SIAM
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

December 6, 2011

MATHEMATICS: AN ENABLING TECHNOLOGY FOR THE 21st CENTURY BIOECONOMY

Society for Industrial and Applied Mathematics Response to Request for Information on the Bioeconomy Blueprint

In support of the Obama Administration's development of a Bioeconomy Blueprint, the Society for Industrial and Applied Mathematics (SIAM) offers below recommendations to promote problem-centric, interdisciplinary research to solve societal challenges in health, energy, environment, and agriculture.

Research is not only part of the bioeconomy, it is the very foundation from which the bioeconomy must build its success, and the Bioeconomy Blueprint should incorporate an aggressive and robust research portfolio to bolster expansion efforts.

As stated in the 2009 National Academy of Sciences report "A New Biology for the 21st Century: Ensuring the United States Leads the Coming Biology Revolution," new information technologies and sciences will be essential to creating a bioeconomy that can tackle societal challenges. The SIAM recommendations identify specific opportunities for the federal government to strengthen research and education at the interface between the life sciences and physical sciences, mathematics, and engineering.

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.

Grand Challenges (Q1)

Q1: Identify one or more grand challenges for the bioeconomy in areas such as health, energy, the environment, and agriculture, and suggest concrete steps that would need to be taken by the Federal government, companies, nonprofit organizations, foundations, and other stakeholders to achieve this goal.

As biology develops as a predictive science, new approaches to information analysis, data, and modeling will be needed to advance our understanding of the natural world in each societal challenge area.

Health: To make a transformational contribution to human health, an understanding of the genotype-phenotype problem, that is, the links between the genotype and phenotype of an organism is essential. Systemic diseases such as cancer are so challenging because they involve processes from the genome level to molecular networks inside a single cell, the tissue level and, finally, the entire organism, **all of which react to the external environment in a coherent fashion**. In fact, environmental influences are known to play a very important role in several disease processes. At the organ scale, recent advances in modeling the dynamics of blood flow in the heart and its connection to intracellular events

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

provide unprecedented new tools for understanding heart disease and its effective treatment.

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, determining the link between genotype and phenotype 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 organisms for efficient biofuel production**.

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

Agriculture: 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. Beyond the individual plant level, modeling will be necessary to be able to predict and control the ecology and spread of crop disease, invasive species, and other agricultural threats.

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 named in the RFI.

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

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

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 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.

Research and Development (Q2-3)

Q2: Constrained Federal budgets require a focus on high-impact research and innovation opportunities. With this in mind, what should be the Federal funding priorities in research, technologies, and infrastructure to provide the foundation for the bioeconomy?

The three themes, described above, that are common to all four challenge areas make clear that mathematics is indeed an important enabling technology for the bioeconomy. **We recommend that any funding programs related to the Bioeconomy Blueprint 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. Nonlinear partial differential equations.
4. Algorithms for high performance computation.
5. Algorithms for new multi-core computer architectures.

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

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.
14. Numerical analysis.

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.)

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:

- 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:

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

- 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.

Q3: What are the critical technical challenges that prevent high throughput approaches from accelerating bioeconomy-related research? What specific research priorities could address those challenges? Are there particular goals that the research community and industry could rally behind (e.g., NIH $1,000 genome initiative)?

To realize the potential of high-throughput molecular approaches, **new computational algorithms as well as new theoretical mathematical and statistical approaches are needed to extract patterns from large volumes of high-dimensional data.** It used to be the case in molecular biology that the expense and difficulty of data generation far outstripped the expense of data analysis. This situation is now reversed. Data analysis now typically consumes as much as 75 percent of project resources. New data generation technologies, such as next generation sequencing, tilt this balance even further in the direction of data analysis. The quantity and complexity of new high throughput data types poses serious challenges to the extraction of information and, ultimately, knowledge from molecular data.

While new data analysis approaches are critical to overcoming the high throughput data deluge, it is also important that centers producing data are equipped and willing to shift to these new approaches. Too often genomics centers lack the knowledge of how to use new techniques and the willingness to integrate or replace current practices. Federal programs should encourage the transfer of research on data analysis to genomics and other high throughput centers.

Workforce development (Q9)

Q9: The majority of doctorate recipients will accept jobs outside of academia. What modifications should be made to professional training programs to better prepare scientists and engineers for private-sector bioeconomy jobs?

Training for mathematicians, scientists, and engineers should explicitly foster skills needed for the bioeconomy workforce. These skills include the ability to work in diverse teams that straddle expertise areas and disciplines, innovative thinking oriented around solving real-world problems, and communication with non-scientists. Federal training programs should encourage university efforts to foster these skills. The NSF Integrative Graduate Education and Research Traineeship Program is an example of a federal graduate training program that encourages mentorship, career development, hands-on experience with innovation and translating research discoveries to solutions for societal challenges. These **best practices should be expanded beyond the frontier interdisciplinary programs that IGERT supports to graduate training across all fields relevant for the bioeconomy.**

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

As one example of a skill needed for the bioeconomy workforce, mathematical scientists and statisticians in both academia and industry will need appropriate awareness of the interdisciplinary research questions central to the bioeconomy challenge areas. Programs should be implemented to ensure this pipeline of mathematicians and statisticians at the undergraduate, graduate, post-doctoral, and early career levels. It will also be critical to train biological scientists with highly developed quantitative skills and interdisciplinary experience.

For example, **mathematicians would benefit from a new program that supports 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 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 would help create a new community of researchers more alert to and equipped to conduct interdisciplinary research. The new NSF Science, Engineering, and Education for Sustainability Fellows is an example program that supports this type of activity.

In order to train students for jobs outside of academia, Federal agencies should also encourage increased collaboration between industry and universities. **Fellowships that allow students to spend part of their graduate careers working in industry or other sectors help create networks between academia and industry, foster real-world learning, and provide students with greater understanding of workforce opportunities beyond the lab. Collaborative research with industry and orienting research towards grand challenges defined in concert with industry help foster student awareness of industry challenges and skills.** The NSF Grant Opportunities for Academic Liaison with Industry (GOALI) program is an example of Federal funding that catalyzes industry student exchanges and research collaborations.

Conclusion

SIAM thanks you for your consideration of these recommendations. We look forward to working with the Administration 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 Bioeconomy Blueprint RFI.