integrated into the theory and practice of software engineering. This research project pursues three objectives: 1) developing a model of software development as a process of on-going communication; 2) providing support for this model through active mechanisms in software tools; and 3) improving the accessibility and acceptance of usability methods by software practitioners. In general, the objectives reflect a process and philosophy of human-centered software development. Active support is being investigated through a mechanism called expectation agents. Expectation agents are software mechanisms that monitor the usage of prototype systems and detect mismatches between actual usage and developers' design assumptions or expectations. This research proceeds along three dimensions: theoretical, technical, and empirical. The research is developing a model of ongoing communication in software development, refining models of the software processes involved in developing interactive systems, and improving understanding of usability methods for workplace settings. In the technical dimension, the research provides new algorithms and representations for capturing and communicating design intent and active mechanisms such as expectation agents to operate over design intent. These mechanisms involve both formal and semi-formal representations. The empirical component includes a critical analysis of the approach as a means of gathering usability information and supporting communication in systems development. This CAREER research project also addresses the software engineering curriculum. First, the work seeks to improve the curricula for educating software engineers in the methods of human-computer interaction. Second, adoption of systematic practice is aided by tools and methods. The research project seeks to meet this need by adapting and, where possible, automating usability methods for software engineering.

Software Architecture, Composition, and Development Environments

Oregon State University; Margaret M Burnett; *NSF Young Investigator (NYI) Award: Visual Programming Languages (Scaling Up)*; (CCR-9457473 A002); $62,500; 12 months.

Visual programming is intuitively very appealing, because it allows the programmer to express relationships among, or transformations to, data simply by sketching them, pointing at them, or demonstrating them – not by translating them into sequences of commands, pointers, and abstract symbols. This removal of many of the concepts traditionally required to program computers, combined with continuous visual feedback, offers the promise of making programming easier and more reliable. Yet, there is a problem that has caused the potential of visual programming to remain largely untapped. Most visual programming languages (VPLs) to date are not easily extended beyond small demonstration programs to adequately handle realistic programming without re-introducing the very concepts they have tried to remove. This problem is called the scaling-up problem. The objective of this research is to improve the scalability of VPLs. The focus is in three directions: (1) developing ways to increase expressive power and code reuse that also maintain the qualities of simplicity and concreteness; (2)

developing a concrete approach to implicit polymorphic types suitable for VPLs; and (3) extending the applicability of VPLs. Research results from investigating these three directions are prototyped in the research system Forms/3, a lazy declarative visual language.

Rice University; Robert S Cartwright and Matthias Felleisen; ***Smart Programming Environments*****; (CCR-9633746); $159,980; 24 months. (Joint support with the Programming Languages and Compilers Program)**

Type safety is an essential feature of advanced programming environments for the rapid development of reliable software, but it is not sufficient. Programmers try to eliminate as many errors as possible by reasoning about the program's behavior. To support this activity, this research seeks to design and construct smart programming environments. A smart programming environment is a system for writing, editing, executing, and analyzing programs. For the latter, the environment includes "soft type checkers", tools for verifying simple program invariants. Most notably, these are the preconditions for each program operation. Two prototypes of soft type checkers for Scheme have been developed, based on different analysis methods. One uses an adaptation of Hindley-Milner type inference (for a non-standard type algebra); the other uses a form of set-based analysis. This research extends the tools, and seeks (1) to develop a new class of soft types that combine the advantages of Hindley-Milner and set-based analysis, (2) to incorporate additional verification technology, (3) to develop alternative interfaces for exploring the verifier's reasoning, and (4) to extend the tools for modular programming. In addition, ways are being explored to transfer the technology to lower-level languages, such as a safe sublanguage of C++ (similar to Java) in the programming environment.

University of Washington; Craig Chambers; ***NSF Young Investigator (NYI) Award: Object-Oriented Language Design and Implementation Research*****; (CCR-9457767 A003); $62,500; 12 months.**

Object-oriented programming promises to make programs easier to write, easier to extend, and easier to reuse. The Cecil project aims to help object-oriented technology deliver on this promise through integrated language design, implementation, and programming environment research. The Cecil language is a purely object-oriented language which is serving as a vehicle for investigating new language features. Current research includes studying more flexible kinds of method lookup based on symmetric multiple dispatching, investigating the effectiveness of advanced static type systems at expressing practical object-oriented programming problems, and developing new module constructs that support application-specific tailoring of objects. Programming environment support for these and other features of object-oriented languages is crucial for effective use of object-oriented programming styles. Advanced language designs and programming styles only impact practice if efficient implementations are available. The Vortex compiler project is investigating whole-program analysis. Interprocedural analysis of object-oriented programs is the foundation of this work, dynamic profile information is treated as an integral component of the compiler, helping to focus the compiler's efforts on profitable optimizations. Incremental recompilation techniques for object-oriented languages help to keep turn-around time short. The Vortex compiler initially targets the Cecil language, with front-ends for other languages such as C++ planned for the future. The Vortex compiler is written in Cecil itself, providing feedback for the language design research and a realistic benchmark for the compiler research.

Michigan State University; Betty Cheng and Mats Per Erik Heimdahl; ***Integrating Formal and Informal Techniques: An Approach to Evolutionary Systems Development*****; (CCR-9633391); $127,162; 36 months.**

A formal method consists of a well-defined specification language and a rigorously defined set of rules that can be used to reason about the specifications expressed in that language. The specification language is used to describe the intended system behavior. In order to facilitate the use of formal methods for software development, particularly for high-assurance systems, the objective of this project is to produce a set of integrated techniques that support the use of formal methods for all phases of software development. Towards this end, several new techniques and tools are being developed that support the application of formal methods for software engineering. These include tools to support correct program derivation from formal specifications, to formally specify requirements analysis and design information, to determine software reuse based on formal specifications, to reverse engineer existing software to obtain formal descriptions, and to statically analyze requirements specifications for completeness and consistency. Formal specifications can be checked for consistency and completeness using automated techniques. However, during the initial phases of a project, it may be difficult to construct formal specifications directly. In contrast, many developers find it more intuitive to create diagrams to model their systems. As a means to bridge the gap between formal and informal approaches to software development, this research investigates the formalization of a commonly used modeling notation, known as the Object Modeling Technique. The notation comprises three complementary diagrams: object (entity-relation) diagrams, state diagrams, and data flow diagrams. These three diagrams, respectively, model the architecture (static), behavior (dynamic), and data flow and services of a given system. The results of the formalization studies will enable the automated generation of formal specifications from diagrams. The next step is to investigate and develop formal semantics for the integration of the three models. Finally, design transformations are studied that can be applied to the diagrams and the corresponding specifications.

University of Illinois – Urbana; Andrew A Chien; *NSF Young Investigator (NYI) Award: Concurrent Object-Oriented Programming Support for Irregular Parallel Applications*; (CCR-9457809 A002); $87,500; 12 months.

For parallel computing to become the vehicle of choice for a broad variety of computing problems, not just regular numeric ones, the expression and exploitation of irregular parallelism must be efficient. This research addresses the major challenge, programmability, by 1) supporting the expression of irregular parallelism with concurrent object-oriented programming languages and 2) supporting the exploitation of irregular parallelism with high performance routing networks and network interfaces. Concurrent object-oriented languages enable programmers to hide concurrency and distribution, allowing incremental parallelization and also confining the complexity arising from irregular parallelism to a small part of the program. However, efficient implementation requires aggressive locality and grain size optimizations in the underlying compiler and runtime system. This project is exploring aggressive type inference and dynamic data structure analysis techniques to enable such program optimizations. Type inference is used to discover program control flow, and dynamic data structure analysis approximates the program storage map. Runtime techniques are also being explored. These techniques are embodied in the publicly available Illinois Concert System. This system is being used to develop a number of parallel application programs.

Indiana University – Bloomington; Daniel P Friedman; *Dynamically Extensible Computational Reflection*; (CCR-9633109); $124,886; 24 months. (Joint support with the ARPA-EVOLUTIONARY DESIGN OF COMPLEX SYSTEMS Program)

Reflection has been explored in an interpreted framework and has tended to assume the availability of particular pieces of the state of a program's interpretation, even the current source code expression. Instead of programs, this research emphasizes computations as the primary elements of reflection. The research has two facets. The first is to formalize and continue the theoretical development of a system for computational reflection that is free from interpreter-based constraints. This system supports the introduction of multiple user-defined reflective properties that can be designed independently and combined in arbitrary fashion. Furthermore, new reflective properties can be introduced after the programs in which they are to be used have already been compiled. The second facet is to explore the many uses of such a system. In addition to the simple, but powerful, uses already explored, the system is expected to be useful for rapid prototyping of language features, collaborative software development, and static analysis of programs for intense optimization. Prior research in programming with first-class contexts has investigated the meaning and use of static meta-level contextual information. The word static here is intended to refer to that information available in the text of a program, without running it. The current research extends the previous work by exploring the meaning and use of dynamic meta-level contextual information, which is that information available in the meta-level of a running process. As in the previous work, this project emphasizes modularity, allowing many different kinds of meta-level knowledge to be utilized simultaneously.

Carnegie Mellon University; David Garlan; *NSF Young Investigator (NYI) Award: Towards An Engineering Basis for Software Architecture*; (CCR-9357792 A003); $62,500; 12 months.

An important goal in software engineering is to describe complex software systems at an architectural level of abstraction. While architectural design has always been a major factor in determining the success of a software system, architectural descriptions are typically understood only in informal, idiomatic ways. Because of this informality architectural designs are typically not well-integrated into the software development environments, but appear (if at all) as corroborative high-level documentation for a system. This research places architectural design on a more solid footing in two complementary ways. First, it provides a formal basis for describing and reasoning about architectural designs. This leads to rules for checking whether architectural designs are well-formed in a way that is analogous to type checking for existing implementation languages. Second, it provides tools for developing style-specific architectural design environments. These tools provide cost-effective ways to generate new environments tailored to the architectural design of specific application families. This grant includes support under the Research Experiences for Undergraduates (REU) Program.

University of Maryland – College Park; Richard R Gerber; *NSF Young Investigator (NYI) Award: Unifying Real-Time Design and Implementation: A Tool-Based Approach*; (CCR-9357850 A003); $25,000; 12 months.

The goal of this research is to develop a tool-based framework for the specification, analysis and implementation of real-time systems. The research includes three principal components: (1) a real-time specification language, (2) analysis techniques for showing that a specification is correct, and (3) a programming language in which the system is ultimately implemented. The specification component centers on the CSR paradigm, in which a real-time system's high-level control structure is entered as a set of Ada-like tasks, which are then compiled into a state-transition model. For the analysis component, an automated toolset based on the technique of finite-state abstraction is developed. The tools include a probabilistic reachability analyzer, a simulator and a model-checker. The language component is based on TCEL, a real-time programming language with CSR-style timing constructs. This feature allows the programmer to embed the real-time requirements within the program itself. A TCEL compiler is being developed, which rearranges the code to meet the inferred timing constraints.

Cornell University; David Gries and R. Tim Teitelbaum; *Software Engineering with a Transformational System*; (CCR-9503319 A001); $90,000; 12 months.

The goal of the Polya project is to increase productivity in the development and maintenance of programs, by placing more emphasis on the user interface within the programming language itself. This research seeks new ways to write programs, to make programs efficient, and to reuse program parts. It is concerned directly with software engineering and rapid prototyping. The programming language Polya incorporates new mechanisms that allow a group of people to define and use notations of their problem domain, so that programs can more closely resemble the algorithms being used. Secondly, Polya has mechanisms that make it easier to describe how variables of high-level types should be implemented, requiring fewer modifications to the algorithm than is the case with more traditional languages. The transform, a new kind of program part, can be reused with relative ease. A prototype implementation of Polya uses the Synthesizer Generator as the platform. This research focuses on three areas: (1) research aimed at a significant extension of the pattern matching capabilities of the Polya transformational approach and exploration of new ideas on reverse engineering; (2) experimentation with Polya, in order to develop a library of useful program parts; (3) practical and theoretical improvements to Polya, its implementation, and related tools such as a source-level debugger.

University of California – San Diego; William G Griswold; *Practical Program Restructuring for Software Evolution*; (CCR-9508745 A001); $72,072; 12 months.

This project is developing and evaluating new techniques for tool-assisted restructuring of large systems. The excessive cost of software evolution is due in large part to inappropriate structure of a system, relative to the changes taking place. Tool-assisted program restructuring is studied for changing a system's structure at lower cost and risk than making changes by hand. In this approach, a tool user selects transformations to apply to chosen program components. The tool checks to make sure that each transformation will preserve the functional behavior of the program. If so, it performs the transformation to the chosen component and updates any related components to preserve consistency. Tool-assisted restructuring has not been demonstrated to be practical in industrial-scale applications. For example, exhaustive interprocedural data flow analysis is required by some tools, but is costly to perform on large programs. Previous restructuring research has not addressed common complicating constructs such as pointer arithmetic, function pointers, comments, textual macros, and conditional compilation. To make the restructuring of large programs practical, this project is developing and adapting techniques to address these problems. Demand-driven data flow is being investigated to overcome many performance problems of exhaustive data flow analysis. Query prioritization and caching of replies are studied to minimize burdensome and potentially incorrect user queries. Testing and semantic differencing may also provide checks of correctness after restructuring. A case study is intended to evaluate the techniques developed and the benefits of large-scale restructuring.

University of Minnesota – Twin Cities; Mats Per Erik Heimdahl; *Integrating Informal and Formal Techniques: An Evolutionary Approach to Systems Development*; (CCR-9615088); $47,374; 36 months.

A formal method consists of a well-defined specification language and a rigorously defined set of rules that can be used to reason about the specifications expressed in that language. The specification language is used to describe the intended system behavior. In order to facilitate the use of formal methods for software development, particularly for high-assurance systems, the objective of this project is to produce a set of integrated techniques that support the use of formal methods for all phases of software development. Towards this end, several new techniques and tools are being developed that support the application of formal methods for software engineering. These include tools to support correct program derivation from formal specifications, to formally specify requirements analysis and design information, to determine software reuse based on formal specifications, to reverse engineer existing software to obtain formal descriptions, and to statically analyze requirements specifications for completeness and consistency. Formal specifications can be checked for consistency and completeness using automated techniques. However, during the initial phases of a project, it may be difficult to construct formal specifications directly. In contrast, many developers find it more intuitive to create diagrams to model their systems. As a means to bridge the gap between formal and informal approaches to software development, this research investigates the formalization of a commonly used modeling notation, known as the Object Modeling Technique. The notation comprises three complementary diagrams: object (entity-relation) diagrams, state diagrams, and data flow diagrams. These three diagrams, respectively, model the architecture (static), behavior (dynamic), and data flow and services of a given system. The results of the formalization studies will enable the automated generation of formal specifications from diagrams. The next step is to investigate and develop formal semantics for the integration of the three models. Finally, design transformations are studied that can be applied to the diagrams and the corresponding specifications.

Yale University; Paul Hudak and Zhong Shao; ***Building Evolutionary Software Through Modular Executable Specifications and Incremental Derivations*****; (CCR-9633390); $150,000; 12 months. (Joint support with the ARPA-EVOLUTIONARY DESIGN OF COMPLEX SYSTEMS Program)**

The first goal of this investigation is a software development strategy in which ultra-modular executable specifications are the starting point of systems design. By ultra-modular is meant a set of specific technologies that allow one to anticipate and realize change, including the use of domain-specific embedded languages to capture high-level, domain-specific semantics, and modular monadic interpreters on which to realize the semantics and associated software tools. Using these techniques a level of modularity in software is obtained that has not been achieved before: namely, modularity at the meta-level of abstract semantics. Second, the research seeks a notion of type-based modularity to capture software architecture. There is a misconception that type systems are contrary to the goals of "dynamic languages," which in turn are viewed as good candidates for building evolvable systems. This research contests this claim. In fact it is expected that modern typed languages such as ML and Haskell can offer even greater support for constructing reliable and efficient evolutionary software. In particular, the design and use of a higher-order module language is studied to capture systems architecture. Finally, a methodology is investigated that supports the notion of incremental derivations. In this methodology, some executable specifications are refined for improved performance, whether by formal methods (program transformation, partial evaluation, etc.) or by informal methods (such as rewriting inner loops in a lower-level language). The key point is that the derivations themselves can be modularized, thus making them suitable candidates for smooth evolution. When formal methods are employed, this also yields a technique for constructing modular proofs of program correctness; yet another innovative aspect of this approach. These three ideas form the cornerstone of the research effort. To evaluate them, the investigation includes the design of a prototype extensible operating system.

University of Massachusetts – Amherst; Barbara Lerner; ***Evolution of Persistent Abstract Data Types*****; (CCR-9504170 A001); $66,387; 12 months.**

Persistent programming languages provide a simple mechanism to make data persistent without the impedance mismatch that occurs when storing data in a database. As a result, persistent programming languages greatly facilitate the development of software that uses persistent data. A serious problem exists when maintaining programs written in persistent programming languages. The storage format for persistent data is derived from the structural representation of the data's type as defined in the program. When the representation is changed, the software is no longer able to access existing persistent data. This research addresses the problems of type evolution. It is developing algorithms to recognize complex type changes, to assist the software engineer in assuring that persistent data conform to modified semantics of the software, and to assist in identifying and updating code affected by structural changes to a type. The approach relies on persistent types defined as abstract data types. Type changes are identified by comparing the interfaces to the types, the structural representations of the types, and constraints attached to the types. The algorithms are being incorporated into a tool that provides computer-aided type evolution support. The primary goal of the tool is to manage the complexity of type evolution and provide assurance that evolution issues are addressed by the software engineer while not restricting the ways in which the software engineer modifies type definitions. A secondary goal is to automate data evolution for a collection of well-understood type changes. The combination of flexibility, automation, and concern with semantic issues is expected to provide advances over existing type evolution support systems. Experimentation is using histories collected from the development of real systems to evaluate the effectiveness of the algorithms.

Northeastern University; Karl J Lieberherr; ***Engineering Adaptive Software*****; (CCR-9402486 A002); $48,500; 0 months. (Joint support with the ARPA-ENG ADAPTIVE SOFTWARE Program)**

This project studies the engineering of adaptive software that adapts without modification to a wide range of requirement changes and that can be reused in unplanned ways. Adaptive programming is at a higher level of abstraction than object-oriented programming. The key idea is that each adaptive program specifies a family of object-oriented programs. Adaptive software subsumes previously developed concepts of class dictionaries and propagation patterns. Adaptive software has two parts: a data part and a functional part. These two parts are loosely coupled through constraints, allowing the same functional part to work with different data parts and the same data part to be matched with different functional parts. The loose coupling yields adaptiveness and allows high-level structuring and reasoning about software. To further understand and facilitate the engineering of adaptive software, this project studies the following topics: (1) Type systems for adaptive software, enabling safety checking at the level of adaptive programs; (2) Reengineering algorithms which abstract higher-level structures from lower-level structures, extending our previous work on abstracting class dictionary graphs from objects and propagation patterns from call graphs; (3) Analysis of adaptive software, addressing resource consumption (such as time and space) at the level of adaptive software; (4) Visual programming, as motivated by the use of graph specifications in adaptive software; (5) Experiments with building an adaptive software library and associated tools.

Stanford University; John McCarthy and Carolyn Talcott; ***Semantic Foundations for Composition and Interoperation of Open Systems*****; (CCR-9633419); $89,431; 36 months.**

The objective of this research is to develop a semantic foundation for semantic composition and interoperation of heterogeneous components of open systems. Composition and interoperation occur at many levels (requirements, specification, code, executables) and along multiple dimensions (functionality, reliability, security, availability). Relations among the various levels and dimensions must be preserved. Existing work on composition and interoperation has largely focused on syntactic structure. It is essential to develop a semantic basis in order to provide for safe, secure, and meaningful combinations. Semantic structuring of complex systems should provide a firm basis for determining what parts are interchangeable under given conditions, and to predict effects of changes: of components on the whole, and of system requirements on requirements for the parts. Metalogical methods from the theory of general logics, and associated techniques of mappings in and across formalisms, are used to achieve formalism-independent semantics for system composition. This supports the multi-dimensionality of a system's levels of description and formalization. Techniques from the theory of Open Mechanized Reasoning Systems (OMRS) as well as work on the formal semantics of actor systems are used to study interoperation aspects. Reflective techniques, particularly reflective logics and multi-model systems are employed extensively. This research is expected to lead to a new technology for composition, interoperation and dynamic evolution of software systems. The research is based on two key technical ideas: module calculi (for integration) and open multi-model systems (for interoperation). The module calculus is formalism-independent, and is intended to have powerful new operations such as generalization. Open multi-model systems formalize the multidimensional aspects of independent interoperating parts. This work is carried out in collaboration with Jose Meseguer of SRI.

SRI International; Jose Meseguer; ***Semantic Foundations for Composition and Interoperation of Open Systems*****; (CCR-9633363); $121,059; 24 months. (Joint support with the ARPA-EVOLUTIONARY DESIGN OF COMPLEX SYSTEMS Program)**

The objective of this research is to develop a semantic foundation for semantic composition and interoperation of heterogeneous components of open systems. Composition and interoperation occur at many levels (requirements, specification, code, executables) and along multiple dimensions (functionality, reliability, security, availability). Relations among the various levels and dimensions must be preserved. Existing work on composition and interoperation has largely focused on syntactic structure. It is essential to develop a semantic basis in order to provide for safe, secure, and meaningful combinations. Semantic structuring of complex systems should provide a firm basis for determining what parts are interchangeable under given conditions, and to predict effects of changes: of components on the whole, and of system requirements on requirements for the parts. Metalogical methods from the theory of general logics, and associated techniques of mappings in and across formalisms, are used to achieve formalism-independent semantics for system composition. This supports the multi-dimensionality of a system's levels of description and formalization. Techniques from the theory of Open Mechanized Reasoning Systems (OMRS) as well as work on the formal semantics of actor systems are used to study interoperation aspects. Reflective techniques, particularly reflective logics and multi-model systems are employed extensively. This research is expected to lead to a new technology for composition, interoperation and dynamic evolution of software systems. The research is based on two key technical ideas: module calculi (for integration) and open multi-model systems (for interoperation). The module calculus is formalism-independent, and is intended to have powerful new operations such as generalization. Open multi-model systems formalize the multidimensional aspects of independent interoperating parts. This work is carried out in collaboration with John McCarthy and Carolyn Talcott of Stanford University.

Stanford University; John C. Mitchell; ***Software Engineering and Programming Languages Workshop; June 12-13, 1996, Boston, MA*****; (CCR-9612496); $11,950; 6 months.**

This award provided travel support for a workshop on the topic of synergy between software engineering and programming language research. The workshop was held June 12-13 in Boston, MA, in juxtaposition with a Computing Research Association and ACM symposium on strategic directions in computing, June 14-15, in the same location. The software engineering and programming language workshop also received partial funding from ARO.

New York University; Robert A Paige; *Improving Productivity of Algorithm and System Implementation in Scaled up Applications*; (CCR-9616993); $50,000; 24 months. (Joint support with the Programming Languages and Compilers Program and the Operating Systems and Systems Software Program)

Improving productivity and reliability of labor-intensive high-performance software is an important scientific problem with obvious economic ramifications. This research is testing the feasibility of a radically new compilation and meta-compilation methodology to improve productivity and reliability for high-performance algorithmic software in the C programming language. The research combines partial evaluation for deeply nested SETL data structures, data structure and algorithm selection techniques, and transformational techniques from relational query optimization. This is leading to a statically typed variant of SETL from which it is expected large-scale C programs can be generated at a high translation rate.

Brown University; Steven P Reiss; *LANDSCAPES: Visions of Programming Environments*; (CCR-9422625 A001); $75,000; 12 months.

This project involves the development of practical programming tools in an integrated and open framework. Tools that support modern programming must deal with large systems, multiple programmers, multiple languages, a mixture of existing and new code, and multiple processes and threads. The work combines a variety of tools in an open, extensible, and powerful setting using control integration and a simple form of data integration based on fragments. Fragments are semantically significant portions of a program artifact: definitions or scopes in a source file, sections of documentation file, rules in a configuration description, etc. The approach to integration identifies such fragments and keeps references to them in a simple and small central database. This information is augmented with the data necessary to identify fragments using database queries, essentially usedef chains and small amounts of tool-specific information. In addition to fragment and control integration, the environment provides common facilities for defining and managing the files and contexts that make up a project, and a common editor for all software artifacts. The project manager serves as a front end for managing version control, configuration management, and multiple programmers. The common editor provides high quality typography along with support for hypertext links and graphics. On top of this environmental framework, this project investigates the development and utility of a variety of programming tools. The research extends visualization tools using a 3D visualization package developed under a previous NSF grant. A single front end would allow the programmer to specify both what information (from a variety of sources) should be visualized and how it should be presented for each particular application. Experiments measure the effectiveness of visualizations for understanding both the static structure and the dynamic behavior of complex systems. A third set of tools involves editing. The project extends an existing editor to allow programmers to view the system as a single electronic document with dynamic and static links connecting the various software artifacts. These links would be defined either by built-in queries (e.g. go to the definition of a name), by user-defined queries, or by static links.

University of Wisconsin - Madison; Thomas Reps; *Semantics-Based Program Manipulation*; (CCR-9625667); $160,386; 24 months.

This research concerns the design and implementation of tools for interactive computer programming. The research focuses on the creation of tools that furnish language-specific, semantics-based, program manipulation operations. The project investigates both program slicing and partial evaluation. In particular, the research studies (1) whether techniques developed for one of the operations can be applied to the other, and (2) whether the two operations can be combined to support more powerful techniques for manipulating programs. The research focuses on two areas. The first is operations for semantics-based program manipulation. Examples include program specialization, symbolic composition, and program generalization. The second is static analysis of programs. Research is conducted on static analysis methods needed in carrying out the above operations. Pointer analysis is of particular importance. In both areas the scope of the work includes development of theoretical foundations, algorithm design, and system implementation. One testbed for the ideas is the prototype C slicer, which currently is capable of slicing large (22,000-line) programs.

Carnegie Mellon University; William L Scherlis; *Fluid Architecture*; (CCR-9504339 A001); $75,000; 12 months.

This project explores the design of systematic techniques to adapt the internal interface and component structure of programs of moderate size or larger. Programming tools based on these techniques could have the effect of increasing the extent to which the internal interface structure of systems can be made more fluid, that is, evolvable and adaptable. One potential application is to reorganizing abstraction boundaries in programs: It is clear in practice that software structural architectures are rarely optimal, but the costs and risks of adapting them are high. Another potential application is to the tailoring of general purpose frameworks and components to specific constrained uses, improving performance. Four questions are addressed: (1) What semantics-based program manipulation techniques are needed to support manipulation of internal software interfaces represented, for example, as encapsulation boundaries in programs? (2) How can soundness and other properties of the program manipulation techniques be established? (3) What are the design considerations for the development of tools that can operate on programs of increasing complexity and size? (4) How can the utility of the techniques be assessed for development and re-engineering of internal software interfaces and associated components?

Kansas State University; David A Schmidt, Matthew Dwyer, Allen Stoughton, and Brian Howard; ***Logical Support for High-Assurance Software Evolution*****; (CCR-9633388); $154,369; 24 months. (Joint support with the ARPA-EVOLUTIONARY DESIGN OF COMPLEX SYSTEMS Program)**

A central problem in the management of large software projects and the maintenance of libraries of reusable software is ensuring the consistency of inter-module connections in the face of continuing system evolution. As bugs are isolated and fixed, capabilities are added or enhanced in response to changes in the user requirements, and support software and hardware is upgraded or replaced, there are myriad opportunities for invalidation of prior assumptions used in code development. In this project, techniques from logic, type theory, and abstract interpretation are employed in the design and implementation of tools to help automate the process of ensuring consistency. A partial-evaluation-driven abstract-interpretation algorithm is used to analyze individual modules for consistency with respect to local specifications, and a linking and configuration manager will be used to check for correspondence between module specifications. Using abstract interpretation, proof generation, and realizability techniques, the manager synthesizes common additional code that may be needed to provide trustworthy linkages. Two aspects of this approach are essential for application to large-scale, evolving systems: (i)link updating must be a local process, so that small changes do not incur large penalties, and (ii)linkage information must be available to the programmer in a user-readable format. Recent experience with building efficient, usable logic tools is applied in the project towards satisfying these requirements.

Florida International University; Geoffrey S Smith; ***Type Systems for Secure Remote Evaluation*****; (CCR-9612176); $118,072; 36 months.**

A remote evaluation system is characterized by a client that sends a program to a compute server for execution on its behalf. The server returns to the client the results of executing the program. This is in contrast to RPC and CGI scripts which are examples of server-side programs, not remote evaluation. Remote evaluation appears to be gaining currency as evidenced by new programming languages emerging to support it. Among them are Telescript, Tcl, Oblique and Java. The download-and-execute paradigm supported by Java and Java-compatible Web browsers is an example of remote evaluation; in this case, however, a Web server sends a Java-enhanced HTML file to a browser which the browser then executes. Server security is a major problem with remote evaluation. This greatly weakens the remote evaluation paradigm. This research investigates the role of programming language design and type systems in ensuring the security of servers in remote evaluation systems. Specifically, the long-term objective is to develop static analyses, within the context of type systems, to ensure that remotely evaluated programs do not violate a server's privacy or integrity. Type soundness theorems are being developed and proved with respect to a formal semantics for a core imperative language. They precisely characterize the security properties of well-typed programs written in the language. The impact of features such as concurrency and exceptions in a programming language on security properties is also investigated. The research is in collaboration with Dennis Volpano of the Naval Postgraduate School.

Massachusetts Institute of Technology; Gerald J Sussman and Olin Shivers; ***Programming-Language Structures for Representing and Optimizing Operating-System Resources*****; (CCR-9633438); $100,000; 24 months.**

This research project seeks to demonstrate that fundamental operating system services and resources can and should be represented directly in the natural linguistic structures of advanced programming languages. A component of this effort is to develop aggressive optimization strategies making this modeling practical and efficient. The central question underlying the research is: how can the benefits of advanced programming languages be brought to bear upon current systems-programming practice? Modern computer applications are most commonly written in C, a language well-known for its lack of support for writing modular, robust programs. As a result, modern software is fragile, error-prone, and difficult to build, maintain or reuse. Modern programming languages, such as the ML family of languages, provide much more support to the programmer for writing large, robust programs. Unfortunately, these languages are not yet practical tools for implementing real-world systems software. This research enables new optimizations and makes available new models to the programmer, allowing systems programs to be written that exploit the expressive power of advanced programming languages. A service kernel provides access to the underlying machine resources in a way that is tightly integrated, both in architecture and in implementation, with advanced languages. The research issue is to investigate the architecture of an operating system that is symbiotic, not with C, but with a modern programming language, and to develop technologies for exploiting the tight integration of the OS and the compiler to provide efficient, high-level services. The kernel is being developed in ML, on two hardware platforms: a bare Intel PC system, and a wireless, mobile "personal digital assistant" with limited on-board resources.

Naval Postgraduate School; Dennis M Volpano; *Type Systems for Secure Remote Evaluation*; (CCR-9612345); $115,093; 36 months.

A remote evaluation system is characterized by a client that sends a program to a compute server for execution on its behalf. The server returns to the client the results of executing the program. This is in contrast to RPC and CGI scripts which are examples of server-side programs, not remote evaluation. Remote evaluation appears to be gaining currency as evidenced by new programming languages emerging to support it. Among them are Telescript, Tcl, Oblique and Java. The download-and-execute paradigm supported by Java and Java-compatible Web browsers is an example of remote evaluation; in this case, however, a Web server sends a Java-enhanced HTML file to a browser which the browser then executes. Server security is a major problem with remote evaluation. This greatly weakens the remote evaluation paradigm. This research investigates the role of programming language design and type systems in ensuring the security of servers in remote evaluation systems. Specifically, the long-term objective is to develop static analyses, within the context of type systems, to ensure that remotely evaluated programs do not violate a server's privacy or integrity. Type soundness theorems are being developed and proved with respect to a formal semantics for a core imperative language. They precisely characterize the security properties of well-typed programs written in the language. The impact of features such as concurrency and exceptions in a programming language on security properties is also investigated. The research is in collaboration with Geoffrey Smith of Florida International University.

SRI International; Richard J Waldinger; *Deductive Composition of Software from Component Libraries*; (CCR-9421098 A001); $68,296; 12 months.

This award supports research in deductive methods, applied to automate the synthesis of software built from component libraries. Software is constructed from relatively large-scale components, rather than from the primitive instructions of a programming language. Theorem-proving techniques are invoked to identify the relevant components and to compose them to solve the problem at hand. The goal is to put computing power in the hands of a broader segment of the population and to increase the productivity and accuracy of the expert practitioner. The software is constructed to meet specifications expressed in a logical or graphical notation. Specifications are elicited from the user via a dialog with a graphical user interface. The specification is phrased as a mathematical theorem, which is proved by an automatic deduction system. The proof is conducted in an application domain theory, which describes the components in the software library, the constructs of the specification language, and the background knowledge necessary to glue the components together to meet the specification. The theorem is proved by the automatic deduction system Snark, which has been implemented for applications in software engineering and artificial intelligence. A program that meets the original specification is extracted from the proof and translated into the desired target language. Software is produced with high assurance of correctness—it is verified by the proof from which it is extracted. This research is expected to make subroutine libraries, component libraries, and software reuse libraries accessible to users who are largely ignorant of the contents of the libraries, their conventions and terminology, or even the target programming language.

George Mason University; Elizabeth L White; *CAREER: General Support for Dynamic Reconfiguration*; (CCR-9625202); $99,887; 24 months.

In recent years, there has been a great deal of attention in the software engineering community on the development of techniques and tools that provide support for the design and implementation of applications to be distributed across networks of workstations. Some of this focus has been on support for initial development and configuration of applications while other research has focused on techniques to support these types of applications at run-time. One novel aspect of workstation environments is that the number of processors available for use tends to grow and shrink over time. This aspect has caused increased interest in dynamic reconfiguration techniques that allow applications to adapt to their changing environment. Much of the current research in dynamic reconfiguration has focused on reconfiguration environments and heterogeneous process migration. This research is developing techniques for supporting dynamic structural and implementation changes. The work has three parts: 1) studying general approaches to implementing dynamic structural and implementation changes to applications, including those changes that require process participation, 2) determining what aspects of this task can be automated, and 3) building tools to support dynamic structural and implementation changes. The most promising approach is one that enables mapping the old global state of an application to a new global state based on the structural changes to be made. This CAREER project includes teaching initiatives geared towards introducing undergraduate students to issues in the design and implementation of distributed applications. The goal is to improve training in an area where many students get inadequate preparation for industry positions that require knowledge in this domain.

University of Colorado – Boulder; Alexander L Wolf; *U.S.-Argentina Cooperative Research on Software Architecture Formal Specification*; (INT-9514202); $6,756; 24 months. (Joint support with the Americas Program)

This U.S.-Argentina collaborative research project will continue the development of a framework for the formal specification and analysis of the behavioral properties of software architectures. Together with Daniel Yankelevich of the Universidad de Buenos Aires in Argentina, the PI at the University of Colorado Boulder will extend his work on a formal specification method based on the Chemical Abstract Machine (CHAM) model. CHAM provides operational semantics which may be more easily understood by a wide audience than competing mathematical formalisms. The collaboration established in this project connects the PI with an existing effort between Argentina and Italy on specification and verification of concurrent systems. It also combines the systems orientation of the U.S. group with the theoretical orientation of the Argentines to understand the potential of the CHAM model for software architecture.

SCIENCE AND TECHNOLOGY CENTERS

Rice University; Ken W Kennedy; ***Science and Technology Center for Research on Parallel Computation*****; (CCR-9120008 A008, A009); $5,384,666; 0 months. (Joint support with the EHR Activities for Women & Girls in Sci,Eng & Math Program)**

Rice University, in collaboration with the California Institute of Technology, Syracuse University, the University of Tennessee, Argonne National Laboratory and Los Alamos National Laboratory, has proposed renewal of the Center for Research on Parallel Computation. The Director of the Center is Ken Kennedy. The goal of the Center is to make parallel computing systems truly usable. As high performance computing systems become progressively more parallel in their logical organization, ease of use comparable to conventional supercomputers becomes essential to high performance computing. The objectives of the Center cover a broad range of issues. The participants will develop theory and algorithms that will be needed to parallelize important scientific applications. They will imbed these algorithms in scalable libraries that can be implemented on different parallel computer architectures. They will demonstrate the utility of these software systems, algorithms and libraries in real science and engineering applications. They will also transfer the knowledge and other byproducts of this research to the science and engineering community and to industry.

Rutgers University; Fred S Roberts; ***Science and Technology Center for Discrete Mathematics and Theoretical Computer Science*****; (CCR-9119999 A010, A011); $2,139,250; 0 months. (Joint support with the Software Engineering Program)**

Rutgers University, in cooperation with Princeton University, AT&T Bell Laboratories (Bell Labs), and Bell Communications Research (Bellcore) has established a Science and Technology Center for Discrete Mathematics and Theoretical Computer Science. The Director of the Center is Professor Daniel Gorenstein of Rutgers University. The Center in Discrete Mathematics and Theoretical Computer Science strengthens the development of these two overlapping areas by focussing research activity in a continuing series of year-long research programs, workshops and conferences. Approximately on hundred researchers from the four collaborating conferences. Approximately one hundred researchers from the four collaborating institutions provide the background coherence and direction for this activity. The core of the research activity is carried out by a vigorous program of visitors (both senior and junior) from other institutions (academic, industrial, government laboratory). Postdoctoral fellows and graduate students are affiliated with the Center for varying lengths of time. Each year, the Center designates specific areas for special attention, inviting leading experts in the designated fields to organize a research program in those areas of concentration. These areas are chosen to take advantage of the expertise of the participating personnel from the collaborating institutions while addressing questions of demonstrable interest to researchers in both the discrete mathematics and theoretical computer science communities. Within these general guidelines, there are several avenues of potential emphasis that involve both serious theoretical inquiry and relevant application. For example, our modern "information society" depends on clear, efficiently designed and reliable telecommunications and communications systems. These are easily described by network theory and can be modeled using linear programming techniques. Transportation problems require similar techniques; and optimization and linear programming, possibly enhanced by parallel computations, are critical in VLSI design and manufacture. Discrete and computational geometry are foundational disciplines for computer graphics and robotics. As the geometries required for the applications become increasingly more complex, research in computational geometry becomes ever more important. Yet there are fundamental questions in discrete and computational geometry that must be addressed before the applications can be realized.

Index of Grantee Institutions

Index of Principal Investigators

Index of Presidential and NSF Young Investigators

Index of Research Initiation Awardees

Index of CAREER Awardees

Index of Research Planning Grant Awardees

Index of Research at Undergraduate Institution (RUI) Awardees