Personal, Background, and Future Goals 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 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 the Goodman Spectrograph, a set of instruments mounted on a telescope in Chile. At 2am on the first night, Dr. Barlow turned to 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 I didn't know what to say, simply telling him that I wasn't scared off. Now, more than five years later, I wish I could tell him that instead of scaring me off **he made astronomy real in a way it had never been, and vastly more alluring.**

Learning Important Tools: I began serious research as soon as I entered college. For my first project, I used archival data my advisor had collected of a particular type of binary star. Part of this research involved data which had to be passed between many different pieces of code written over the years; a common occurrence in modern astronomy. While this could be done manually, my advisor asked me to automate the process. In doing this I learned how to build small pieces of code that tie larger programs together, also known as scripts. **Scripts are an astonishingly valuable tool and *all* of my subsequent work has incorporated the use of progressively more complex and efficient scripts.** I presented this work at the 227th Meeting of the American Astronomical Society (AAS) and later contributed to a publication using this data (Vos et al. 2019)

While at the AAS, I met a student named Micheal Tucker whose work was applicable to the study of hot subdwarf B stars (sdBs). These stars are the helium fusing remains of stripped red giants and the main focus of the HPU astrophysics research group. With this in mind, Dr. Barlow recommended that I apply to the Space Telescope Science Institute's (STScI) Summer Astronomy Student Program, the same program at which Micheal had conducted his work. In my application, I specifically proposed to extend Micheal's work to sdBs. I was accepted to the program, and over the following year, my REU advisor and I investigated the feasibility of using software he had developed (gPhoton) to identify and study pulsating hot sdBs. **I was able to take on this project due to the experience I had developed working with Python in my previous project.**

Learning The Scientific Process: The STScI summer program placed an emphasis on student-led research, so my REU advisor acted as a guiding hand, giving me generous leeway to work on the project as I saw fit. I started by using the skills I had established during my first project to write a set of tools that would help automate the process of analyzing the hundreds of thousands of individual data points we had to work with. **This was my first exposure to large datasets.** In collaboration with Dr. Barlow I discovered that a known star was actually a pulsating sdB. Despite our discovery, we also were able to make the more general statement that gPhoton was not well suited for our task. For the first time I was able to lead the publication of an article in a peer reviewed journal detailing these findings (Boudreaux et al. 2017). **Leading the writing, editing, submission, and referee response for this paper began to teach me how to meaningfully contribute to astronomy as a researcher, rather than just as a student.**

Learning Academic Independence: During my previous project, I became frustrated working with large data sets. I considered sifting through tens of thousands of data points an inefficient use of time. Consequently, I asked my advisor 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 heavily automating the process. **This project would, if successful, significantly reduce the observational costs typically associated with the discoveries of variable stars**. 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, he would support me. I taught myself both the theoretical basis for, and practical applications of, neural networks. Networks need data to learn from; therefore, I developed tools from scratch to generate synthetic data. I then trained and tested models against these data. I honed these models' parameters until I found which ones performed best. 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 that I developed an individualized major in computational physics that combined all of the normal physics degree coursework with large portions of both a computer science and a math degree.

I presented my work on neural networks at a conference on sdBs in Krakow, Poland, during the summer of 2017. While at the conference I met many of the experts whose papers I had been reading for the last two years. One these people was a postdoctoral fellow at CalTech, Thomas Kupfer. Thomas had very similar research interests to those of our group, and he was intrigued by the preliminary results that I presented. Up until that point, I had only worked with the synthetic data that 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 traveled to Pasadena for a short stint as a visiting researcher to apply my deep learning models to those data. **This experience was my first independent collaboration with another scientist in my field.**

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

From Learning To A Career: For the summer of 2018 I applied and was accepted to an REU program at the Harvard Smithsonian Center for Astrophysics (CfA). While at the CfA I worked on a project studying globular clusters using numerical simulations, a topic quite divorced from my research up to that point. That summer studying globular clusters led me to apply to, and eventually choose, the same graduate program my REU advisor that had attended - Dartmouth College's Department of Physics and Astronomy. Dartmouth provided opportunities to study globular clusters, to use and develop numerical tools, and to advance my understanding of observational astronomy. For all of these reasons, at Dartmouth I have chosen to work with Dr. Brian Chaboyer. We are studying, through the use of the Dartmouth Stellar Evolution Program (DSEP), the chemical compositions of globular clusters.

Intellectual Merit: Computational astrophysics is a field that has become essential to our understanding of the universe for two main reasons: (i), astronomers are limited in their studies by the light that happens to fall upon Earth; and (ii), the mathematics that underpin our models of the universe do not tend to have analytic solutions. Computers partially mitigate both of these problems simultaneously, allowing astronomers to run progressively more complex and complete simulations of natural phenomena, using observations as boundary conditions. For example, though we cannot directly observe the interior structure of a star, numerical models of the stellar structure equations allow us to make robust predictions about that structure. Those structural predictions then imply certain surface conditions, which can be observed, and used to validate the predictions.

A key element, then, when framing the importance of computers is the understanding that they are useful only insofar as they *complement* observation. My interests are in computational astrophysics but much of my background is in observational astronomy. This was and is an intentional choice made to strengthen my intellectual merit as an astronomer. When I started graduate school at Dartmouth I was given the opportunity to work with either an observer, Dr. Elisabeth Newton, or a computationalist, Dr. Brian Charboyer. **I chose to work with the observer because, to understand and apply computational tools as I eventually hoped to, I needed a stronger observational background.** I have now nearly completed that first-year project, more than doubling the number of constraints on a relation between the rotational period and magnetic field strength of cool stars, and now I am in the final stages of preparing to submit a manuscript for publication. Going forward, to mesh more closely with my computational research interests, I am transitioning to work with a new advisor, Dr. Charboyer, on a project that is more computational. The last year, however, will be invaluable for the deeper understanding of observations they gave me.

Broader Impacts: The overarching focus of my research has been, and continues to be, making astrophysics more accessible to researchers, and specifically to those working with big data-sets and who might not have access to large and expensive computational facilities. I have worked on this problem over many years, from clearing the way for astronomers to use gPhoton, to building tools allowing for the quick and easy generation of large amounts of synthetic data, to building software which identifies variable sdBs faster using deep neural networks, and finally to studying how neural networks can be used to speed up n-body simulations. **I will continue working with this theme into my thesis research as I study ways to make the stellar evolution code DSEP run many instances concurrently using graphics cards.**

Through my research, I have built up significant experience with various computer languages. Since I was a sophomore, I have used this experience to tutor my peers in Python as well as other languages. One example of this experience was teaching multiple Python master classes covering general data structures, machine learning, and data visualization to both students and professors. As a TA I have emphasized computational tools to my students. In Introductory Mechanics I helped re-frame labs around the numerical analysis of data, and as a TA for Introductory Astronomy I helped to rewrite labs to adapt to the COVID-19 pandemic by having students use software, such as Stellarium, in lieu of in-person lab work.

Conclusion: The research I have done since that night with Dr. Barlow at UNC has not just shaped my experiences as a student, it has also been the driving factor in my life. Every project that I have worked on since has provided me with more evidence of why I want to continue in this field and earn my doctoral degree. Through my past research experience, I have learned how to think like a scientist, develop and use computational tools for that science, and lead and eventually publish a research project. Going forward I am committed to continuing and expanding my research, incorporating new advances in numerical methods and new observations of the universe. At the same time, some of my most rewarding experiences have come from sharing what I have learned with others, from helping younger students with their first research projects to showing my peers different and better ways to program. If my senior experience made concrete my desire to be an astronomer, the experiences that I have had over the past five years have made concrete my desire to be a professor --- to teach, to research, and to generally pass along knowledge of astronomy to a new generation of students.

References

[1] Vos, J., et al. *Monthly Notices of The Royal Astronomical Society* 482 (2019): 4592.
[2] Boudreaux, Thomas M., et al. *The Astrophysical Journal* 845.2 (2017): 171.
[3] Boudreaux, Thomas M. *Open Astronomy* 26.1 (2017): 258-269.

Using Graphics Cards to Quantify Uncertainties in Models of Stellar Evolution.

Intellectual Merit: Stars remain an important probe of astrophysical phenomena due to processes that drive their individual evolution, and because they contribute, through enrichment of the interstellar medium, to the evolution of larger structures [1]. While many physical properties of a star may be observed or inferred, peering past a star's outermost layers quickly becomes impossible. Numerical models provide astronomers with a way of studying the insides of stars without actually seeing them. These models allow astronomers to infer such things as the integrate ages of distant galaxies or the masses of stars which host exoplanets. The most generally applicable numerical model of a star is based on integrating four coupled ordinary differential equations that, when used in concert with an equation of state, energy generation rate, and opacity, describe the temperature, pressure, density, and luminosity at any radius within a star. These are known as the equations of stellar structure.

The Dartmouth Stellar Evolution Program (DSEP) solves the linearized equations of stellar structure in one dimension, specifically for the evolution of low mass (M < 5 solar masses) stars. Integrated models for stars with similar ages and compositions, known individually as isochrones, from DSEP are well established in the community; however, studies tend to use a single isochrone, or at best two isochrones, when classifying stars. This limits researchers' ability to quantify uncertainties associated with the models through methods such as Monte Carlo simulations. Despite DSEP's optimization, which only takes minutes to evolve most models over ~10 billion years, it is currently infeasible to generate the number of isochrones needed for a Monte Carlo simulation without months of time on a supercomputer. **I propose to reduce the costs of quantifying uncertainties in isochrones by modifying DSEP so that many isochrones may be generated concurrently, thus reducing the required time to generate tens of thousands of isochrones from months to days.**

Why DSEP: While other stellar evolution programs, such as the widely used program MESE [2], consider a more complex handling of nuclear reaction rate calculations, and are more generally applicable to a wider range of stellar masses than DSEP, DSEP has certain advantages over these other programs that make it well suited for this task. For one, DSEP has a *much* smaller memory footprint than MESA, and memory is a major bottleneck here. Additionally, each individual instance of DSEP runs much faster than an instance of MESA, as DSEP can make some simplifying assumptions due to its focus only on models with initial masses between 0.1 and 5 solar masses. Despite this trade off, the grid of isochrones generated by DSEP [3], has been heavily cited since its initial release in 2008, proving that there is a place for a code as specific as DSEP. Finally, DSEP, unlike MESA, has a small enough codebase that it can realistically be modified on the time scales of a graduate program.

Isochrone Uncertainties: It is essential to have grids of isochrones dense enough to run Monte Carlo simulations for two reasons: (i), there are physical uncertainties in the measurements used as inputs when evolving a stellar model, such as those of opacity or nuclear reaction rate, that are not currently accounted for when evolving isochrones; and (ii), without a dense grid of models it is impossible to accurately know how small variations in the input parameter affect final results. The Dartmouth Stellar Evolution Group currently has funding to study the feasibility of generating just such a grid of models; however, the proposal calls for an eventual 4.5 months of time on an NSF XSEDE computer. While this is not an unreasonable amount of time on a supercomputer, it does place the generation of Monte Carlo isochrones out of the reach of those without ready access to a such a computer. This is where graphics processing units (GPUs) may permit significant advancement.

GPU Computing: GPUs operate fundamentally differently than central processing units (CPU). In contrast to a CPU's handful of monolithic compute "cores," GPUs have thousands to tens of thousands of comparatively weak compute cores, known as streaming processors. Streaming processors, while individually slower than a CPU core, can all run in parallel. This means that if a program can be made either to run portions of itself concurrently, or full instances of itself (packaged within a single controller process) concurrently, the total execution time can be drastically reduced.

At Dartmouth, we have a high-performance computer cluster, Discovery, containing two Nvidia K80 GPUs. Each of these has ~5000 cores and 24 gigabytes of memory and can be programmed using CUDA, a language developed by Nvidia. DSEP has a memory footprint of ~50 MB placing a worst-case upper limit on the number of concurrent instances that can be run on the K80 GPUs in Discovery at ~500. **For one K80, running at 0.56 GHz, this work would decrease the run time required for the generation of 10,000 isochrones from ~70 days, to approximately one day.**

Research Plan: This project will be broken down into two main phases: (i), modifying DSEP to use programming conventions in line with what Nvidia's CUDA software requires; and (ii), modifying DSEP to actually run on a GPU. The former is necessary as DSEP currently makes heavy use of global variables; a data model not compatible with CUDA. In modifying DSEP's data model away from global variables, its code will become more similar to C. This is an important precursor to the second phase of the project. CUDA is a C-like language; therefore, much of this work will involve recasting DSEP's subroutines as far more C-like objects.

I will spend the first year and a half modifying DSEP away from global variables; this is the bulk of the programming work. I anticipate that the second stage, converting DSEP subroutines into CUDA functions, will take half a year. I will reimplement one subroutine at a time in CUDA, testing each as I go against a mainline branch of DSEP. When all subroutines are reimplemented I will write a “controller” program to distribute the calculation of isochrones to separate cores on the GPU. My significant past experience with CUDA and C will allow me to conduct this work. Finally, I will spend the last portion of grant funding generating a new, publicly available grid of isochrones, suitable for Monte Carlo simulations, using this version of DSEP.

Broader Impacts: Computers are one of the most democratizing tools in astronomy. While large research-class telescopes will forever be scarce, almost every student and professor today have access to a computer. It is thus essential that continuing efforts be made not just in the development of software that can run on clusters or government-funded supercomputers, but also that software be developed for personal computers that can reach more people to support new, effective research. This modified version of DSEP will be portable to GPUs at many price points. A mid- to high-end consumer GPU costs less than 1000 dollars, a sum which is in reach of many smaller liberal arts colleges' departmental budgets, as opposed to the tens of thousands of dollars that a full compute cluster would cost. Additionally, days of GPU time have a vastly smaller energy footprint than months of supercomputer time. **This work will open the door for a far more diverse and larger population of scientists to generate their own isochrones in a more energy efficient manner.**

Observations of globular clusters show that they are composed of multiple populations of stars and, by fitting observations to isochrones, we believe that the populations vary in Helium abundance, with the younger stars being enhanced in Helium [4]. However, due to the lack of dense isochrone grids, we are unable to robustly quantify how certain those enhanced abundances are. The Monte Carlo grids generated here will allow us to quantify these uncertainties. Another important question that may be resolved is the sensitivity of isochrones to mixing length. Recent work shows that adopting a solar calibrated mixing length is insufficient for other spectral class [5]; however, there are no quantified uncertainty bounds known yet. Through this work I will address these problems and allow other researchers to answer more, as yet unknown, questions.

References

[1] Genzel, R., et al. *The Astrophysical Journal* 800.1 (2015): 20.
[2] Paxton, Bill, et al. *The Astrophysical Journal Supplement Series* 192.1 (2010): 3.
[3] Dotter, Aaron, et al. *The Astrophysical Journal Supplement Series* 178.1 (2008): 89.
[4] Bastian, Nate, et al. *Annual Review of Astronomy and Astrophysics* 56 (2018): 83-136.
[5] Pinheiro, F. J. G., et al. *Monthly Notices of the Royal Astronomical Society* 433.4 (2013): 2893-2899.