RESEARCH INTERESTS

In the last years the scope of theoretical physics has grown enormously, and its methods have led to successes in a variety of fields including, in particular, biological physics. While some of my interests belong to more traditional areas of soft condensed matter theory, it is the field of theoretical biological physics which is the most interesting to me, and the one which I will focus on here.

The biophysics research program that I have started as a graduate student at Princeton, am now developing as a post–doctoral scientist, and intend to continue as a faculty member is centered on an interface between theoretical physics, information theory, computer science, statistics, and biology. To this interface theoretical physics brings its way of thinking and the reductionist's worldview, its powerful mathematical methods, and the desire to create universal, deductive explanations of the observed phenomena. Biology asks the questions. And the other fields provide the language, the frameworks, and the toolkits that let physics answer these questions.

A critical ingredient of this program is working in an environment where physicists and mathematicians work with biologists as a single, synergistic, well–interacting team. On the other hand, I believe that my research goals and my long term scientific career are better served by being at a physics department. This is precisely why a position at the Biocomplexity Institute, which will allow for interdisciplinary collaborations, and an appointment at the Department of Physics are very appealing to me.

Let me describe my research interests in some detail to illuminate both of these points.

Roughly, there are three major branches of biological physics: molecular biophysics, systems biology (including systems neuroscience), and population dynamics (which may include evolutionary processes as well). Many theoretical biology programs focus mainly on the first and sometimes on the last of these branches, leaving the second one behind. However, to me biology is interesting precisely because of how all the parts work together as a single coherent organism. So my current and future research is and will be focused mostly on quantitative understanding of systems' behaviors. In additions to the standard physical methods of statistical mechanics, the tools for doing this come from the fields of information and learning theories and Bayesian statistics.

Treating a biological system as an information processing device, one may try to pose questions about *what* the system is doing, the necessary precursor to the *how* question usually asked by molecular biology. One can ask questions about the optimality of the design, see if systems perform well in an absolute sense, and search for optimization design principles in biology. Among other things, this will lead to understanding which features of biological systems are accidental, and which ones are constrained uniquely by requirements of optimal performance for some important natural tasks.

An advantage of the informatics approach is the ability to stay model and detail independent. Thus we can use the same tools to analyze neuronal, biochemical, and genetic networks. My previous work has focused on the neuronal side. There we have answered and are still answering some interesting questions about the neural code, adaptation, brain's generalization and learning abilities, and the reason for the topographical map layout of the cortex. Now I also devote time to studying biochemical and transcription regulatory networks. Here we have made some progress in analyzing the properties and purpose of noise in biochemical networks and in understanding the systematic structure of genetic regulatory pathways in yeasts using gene expression microarrays data. If hired by the Biocomplexity Institute and the Department of Physics, I will continue further research in these fields.

From the previous paragraphs it is clear that my biological physics interests define and are strongly dependent on the research I do in statistical inference and information theory. Here I have worked on learning in nonparameteric and severely undersampled parametric models in the framework of Bayesian inference, which is well suited to be analyzed by methods of Statistical Mechanics. We have generalized Bayesian model selection to the nonparameteric case using techniques of Quantum Field Theory and developed what is now considered to be the state of the art estimator of entropies of undersampled probability distributions. Further, I have participated in establishing of and will continue to develop the information theoretic approach to feature selection: compressing data while preserving the relevant information in them. These techniques are indispensable for many theoretical biology applications, such as genetic regulatory network inference, or better understanding the neural code. They also lead to surprising insights into seemingly unrelated problems like explaining critical periods in learning in animals. While this information theory research has been quite successful, there are still many biologically and physically interesting problems left for my next appointment.

The list of possible directions in my research program is much longer that I can detail here. In addition to the most exciting branches described above, I am also interested in, for example, related questions that arise from analysis of complex systems by means of the information theoretic approach based on identifying predictive features of data, which was co–developed by me. Further, I participate in theoretical analysis of magnetostatic properties of type II superconductors, which is of great relevance to the Gravity Probe B and the Satellite Test of Equivalence Principle experiments. Finally, recently I got involved in studying pattern formation and diffusive transport in a dense medium (e. g., inside a cell) by means of the field theory approach to stochastic particle dynamics.

It is important to emphasize once again that, I believe, substantial progress with many of these problems is possible only in a multidisciplinary, yet physically sophisticated environment. I hope to find just such an environment at the Department of Physics and the Biocomplexity Institute at the Indiana University

TEACHING INTERESTS

I have enjoyed several opportunities to teach in my career. These include short lecture series on Statistical Inference at NYU and UCSB, which required development of original curricula, three summers as a faculty member for the Methods in Computational Neuroscience summer course at Woods Hole Marine Biological Laboratory, experience as a Teaching Assistant and a laboratory instructor at SFSU and Princeton, and advising undergraduate and junior graduate students in research collaborations.

I genuinely like teaching, and I am open to and excited about teaching standard courses in the physics and theoretical physics curriculum based on the Department's requirements and needs. However, having worked in the areas of theoretical biophysics, information theory, and machine learning, I will be most interested in teaching opportunities for advanced or special–topics courses in biophysics, statistical learning, and related fields. While targeted to senior undergraduate or junior graduate physics students, these courses will also be useful to mathematicians, computer scientists, biologists—all those who wants to use quantitative methods to understand the design and function of biological systems. The following topics for courses seem to be of great overlap with my research experiences.

INFORMATION THEORY AND STATISTICAL INFERENCE

Many computer science and statistics departments now offer such courses with the aim of introducing the students to new data analysis methods emerging from machine learning. *The Elements of Statistical Learning* by T. Hastie et al. is a usual textbook. While hands–on experience with different algorithms is important, this is not the goal of my envisioned course. Instead I would like to make sure that the students in this undergraduate–level course will see the field not as seemingly unrelated computational techniques, but as a coherent subject, where similarities of the methods are emphasized, and the general requirements for any learning to be successful are studied. Thus a textbook *Information Theory, Inference, and Learning Algorithms* by D. J. C. MacKay will also be used, as well as my own notes from the short courses I developed and presented at NYU and UCSB.

INFORMATION THEORY, LEARNING, AND COGNITION

This course will have some overlap with the above; however, its main goal will not be the learning theory itself, but rather its application to understanding how we think. I will build this course based on the lectures *Thinking about the brain* by W. Bialek (Les Houches Summer School, 2001), and on my own notes on the subject. As models needed to quantify the phenomena may involve elaborate mathematics, the course will be aimed at graduate students, but it should be accessible to advanced undergraduates as well.

INTRODUCTION TO BIOPHYSICAL MODELING

There are many areas in biology where simple quantitative models provide clear answers to biologically important questions. These include bacterial chemotaxis, protein folding, problems in adhesion, polymer and membrane dynamics, protein–DNA interactions, pattern formation, biological scalings, population dynamics, and many others. Currently, there are no textbooks covering such a wide range of modern topics, though, for example, *Biological Physics: Energy, Information, Life* by P. Nelson is very helpful. Thus the course will be a graduate level theoretical and computational course based on original research articles. Since a crucial aspect of a graduate education is the transition from a student to an active researcher, the course will be of a seminar type, with many topics presented by the students themselves.

INTRODUCTION TO COMPUTATIONAL NEUROSCIENCE

Computational Neuroscience is arguably the branch of biology in which quantitative methods have been the most successful in asking and answering the right questions. A course that introduces students to standard tools in neurophysiological modeling, but is mostly focused on the understanding of physical, mathematical, and experimental foundations behind such models would be of great benefit not only to students interested in neuroscience but to students interested in quantitative biology in general. Excellent textbooks such as *Theoretical Neuroscience*, P. Dayan and L. Abbott and *Spikes* by Rieke et al. are available to guide the presentation.