Thomas M. Boudreaux - Personal Statement

Inspiration — My first research experience took place during the spring semester of my senior year in high school. I reached out to Dr. Barlow, an astrophysics professor at the university I was set to attend that fall, High Point University (HPU), and asked if I could work with him on a project as a way to fulfill my high school's senior-year intern experience. He graciously allowed me to join him for two nights at UNC Chapel Hill working in a "remote-observing room," where we operated a set of instruments mounted on a telescope in Chile. I remember, around 2am on the first night, Dr. Barlow turned too me and said, "I hope I haven't scared you off with all this stuff." The "stuff" to which he was referring was the flood of astronomy information — a crash course in extreme horizontal branch evolution of stars — he had been explaining all night. Back then, when I was exhausted, and a little intimidated, I didn't really know what to say. I told him that I wasn't scared off, and thanked him again for allowing me to tag along. Now, nearly four years later, I wish I could go back and tell him that instead of scaring me off **he made astronomy real in a way it had never been, and vastly more alluring to a younger me.**

Learning Tools — I began serious research my first semester of college. For my first project, I used archival data my advisor had collected of a particular type of binary star. The data had to be passed through a reduction pipeline to a separate analysis pipeline which would allow us to extract data on the orbital velocity of the stars in the system. I built scripts using the Python I was teaching myself to glue together different parts of these pipelines and move data between them. **Scripts were an astonishingly valuable tool I learned through this research. My subsequent work has incorporated use of progressively more complex scripts.** I eventually presented this work at the 227th Meeting of the American Astronomical Society (AAS). A portion of the data we collected in the process of this work is included in a paper which has recently been submitted for publication.

While at the AAS, I met a student named Micheal Tucker. Micheal's work was conducted as an intern at the Space Telescope Science Institute (STScI) using a software suite called *gPhoton* to look for new pulsating white dwarfs. The method Micheal presented seemed applicable to the study of hot subdwarf B stars (sdBs), the helium fusing remains of stripped red giants and the main focus of the HPU astrophysics research group. With this in mind, my advisor recommended that I apply to STScI, and I did. In my application, I specifically proposed to extend Micheal's work to hot subdwarf stars, which can exhibit pulsations similar in nature to white dwarfs. I was accepted to the program, and over the following year, my REU advisor and I investigated the feasibility of using *gPhoton* to identify and study pulsating hot subdwarf B stars. **I was able to take on this project due to the skills I had developed working with Python in my previous project**

Learning Data — The organizers of the STScI summer program placed an emphasis on students leading the research, so my REU advisor gave me generous leeway to work on the project as I saw fit, acting as a guiding hand. I started by using the skills I had established during my first project to write a set of tools; these tools would read a list of all known sdBs and output files in the form *gPhoton* expected. I then automated the process of calling *gPhoton* on these files. **This was my first experience with very large data sets.** Once *gPhoton* had returned all of the light curves from the GALEX mission database, I wrote a series of data visualization and analysis tools to analyze these data. This work took all summer, so when the program at STScI concluded, we did not have many firm results to present.

I continued the work I had started at STScI at my home institution. We used the large number of data points to make statistical statements regarding the usefulness of *gPhoton* for this line of work. While we did manage to discover one new pulsating star and rediscover five already known pulsating stars, we eventually determined that, because of the extremely high noise band, *gPhoton* was not a tool that made sense to use when looking for new pulsating sdBs. While we were still finalizing our results, my advisor asked me to author a paper on our findings. **I led the team that wrote and published the peer reviewed paper, a process through which I learned the skills necessary to take research results and convey them to the rest of the community** (Boudreaux et. al. 2017).

Learning Research — During the GALEX project, I became frustrated with working with large data sets. Sifting through tens of thousands of light curves seemed to me an inefficient use of time. Instead, I went to my advisor and asked if I could investigate the uses of deep neural networks for the rapid and automated classification of pulsating sdBs — the same fundamental task as my previous research, but now focused on automating the process. **This project would, if successful, significantly reduce the observational costs typically associated with the discoveries of variable stars** — possibly in fields far divorced from pulsating sdBs. Despite my advisor having little exposure to much of what I would be doing, he agreed that, if I were willing to take the lead, I could move forward with this project. Using resources available online, I worked through the derivations of forward and backward propagation, the mathematical underpinnings of deep learning, before moving to neural network tutorials. Over the span of a month and a half, I taught myself the basics of building and training neural networks. From there, I learned how to build, train, and apply neural network based models to light-curves. I developed tools from scratch to generate synthetic training data for the models. After generating the data, I trained models, tested them, compared their results against other neural architectures, and honed in on some of the best performing sets of models. This was the first project for which I took an initial concept and drove it all the way through to execution and results. I found I had such an affinity for working with neural networks, I developed a individualized major in computational physics that was approved by the administration. This major combines all of the normal physics degree coursework along with large portions of both a computer science and a math degree.

I presented my work on deep neural networks at a conference on sdBs in Kraków, Poland the summer leading into my junior year. While at the conference, I met experts whose papers I had been reading for the last two years, and felt once again that my choice to pursue astronomy was the correct one. One of the people I met at the conference was a postdoctoral fellow at CalTech, Thomas Kupfer. Thomas had very similar research interests to those of our research group, and he was intrigued by the preliminary results I presented. Up until that point, I had only worked with synthetic data I generated. Thomas offered me the chance to work with vast troves of archival data from the Intermediate Palomar Transient Factory (iPTF). In March of 2018, I travelled to Pasadena for a week-long stint as a visiting researcher to apply my deep learning models to those data.

After the Kraków conference, I had the opportunity to write a paper for a special issue of a journal organized by the conference directors. I wrote and submitted my single-author article to the journal and after two rounds of referee reports, my paper was published in *Open Astronomy* (Boudreaux 2017). **This paper represented one of the earliest demonstrated applications of deep learning to time-series data classification in astronomy.**

Current Research — Near the end of my junior year, I applied and was accepted into a summer REU program at the Harvard Smithsonian Astrophysical Observatory (SAO), where I spent ten weeks in Cambridge working with an SAO staff researcher.

Over the summer, I applied the programming and computer science knowledge I had developed from my past research projects to an entirely new domain: globular cluster evolution. With the help of my REU advisor, I ran n-body simulations of small, dense globular clusters, studying how the initial fraction of binary stars in the cluster affects expansion over the course of its life. There were challenges associated with getting the time we needed on the high-performance computer cluster at Harvard; however, we did eventually manage to run a limited number of simulations. When I returned to my home institution to start my senior year, I continued this research, and now I am now developing an analytic and observational framework to understand the simulation-based results from over the summer. Additionally, the challenges my REU advisor and I faced getting time on the high performance computer cluster inspire my proposal to search for ways to make the study of n-body systems more accessible to researchers with less powerful computer hardware, specifically study the applications of neural networks to n-body systems.

The skills and experiences I have acquired over the course of my research have progressively inspired me to propose a research project using deep-learning to time-evolve n-body systems. I first developed the idea for this project when I was working to classify sdBs with neural networks. Working at the SAO filled in the final set of skills I would need to accomplish this goal.

Broader Impact — The overarching focus of my research has been and continues to be making astrophysics more accessible to researchers, specifically those working with large data-sets. I have worked on this problem over multiple years, from clearing the way for astronomers to use gPhoton, to building tools to allow for the quick and easy generation of large amounts of synthetic data, to finally building tools which allow variable sdBs to be identified faster using deep neural networks. **I will continue this theme into graduate school by studying how deep neural networks may allow for faster time evolution of n-body systems.**

Through research, I built up a significant knowledge base in Python. Given that Python was the language of choice at my institution, I decided to hold two Python master classes to teach other students some of the tools and skills I had developed. I taught a master class on general data science with Python to approximately ten students and two professors, and then later in the semester taught a follow up master class, an introduction to machine learning with Python, to the same group of people. Additionally, I have hosted lectures in the computational physics course on animating graphs using Python. These experiences disseminating what I have learned about Python to the rest of my department, along with my time as a supplemental instructor for Calculus Three, serves as experience with instructing and cemented my desire to become a professor.

Conclusion — The research I have done these past three years hasn't just shaped my experience as an undergraduate, it has been the driving factor in my life. Every project that I worked on has provided me with more evidence of why I want to continue in this field and earn my doctoral degree. I am committed to continuing and expanding my research and at the same time, some of my best experiences have come from passing on what I have learned to others, from helping the underclassmen in the physics department with their work and early research projects to showing the other students at SAO different ways to code. If my senior experience made concrete my desire to be an astronomer, the experiences I have had over the past three years have made concrete my desire to be a professor, teaching and researching.

References

[1] Boudreaux, Thomas M., et al. *The Astrophysical Journal* 845.2 (2017): 171.

[2] Boudreaux, Thomas M. *Open Astronomy* 26.1 (2017): 258-269.

Thomas M. Boudreaux - Research Proposal

Background & Intellectual Merit — N-body systems larger than two-body systems are essential to our understanding of the universe because they allow us to test models of galaxy formation (Somerville et al. 2015), planetary formation (Liu et al. 2011), and cosmological structure (Leo et al. 2018). Although researchers have long recognized these systems were important, they lacked the resources to solve for their dynamics. Computers have addressed the so called "n-body problem" by enabling a better understanding how these systems behave and, as a result, more accurate and predictive statements may be made regarding the evolution of n-body systems without requiring a closed-form solution. Direct numerical integration over some arbitrary time has been the main tool used when analyzing n-body systems. While this integration is generally a precise method, the costs of precision are substantial, as numerical integration can be very slow when many particles are involved in a simulation. Due to the high time and material cost of running large numerical integration, the study of many n-body systems has been off-limits to researchers at smaller institutions without access to larger computers. **I propose to study whether deep neural networks can make predictive statements regarding the time evolution of an n–body system more quickly and with at least equivalent accuracy to traditional direct numerical integration schemes**.

The work I have done with large datasets, deep learning, and n-body systems has prepared me to develop this system which may reduce the computational complexity of solving n-body systems. Additionally, a set of recent advancements in the understanding of neural networks indicates the feasibility of this work (Srivastava et al. 2014). The remainder of this proposal is divided as follows. Stage I discusses using neural networks to time-evolve simple two-body gravitationally-bound systems. Stage II extends this work from two bodies to an arbitrary n number of gravitationally-bound bodies. Finally, Stage III extends not the number of bodies, but the complexity of the interaction between bodies, thus expanding the results of this work for use in more fields of astrophysics and other fields in the natural sciences.

Stage I: Keplerian Systems — Two body systems are the only systems for which a closed-form time evolution solution exists; therefore, they are an ideal candidate for the initial testing of neural networks. **I will first investigate the levels of accuracy and precision deep neural networks can achieve when classifying two body systems**. Understanding how different network architectures perform in this simple case will be essential to understanding how they will generalize to the more complex case. Additionally, this will serve as a test bed for both the network architecture, and the generalization issues that often plague neural networks.

Using the well known equations governing two body evolution, I will quickly generate large amounts of accurate training data consisting of the initial and final positions, masses, and velocities of both bodies. Based on my past experience using neural networks to classify variable stars, I will construct a set of networks with varying architectures using the well optimized *TensorFlow*[1] code. These will consist of multilayer perceptrons, convolutional neural networks, and recurrent neural networks. Each network will be embedded into a grid-search hyperparameter tuning algorithm and will be provided fifty percent of generated data to train on. After training, 25 percent of the remaining, unused data will be used to compare different networks directly to one another through k–fold cross validation. The best performing networks will be selected and have their accuracy tested against the final remaining 25 percent of generated data. I will complete this stage of the work over my first two years of graduate study.

Stage II: Large N–body systems Where two-body systems can be solved analytically, n-body systems cannot. Additionally, the numerical integration techniques which have been used scale their complexity at non-linear rates. The chaotic nature of n-body systems has traditionally been seen as a barrier to the use of techniques other than numerical integration; however, it has been shown recently that neural networks can make predictive statements regarding the future state of chaotic systems (Pathak et al. 2018, Maathuis et al. 2017).

[1] https://www.tensorflow.org/

In this stage of the work, **I will investigate how to train a network to work with progressively larger numbers of particles**. The computational costs of training these networks will come predominately from the generation of the data necessary to train them. I will use the mature code *N-body6++GPU* (Wang et al. 2015) to generate training data in the same form as the two-body training data. Harvard University's Department of Astronomy is the ideal location to work on this research, due mainly to the available computational resources and its close ties to the Center for Astrophysics (CFA). The faculty working in the Institute for Theory and Computation at the CFA, who have significant experience working with simulations of n-body systems, will provide the support necessary to conduct this stage of work. Additionally, I will make use of the two graphical processing unit partitions on Harvard FAS's Odyssey high performance computing cluster to run *N-body6++GPU*. My work at an NSF-funded REU program using *N-body6++GPU* has prepared me for this stage of work, which I anticipate will take two years. I will start work on this stage while completing Stage I, finalizing results during my third year of graduate study.

Stage III: Complex Interactions — In the third stage of this work, taking place over my final two years of study, I will investigate whether a neural network might be effective in an n-body system where there is more than one force regulating the evolution of the system. Here, more complex integration programs, such as the soon-to-be-released *GADGET4*[2] will facilitate running new sets of simulations where effects other than gravity are considered, such as stellar feedback mechanisms and hydro–dynamical interactions. Once data are generated, the same training procedure will be applied as in the first two stages of this project. **This stage serves to generalize the work I will have done by accounting for the interactions important in n-body systems pertinent to fields other than astronomy such as biology and chemistry.** At the conclusion of this stage, the body of my doctoral candidate work will be complete.

Broader Impacts — N-body systems, whether gravitationally bound or not, are essential for understanding a whole host of scientific problems. This work will primarily focus on applying deep learning to those n-body systems relevant to Galactic astronomy. However, the applications of the proposed work are neither limited to Galactic astronomy, nor astronomy in general. From electromagnetically-regulated interactions between proteins in cells to the large-scale structures that form filaments between galaxies, n-body systems are present over many orders of magnitude of scale in the universe. I have presented a method here to significantly reduce the time needed to evolve n-body systems. The trained models resulting from this work will allow more researchers, including those at smaller institutions who may not have access to the powerful computational resources of those at larger institutions, to test models against n-body systems more quickly. By opening the in-depth study of n-body systems to more universities not only can more senior researchers contribute to the literature, more students can be exposed to this essential field early on in their careers. Generally then, this work serves to accelerate the pace of scientific discovery.

Conclusion — Modeling n-body systems is an essential component of understating many phenomenon in astronomy and in other fields. Here I have argued that neural networks will allow for great headway to be made in the study of n-body systems by reducing the cost associated with their modeling. Consequently, more researchers will be able to both contribute and introduce students to the field.

References

[1] Rachel S Somerville, et al. *Annual Review of Astronomy and Astrophysics*, 53:51–113, 2015.

[2] H. Liu, et al. *The Astrophysical Journal*, 732:66, May 2011.

[3] M. Leo, et al. *ArXiv e-prints*, July 2018.

[4] Nitish Srivastava, et al. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[5] Jaideep Pathak, et al. *Phys. Rev. Lett.*, 120:024102, Jan 2018.

[6] Henry Maathuis, et al. *29th Benelux Conference on Artificial Intelligence*, November 8–9, Groningen, volume 3, page 326, 2017.

[7] Long Wang, et al. *Monthly Notices of the Royal Astronomical Society*, 450(4):4070–4080, 2015.

[2] https://wwwmpa.mpa-garching.mpg.de/gadget/