STATEMENT OF RESEARCH INTERESTS

Over the past few years, my research interests focus on automata processing, data mining, high performance computing, compressive sensing, radar imaging, computational electromagnetics, electromagnetic scattering and inverse scattering, and medical image processing. As a Research Assistant at the Electromagnetic Wave Propagation and Remote Sensing Laboratory at the Catholic University of America (CUA) and a Research Associate at the University of Virginia (UVA) in the Department of Computer Science, I have had more than six years of research experience in different projects that span a wide range of applications.

In the first year of my PhD study, I was in charge of investigating the interferometric imaging through random media on hardware-accelerated computing platforms. Terahertz frequencies are being studied for many security imaging purposes due to their ability to penetrate most clothing while still reflecting off metallic objects. Furthermore, they are non-ionizing unlike microwave frequencies, making them safe for use on people and can be compact compared to microwave frequencies. In this research, we considered interferometric imaging at terahertz frequencies. Interferometric imaging relies on the complex correlations between each pair of sensors in a detector array. The image is constructed based on these correlations using Fourier transform. This technique can be applied for detecting concealed objects such as weapons and/or explosives, etc. behind cluttered media. However, such systems require intensive computations that could potentially make it impractical for security applications, which generally require real time or near real time imaging. This can be accomplished with the aid of parallel hardware accelerators since the correlations between detector pairs can be processed independent of each other. In our investigation, we addressed the challenge of terahertz interferometric imaging in potential security applications: near real-time imaging utilizing hardware-accelerated computing platforms, such as graphics processing units (GPUs) and field programmable gate arrays (FPGAs) for parallelizing the image construction algorithm.

The second project I was involved is the accurate solution of large-scale electromagnetic problems using the fast multipole method (FMM) on a heterogeneous CPU-GPU architecture. Accurate computation of electromagnetic scattering from large-scale objects is a challenging problem in the areas of target tracking, through-wall imaging, and many other applications since it requires full-wave analysis which is computationally intensive and memory-consuming. The Fast Multipole Method (FMM) is one of the well-known techniques developed to reduce this burden of full-wave techniques without a significant loss of accuracy. In our preliminary effort, the single-level FMM was implemented on GPU clusters and a good speedup over the CPU implementation was achieved. However, our previous work focused only on GPUs which can accommodate a small problem size due to the limitation of GPU memory. Hence, I developed an efficient hybrid framework of single-level FMM for large-scale scattering problems on a heterogeneous CPU-GPU cluster to overcome this limitation. By taking advantage of CPU and GPU computational capabilities and resources, this framework allows solving for a larger problem size by an order of magnitude (up to 10 million of unknowns) than our previous implementation. The framework is developed by combining MVAPICH2 (an implementation of MPI), POSIX pthreads and CUDA model for the CPU-GPU parallel programming. The FMM itself has multiple stages ranging from

near-field calculation to aggregation, translation and disaggregation. Each of these stages has different computational and memory requirement. The proposed parallelization scheme achieves optimally balanced workload distributions for those stages among the multi-core CPUs and GPU with a minimal communication overhead. The framework I developed have been successfully applied for electromagnetic modelling of electrically large 3-D structures such as printed antennas, for radar cross section calculation for arbitrarily shaped objects. In my future work, an extension of this framework will be addressed for the multi-level FMM on heterogeneous CPU-GPU clusters.

For the last three years at the CUA, I was working in the field of detection and tracking of human vital signs through the wall with the aids of compressive sensing and hardware-accelerated computation. Detection and tracking of human vital signs (motion, heart rate, respiration rate, etc.) in cluttered environments have been receiving much attention in numerous applications, such as disaster search-and-rescue, healthcare, law enforcement, and urban warfare. For instance, it is essential to search for human beings trapped in debris, wreckage of collapsed buildings after an earthquake or trapped in buildings on fire. Several different technologies have been developed for this purpose, including X-ray, optical, infrared, radio frequency (RF), and ultrasound. Of all these technologies, RF sensors have the unique abilities of operating day-and-night, under all weather conditions and in through-wall environments. General speaking, the RF-based system illuminates electromagnetic waves into an area/scene and examines the reflected signal in order to acquire the information of targets. There have been many approaches being used for this purpose which are mainly focused on continuous wave (CW) Doppler radar, impulse ultra-wide band (UWB) radar, or stepped-frequency (SF) CW radar. UWB and SFCW radar systems are both superior to CW radar systems due to their capability of localization and multiple subject monitoring. Nevertheless, SFCW radar possesses several advantages over the UWB radar system such as high reliability, stability, and relatively easy implementation. The stepped-frequency radar, however, suffers from long data acquisition time, which is not suitable for time-sensitive applications, since it requires a large bandwidth and large number of antenna elements to achieve high resolution. Therefore, I proposed a compressive sensing (CS) framework for the stepped-frequency continuous wave (SFCW) radar in which the sparse signal, i.e. the target space, is recovered from far fewer samples than required by the Shannon-Nyquist sampling theorem, through optimization. This framework enables a significant decrease of data acquisition time, and reduces the amount of data to be processed in the SFCW radar system. However, most CS reconstruction algorithms are computationally intensive that could potentially make them impractical for real-time applications. Therefore, I also implemented the CS reconstruction algorithms on hardware-accelerated platforms, such as GPUs, due to their powerful floating computational capabilities and massively parallel processor architecture. Furthermore, a novel scattering model which involves wave propagation and signal reception was also developed. This model enables an accurate representation of the scattering fields of the targets in a cluttered environment. However, the system has been validated on a simulation platform only. Practical situations are more complex where we need to consider several factors such as antenna gain, antenna radiation pattern, coupling effects among antenna elements, etc. The response latency of the CS reconstruction algorithms also needs to be verified and improved (if necessary) with the real-time system. The above issues will be addressed in my future research.

As a Research Associate at UVA since 2015, I have been involved in building a framework for performance evaluation of regular expression processing engines. Regular expression matching plays an important role in a variety of applications, such as bioinformatics (e.g. for analyzing genome sequence), data mining (e.g. for discovering sets of frequently items in large databases), network inspection (e.g. for detecting worm, analyzing traffic), etc. It is a challenging problem when dealing with large data and complex regular expression rulesets because pattern matching is a computationally intensive operation. Rulesets of large, complex regular expressions can produce large, complex automata, which requires massively high memory bandwidth with low-latency access to compute efficiently. In this research, we investigate performance of state-of-the-art regular expression processing engines across five main categories of computer architectures, i.e. CPU, XeonPhi, GPU, FPGA, Automata Processor (the industry's first non-Von Neumann computing architecture leveraging the parallelism found in DRAM technology). Performance evaluation is explored on many dimensions: “complexity” of regular expressions, the number of regular expressions, and multiple packets (streams) processing capability. This research aims to evaluate how different types of regular expressions perform on different architectures and identify architectural/computational advantages and bottlenecks for each of the architectures. This work is also extended to a wider range of automata-based applications rather than just limited to only regular-expression-based applications. My future research plan would include more in-depth analysis of micro-architectural bottlenecks to automata processing within individual hardware architectures. It is essential to investigate changes in performance of automata applications through generations of these hardware architectures as well as to explore automata processing engines on other architectures rather than CPU, GPU, XeonPhi, FPGA, and AP. Furthermore, I would continue to evaluate relative power efficiency of each of these architectures, instead of only performance. These findings may help motivate other researchers to further investigate the most efficient regular expression processing methods, especially in heterogeneous environments that involve combinations of these architectures.

My future research plan will be built upon my past and present research. One area of investigation in which I am currently working on at UVA and plan to further pursue can be broadly labeled “data mining on FPGAs”. Over the past few years, FPGA has demonstrated its incredible potential in variety of applications with high performance and energy efficiency compared to other computing platforms due to their very high computational density and ability to provide computing solutions tailored to each individual algorithm. This makes FPGAs fascinating choices for datacenters, where energy efficiency and power provisioning are critical. However, programming FPGA remains difficult (which is mainly at the register-transfer level) to many new users. Recent advances in high-level synthesis (HLS) allow programmers to use more conventional languages such as C/C++ and OpenCL on FPGAs. Typical examples are Intel FPGA SDK for OpenCL and Xilinx SDAccel. However, best practices for high-level programming, which enables efficient FPGA implementations, have not yet been thoroughly developed. This research has two main goals: (i) identify application areas where FPGAs have dramatically better performance than other architectures, e.g. CPUs and GPUs; and (ii) develop best-practice FPGA development guidelines for HLS C/C++ and OpenCL. Specifically, we aim to develop FPGA codes implementing association rule mining (ARM) which is a big data application of machine learning. ARM is a computationally-intensive data mining technique which identifies strong and interesting relations

among data items in databases. Four types of pattern mining methods, namely frequent itemset mining (in which the order of items in a transaction doesn't matter), sequential pattern mining (in which order matters), frequent tree mining (in which data have tree relationships), and frequent graph mining (in which data have graph relationships) will be developed. All these mining methods are widely used in industry and academia, so libraries of ready-to-use FPGA solutions, as well as in-depth micro-architectural benchmarks would be desirable and could benefit a large available market.

Furthermore, I have been teaming with professors from the Department of Electrical Engineering and Computer Science at the CUA, and the Department of Electrical and Computer Engineering at the University of Tennessee, Knoxville to coordinate and advance further my findings from the research I conducted at the CUA on detection and tracking of human vital signs in cluttered environments. A more realistic, complex indoor/outdoor environment consisting of multiple scatterers and walls with inhomogeneity will be investigated. Due to the complexity of the problem, new compressive sensing reconstruction algorithms are necessary for accurate and efficient solution. A combination of hardware and software solutions is developed to achieve this objective. We also further extend the proposed system to health monitoring applications, such as fall detection for elderly people.