Abu Asaduzzaman | Research Statement

My current research focus includes high performance computer/computing (HPC) systems, machine learning (ML), HPC and ML in healthcare technology, and performance and power evaluation. I have been conducting innovative research to develop power-aware HPC systems since I was a PhD student at Florida Atlantic University (FAU). As a principle investigator (PI) at Wichita State University (WSU), I have received more than 15 research grants from external agencies including National Science Foundation (NSF), U.S. Department of Energy (DOE) Argonne National Laboratory (ANL), NVIDIA Corporation, and NetApp, Inc. I also have received several research grants from the WSU internal sources. I have published more than 24 refereed journal articles, three book chapters, more than 90 peer-reviewed conference papers, and more than ten technical articles out of my research work. I have been serving the related research communities as a panel reviewer, panel presenter, invited speaker, journal reviewer, conference organizer, technical/international program committee member, and numerous volunteer activities inside/outside my university.

In my Computer Architecture and Parallel Programming Laboratory (CAPPLab) at WSU, I have been conducting several important research projects. In the project funded by ANL (08/2022-12/2023), we model business applications on Exascale machines using the Portable, Extensible Toolkit for Scientific Computation (PETSc) and the Data Management Network (DMNetwork) libraries that are of interest to ANL and DOE. We generate realistic traffic data in Hierarchical Data Format 5 (HDF5) for selected cities/regions by using SUMO netconvert (a network simulator) and develop/update PETSc subroutines with a focus on parallel data processing. The resulting code development is included into the PETSc DMNetwork libraries benefiting the scientific community in large. In the project funded by DOE Visiting Faculty Program (VFP) at Lawrence Berkeley National Laboratory (LBNL) (06/2020-08/2021), we explore hardware acceleration through a fast-reliable methodology for generating graphic design system (GDS) files using Verilog code files of matrix-matrix multiplication (MMM). We synthesize the Pythonic Register-Transfer Level (PyRTL) generated Verilog code to solve general MMM problems using the OpenROAD toolchain to evaluate the impact of hardware acceleration on area and power consumption. The RTL-to-GDS flow on the OpenROAD nangate45 platform for various MMM Verilog files helps assess the chip area and power consumption with the matrix dimensions. The research outcomes are adopted by the Computer Architecture Group (CAG) of LBNL Applied Mathematics and Computational Research Division benefiting the research community in large.

I was first involved with academic research activities during my undergraduate study at Bangladesh University of Engineering and Technology (BUET), an ABET accredited institution. I developed a computer simulation platform using FORTRAN language to facilitate *"Computer Aided Design of Batwing Antenna Arrays by Standing Wave Modeling of Current Distribution."* This research work was published as a BUET Technical Report and it was accepted in partial fulfillment of the requirements for the BS degree in Electrical Engineering. I was so inspired by my undergraduate research work that I decided to go with Thesis option for my MS degree in Computer Engineering at FAU.

My MS thesis *evaluates memory latency of cluster-based cache-coherent multiprocessor systems with different interconnection topologies*. We focus on a cluster-based architecture which is a variation of Stanford Directory Architecture for SHared memory (DASH) architecture. In this architecture, snoopy protocol is used inside each cluster (as the number of processors per cluster is small). Clusters are connected using directory-based scheme through an interconnection network to make the system scalable. Trace-driven simulation has been developed using three different network topologies (ring, mesh, and hypercube). The overall memory latency has been evaluated by running a representative set of SPLASH-2 applications. Simulation results show that the cluster-based multiprocessor system with hypercube topology outperforms those with mesh and ring topologies. This work is published in the Elsevier Journal of Computers & Electrical Engineering.

My PhD dissertation *"Cache Optimization for Real- Time Embedded Systems"* demonstrates several effective solutions to select the right cache parameters to improve performance, decrease total power consumption, and significantly enhance execution time predictability. Our proposed Miss Table based cache locking scheme with victim cache is proven to be very effective for both single-core and multicore high-performance computer architectures running real-time applications. In this research work, we show that even though fundamentally memory is slower than the processing core and power-hungry cache introduces additional unpredictability, various cache-memory techniques including cache locking and cache optimization can be implemented to improve the performance and make the system more predictable and power-efficient.

During my PhD study, I worked on two Motorola-FAU Research Grants – Executable process Flow and One Pass to Production under the supervision of my PhD advisor. We developed simulation platform using VisualSim (a popular modeling and simulation tool) to evaluate embedded architectures for Motorola iDEN Group. We presented our work at Motorola and various conferences. Our work was appreciated by Motorola and was published in various journals including Springer Journal of Multimedia Tools and Applications (MTAP) and conference proceedings sponsored by the Institute of Electrical and Electronics Engineers (IEEE). I presented part of my research work in the 2005 FAU Graduate Research Competition and stood first (in oral presentation).

At WSU, I have served as the PI of Kansas NSF EPSCoR First Award research grant (2013-2014). My responsibilities include inventing novel task/data regrouping-based parallel solutions for multicore systems. The Compute Unified Device Architecture (CUDA) C/C++ parallel programming solutions developed in this project are being used by several university and industry research groups. I have served as the PI of NetApp Network File System (NFS) Connector project (2015-2016). My responsibilities in this NetApp project include developing NFS Connectors for Apache Spark systems and integrating project findings into educational materials. Other grants where I served as the PI include WSU URCA (2019-2020 and 2014-2015) and Wiktronics Embedded Systems Project (2014-2015). My research laboratory earned the NVIDIA Graphics Processing Unit (GPU) Research Center at WSU award (2015) and NVIDIA CUDA Teaching Center at WSU award (2012). Through various grants, I have supported more than 50 graduate and undergraduate students. I have supervised more than 15 PhD, 18 MS Thesis, 12 MS Project, and 16 undergraduate research students.

I have obtained U.S. Patents, published book chapters, and made Invited Guest Speaker presentations in the U.S., Japan, Turkey, Sri Lanka, and Bangladesh.

The continuous chase for high performance computer/computing systems (that are capable of sensing, monitoring, controlling, and communicating) has made the contemporary computing systems exceedingly complex (with heterogeneous, parallel, concurrent, and distributed subsystems). The complexity comes from both hardware and software, making it difficult, but not impossible, to further improving the performance to power ratio. Due to its tremendous potential, multicore architecture is being deployed in all sorts of modern computing devices. Similarly, more applications are being developed using multithreading techniques. We are developing HPC models with a communication-aware cache-mediator suitable for multithreaded applications to investigate energy-efficient scalable performance of multicore/many-core systems. We are also applying ML and deep learning techniques to analyze heterogeneous systems for real-time edge computing.

In conclusion, I am eager to continue working with students and collaborators from various institutions, write proposals to bring external grant money, and make significant contributions with researching, publishing, and supervising students. From five years from now, I picture myself as a distinguished research professor at *Wichita State University*.