I have always wanted to understand how things work. My favorite books as a child dissected cars, computers, the human body, page by page, explaining each component with vivid illustrations. Today, I continue this habit by collaborating with experts in a wide range of domains, from molecular biology to political science to mechanical engineering. More than ever before, interdisciplinary research is essential to solving the toughest problems in science. Researchers with different backgrounds have remarkably different perspectives on similar problems, and techniques from one field can shed light on interesting solutions in other fields. My goal as a researcher is to ***leverage my mathematical understanding and computational skills to solve problems in specific domains***. I have applied these skills in fields as diverse as worker's compensation insurance, human interface design, nuclear physics, and evolutionary biology, and have learned an incredible amount in the process.

**High School** Starting after tenth grade, I spent my summers and many of my afternoons in the Romalis lab at the Princeton University physics department. I developed applications and improvements for the highly sensitive atomic magnetometers the group studied.

I spent the first summer in the lab helping develop a magnetic encephalography (MEG) system to measure brain activity. I significantly improved the efficiency of the analysis software, which substantially sped up the group's progress on the research.

I also spent two summers and a school year designing and building optical multi-pass cells that increased the sensitivity of the magnetometers. This work showed me that I could play an important role in experimental research by developing software that made it easier to design physical systems. I found it very motivating to turn the abstractions represented by my software into a working experimental setup that was useful for real-world applications.

Intellectual Merit I presented my work on multi-pass cells at an APS meeting in Houston in 2010. I published my findings in Physical Review A (Li, Vachaspati, et al., 2011), which has been cited 13 times, including in high-impact journals such as Physical Review Letters.

The software I developed to optimize parameters for multi-pass cells continues to be used in the Romalis lab, and is useful even with recent advances in the fabrication of multi-pass cells.

Broader Impact Atomic magnetometers have a number of advantages over the SQUIDs, which are the most commonly used high-sensitivity magnetometers. They have the potential to be cheaper, less prone to failure, and more accurate. Multi-pass technology allows smaller, more sensitive sensors useful for geological sampling, flow rate sensing, and fundamental physics research, among other applications.

**Undergraduate** Largely because of my positive experience with research in high school, I decided to pursue physics for my undergraduate degree at MIT. At the same time, I took a lot of computer science classes and worked on software-related jobs and internships.

Therefore, I wanted to find a research project that answered questions about physics, but relied heavily on computing as a tool. Professor Will Detmold's research in the Center for Theoretical Physics developing algorithms for lattice quantum chromodynamics (QCD) was exactly what I was looking for. Here, I developed software that substantially decreased the time needed to perform certain calculations necessary for investigating systems containing large numbers of quarks.

I learned a lot about physics and computational science here, and I learned two important lessons about doing research.

First, I learned that I don't have to be an expert in everything about my research. I hadn't taken any quantum field theory classes, which would have provided the theoretical foundation for much of my work. Instead, I got things done by focusing on a project that played to my strengths in algorithm analysis and software design.

Second, I began to understand the level of tenacity needed to succeed in academia. I spent several months developing algorithms and software that I hoped would use an elegant theory to speed up a large class of computations. Eventually, though, I realized that there were fundamental problems with my work, and the theory would not work. Although I was discouraged, I pushed on, and I discovered that I could use clever numerics to speed up a different part of the problem I was studying. I was disappointed that I had to discard the theory that I had spent so much effort developing, but it was necessary to achieve practical results.

<u>Intellectual Merit</u> I published these findings in my undergraduate thesis, and presented at the LATTICE 2014 meeting at Columbia University. The techniques I developed are still being improved and used for analyses of large datasets by the group I was in.

<u>Broader Impact</u> QCD is the study of the strong nuclear force, and lattice QCD is the only way to understand this at low energies. My research provides a method to analyze systems consisting of several particles, which is crucial for understanding the nuclei of larger atoms. Of course, QCD is critical for our understanding of fundamental physics, and it also has applications to astrophysics, fusion power, and cosmology.

While at MIT, I designed and taught summer classes in computer science for high schoolers as part of the SPLASH program. In the summer of 2013, I taught an introductory computer architecture class to 20 high school students. The class covered the theory of computer architecture and had a lab component where students designed and built logic circuits. It was gratifying to be able to share some of the knowledge I had gained, and to be able to expose the students to the same sense of wonder and understanding that I experienced when I first encountered these topics.

In 2014, I co-taught a course that introduced theoretical computer science to 15 high school and middle school students. We adopted a pedagogical approach that presented problems to the students and encouraged them to find solutions in groups and share them with classmates. Since the students had a variety of backgrounds and interests, this strategy encouraged the type of collaborative and interdisciplinary behavior that I find so important in my own work.

**Graduate** I knew I wanted to go to grad school to pursue a career in academia, and UIUC's program in computer science appealed to me because of the strong tradition of interdisciplinary collaboration at the university. After about a semester of searching, I decided to work with Professor Tandy Warnow, studying the development of methods for phylogenetic inference. Given DNA sequences from a number of organisms, the goal is to reconstruct the species tree that represents their evolutionary history. This is an extremely interdisciplinary topic, requiring a combination of traditional algorithm development, statistics and machine learning, high performance computing, and understanding of the underlying biological processes. Recent increases in genetic sequencing capability have made the development of fast, accurate methods critical for biological analyses.

I recently led the development of a software package called ASTRID [1], which is one of the few scalable methods that can accurately estimate species tree based on evolutionary trees for individual genes. ASTRID's accuracy is similar to ASTRAL [2], which is the most accurate existing algorithm, and it is substantially faster than ASTRAL. We re-analyzed a genome-scale dataset [3] with 14,446 gene trees for 48 bird species, using ASTRID, in just 3.5 minutes. In comparison, ASTRAL took 34 hours to analyze the same dataset. ASTRID found a slightly different tree than ASTRAL, and on simulated datasets based on this dataset, ASTRID had higher accuracy.

I am working on a number of ongoing projects, developing methods and analyzing biological datasets. ASTRAL and ASTRID are the two fastest and most accurate species tree estimation methods, and I am working on ways to improve their accuracy and performance. ASTRAL is a constrained-search algorithm that determines its search space based on its input, so I am developing techniques to improve the search space selection, especially in difficult cases when gene trees are missing data. Many biological datasets have this characteristic, so it is critically important for ASTRAL to perform well in these cases. I am also developing extensions to ASTRAL's constrained search algorithm to allow for a broad range of optimization functions. This will expand this technique beyond species tree estimation to a wider range of phylogenetic estimation tasks that do not currently have efficient and accurate solvers.

Furthermore, I am collaborating with a number of biological groups at UIUC and elsewhere to analyze data and develop solutions for new datasets. I am using tools I have developed, including ASTRID, to build trees for the 1kp (Thousand Plant Transcriptome) project (onekp.com), as well as the Hemipteroid Insect Phylogeny project. I am also working with a group studying the origins of the hepatitis B virus, which provides insights into treatment as well as into historical human migration patterns.

**Outreach** An important part of my research is working with biologists to discover their needs, as well as guiding them in using the most useful methods to solve their problems. I have taught workshops about phylogenetic estimation to several groups of biologists, which have resulted in increased adoption of fast and accurate methods. Interactions like these are critical for the field not only because they help make sure that biological analyses use the right tools, but also because they provide insight into the needs of biologists so I can tailor my tools to their needs.

**Future Goals** I plan to continue an academic career in method development for bioinformatics and genomics. I don't know if I will remain in phylogenetics, but I am confident that the skills I am developing will help me solve a wide range of problems. I will keep working in interdisciplinary collaborations, seeking opportunities where I can most effectively leverage my skills to provide the greatest scientific and social benefit. I plan to expand my outreach activities to undergraduate and junior graduate students over the next three years in preparation for this career, as mentoring students will be critical to my success as a professor and researcher. With the support of the NSF, I will be able to make lasting contributions to computer science, biology, and to the lives of the students I teach and mentor.

[1] P. Vachaspati, T. Warnow, *BMC Genomics* **16**, S3 (2015).
[2] S. Mirarab, T. Warnow, *Bioinformatics* **31**, i44 (2015).
[3] E. D. Jarvis, *et al.*, *Science* **346**, 1320 (2014).