Graduate Research Plan – Kristi Belcher

Power and Energy Efficiency Optimization of Accelerators for Mobile Devices

Keywords: mobile GPUs, energy efficiency, accelerators

Introduction

Heart disease is the leading cause of death in the United States, killing over 600,000 people annually. Persistent high blood pressure is one of the most common signs of developing heart disease, and early detection via personal health monitoring from mobile devices could save many lives [3]. In an average month, over 4000 emergency vehicles take 12 minutes or longer to arrive on scene; real-time traffic rerouting using data sourced from mobile phones could drastically reduce these response times [7]. In addition to such life-or-death situations, the desire to monitor the activities of children, pets, or an empty living space has given rise to mobile applications expected to perform a broad range of tasks from turning on lights or surveilling a room to alerting authorities. With the continuing shift to mobile computing, these scenarios will relentlessly increase the demand for efficient, sophisticated, and convenient mobile applications that have the ability to fulfill a near infinite number of personal needs. Some of these applications will require copious amounts of data and processing power to quickly obtain a sufficiently complete picture of the problem and formulate a response. This will necessitate mobile accelerators that are equipped to perform such tasks within a reasonable amount of time. Additionally, with limited battery capacity on devices, the requirement for efficient use of mobile accelerators is paramount. By developing optimization techniques for mobile accelerators to efficiently run time-critical and computation-heavy applications that can notify users when their blood pressure is too high, provide real-time traffic routing, or continuously surveil areas for targeted activity – among other possibilities – my research will contribute to leading-edge technological advances in computing capabilities in the mobile realm.

Research Plan

Previous work suggests that there are many software optimizations for increasing the energy and power efficiency of accelerators, such as GPUs, even for complex algorithms [1]. Other work provides insight on how to reduce power and energy consumption on mobile accelerators, in this case, mobile GPUs [2, 4], but for specific applications. Due to the unique hardware of these chips, it can be difficult to optimize applications for efficiency. Since my previous research experience with multi-GPU parallelization successfully resulted in a general technique for optimizing algorithms, I plan to use a similar approach in my graduate research. As I have worked with GPUs in the past, I will focus on mobile GPUs and provide results that can be expanded to other types of mobile accelerators like DSPs and FPGAs. My research goal is to use computation-heavy algorithms of varying complexity (such as the Single Source Single Sink Evacuation Route Planner (SSEP) or the Capacity Constrained Route Planner (CCRP) [6]) to investigate and ultimately understand how to optimize them for mobile accelerators, ultimately creating general optimization strategies for designing mobile-accelerator-friendly algorithms. Since prior work has provided information on how to employ source-code optimizations for lowering the power and energy consumption of GPUs [1], I will refer to these findings as a basis for starting my own research, which will look for similar optimizations, but for mobile accelerators. Additionally, I will perform a detailed analysis of how different applications, whether compute or memory bound, behave on accelerators and determine the best strategies to increase performance and efficiency in the face of these bottlenecks. The proposed research will result in general strategies and techniques for implementing these algorithms on mobile accelerators that can then be used by software developers and hardware designers as well as for building automated programming tools. These results will allow for widespread adoption and added functionality, thereby increasing the benefit of my research to the community.

Anticipated Results

My research will yield an accurate model of the behavior and consumption patterns of mobile accelerators that will enable the use of mobile devices to run increasingly diverse algorithms. Hence, my work will create an avenue to overcome the current obstacles of porting algorithms for which it is difficult to achieve efficiency and performance on mobile accelerators. Additionally, my research will yield best practices for future work with mobile accelerators where automation can be added to help the programmers adjust their code to enhance efficiency. My results can then be applied to other, future real-world mobile applications to help improve the lives of users.

Intellectual Merit

Accelerators have demonstrated the ability to efficiently and cost-effectively run computation-heavy algorithms as compared to CPUs [1]. However, with the broadening utility of accelerators to mobile applications, it is essential to find general, simple techniques to ensure portability of these applications. Optimizing an algorithm for efficiency on mobile accelerators is already a popular subject of research [2, 4, 5], but it does not focus on creating simple, yet general techniques for energy and power optimizations for any algorithm on mobile GPUs. My research will advance our knowledge of computer science by capturing the power and energy behavior of mobile accelerators running time-critical and computation-heavy algorithms. As a research assistant in the Efficient Computing Laboratory (ECL) at Texas State University, I have access to many top-of-the-line GPUs donated by NVIDIA. Additionally, with my previous experience of working with supercomputers at the Texas Advanced Computing Center (TACC), I have the background experience necessary to make my research plan feasible and succeed in graduate school.

Broader Impacts

With the rise of mobile platforms, there is a need to increase the potential of mobile accelerators to improve computational efficiency. My research will be a crucial step toward understanding how to efficiently utilize accelerators so that mobile devices can run time-critical, computation-heavy applications that may save lives while also being able to deliver normal functionality. My work will contribute to providing the U.S. with a competitive advantage in the development of cutting-edge mobile technology. I will also expand parallel computing research knowledge by making the results of my research available as open-source resources and encouraging the scientific community to access and utilize my findings. Additionally, through outreach at conferences and efforts to educate others about the findings of my research, I can positively impact the research community, especially women in STEM.

[1] J. Coplin and M. Burtscher. Energy, Power, and Performance Characterization of GPGPU Benchmark Programs. Twelfth Workshop on High-Performance, Power-Aware Computing. May 2016.

[2] K-T. Cheng, Y-C. Wang. Using Mobile GPU for General-Purpose Computing – A Case Study of Face Recognition on SmartPhones. 2011 International IEEE Symposium on VLSI Design, Automation, and Test. June 2011.

[3] "Leading Causes of Death." Centers for Disease Control and Prevention. Web. 19 Oct. 2016.

[4] M. B. Lopez, H. Nykanen, J. Hannuksela, O. Silven, M. Vehvilainen. Accelerating Image Recognition on Mobile Devices Using GPGPU. January 2011.

[5] G. Wang, Y. Xiong, J. Yun, J. Cavallaro. Accelerating Computer Vision Algorithms Using OpenCL Framework on the Mobile GPU – A Case Study. 2013 IEEE International Conference on ICASSP. October 2013.

[6] G. Mishra, S. Mazumdar, A. Pal. Improved Algorithms for The Evacuation Route Planning Problem. 2015 International Conference on Combinatorial Optimization and Applications. 2015.

[7] "FEMS Response Time Metrics." DC.gov. Fire and EMS Department, 2015. Web. 21 Oct. 2016.