Final Report to

U.S. Department of Energy

A Broad Research Project on the Sciences of Complexity

Award DE-FG03-94ER61951
9/15/94 to 11/15/99

Santa Fe Institute
1399 Hyde Park Road
Santa Fe, New Mexico 87501
www.santafe.edu

No Unexpended Funds

DISCLAIMER

A Broad Research Program in the Sciences of Complexity

Previous DOE support for "A Broad Research Program in the Sciences of Complexity" permitted the Santa Fe Institute to initiate new collaborative research within its Integrative Core activities as well as to host visitors to participate in research on specific topics that serve as motivation and testing-ground for the study of general principles of complex systems. The critical aspect of this support is its effectiveness in seeding new areas of research. Indeed, this Integrative Core has been the birthplace of dozens of projects that later became more specifically focused and then won direct grant support independent of the core grants. But at early stages most of this multidisciplinary research was unable to win grant support as individual projects--both because it did not match well with existing grant program guidelines, and because the amount of funding needed was often too modest to justify a formal proposal to an agency. In fact, one of the attributes of core support has been that it permitted SFI to encourage high-risk activities because the cost was quite low. What is significant is how many of those initial efforts have been productive in the SFI environment.

Many of SFI's current research foci began with a short visit from a researcher new to the SFI community, or as small working groups that brought together carefully selected experts from a variety of fields. As mentioned above, many of the ensuing research projects are now being supported by other funding agencies or private foundations. Some of these successes are described here for illustrative purposes.

- DOE support for the Integrative Core research at SFI led to a program "Foundations of Complex Adaptive Systems" supported by DARPA (for a total of $2.2M over the period 1995--98) focusing on distributed computing, self-repair mechanisms, biodefense, and design of autonomous robots. Another offshoot of the program is the DARPA-sponsored project led by Stephanie Forrest (computer science, UNM, SFI External Faculty, currently interim VP for Academic Affairs) on the design of a computer virus defense system based on principles of the immune system.

- The core program in adaptive computation led to a study on "Evolving Cellular Automata" by SFI Research Professor Jim Crutchfield and Melanie Mitchell (LANL, SFI External Faculty member). The work has been subsequently supported by the NSF CISE Directorate, and the NSF Directorate of Mathematical, Information and Computational Sciences. The effort now involves two SFI postdoctoral fellows (including a NSF postdoc fellow), and Crutchfield and Mitchell have served as mentors for five NSF REU summer interns over the past two years.

- The adaptive computation effort has led to the active involvement in SFI research of many researchers in the areas of cognitive psychology, neurobiology, parallel computing, and robotics. The participation of Mitchel Resnick (MIT Media Lab) led to the initiation in summer 1998 of a new program in complex systems for Santa Fe secondary school students.

- The core research on "Computation, Dynamics, and Inference," coordinated by Jim Crutchfield, is now being funded by Intel Corporation

- The research collaboration between Jim Brown (ecology, UNM, SFI External Faculty) and Geoffrey West (physics, LANL) on allometric laws in plants and mammals was initiated by SFI, and supported by DOE core funds. Brown's graduate student Brian Enquist was awarded a NSF postdoctoral fellowship, and is now at NCEAS in Santa Barbara. West has recently been awarded a grant from the NSF Physics Division for allometric research; that grant is being subcontracted through Los Alamos National Laboratory. Two workshops "Scaling in Biology and Ecology" and ``Energy Flow in Ecosystems" have addressed some of the implications of the allometric research for biological mechanisms and ecological interactions, and were partially funded by the Thaw Charitable Foundation.

- Core research funds provided initial support for research in pediatric AIDS that has led to a grant from the Elizabeth Glaser Foundation. SFI served as a cosponsor of an international conference on AIDS, with special sessions held at SFI on phylogeny and population dynamics of the HIV virus. The research effort on pediatric AIDS and HIV is proceeding in collaboration with the Theoretical Biology group at LANL.

- The Graduate Workshop in Computational Economics was for several years funded by Coopers and Lybrand. The workshop organizers---Scott Page (U. Michigan) and John Miller (decision sciences, Carnegie Mellon, SFI External Faculty)---began their involvement with SFI in core-sponsored research and educational activities. John Miller was SFI's first Postdoctoral Fellow (supported by core funds 1987--89) and attended the 1988 core-supported Complex Systems Summer School. Scott Page is an alumnus of the 1990 core-supported School.

- Core funds provided initial seed money for the Swarm software development effort. Swarm has led to the establishment of several major software projects at other institutions, including a $2M project at SAIC funded by DoD on battlefield awareness, as well as a large number of university research projects in ecological studies.

- Core support for the SFI economics program is being matched by funds from McKinsey and Company.

- DOE support for SFI's Integrative Core program has led to research projects that have attracted the interest and financial support of a large number of commercial research organizations. SFI's Business Network, with over 55 members, now serves as a major source of funding for the Institute's basic research agenda.

- DOE support for research efforts in agent-based modeling, applications to industrial modeling, and analysis of biological data has led to the establishment of four for-profit companies--Prediction Company, Swarm

Corporation, Intelligize, and Bios Group---that have spun off from SFI's basic research programs.

The remainder of this report is organized as follows: we present results on **computational biology** including work on the sequence-structure mapping for RNA, the immune system, and the evolution of redundant gene functions. We review research in the area of **biodiversity and ecosystem research** including a model for extinctions that demonstrates that critical phenomena are *not* necessary to explain fossil data; and derivation of allometric scaling properties for mammalian cardiovascular and respiratory systems, and for plant vascular systems. Finally, we review results of research in **advanced computing and simulation**, including derivation of exact quantitative results for the "punctuated equilibrium" behavior of populations of strategies evolving according to certain kinds of genetic algorithms; rigorous results for computation in dynamical systems; and the development of a general-purpose software platform (Swarm) for simulations of interacting agents.

Results of DOE Support: Computational Biology

In line with an increasing emphasis at SFI on analysis of empirical data, the computational biology effort has focused over the past year on studies combining simple modeling with intensive data analysis for the study of sequence and structure data. Work has progressed in topics including RNA structures, models of the immune system, and redundant gene functions.

DOE funds have been used in particular to support research in these areas that relate directly to interactions with the environment and the effects of environmental stress.

Structure in RNA and Proteins

In order to make evolutionary biology (as opposed to evolutionary computation) accessible to systematic studies, one needs (i) to reduce generation times in order to speed up evolution, (ii) to minimize complexity of phenotypes in order to allow for an analysis of genotype-phenotype relations, and (iii) to shorten genotype lengths in order to keep possible diversity below a certain limit. All three conditions are fulfilled by test-tube experiments on RNA molecules. This experimental approach showed that populations of replicating molecules adapt to environmental changes by means of a Darwinian mechanism of variation and selection just as cells or multicellular organisms do. In addition it was successful in the design of RNA molecules with predefined properties; for example, molecules binding to predefined targets, catalytic RNA molecules called ribozymes with novel catalytic functions, as well as RNA molecules that are resistant to cleavage by specific enzymes (RNases).

The research team of Peter Schuster (biology, Vienna, SFI External Faculty), Peter Stadler (biology, Vienna, SFI External Faculty), and Walter Fontana (chemistry, SFI Faculty) used theoretical and extensive computational methods to study RNA phenotypes; i.e., structures derived from their genotypes through "folding." These investigations on RNA secondary structures revealed four fundamental features of genotype-phenotype mappings: (1) the numbers of RNA sequences exceeds the number of structures by several orders of magnitude and hence, sequence-structure maps are many to one. (2) Relatively few common structures are contrasted by many rare structures which usually play no role in evolution. In the limit of long chains we find that almost all sequences are folding into a tiny fraction of all structures. (3) In order to find a given sequence forming any of the common structures one need not explore whole sequence space. It is sufficient to search a relatively small spherical neighborhood of any arbitrarily chosen reference sequence (shape space covering). (4) Common structures are characterized by a high degree of neutrality expressed as the fraction of nearest neighbors that are neutral with respect to structure formation. The sets of sequences folding into them form extended neutral networks in sequence space.

The essential properties of the fitness landscape obtained by evaluating the phenotypes are inherited from the genotype-phenotype mapping. Neutral mutation at phenotype level translates into fitness (near-) neutrality. A necessary prerequisite for shape space covering is the fact that a substantial fraction of mutations leads to significantly different phenotypes. Hence a fraction of mutants will have very different fitness values. This type of landscape is completely different from the isotropic, and relatively smooth, landscapes arising from combinatorial optimization problems (such as the TSP) or spin glass Hamiltonians. Not surprisingly, the dynamics of biological adaptation or the performance of optimization heuristics is determined by the qualitative features of a landscape, such as ruggedness, neutrality, and (an)isotropy. An algebraic theory of landscapes constitutes a new and promising avenue for approaching the relation of landscape structure and evolutionary dynamics.

Recent developments introduced genotype-phenotype relations into the theory of molecular evolution. The model has been implemented for computer simulations which are based on (realistic) rules derived from biophysical chemistry. Results of these computer-mediated studies on evolution are (i) Optimization occurs usually in steps: short adaptive phases of fast and substantial improvements are interrupted by long quasistatic periods of almost constant fitness which correspond to neutral evolution. (ii) Neutral evolution corresponds to a diffusion process in sequence space with a diffusion constant that is proportional to the mutation rate. (iii) Probabilities of success and efficiency of evolution depend strongly on existence and extension of neutral networks which enable populations to drift towards regions of higher fitness in sequence space. As shown in a recent paper, the course of evolution can be traced down to the molecular level and relay series of phenotypes leading to the target structure can be reconstructed from computer simulations. The analysis of relay series provides direct insights into the evolutionary process and allows us, for example, to explain the causes of punctuation and the lengths of neutral drift phases. Evolutionary jumps are initiated by major structural transitions in the RNA. Relay series can also be studied experimentally: Replication propagating through wave fronts in capillaries leaves behind a recording of the history which is, at least, a documentation of the most frequent phenotypes and genotypes.

Going beyond RNA, SFI External Faculty members Alan Lapedes and Rob Farber (biophysics, LANL), and Peter Stadler (biophysics, Vienna) have used knowledge-based potentials to explore the protein sequence-structure map. The number of possible sequences is dauntingly high, but the repertoire of stable native folds seems to be highly restricted or even vanishingly small. Focusing on the global features of the sequence-structure map of polypeptides, the questions will be of a statistical nature, dealing with ensembles of sequences and structures rather than targeting a particular sequence or structure of a particular protein: How many sequences adopt the same fold? How are these sequences distributed in sequence space? How similar are sequences that fold into very different shapes? Results indicate: (1) Neutral networks of mutations exist in protein sequence space; (2) The distances of closest approach for proteins are surprisingly small: it is theoretically possible to construct sequences that fold to one shape, and for which a few amino acid changes will cause it to fold to an entirely different shape; (3) The same qualitative results are obtained from computer experiments using two very different potential functions.

Finally, generalizing some of the above results, Mark Newman (physics, SFI Faculty) and Robin Engelhardt (Univ. of Copenhagen) have introduced a model of evolution on a fitness landscape possessing a tunable degree of neutrality. The model allows them to study the general properties of molecular species undergoing neutral evolution, including the occurrence of "common" structures and the formation of percolating neutral networks. The model also generates a number of new results. In particular, Newman and Engelhardt show that the maximum fitness attained during the adaptive walk of a population evolving on such a fitness landscape increases with increasing degree of neutrality, and is directly related to the fitness of the most fit percolating network. Peter Stadler (Vienna, SFI External Faculty) and Wim Hordijk (SFI Postdoctoral Researcher) have used the model as a test-bed for new measures for quantifying landscape structure which they are developing, and Ricard Solé (ecology, Barcelona, SFI External Faculty) has used it to construct a computer model of Red Queen effects and critical dynamics in the evolution of RNA viruses.

An intriguing result in the previous grant period, obtained by Peter Stadler (Vienna, SFI External Faculty), F.X. Heinz and C. Mandl (virology, Vienna), Martijn Huynen (EMBL Heidelberg, former SFI postdoc), Ivo Hofacker (UIUC), and Paul Stolorz (JPL/NASA/CalTech), been the identification of functional secondary structures in viral RNA genomes. The genome of a typical RNA virus contains on the order of 10000 bases, mostly coding for viral proteins. Non-coding regions on both ends of the genomic RNA are important for regulating the viral life cycle. Additional regulatory sites are present within the coding region in some viruses. These regions have in general distinctive secondary features that are conserved among related viruses.

For most groups of viruses only a few complete genomic sequences are known. Massively parallel computers can be used however to predict the secondary structures, and even the matrix of all base pairing probabilities, from the sequence using recursive algorithms based on thermodynamic considerations. The sequences are far too long, however, to expect a perfect prediction of the structure based on thermodynamic considerations. The predicted structures of closely related viral genomes are therefore compared in order to detect common structural elements among related viruses. Only these conserved structures are likely to be of functional importance. Compensatory mutations in the underlying sequences can then be used for checking the reliability of the predicted substructures. The methods have been applied successfully to lentivirus sequences (such as HIV) and to flaviviruses (such as Yellow Fever, Dengue and Tick-Borne Encephalitis), thereby identifying conserved and non-conserved secondary structure elements in this group. Considering that the genomic RNA sequences of these viruses are highly divergent, conservation of secondary structure elements strongly suggests functional constraints on their evolution.

Immune System Modeling

A collaboration between Stephanie Forrest and Alan Perelson (immunology, LANL, SFI External Faculty) has focused on models of the immune system from an information-processing perspective. Their models represent the complex chemistry of antibody/antigen recognition in highly simplified form as binary string matching. In work supported by SFI using funds from DOE and the Sullivan Foundation, Forrest's Derek Smith has developed a new model of cross-reactivity in the immune system, based on the observation that several disparate phenomena--previously treated separately by immunologists---can be viewed as different regimes of an associative memory. Cross-reactivity is the reason that vaccines can be successful when the epidemic strain of a virus is slightly different from the vaccination strain.

The model has been validated against available experimental data, and applied to the problem of annual vaccination against influenza, a problem of great interest to the Centers for Disease Control (CDC). In populations of repeat vacinees, vaccine effectiveness sometimes declines dramatically from that observed with first-time vacinees. Although this phenomenon has been documented for over twenty years, there has never been a satisfactory explanation. The model makes predictions about the conditions under which vaccine effectiveness would decline, and suggests how vaccine strains could be selected to minimize this phenomenon. Smith is now working with CDC in designing laboratory experiments to test the predictions.

In the area of viral treatment strategies, the team led by SFI External Faculty members Alan Perelson, Avidan Neumann (Bar-Ilan), and Rob de Boer (Netherlands) has focused on analysis of hepatitis C virus (HCV) infection. They have developed a model for the dynamics of HCV and the anti-viral effect of interferon, and analyzed the viral decline in 23 patients during therapy. Analysis indicates that the major initial effect of IFN is to block virion production or release, with blocking efficacy of 81%, 95%, and 96% for daily doses of 5, 10, and 15 mIU, respectively. The estimated virion half-life was, on average, 2.7 hours, with pre-treatment production and clearance of $\$10^{12}$ virions/day.

Results indicate that IFN doses of 10 and 15 mIU daily have significantly better early antiviral efficacy than a dose of 5 mIU daily, and based on the team's previous findings, also higher efficacy than the standard dose of 3 mIU. Studies with larger number of patients are necessary to understand the importance of this early anti-viral efficacy, manifested as the drop in viral load after 2 days of therapy, as well as the rate of the second phase decline on the long-term success of treatment. Uncovering the rapid dynamics of HCV has implications for the possible emergence of viral resistance to new therapeutic agents, such as the protease inhibitors for HCV that are currently being designed, for assessing the possibility of viral eradication, and for managing patient treatment.

Current research on the immune system focuses on its ability to use molecular mechanisms such as somatic mutation in combination with the cellular level processes of clonal selection and affinity maturation to enable the collective behavior of recognition and response to a foreign pathogen. The fundamental question, in other words, is to understand the evolutionary mechanisms that enable the immune system with limited genetic resources to attain very broad coverage of the set of pathogens.

The work is examining the roles of both germline and somatic diversification mechanisms in the evolution of immune systems, with preliminary results indicating that the selection pressure for ``mutability" of immune receptors under somatic hypermutation is at least as strong as selection pressures for germline diversity. One specific goal of the research is to develop models and data analysis techniques that will help identify the properties of the specific mutator mechanism that governs somatic diversification of antibody genes.

A July 1999 workshop on "Effector Choice in Distributed Systems"--organized by Irun Cohen (immunology, Weizmann), Stephanie Forrest (computer science, UNM), Alan Perelson (immunology, LANL), Lee Segel (applied math, Weizmann), and Alan Sher (parasitic diseases, NIH)--will bring experimental and theoretical immunologists together with researchers in distributed computational systems to focus on "cognitive" capabilities of the immune system and the biological mechanisms of effector choice. The workshop will explore the effectiveness of the immune system in protecting the host from pathogen-induced harm (including harm by the immune system) via cognizance of context.

Genomic Studies in Model Organisms

In the past two years, SFI has established an exploratory research program in simple microbial systems with the objective of understand organizational and functional principles of genomes both in model organisms and in humans. The program has led to several collaborative efforts that will be the basis of a major research thrust at SFI in the coming years.

In October 1997, a workshop on "Cellular Computation and Decision-making" was held at SFI organized by John Holland (Computer Science, University of Michigan, and SFI External Faculty), Jim Shapiro (Department of Bichemistry and Molecular Biology, University of Chicago), and Erica Jen (SFI). The purpose of the workshop was to bring together the concepts and approaches of two fields--molecular biology and theoretical computer science--for the study of molecular computing/decision-making networks in cells, and to establish the foundations for a future research thrust at SFI in this area. The focus was on the experimental data and theoretical analysis for molecular networks in a small number of well-analyzed microbial systems: specific examples included the molecular switch that controls phage lambda repression, bacterial chemotaxis, bacterial sporulation, yeast response to mating pheromones, and the yeast cell cycle.

A specific goal of the workshop and the ensuing research collaborations was to use the experimental information on real molecular networks in microbes to derive novel computational representations and models with predictive powers. Holland and Stephanie Forrest (computer science, UNM, SFI External Faculty) developed a prototypical model for the phage lambda switch based on what is known in the genetic algorithms community as the "classifier system" approach. The model incorporates novel theoretical features including rule-based update dynamics, and networks of adaptive agents interacting with themselves and with their environments. The proposed modeling framework has the potential for representing the analog processing capabilities of proteins that interact with multiple ligands, and for translating experimentally-determined molecular properties, such as binding and catalytic constants, into components of computational algorithms. A theoretically intriguing question raised by the model is the possibility its use in exploring the emergence and evolutionary development of molecular processes constituting the phage lambda switch.

As another result of the "Cellular Decisionmaking" workshop, one of the participants--Lee Hartwell, Director of the Hutchinson Cancer Research Center and a leading experimental cell biologist--joined the SFI Science Board in March 1998. During the past year, Hartwell also proposed the establishment of an exchange program between the SFI and the Hutchinson Cancer Research Center to encourage collaborations on a postdoctoral level between theorists and experimentalists. The exchange program has now been approved by both institutions. A planning meeting scheduled for November 1998 will be attended by a small set of experimentalists--including Barry Stoddard (structural biology), Stephen Tapscott (clinical genetics), Mark Roth (structural biology), and Lee Hartwell (cell biology)--representing the Hutchinson, and a small set of theorists--including Marc Feldman (population genetics, Stanford, SFI Science Board), Stan Leibler (biophysics, Princeton), Richard Lewontin (evolutionary biology, Harvard, SFI Science Board), George Oster (biophysics, Berkeley, SFI Science Board), and Andreas Wagner (biology, UNM)--representing the SFI.

A number of collaborative efforts were mapped out at the November 1998 meeting, and are being pursued. A major topic will be the role of natural variation in populations, and the mechanisms used by cells to tolerate enormous polymorphism without major phenotypic consequences. SFI theorists and Hutchinson experimentalists plan to define the population genetics issues that can now be approached experimentally in organisms like yeast.

A second topic for collaboration will be the question of functional polymorphisms within genes. SFI researchers are bringing the perspectives of evolutionary biology and population biology to questions of mutation rates, representation of mutations in a population, and the constraints on mutations during evolution. Hutchinson researchers are interested on the other hand in structure/function question for DNA, RNA, and proteins. Combining an SFI approach to identifying functionally significant polymorphisms in a specific gene, both within a population and across evolution, with a Hutchinson structure/function analysis may begin to reveal the basis for transitions in the evolution of specific families of genes. Conversely, identifying polymorphisms associated with epochal events could be a tool for identifying important functional variants.

Results of Prior DOE Support: Biodiversity and Ecosystem Research
The work in this area has addressed basic issues in the origins of biodiversity and ecological complexity. Areas in which major results have been obtained include the relative importance of internal dynamics versus external stress in leading to mass extinctions, and the origins and implications of allometric scaling laws for biological organisms.

DOE funds have in particular supported the components of this research that relate directly to the response of ecosystems to environmental stress, and to the implications of scaling laws for resource requirements and resource use of organisms.

Models of Extinction
Extinction has played an important role in the development of life on the Earth. Of all the species that have lived here since life first appeared, somewhere between one in a thousand and one in a hundred is still living today. All the others, the vast majority, became extinct, typically within one to ten million years of their first appearance. What is the cause of this high extinction rate? Is it an inevitable part of the process of evolution or merely a result of the particular environment found on the Earth? Has extinction helped further evolution by removing less fit species from the ecosystem, as Darwin believed, or is it a random process which affects all species equally? What signatures has extinction left in the fossil record, and how much can we tell about its causes and effects from fossil data? Efforts to understand all of these questions are under way at SFI.

It is important first to understand the nature of extinction in the fossil record, before one can undertake any modeling efforts to try to explain that extinction. In recent years a number of claims have been made about apparent trends in the fossil record of extinction which might point to interesting underlying dynamical processes in the biosphere. Stuart Kauffman (biology, Bios, SFI External Faculty), for example, has argued that the distribution of the sizes of extinction events approximately follows a power-law form, taken by some to be indicative of criticality in the processes giving rise to extinction. Sneppen and others have suggested that the distribution of the lifetimes of genera may also be a power law, and Ricard Solé (ecology, Barcelona, SFI External Faculty) have presented evidence that the power spectrum of extinction intensity during the Phanerozoic has a 1/f form.

Using a comprehensive database of fossil species compiled by the late Jack Sepkoski (paleontology, Chicago), Mark Newman (physics, SFI Faculty) and Gunther Eble (paleontology, SFI/Smithsonian Postdoctoral Fellow) have been investigating these and other trends using methods such as Monte Carlo simulation and Fourier transform methods which, although common in the physical sciences, have not been much used in paleontology previously. Initial results indicate some interesting conclusions. Some of the trends noted in the fossil record are probably real. Newman, for instance, has used a comparison of Monte Carlo simulations of genus survivorship with actual survivorship curves from the Sepkoski database to deduce a so-called "kill curve" for genus extinction. From this he has shown that the distribution of the sizes of extinction events in the fossil record is indeed compatible with a power-law form, as Kauffman suggested. On the other hand, Newman and Eble have looked at the power spectrum of extinction (which Sole et al. suggested possessed a 1/f form) and found that on careful analysis the spectrum in fact probably has two separate regimes, one of exponential decay with a typical timescale similar to the mean timescale for biotic turnover in the fossil record, and other of rapid power-law decay which is an artifact of timescale interpolation related to the Porod law of scattering theory. Newman and Eble are also investigating the possibility that some other characteristic behaviors seen in the fossil record may be artifacts of the gradual decline in background extinction during the Phanerozoic eon. They have shown that the cumulative extinction of families over geological time follows a logarithmic growth law similar to that proposed by Kauffman and Simon Levin (Princeton, SFI Science Board) and more recently by Paolo Sibani. This corresponds to a decay in the extinction rate which goes as 1/t, and this decay may be responsible for a variety of systematic biases in previous analyses.

Turning to the evolutionary and environmental processes which drive extinction, there has in the last ten years been considerable interest in modeling the effects of large-scale coevolution on ecosystems. Organisms coevolve because of interactions between them---predation, competition, parasitism and other effects can force one species to evolve as the result of the evolution of another. It has been argued by Per Bak (physics, NORDITA, SFI External Faculty), for example, that co-evolution can give rise to a "domino" effect, in which the initial mutation of one species causes a wave, or avalanche of evolution to travel across the ecosystem. It is possible that such collective evolution effects, because of their size, could be the dominant mode in the evolution of ecosystems; large coevolutionary avalanches may well be rare, but some theories suggest that they are still frequent enough that any particular species is more likely to be touched by one such than to undergo a spontaneous mutation of its own.

In addition, the possible existence of coevolutionary avalanches raises intriguing questions about the way in which ecosystems organize themselves and about the closely related issue of extinction rates. Chief amongst these questions is that of self-organized criticality in evolution, which arose first in the work of Kauffman, and of Bak and Kim Sneppen. A number of authors have attempted to explain the nature of the fossil record with these theories. There are however a number of open questions. First, the existence of these avalanches has never been empirically confirmed. Secondly, the actual mechanism by which coevolution could lead to avalanches of extinction has never been identified. Thirdly, there is the problem of determining the time and spatial spaces for these conjectured coevolutionary avalanches.

A contrasting view, shared by much of the paleontological community and currently being explored at SFI, is that extinction is caused by external perturbations on the ecosystem. An argument in favor of external perturbations as a principal causal factor in extinction comes from a new model developed by Newman and Sneppen. In contrast to the models based on self-organized criticality, the dynamics of Newman/Sneppen model evolve solely in response to differing environmental stresses on species. Interestingly, the model shares with the other models the feature that it gives reasonable qualitative agreement with the power-law distribution of the data. The work indicates that it is {\it not} necessary to invoke critical processes or other internal dynamical effects to account for the observed fossil extinction record.

Ongoing research will focus on a search for characteristics in the fossil data that would constitute distinctive signatures of either internal coevolutionary interactions or external environmental stresses. One possible signature for external environmental stress has been suggested by Newman--the so-called "aftershock" extinctions.

A workshop organized by Newman was held in February 1997 to explore diverse approaches to modeling the extinction data; since then, Newman has set up a working group for testing the range of theoretical models against experimental data provided by Sepkoski.

Scaling Laws in Biology and Ecology

In a collaboration initiated by SFI, Geoffrey West (physics, LANL), Jim Brown (ecology, UNM, SFI External Faculty), and Brian Enquist (ecology, SFI Postdoctoral Fellow are pursuing studies of scaling in biology. West and his collaborators have developed a general for the origin of allometric scaling relationships in biology. They claim to have solved a long-standing problem: why metabolic rate scales as the 3/4 power of body mass. The model assumes that the distribution of materials within organisms is the rate limiting process and derives the properties of fractal-like distribution networks based on basic physical and biological principles. The model makes many predictions that can be tested by comparisons with values in the literature, or by performing new measurements. The published version accurately predicts many features of mammalian cardiovascular and respiratory systems in the biomedical literature, and a recently completed manuscript makes similar predictions for the structure and function of trees in the forestry literature. Most of the predicted scaling relationships have exponents that are simple multiples of 1/4, due in part to the central role of metabolism in organismal biology and in part to the requirement that organisms have an integrated solution to the problem of maintaining viable structures and functions in the face of widely varying body size.

The research results provide a general theoretical basis for biological allometry, which has until now been a discipline rich in data and descriptive studies but lacking conceptual unification. Current research uses the model and the central role of metabolism to explore scaling at other levels of biological organization: both above and below the individual organism. On the one hand, variation in body size is one of the most important axes of biological diversity. Living organisms vary in size by an incredible 21 orders of magnitude, from 10^{13} g microbes to $\$10^{8}$ g whales. There is a rich literature showing relationships between body size and life history traits (rate of growth and development, lifespan, litter or clutch size, and population age structure), ecological attributes (intrinsic rate of population growth, territory or home range size, population density, and species diversity), and evolutionary phenomena (speciation and extinction rates). Enquist et al. have recently used the plant model to derive a resource-based thinning law (population density as a function of plant size) which predicts that productivity of ecosystems is independent of the size of the dominant plants. They are currently trying to extend the results and to provide a theoretical explanation for the relationships between body size, abundance, and species diversity seen in nearly all kinds of organisms.

In collaboration with W.H. Woodruff, a biochemist at Los Alamos National Laboratory, the team of West, Brown, and Enquist has started to explore problems of scaling of structure and function at the cellular and molecular levels. It appears that the scaling of metabolic rate with size of "unit" holds not only for individual mammals from shrews to whales, but also within mammals, from the whole body down to cellular organelles and even metabolic molecules, so that rates of activity of a mitochondrion and a terminal oxida se fall right on an extension of the mammalian scaling line. These models should generate testable predictions about the structures and dynamics of material transport and energy transformation at levels from organismal to molecular.

Finally, West, Brown, and Enquist are interested in the implications of their work to other complex systems in which energy or materials are distributed through branching networks, both natural (e.g., streams) and human (e.g., water pipes and electrical and telephone grids).

Other Results

A collaboration in the last two years between Tim Keitt (ecology, NCEAS) and Eugene Stanley (physics, Boston University)---initiated during a visit by Stanley to SFI under the auspices of the Economics program---has led to results in applying techniques from statistical physics to the analysis of biological population dynamics. Keitt and Stanley undertook a study of dynamical scaling in an ecological system for which considerable data exist over a 31-year period: breeding bird populations across North America. The distribution of year-to-year population fluctuations is found to be leptokurtic and strongly non-Gaussian. The tails of the distribution are power-law, consistent with the possibility that changes in population size have a fractal dynamical structure, and the dynamics resembles a process known as a Lévy flight. These results suggest the possibility that bird populations in North America are strongly influenced by climate fluctuation and hence never saturate the available resources (e.g., the severe 1976-77 winter in eastern North America caused widespread declines of resident species.

Other research projects resulted in the development of methods for quantifying habitat connectivity at multiple scales, and assigning conservation priority to habitat patches based on their contribution to connectivity. It was shown that percolation theory could be used to quantify connectivity at multiple scales from empirical landscape data. The results indicate that connectivity of landscapes is highly scale dependent, exhibiting a marked transition at a characteristic distance and varying significantly for organisms with different dispersal behavior. More importantly, it is shown that the sensitivity and importance of landscape pattern is also scale-dependent peaking at scales associated with the percolation transition. The sensitivity analysis permits the identification of critical "stepping stone" patches that when removed from the landscape cause large changes in connectivity.

Three workshops were held in the previous grant period (with partial support from the Thaw Charitable Foundation) on topics relating to biodiversity, originations, allometric laws, and other topics of ecological complexity. The first workshop "Universal Phenomena in Ecology?" held in March of 1997 addressed the question of whether ecological systems organize themselves into more or less discrete assemblages of co-adapted species, or whether species are distributed continuously and independently in space and time. The second workshop "Scaling in Biology and Ecology" held in October 1997 addressed general scaling behaviors in organisms and ecosystems, and the third workshop "Energy Flow in Ecological Systems" focused on the effect of flows of energy and materials on the structure and dynamics of ecosystems.

Results of Prior DOE Support: Advanced Computing and Simulation
A major research area over the past grant period has been the development of novel computational approaches that incorporate information-processing features of biological and social systems.

DOE funds have supported the components of this research that relate directly to the design of computational search and optimization algorithms.

Evolutionary Computation
A section above presented results pertaining to epochal evolution, neutral genetic variation, and finite population size in the context of structural transitions of RNA molecules. In this section, similar results are presented for tuning the search parameters of evolutionary computation.

Researchers at SFI have adapted techniques from statistical mechanics and stochastic process theory to develop a new mathematical approach to understanding the dynamics of genetic algorithms. The starting point is a simple genetic algorithm using mutation only (no crossover) working on a class of fitness landscapes that were believed to be easy for the genetic algorithm---one of the so-called "Royal Road" functions. It turned out that the GA's performance was much richer than originally conjectured.

The first step was to construct a generation operator that captures the action of the GA on a population fitness distribution. Due to a judicious choice of projection onto macroscopic "phenotypic" coordinates, the eigenvectors of this operator can be computed, yielding the operator's fixed points. These fixed points translate into the evolutionary "epoch"--periods of stasis in the population mean fitness, punctuated by "innovations"--jumps in mean fitness. It has been proved that only the last epoch is stable; moreover, the temporal dynamics for the GA's mean fitness in the infinite population case have been solved analytically. The theory has made predictions that match experimental data very well. A new analytic method, using both statistical mechanics and the technique of path integrals, has been developed for calculating the epoch lengths as a function of population size and mutation rate. These results represent an "exact solution" for the GA in the sense of predicting its behavior on the Royal Road class of fitness landscapes. This is, to our knowledge, the first example of a predictive theory for a wide class of evolutionary search problems. Crutchfield and Erik van Nimwegen (SFI Postdoctoral Researcher) have adapted the theory to analytically estimate the computational effort required to reach a global fitness optimum. The results include estimates of the mutation rate and population size that lead the evolutionary search to operate more effectively.

An outgrowth of this research was a study of the evolutionary advantages of recombination. In work inspired by interactions at SFI with computer scientist John Holland, Marc Feldman (biology, Stanford, SFI External Faculty) and collaborators calculated the average waiting time until a new genotypic combination first appears in a haploid population. The link with the SFI adaptive computation program was the interest in obtaining estimates for the time until appearance of the first chromosome of a specific desired genotype under evolution of populations of strategies using genetic algorithms. A related result was a generalization by Freddie Christiansen and Feldman of John Holland's schemata theorem for genetic algorithms: the result provides a lower bound in the context of population genetics on the expected changes in gamete frequencies as functions of the genetic operators and the adaptive values.

Computation in Dynamical Systems

In the past two years, researchers Cris Moore (physics, SFI Research Professor), Jim Crutchfield (physics, SFI Faculty), David Griffeath (mathematics, Wisconsin, SFI External Faculty), Mark Newman (physics, SFI Faculty), Mats Nordahl (physics, Chalmers, SFI External Faculty), and Kristian Lindgren (physics, Chalmers, SFI External Faculty) have made significant progress in the rigorous analysis of the computation and information-theoretical properties of dynamical systems that serve as the basis of novel forms of computation.

Spin glasses typically present analytical and computational challenges to statistical physicists because of their inherent randomness and slow dynamics. Moore and Newman have shown, however, that spin glasses need not be random at all. They recently introduced a simple two-dimensional spin model with short-range interactions which shows glassy behavior despite a Hamiltonian that is completely homogeneous and possesses no randomness. The model has been solved exactly for both its static properties and its lowtemperature dynamics. This is one of the first glassy models for which the dynamics can be fully understood from a solution for the distribution of the energies and barriers for the lowest-lying excitations. Moore and Newman have also performed extensive Monte Carlo simulations of the model and are currently investigating its aging behavior.

Other results in this area include the analytical calculation of excess entropy and statistical complexity for one-dimensional Ising spin systems, use of excess entropy as a generalized order parameter that outperforms traditional statistical mechanical observables (e.g., correlation length and structure functions) at detecting patterns in spin systems.

A second set of results identifies the classes of cellular automata (CA) that can be predicted quickly; for instance, in $O(\log t)$ or $O(\log^2 t)$ steps on a parallel computer, rather than the t steps it would take to simulate the CA explicitly. A wide range of nonlinear rules can be predicted in this way, including some with diffusing defects that show pseudo-random motion. Moore and collaborators further used the theory of non-associative algebras to solve the problem of when "permutive" CAs, where the output is a one-to-one function of its leftmost and rightmost inputs, can be predicted quickly.

Similar results have been obtained for systems of direct physical interest. The question for such systems is to determine whether there is a fast prediction technique, or whether the system is *P*-complete, in which case no short-cut exists unless widely held beliefs in computer science are false. Griffeath and Moore have proved that a prototypical 2-D growth rule is *P*-complete, and Nordahl and Moore proved the same for the HPP and FHP lattice gases in 2-D (the latter being the basis of novel numerical methods for fluid turbulence). It has now been shown that majority-voting rules, zero-temperature Ising models, and sandpiles are all *P*-complete in three or more dimensions. In 1-D, they can all be predicted quickly; this is non-trivial for sandpiles, but Nilsson and Moore found an algorithm that runs in $\acute{O}(\log^3 n)$ parallel time. This leaves the 2-D case of these systems as a tantalizing open problem, since no fast algorithms have been found for these systems, nor have they been shown to be *P*-complete.

Finally, Crutchfield and Moore have defined quantum versions of finite-state and push-down automata, two of the lowest-level architectures of computing devices, which recognize regular and context-free languages, respectively. They proved that while quantum finite-state automata can be simulated by classical stochastic ones with more states, quantum push-down automata are strictly more powerful than classical ones. That is, there are quantum context-free languages that are not classically context-free; i.e., they are more complex than the classical case. This is one of very few provable separations between quantum and classical complexity classes.

A Modeling Platform for Agent-Based Simulations
A major result of prior core funding has been the development of software tools designed for the study of collections of interacting, autonomous agents. The behavior of systems is determined by the strategies adopted by the individual

agents in response to interactions with other agents, and often, with a changing environment or with an environment that is being changed by the agents themselves. Agents may be either homogeneous or, more realistically, heterogeneous in their structure, function, and information-processing capabilities. Thus, depending on the context, the agents in these networks may represent genes involved in regulatory networks, individual species in ecological systems, individual members of social insect colonies, individual consumers in an economy, or any of a wide variety of other examples. The approach represents a powerful modeling tool for incorporating realistic data--differences in individual characteristics, details of spatial distributions, and features of network connectivities--difficult to address using traditional theoretical or computational approaches.

The "Swarm" platform is a freely-redistributable software tool whose development was supported by DOE through 1997. In the previous grant period 1996--98, the Swarm effort moved from initial design stages through an early beta-version to a mature, stable, fully documented Version 1.3 (available on both UNIX and Windows platforms) that is now being used by over 400 researchers. The original design and development was led by Chris Langton (Swarm Corporation, former SFI Faculty), and is now managed by an Advisory Board consisting of Swarm users and developers. Usage statistics show continuing steady growth. A SwarmFest Users Workshop held at SFI in March 1998 was attended by over 75 individuals (who paid their own expenses to attend). The most recent SwarmFest Users Workshop was held in March 1999 at UCLA and was attended by over 125 users.

Specific problems being modeled by Swarm include population mobility and the growth of highway systems; a spatial model of an artificial chemistry developed by Fontana and Buss; the spread of forest fires; mosquito breeding and feeding patterns in the context of developing an anti-malaria strategy for Western Africa; and several large-scale simulations sponsored by the DoD in battlefield awareness and network monitoring.

A major new linkage for the Swarm effort has been the simulation and analysis of physical and ecological systems for which there exists large-scale spatial data sets, especially GIS data, with a special emphasis on issues of novel architectures and algorithms required to synthesize GIS tools with dynamic modeling capabilities. An SFI workshop was held in March 1998 at which representatives from several large-scale ecological simulation groups--including the ATLAS Everglades, Chesapeake Bay, Columbia River, James San Jacinto Mountain Reserve, and USDA Long Term Ecological Research Stations--met to review the technology developed to date, and to coordinate efforts in dynamic spatial modeling. At the 1999 SwarmFest Users meeting, one of workshop topics focused on the development of a GIS-CA class for the Swarm libraries.

Data on Publications

The Institute monitors publications by SFI researchers in the scientific literature in two ways. The first is through self-reports by the authors to the Web-based SFI Bibliography. In addition, SFI subscribes to the reporting services of the Institute for Scientific Information (ISI) which produces Science Citation Index and compiles the contents of about 16,000 journals, all of which are peer-reviewed. Listings from ISI include only those articles for which SFI is listed as a primary or secondary affiliation of one of the authors.

On the basis of these reporting methods, SFI has compiled a bibliography that indicates that in the period 1996--1998, SFI has contributed directly to over 500 scientific publications. Of these, about more than articles have appeared or have been accepted to date in the peer-reviewed literature. At rough estimate residential and visitor research here during the same time period total approximately 60 person-years of effort. See http://www.santafe.edu/sfi/research/indexAllpubs.html.

Education and Outreach

SFI makes a continuing contribution to the national research and training infrastructure through its innovative work in the study of complex adaptive systems and its success in seeding new interdisciplinary research networks. Since its founding the Institute has been committed to speeding the infusion of its approach into traditional science practice and to promoting understanding of the implications of this kind of interdisciplinary research. With educational programs ranging from the secondary to postgraduate level, a growing network of institutional interactions, and aggressive outreach strategies for a range of constituencies, SFI pursues this commitment on many fronts.

Schools

The goal of the Complex Systems Summer School is to provide graduate students, postdoctoral fellows, and active research scientists with an introduction to the study of complex behavior in mathematical, physical, and living systems. Even after eleven years of CSSS events, there are virtually no other programs that comprehensively present the concepts and techniques of complex systems, together with the work in progress of researchers applying these ideas to specific problems. It is especially difficult for graduate students in any one institution to be exposed to more than a small segment of the intellectual activity in this area.

The annual program consists of short courses (which vary from year to year) together with seminars on selected topics. Approximately 50 scientists, predominantly graduate students, but also postdoctoral fellows and senior researchers, attend the month-long school. Annual themes vary but typically include such topics as (i) stochastic processes in physical, chemical and biological systems; (ii) pattern formation in physical, chemical, and biological systems; (iii) cognition and computational approaches to brain and cognitive function; (iv) adaptive computation; and (v) the economy as a complex adaptive system.

The annual Graduate Workshop in Computational Economics brings together a group of advanced graduate students and a small faculty for an intensive two-week study of computational economics. The program is co-chaired by John Miller (Social and Decision Sciences, Carnegie Mellon) and Scott Page (U. Michigan). The workshop consists of lectures by faculty, special topic seminars by members of the Santa Fe Institute, and presentations of work in progress by graduate student participants. The primary goal of the summer workshop is to assist graduate students pursuing research agendas which include a computational component. A significant portion of the workshop is devoted to analyzing and improving research being conducted by the graduate student participants.

Postdoctoral, Graduate, and Undergraduate Student Programs

Work of SFI's Postdoctoral Fellows is at the core of the Institute's research; much of this work is described elsewhere as part of ongoing or prospective projects. Typically, about eight Fellows are in residence fulltime at SFI, and frequently several other visiting postdoctoral fellows are in short-term residence as well. The duration of the typical postdoctoral fellow appointment is two years.

During the grant period September 1994 to 1999, 18 postdoctoral fellows are or have been in residence in concurrent and/or overlapping appointments comprising a total of approximately 35 person years. They represent the fields of theoretical biology, physics, computer science, ecology, computational mathematics, linguistics, molecular genetics, theoretical mechanics, immunology and paleontology. During the 1995-1999 period, their publications in refereed journals resulting from research during this grant period number 101 articles. Of these 18 individuals, ten have now completed their tenures at SFI: three have moved on to tenure-track faculty positions; three hold research positions in industry; and four are currently at national laboratories and research centers.

To encourage networking and collaborative research for these young scholars, the SFI Science Steering Committee resolved at its November 1997 meeting to ask SFI Science Board and External Faculty members to take the initiative, as part of their responsibilities at SFI, in establishing and maintaining collaborative research projects with SFI postdocs. SFI now provides travel money to support extended research visits to offsite affiliates. SFI is also strengthening the experimental and empirical component of postdoctoral research by providing support to postdocs to participate in offsite experimental and data collection projects. Eric Bonabeau, for example, has devoted a significant portion of his tenure at SFI to participating in experimental design, specimen collection, and experiments on foraging, dead body collection, and task allocation in ant colonies, and sociogenesis in social wasps. SFI advertises that special consideration will be given in the selection process to postdoctoral applicants who wish to incorporate offsite experimental work into their SFI residencies.

SFI also provides longer-term interdisciplinary research opportunities for graduate students. These students may accompany senior scientists visiting SFI or, with the agreement of their home institution, conduct thesis research and writing in residence as part of one of the SFI research initiatives. The number of graduate students at SFI fluctuates, with an average of 3--5 in residence for longterm projects and as many as 10--20 on hand when summer programs are in session. As with the postdoctoral fellowship program, SFI encourages graduate students to develop collaborative relations with its offsite affiliates in addition to their own thesis supervisors.

Through partial support from the National Science Foundation (in addition to SFI core funding) the Institute has for the last six years supported summer research residencies by undergraduates. Work is on a project mutually designed by student and mentor centering on some aspect of the computational properties of complex systems. Each intern participates in the variety of colloquia, seminars and informal discussion groups that occur continually.

Secondary School Program
Last summer the Santa Fe Institute, the Massachusetts Institute of Technology Media Lab, and the University of Maine combined forces to help Santa Fe high school students and teachers from a variety of disciplines learn about modeling and complexity. The two-week StarLogo Community of Learners Workshop brought together twenty-one local participants and five workshop leaders from the three institutions for a hands-on experience in model building. Participants in the workshop explored a variety of social and scientific models, including predator-prey dynamics, population growth, and traffic patterns, using StarLogo and a variety of participatory activities. This summer SFI will again offer student and teacher workshops and will also sponsor an ongoing academic-year project.

Finally, the SFI established in 1998 a new Fellows-at-Large program. The program is aimed at supporting the research efforts of advanced graduate students, postdoctoral fellows, and junior faculty in the area of complex systems, and promoting the establishment of such research agendas in the individuals' home institutions. The Fellows-at-Large program will provide funds for the Fellows to invite SFI-affiliated researchers to visit the Fellows' home institutions for research talks and short-term collaborations. In addition, the SFI Fellows-at-Large will be invited to two group meetings to be held at SFI (one meeting at the beginning of their tenure, and one at the conclusion of their tenure) to interact with each other, with outgoing Fellows, and with SFI visitors and research staff. The fellowships will be one year in duration. Five fellows have been selected in 1999. Brian Billman is an Assistant Professor of Anthropology at the University of North Carolina at Chapel Hill and Director of Latin American Studies for the Center for Indigenous Studies in the Americas (CISA). Scott de Marchi, Assistant Professor of Political Science at Washington University, specializes in the fields of computational economics, game theory, the Presidency, and voting behavior. Princeton's Laura Landweber's main research interest is the evolution of biological information processing. Filippo Menczer teaches courses in Management Information Systems at the University of Iowa where he is an Assistant Professor of Management Sciences. Assistant Professor of Kinesiology at Pennsylvania State University, Dagmar Sternad's work focuses on exploring the laws of biological movement control.

Public Lectures
For the past decade the Institute has hosted monthly public lectures for the community on a wide range of topics within the science of complexity. These talks are free, and draw audiences of between 200 and 700. Since 1996 every one of the monthly lectures has been financially underwritten by a variety of local businesses, testament to the value of the series to the community.