Statement of Purpose

Vineetha Govindaraj

am applying to the PhD program in Electrical engineering and Computer Science department at the University of California Berkeley and I am interested in pursuing research in the field of Computer Architecture. I am a second-year graduate student at Pennsylvania State University in Computer Science and Engineering. I have had a well rounded academic research and industry experience during the course of my professional career and would like to continue solving impactful research problems as a PhD student at Berkeley.

Since childhood, I've been drawn to technology through a natural curiosity to understand how things work. This manifested as an eagerness to develop circuits ranging from making an electromagnet to modelling a propeller car. As a first step towards this, I decided to take up an engineering career. My inclination towards science and mathematics helped me secure the top 1 percentile in the national level engineering entrance examination in India. This provided me with an opportunity to pursue an engineering degree at the prestigious **National Institute of Technology (NIT)**. I opted for Electronics and Communication Engineering in NIT, Trichy (NIT-T) intending to demystify complex circuits.

A well rounded academic curriculum at NIT-T helped me develop a broad sense of understanding in Electronics and Communication Engineering (ECE). To further my knowledge, research internships at **Indian Institute of Technology, Madras(IIT–M)** and **Indian Institute of Science(IISc)** during my sophomore and junior years respectively exposed me to open research problems in computer architecture. This was instrumental in channelling my interest in taking up research problems in computer architecture as a part of my capstone project. I started working on improving the performance of interconnects present in multicore processors. After several weeks of motivational experiments and literature survey, I designed an experimental setup with which I started measuring the performance of multi-threaded scale-out workloads when executed on multi-core processors under different interconnect topologies. While experimenting with several interconnect topologies and keeping power, performance and area as constraints, I was able to conclude with experimental evidence that a Flattened Butterfly topology was much more efficient compared to a Mesh or a star topology for executing scale-out workloads. Through this project, I learnt how different microarchitectural designs affect application performance at varying levels. The feedback and assistance from Prof. N. Ramasubramanian of the Department of Computer Science at NIT-T, helped scope this project to a great extent. Graduating from my program, I was motivated to take up much more challenging problems in Computer Architecture.

After my undergraduate studies, I took up a industry job at Freescale Semiconductors as a Verification Engineer. During my brief tenure at Freescale, I was involved in verifying in house System On Chip designs (SoCs) that integrated processors, peripherals and controllers. Following that, I took up another challenging role at Samsung Semiconductors, where I was working on Solid State Drives (SSD). Specifically, I was working on designing the Physical Layer of an in-house PCIe Verification IP (VIP). This was an intellectually stimulating project where I learnt how different kinds of packets are encapsulated at each layer of PCIe. This motivated me to delve deeper into understanding the PCIe and NVMe subsystems that played an instrumental role in boosting the performance of SSD cards.

My interest towares academic research motivated me to pursue a Masters degree in Computer Science and Engineering at Pennsylvania State University. At Penn State, I started collaborating with professors **Vijaykrishnan Narayanan**, **Mahmut Kandemir** and **Jack Sampson**. Initially, I was assigned to work on building an in-house Non-Volatile Processor (NVP) simulator based out of Gem5 microarchitectural simulator. I implemented a NVP simulator that tried to mimic the functionality of applications when executed in an energy harvested processor infused with microarchitectural save and restore states. This task exposed me to state-of-the-art microarchitecture designs housed on processors that utilize energy harvested power sources. Furthermore, it got me acquainted to cycle-accurate microarchitectural simulators. This enabled me to rapidly implement and try out several ideas that I had in my subsequent research projects.

As my work in NVP research continued, I was eager to explore other research directions simultaneously. Hence, I got involved in a research project on approximate computing where I was extensively collaborating with other senior graduate students. The fundamental goal of our project was to introduce language and

runtime extensions that allow specific regions of a program in which dataflow dependencies can be approximated. One of my major contributions in this project was towards identifying and designing workloads where regions of programs can be approximated. The main challenge in this task was to identify execution patterns in regions within workloads that when approximated can show significant improvement in execution time without a major loss in accuracy. After being on-boarded to this project, I was able to quickly understand and comprehend the framework and develop suitable applications that can be approximated. My prior knowledge in design verification proved to be extremely useful in developing test cases aimed to stress test the framework. Our framework had huge performance gains which resulted in submitting our findings to a top tier programming languages conference (PLDI). This research collaboration was an immense learning experience for me. Firstly, from having to learn approximate computing, I learned to teach myself from technical papers and textbooks. Following that, I closely observed the efforts involved while presenting academic research coherently and how key research ideas are manifested to manuscripts.

With the experiences gained from earlier research projects, I was keen towards carving out a new research problem by myself as a part of my Master's thesis. I always believed that the solution to many research questions is not contained within a single field. I experienced this while discussing the challenges present in cloud computing at one of the classes that enrolled with Prof. Mahmut Kandameir. In my class, I learnt the existence of energy dis-proportionality among datacenters that house latency-critical workloads. I found this problem very intriguing and my curiosity let me to further explore this area.

As a first step towards studying this, I started setting up a popular datacenter workload, `memcached` on a small server cluster. With my experimental setup, I tried to observe the characteristics of the server under different loads scenarios. From this, I witnessed that the CPU utilization was extremely low for small loads and increased gradually as server loads increase. However, I observed that the CPU utilization was still low even at peak loads. At this point, I realised that power management techniques need to be fundamentally different for modern datacenters executing user facing workloads in order to curb energy dis-proportionality. This led me to pursue a research problem that studied the effect of power management schemes on datacenter servers. During the course of this project, I learnt how deep sleep states are disabled in production data centers to meet the Service Level Objectives (SLO). I did a series of experiments on the Memcached servers to understand the effect of SLO violations when deep sleep states are enabled. With that, I started to breakdown the steps involved in entering and exiting deep sleep states and narrowing down on the critical stages which cause excessive performance overheads. These motivational experiments along with the feedback from faculty and collaborators from industries, I was able to hone in to a well defined research problem.

The first step that I undertook towards solving this problem was to create a microarchitectural simulation infrastructure that mimics the performance of applications executing in datacenter clusters. This simulation infrastructure had proved to be extremely useful for rapidly implementing and trying our new ideas where there existed limitations on designing real system experiments. However, while building this infrastructure, I faced a major challenge where I had to run a memcached server connected by a distributed system of multiple clients on an architectural simulator. After weeks of understanding and debugging multiple open source tools and softwares, I was able to set up micro-architectural simulator-based out of gem5, called dist-gem5, customised for my experimental setup. I then formulated key changes in existing deep sleep states. From this, I was able to show how emerging technological advances such as NVSRAM can be leveraged to realise deep sleep state that would save energy without disrupting performance. I prepared my final results by evaluating the proposed techniques on datacenter workloads. I submitted my findings in Design and Automation Conference (DAC-2020). This particular project was a well-rounded learning experience in terms of what goes on in academic research. It was also a new impetus to pursue further research after my Master's degree.