NSF GRFP Fellowship Report

2017-18 Richard Clark Fitzgerald

Intellectual Merit

My research focuses on code analysis to increase the performance of the R language for statistical computing. I increase the performance by automating parallel code generation so that the computer(s) can execute multiple instructions simultaneously. This allows researchers to scale up analyses to ask new questions on larger data sets. The current state of the art is for *users* to modify their code when they need to run code in parallel. I'm working on a *system* capable of doing this modification for them automatically. This requires careful analysis of the language as an object using metaprogramming techniques. With such a system a user can take potentially sophisticated custom data analysis code written in R and generate correct and efficient parallel code specialized for different platforms.

In June 2017 I passed my PhD oral qualifying exam by presenting a talk titled "Parallel Computing Through Code Analysis". I described a motivating application to traffic engineering of a qualitatively new kind of analysis that requires processing 3 billion data points. In the fall I completed this analysis, demonstrating how to generate code to efficiently combine several powerful technologies. I'm currently focusing on how to generalize the code analysis techniques. Along the way I've learned much about parallel computing and was able to contribute to an article "The R Language: A Powerful Tool for Taming Big Data" which will be published in the "Encyclopedia of Big Data Technologies" this year.

Broader Impacts

I've made my results available to a broader audience through software and a blog. This past year I've written about both basic and research topics in data intensive computing in 25 posts in my blog "Data Programming" which can be found at http://clarkfitzg.github.io/. I've created prototype software which is available publicly in the `autoparallel` package at https://github.com/clarkfitzg/autoparallel.

During the last year I've volunteered as a tutor, seminar organizer, and programming instructor. 2017 marked the opening of the new UC Davis Veterans Success Center. I continue to work closely with the coordinator, Earl Raehsler, as a weekly tutor in mathematics and statistics for fellow student veterans. Last fall I took another turn as the organizer of the statistics student seminar that a fellow student and I initiated in 2016. We successfully transitioned the organizer role between students, and the seminar is now in its 5th quarter. In 2018 I attended a 2 day Software Carpentry (SWC) instructor training to become a certified instructor. SWC's mission is to "Teach basic lab skills for research computing". I co-instructed a 2 day workshop catered towards graduate students in the life sciences, and look forward to leading more workshops in the future.