SOCIETY for INDUSTRIAL and APPLIED MATHEMATICS

3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA
Phone +1-215-382-9800
Fax +1-215-386-7999
www.siam.org • siam@siam.org

May 25, 2010

MATHEMATICS: AN ENABLING TECHNOLOGY FOR THE NEW BIOLOGY
Society for Industrial and Applied Mathematics Response to the "New Biology" Report

In support of federal agencies' consideration of responses to the 2009 National Academy of Sciences report on "A New Biology for the 21st Century" and other reports on the interface between the life sciences and physical sciences, mathematics, and engineering, the Society for Industrial and Applied Mathematics (SIAM) is offering recommendations on specific opportunities for the federal government to strengthen research and education at this interface.

SIAM is an international community with approximately 13,000 members from academia, industry, and government. Our members, from many different disciplines, have a common interest in applying mathematics in partnership with computational science towards solving real-world problems.

MATHEMATICAL SCIENCES, INFORMATION, AND THE NEW BIOLOGY

The 2009 National Academy of Sciences report "A New Biology for the 21st Century: Ensuring the United States Leads the Coming Biology Revolution" proposes a national initiative to promote the New Biology that focuses on problem-centric, interdisciplinary research in the life sciences to solve societal challenges in Health, Food, Energy, and Environment. A central finding of the report is that new information technologies and sciences will be essential to achieving the New Biology and meeting these challenges.

For each challenge area, the report outlines how biology can contribute and which research and technological needs must be met in order to do so. **In each area, new approaches to information analysis, data, and modeling will be needed to advance our understanding of the natural world,** as biology develops as a predictive science**.**

Food: In order to help ensure a sustainable and responsibly grown food supply, particularly in light of the changing global climate, one of the challenges is to understand and quantify how plants grow and interact with their environment. This involves characterizing the relationship between genotype and phenotype, a fundamental problem in biology. **At the genome level biology is essentially digital, and genetic sequence information is translated into dazzlingly complex interacting networks of genes, proteins, and metabolites, making up cellular function**. Cells organize into tissues, which, in turn form the whole plant. Functioning of the cellular networks is directly influenced by features of the environment the plant finds itself in, such as climate, resource availability, and microbial communities.

Environment: In order to sustain ecosystem functions in the face of rapid change, we need to be able to monitor multiple heterogeneous variables spanning a range of temporal and spatial scales. The **vast amount of data so collected needs to be integrated and used to construct unifying mathematical models that help guide environmental policy, and have the predictive capability to assess consequences of informed intervention**. Here too, the models need to

3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA
Phone +1-215-382-9800
Fax +1-215-386-7999
www.siam.org • siam@siam.org

integrate interconnected networks and systems of complex systems at vastly different scales, all affected by a common environment.

Energy: In order to expand sustainable alternatives to fossil fuels, new approaches beyond ethanol derived from corn must be developed. Microbial biocatalysis is a promising direction. In order to make it a reality, solving the genotype-phenotype problem will lead to the capability to engineer microbes from standard DNA modules that perform a specified metabolic function. Another promising approach is to engineer plants with molecular networks that produce more leaves and fruit without using additional fertilizer, thereby increasing energy production through photosynthesis. **With predictive models of the intertwined gene, protein, and metabolic networks, it becomes feasible to engineer and optimize the organism for efficient biofuel production**.

Health: To make a transformational contribution to human health, solution of the genotype-phenotype problem will contribute to **integrating genomics information with complex genetic, protein, and metabolic networks, on up to the tissue and organism levels, all of which react to the external environment**. In fact, environmental influences are known to play a very important role in several important diseases, such as cancer and neurological disorders.

The importance of developing better modeling, computational, statistical, and analytical tools to enable a better understanding of biological systems and detailed discussion of the potential impact and key problems are also described in the 2005 National Research Council report "Mathematics and 21st Century Biology."

COMMON THEMES

Three common themes emerge from the challenges described in the report.

1. **All four challenges require the construction and analysis of predictive mathematical models of large, nonlinear dynamic networks that span several spatial and temporal scales.** Understanding and manipulating these systems will require large, multi-scale, nonlinear, and hybrid models. Existing simulation and analysis tools for such models are in their infancy, or nonexistent in some cases. For instance, an increasingly popular modeling paradigm for complex networks in fields ranging from molecular biology to ecology is agent-based modeling, which captures the important feature of many complex systems that global behavior emerges from local interactions. Very few analysis tools exist for such models. For many applications it is desirable to use models to predict how interventions on one level will impact biological systems on other levels, such as in the development of therapeutics. This process requires control approaches, but for the systems at the heart of the New Biology challenge areas, it is sometimes difficult or impossible to apply existing control theoretic approaches.

2. In all problem areas **high performance computation will play a crucial role, from simulating complex multi-scale models to analyzing sequence data**, e.g., multiple sequence alignment. **This will require new breakthroughs in algorithm development**, since we cannot

3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA
Phone +1-215-382-9800
Fax +1-215-386-7999
www.siam.org • siam@siam.org

expect significant increases in clock speed due to silicon technology. Performance improvements in computation will come from more cores on a chip. This means significant changes in algorithms to take advantage of parallelism on the chip as well as parallelism between computational nodes comprised of multiple chips. In order to achieve high rates of performance, algorithms that minimize data movement, possibly at the expense of doing additional computations, will be the most efficient. Algorithm developers will need to take these facts into account as they develop multi-scale, multi-physics algorithms.

3. In all four challenge areas we face **ever-growing data volumes**, from DNA sequence data to satellite surveillance data. These data need to be stored in databases that are easily accessible and searchable, requiring increasingly sophisticated and scalable data mining algorithms. In addition, the data from heterogeneous sources need to be integrated, within databases as well as within models. Once accessible in databases, the typically high dimensional data sets need to be analyzed using statistical methods. In order to meet these challenges, new tools from multivariate statistics and discrete mathematics are needed, in particular graph theory and combinatorics.

BIOLOGY TO INFORM MATHEMATICAL RESEARCH

As happened with physics in the last century, we can expect that an increasingly strong feedback loop will develop between biology and the computational disciplines that now serve as tools, such as mathematics, statistics, computer science, and engineering. For instance, the National Science Foundation is already capitalizing on this feedback with its program “Quantum and Biologically Inspired Computing.” We mention here two more examples.

It is well appreciated that the human immune system has important lessons to teach us about computer security. But the immune system is also a vast distributed information-processing network. Once we understand its design principles well enough to build mathematical models capturing its key capabilities we will be able to transfer these principles to engineered networks. The immune system’s complexity and the multiple spatial and temporal scales involved offer several mathematical and computational challenges that can only be overcome by fundamental breakthroughs in these fields.

As another example, it is observed frequently by experimentalists that after engineering an organism with a gene deletion, even an apparently essential one, its phenotype remains unchanged. That is, the organism is robust to many such changes and can remodel its molecular networks after a change in its genome to maintain function. The underlying fundamental problem of understanding the genotype-phenotype relationship is mirrored by the analogous mathematical problem, namely understanding the relationship between the structure of a dynamical system and its resulting dynamics. This problem is still largely unsolved and poorly understood. Biological insights about the sources of this robustness in organisms can help generate hypotheses about solutions to the corresponding mathematical problem.

3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA
Phone +1-215-382-9800
Fax +1-215-386-7999
www.siam.org • siam@siam.org

RECOMMENDATIONS — RESEARCH AREAS

This analysis makes clear that mathematics is indeed an important enabling technology for the New Biology. **We recommend that any funding programs related to the New Biology initiative provide support for mathematical research related to the problems identified above in the following areas:**

1. Complex networks, both in the graph-theoretic sense and in the dynamical systems sense.
2. Multi-scale modeling and simulation, including computational science research to enable new approaches.
3. Systems of partial differential equations.
4. Algorithms for high performance computation.
5. Algorithms for new multi-core computer architectures.
6. Multivariate statistics.
7. Dynamical systems.
8. Hybrid models.
9. Control theory.
10. Combinatorics and graph theory.
11. Data mining algorithms.
12. New methodologies for modeling complex stochastic biological systems.
13. Quantification of model uncertainty.

In addition to research in these areas, it is becoming increasingly clear that there is much untapped potential in mathematical fields that are not traditionally considered as applied. Good examples are recent applications of algebraic geometry to biological problems and the use of methods from algebraic topology for high dimensional data analysis. (Within SIAM, recognition of these emerging opportunities has led to the establishment of a new SIAM Activity Group in Algebraic Geometry.)

RECOMMENDATIONS – RESEARCH SUPPORT MECHANISMS

To support the research areas outlined above, programs at individual agencies and interagency initiatives will be needed. Specifically, **an array of complementary approaches will be needed – from those that focus on building expertise in a single topic area, often at a single agency, to application-driven programs that combine mission agency's user communities and discipline-organized research programs.** Agencies likely to have relevant expertise, communities, programs, and missions include: the National Science Foundation (NSF), the National Institutes of Health (NIH), the Department of Energy (DOE), the U.S. Department of Agriculture (USDA), the Department of Defense (DOD), the Environmental Protection Agency (EPA), the Department of Homeland Security (DHS), and others.

There are a number of existing programs that effectively support research at the interface of mathematics and the life sciences. These programs could be expanded or used as models for the establishment of new programs. Examples of existing programs include:

3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA
Phone +1-215-382-9800
Fax +1-215-386-7999
www.siam.org • siam@siam.org

- NSF-NIH collaborations, such as the long-standing NSF Division of Mathematical Sciences (DMS) program with the NIH National Institute of General Medical Sciences on applications of mathematics to biomedicine and the new NIH-NSF programs at the Interface of the Life and Physical Sciences.
- NSF-DOD collaborations, such as the recently-established NSF DMS program together with the Defense Threat Reduction Agency to develop the next generation of mathematical and statistical algorithms and methodologies in sensor systems for the detection of chemical and biological materials, and the NSF program under development with the U.S. Army to develop mathematical algorithms to integrate and analyze heterogeneous battlefield sensor data.

Mechanisms should be available to support a variety of sizes of research projects, from individual investigators to center-scale collaborations. Examples of multi-agency and single-agency center-scale initiatives in this area include:

- The National Institute for Mathematical and Biological Synthesis (NIMBioS), jointly supported by the NSF Biological Sciences Directorate and DMS, together with USDA and DHS.
- NSF DMS also supports the Mathematical Biosciences Institute (MBI) at the Ohio State University.

Both institutes focus on research at the interface between mathematics and biology and foster interactions between mathematical scientists and bioscientists.

RECOMMENDATIONS – WORKFORCE DEVELOPMENT, NETWORKING, AND OUTREACH

Mathematical scientists and statisticians with appropriate awareness of the interdisciplinary research questions central to the New Biology will be critical to moving forward in the challenge areas. Programs should be implemented to ensure this pipeline of mathematicians and statisticians at the undergraduate, graduate, post-doctoral, and early career levels. The New Biology report stresses the need to train scientists with highly developed quantitative skills and interdisciplinary experience. While the report mainly focuses on programs to train biologists, it will be also be critical to train mathematical scientists and scientists in other disciplines that can contribute to advancing the New Biology.

For example, **mathematicians would benefit from a new program that supported university efforts to develop curricula and programs focused on horizontal integration of mathematics training with other disciplines.**

In addition to programs that support research activities, federal agencies should focus on raising awareness in the biological and mathematical communities about science at the interface and facilitating cross-disciplinary collaborations, as creating research teams and partnerships across disciplines takes more time and conversation than building teams of people who are within a discipline and share a common culture. In addition, outreach within each community about interesting results in one discipline that may potentially be relevant to problems in the other

3600 Market Street, 6th Floor
Philadelphia, PA 19104-2688 USA
Phone +1-215-382-9800
Fax +1-215-386-7999
www.siam.org • siam@siam.org

discipline could have a significant impact (i.e. the discovery of applications of algebraic geometry to biological problems mentioned above). Such unexpected linkages can bring very high returns, and their development should be systematically fostered and supported.

To accomplish the above goals, programs that support network creation, workshops, travel, and summer programs, would be useful. "Sabbatical" cross-disciplinary opportunities for researchers, post-doctoral students, and graduate students also might be effective in creating a new community of researchers more alert to and equipped to conduct interdisciplinary research.

IMPACT BEYOND THE NEW BIOLOGY

The **mathematical research needed to enable progress on the New Biology challenges will at the same time benefit several other fields, as described earlier, in particular other complex networks such as the power grid and computer networks.** Therefore, increased support for the mathematics areas listed above will have a substantially broader impact. Furthermore, understanding design principles of complex biological networks can provide important insights about the possible design of engineered systems, especially if this understanding can be quantified in mathematical language. For instance, gene regulatory networks have to be very robust to many extracellular perturbations, but at the same time have to be very sensitive and respond very quickly to certain changes in the extracellular environment. Understanding the basic engineering and physical principles underlying this network design could have benefits in designing a variety of engineered control systems.

SIAM thanks you for your consideration of these recommendations. We look forward to working with you to help define and implement the mathematics and information sciences programs needed to take full advantage of research at the mathematics/biological sciences interface and move us toward solving societal challenges such as those outlined in the New Biology report.