Tanzima Islam

Assistant Professor
Texas State University

601 University Drive
San Marcos, TX, 78666
tanzima@txstate.edu
www.tanzimaislam.com

Research Objective

My research involves both modeling compute performance and designing scalable data movement strategies for large-scale applications in HPC environments. In the short-term, I am interested in leveraging data-driven analysis approach for comparative modeling with a focus on HPC co-design. My long-term research objective revolves around developing software solutions to ensure scalable performance under resilience and power constraints for large-scale systems.

Education

2013 **Ph.D., Computer Engineering**, *Purdue University*.
Reliable and scalable checkpointing systems for distributed computing environments.
Co-advisors: Saurabh Bagchi, Rudolf Eigenmann.

2006 **B.Sc., Computer Science & Engineering**, *Bangladesh University of Engineering & Technology*.

Professional Experience

2019-Present **Assistant Professor**, *Department of Computer Science*, Texas State University.

Summer 2019 **Visiting Scholar**, *Lawrence Berkeley National Laboratory*, Computation Directorate.

2017-2019 **Assistant Professor**, *Department of Computer Science*, Western Washington University.

2013-2017 **Postdoctoral Research Staff Member**, *Lawrence Livermore National Laboratory*, Center for Applied Scientific Computing.
Developed machine-learning techniques for performance analysis.

2006-2007 **Member, Research and Development**, *Commlink Info Tech Ltd.*, Bangladesh.
Developed software for a service-independent telecommunication network (Intelligent Network).

Awards & Honors

2016, 2015, 2014 Best Poster Award, Lawrence Livermore National Laboratory Annual Scholars Poster Symposium.

2014 **LLNL Director's Science & Technology Award.**

2014 2^{nd} Place Winner, LLNL Computation Directorate Postdoctoral Poster Symposium.

2012, 2009 Best Student Paper Nominations, International Conference for High Performance Computing, Networking, Storage and Analysis (SC).

2010 2^{nd} Place Winner, ACM Student Research Competition, Grace Hopper Celebration of Women in Computing.

Travel Awards Sustainable Research Pathways Program at Lawrence Berkeley National Laboratory'18.
CRA-W Career Mentoring Workshop at Phoenix, AZ'18.
Google Computer Science Grad Forum'12.
SC'09—'11, HPDC'12.

Research and Other Funding

Summer 2019 **PI**, Proxy Application Validation for Exascale Co-design, Visiting Faculty at Lawrence Berkeley National Laboratory, $40K.

2018, 2019 **PI**, Parallel Computing course, Time allocation grant from XSEDE, 100,000 core-hours.

2018 **Co-PI**, Scientific Data Visualization course development, Office of Research and Sponsored Programs at Western Washington University, $12K.

2016 **PI**, VERITAS for Understanding Performance Evolution during Code Development, Linking Exploratory Application Research to Next-gen Development at Lawrence Livermore National Laboratory, $200K/1year.

Publications

Peer Reviewed Conference & Journal Papers

[1] Tapasya Patki, Jayaraman J. Thiagarajan, Alexis Ayala, and **Tanzima Islam**. Performance optimality or reproducibility: that is the question. In *International Conference for High Performance Computing, Networking, Storage and Analysis (SC)*, November 17-22 2019. Acceptance rate: 20%.

[2] Jayaraman J. Thiagarajan, Rushil Anirudh, Bhavya Kailkhura, Nikhil Jain, **Tanzima Islam**, Abhinav Bhatele, Jae-Seung Yeom, and Todd Gamblin. PADDLE: Performance Analysis using a Data-driven Learning Environment. In *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, May 21-25 2018. Acceptance rate: 24.5%.

[3] Teng Wang, Adam Moody, Yeh Zhu, Kathryn Mohror, Kento Sato, **Tanzima Islam**, and Waikuan Yu. MetaKV: A Key-Value Store for Metadata Management of Distributed Burst Buffers. In *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pages 1174–1183, May 2017. Acceptance rate: 22%.

[4] **Tanzima Islam**, Kathryn Mohror, and Martin Schulz. Exploring the MPI Tool Information Interface: Features and Capabilities. In *The International Journal of High Performance Computing Applications (IJHPCA)*, volume 30, pages 212–222, 2016.

[5] Tania Banerjee, Jason Hackl, Mrugesh Shringarpure, **Tanzima Islam**, S Balachandar, Thomas Jackson, and Sanjay Ranka. CMT-Bone — A Proxy Application for Compressible Multiphase Turbulent Flows. In *IEEE 23rd International Conference on High Performance Computing (HiPC)*, pages 173–182, Dec 2016. Acceptance rate: 23%.

[6] **Tanzima Islam**, Jayaraman J. Thiagarajan, Abhinav Bhatele, Martin Schulz, and Todd Gamblin. A Machine-Learning Framework for Performance Coverage Analysis of Proxy Applications. In *International Conference for High Performance Computing, Networking, Storage and Analysis (SC)*, Salt Lake City, UT, November 13-18 2016. Acceptance rate: 23%.

[7] Lee Savoie, David K. Lowenthal, Bronis R. de Supinski, **Tanzima Islam**, Kathryn Mohror, Barry Rountree, and Martin Schulz. I/O Aware Power Shifting. In *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pages 740–749, May 2016. Acceptance rate: 23%.

[8] Anup Mohan, Thomas Hacker, Gregory P. Rodgers, and **Tanzima Islam**. Batchsubmit: A high-volume Batch Submission System for Earthquake Engineering Simulation. In *Concurrency and Computation: Practice and Experience*, volume 26, pages 2240–2252. Wiley Online Library, 2014.

[9] **Tanzima Islam**, Saurabh Bagchi, and Rudolf Eigenmann. Reliable and Efficient Distributed Checkpointing System for Grid Environments. In *Journal of Grid Computing (JoGC)*, volume 12, pages 593–613, Dec 2014.

[10] **Tanzima Islam**, Kathryn Mohror, Saurabh Bagchi, Adam Moody, Bronis R De Supinski, and Rudolf Eigenmann. McrEngine: A Scalable Checkpointing System Using Data-Aware Aggregation and Compression. In *Scientific Programming*, volume 21, pages 149–163. Hindawi, 2013.

* [11] **Tanzima Islam**, Kathryn Mohror, Saurabh Bagchi, Adam Moody, Bronis R. de Supinski, and Rudolf Eigenmann. McrEngine: A Scalable Checkpointing System Using Data-aware Aggregation and Compression. In *International Conference on High Performance Computing, Networking, Storage and Analysis (SC)*, pages 17:1–17:11, Los Alamitos, CA, USA, 2012. IEEE Computer Society Press. Acceptance rate: 20%. **Best student paper nomination.**

* [12] **Tanzima Islam**, Saurabh Bagchi, and Rudolf Eigenmann. FALCON: A System for Reliable Checkpoint Recovery in Shared Grid Environments. In *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis (SC)*, pages 1–12, 2009. **Best student paper nomination.**

[13] Hemayet Hossain, Mostofa Ahmed, Abdullah Al-Nayeem, **Tanzima Islam**, and Md Mostofa Akbar. gpNoCSim - A General Purpose Simulator for Network-On-Chip. pages 254–257, March 2007.

Workshop Papers

[14] **Tanzima Islam**, Alexis Ayala, Quentin Jensen, and Khaled Ibrahim. Towards A Programmable Analysis and Visualization Framework for Interactive Performance Analytics. In *Workshop on Programming and Performance Visualization Tools held in conjunction with International Conference for High Performance Computing, Networking, Storage and Analysis (SC)*, Denver, CO, November 16-23 2019. IEEE.

[15] Nicholas Majeske, Filip Jagodzinski, Brian Hutchinson, and **Tanzima Islam**. Low Rank Smoothed Sampling Methods for Identifying Impactful Pairwise Mutations. In *International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pages 681–686. ACM, 2018.

[16] Aiman Fang, Ignacio Laguna, Kento Sato, **Tanzima Islam**, and Kathryn Mohror. Fault Tolerance Assistant (FTA): An Exception Handling Programming Model for MPI Applications. 2015.

[17] **Tanzima Islam**, Kathryn Mohror, and Martin Schulz. Exploring the Capabilities of the New MPI_T Interface. In *Proceedings of the 21st European MPI Users' Group Meeting*, page 91. ACM, 2014.

[18] John Tramm, Andrew Siegel, **Tanzima Islam**, and Martin Schulz. XSBench-the Development and Verification of a Performance Abstraction for Monte Carlo Reactor Analysis. *The Role of Reactor Physics toward a Sustainable Future (PHYSOR)*, 2014.

Ph.D. Dissertation

[19] **Tanzima Islam**. *Reliable and scalable checkpointing systems for distributed computing environments*. PhD thesis, Purdue University, West Lafayette, IN, May 2013.

Research Posters

[20] Logan Moody, Nathan Pinnow, Michael Lam, Harshitha Menon, Markus Schordan, Scott G. Lloyd, and **Tanzima Islam**. Automatic Generation of Mixed-Prevision Programs. In *International Conference for High Performance Computing, Networking, Storage and Analysis (SC'18)*, 2018.

[21] Simone Smarr, **Tanzima Islam**, and Yolanda Rankin. Modular Extensible Framework for Performance Comparative Analysis. In *ACM Richard Tapia Celebration of Diversity in Computing Conference*, 2015.

[22] Xiang Ni, **Tanzima Islam**, Kathryn Mohror, Adam Moody, and Laxmikant V. Kale. Lossy Compression for Checkpointing: Fallible or Feasible. In *International Conference for High Performance Computing, Networking, Storage and Analysis (SC'14)*, 2014.

* [23] **Tanzima Islam**, Saurabh Bagchi, and Rudolf Eigenmann. Harnessing multiple cores for efficient checkpointing in grid systems. In *ACM Student Research Competition (SRC) at Anita Borg Institute Grace Hopper Celebration of Women in Computing*, 2010. **ACM Student Research Competition 2^{nd} prize winner.**

[24] **Tanzima Islam**, Kathryn Mohror, Adam Moody, Bronis de Supinski, Saurabh Bagchi, and Rudolf Eigenmann. Data-Aware Inter-Process Checkpoint Compression. In *International Conference for High Performance Computing, Networking, Storage and Analysis (SC'10)*, 2010.

[25] Mohammad S. Hossain, **Tanzima Islam**, Saurabh Bagchi, and Vijay Raghunathan. Fast and Collaborative Interference Avoidance for Wireless Medical Devices. In *International Conference on Dependable Systems and Networks (DSN)*, 2009.

Professional Activities

Organizer	Vice-Chair, Performance track, International Conference on Parallel Processing, 2019.
Technical Program Committees	International Conference for High Performance Computing, Networking, Storage and Analysis (SC) '17–'19 (Performance track). SC'19 (Posters). SC'18 (HPC for Undergrad), The Platform for Advanced Scientific Computing (PASC) '19, International Conference on Parallel Processing (ICPP) '18, International Symposium on Computer Architecture and High Performance Computing (SBAD-PAD) '13–'14.
Peer Reviewing	IEEE Transactions on Parallel and Distributed Systems (TPDS)'19, International Conference on Parallel Computing (ParCo)'19, Journal of Grid Computing (JoGC)17–'19, International Journal of High Performance Computing (IJHPCA)'18, IEEE International Parallel & Distributed Programming Symposium (IPDPS)'12, Symposium on Principles and Practice of Parallel Programming (PPoPP)'11, Dependable Systems and Networks (DSN) '10.
Memberships	ACM

Invited Talks and Panels

May, 2019	**The NIMBioS Workshop on Scientific Collaboration Enabled by High Performance Computing**, *Scalable I/O Performance for Scientific Applications–Challenges and Potentials*, Knoxville, TN.
December, 2018	**Lawrence Berkeley National Laboratory**, *Understanding the Performance Portability Challenges and Opportunities using Machine Learning*, Berkeley, CA.
November, 2018	**HPC for Undergraduates**, *International Conference for High Performance Computing, Networking, Storage and Analysis*, Dallas, TX.
July, 2018	**The Platform for Advanced Scientific Computing Conference**, *Machine Learning Framework for Performance Coverage Analysis of Proxy Applications*, ACM and the Swiss National Supercomputing Center (CSC), Basel, Switzerland.

February, 2015 **JOWOG-34**, *Proxy Application Validation using Machine Learning (Veritas)*, Sandia National Laboratories, Albuquerque, NM.

Teaching Experience

2017 - Present Parallel Computing, Data Structures, Computer Networks, Scientific Data Visualization (co-instructor).

Outreach and Mentoring Students

2014-Present **Co-founder**, *Bangladeshi Women in Computer Science and Engineering*, Dhaka, Bangladesh, First research, networking, and mentoring platform for female Computer Science and Engineering students of Bangladesh.

2017–Present **Research advisor**, *Undergraduate (7) and graduate (3) students*, Western Washington University.

2017-Present **Faculty advisor**, *ACM-W chapter at Western Washington University*.

2014–2016 **Intern supervisor**, *Undergraduate, M.S. and Ph.D. students from University of Hamburg, University of Illinois Urbana-Champaign, Spelman College, University of California San Diego*.

2011–2013 **Student mentor**, *Undergrad, M.S., and Ph.D. students*, Purdue University.

Five Most Significant Research Projects

[Numbers] refer to items listed in the Publications section earlier

2018-Present **Performance-Aware Application Development using Machine Learning**
The objective of this research is to predict performance of applications **before** it is fully written to guide programmers develop better performant code. Specifically, we investigate various machine-learning techniques to correlate the structural information of applications such as Abstract Syntax Trees to the changes in their performances [20]. This research will make the application development process be performance-aware and will reduce the need for thousands of developer-hour spent afterwards in performane optimization.

2018-Present **Scalable Framework for Mutant Generation in Proteins**
The objective of this research is to facilitate the in-silico protein mutation generation with exhaustive 2 and 3 amino acid substitutions. In that direction, we will develop a scalable parallel framework for mutation generation, and scalable algorithms for in-situ analysis. Most notable, we have developed machine-learning algorithms to gain insights into the resilience of protein structures subjected to a gamut of conditions [15]. In this project, we will develop software technologies to enable discoveries in the field of precision medicine.

2013-2017 **Performance Analysis Techniques for Effective Co-design in HPC**
In this project, we have developed machine-learning models to quantitatively compare the performance of proxy applications to their production counterparts. Application teams in several organizations such as LLNL, the NSF PSAAP Center for Compressible Multiphase Turbulence, and Atomic Weapons Establishment (AWE) are using the machine-learning frameworks developed in this project (Veritas [6], PADDLE [2]). The most notable success of this research has been informing the NSF co-design center about a performance problem that existed in their proxy for the Gordon Bell winning application Nek5000 [5].

2010-2012 **Scalable Checkpointing System for HPC**
We have developed a scalable checkpointing system for HPC environments. The contributions include an algorithm for compressing non-uniform, mixed-precision scientific data, and a scalable framework that can be readily deployed in any HPC system without user intervention. The novel algorithm disrupted the long-standing notion in the data compression community by compressing scientific data up to **4x** compared to their original size, which results in up to **67% reduction** in recovery time and up to **87% reduction** in checkpointing time. This project resulted in several publications [10, 11], a best student paper nomination at SC'12, and received the "Director's Science & Technology Award" at LLNL for its impact in significantly improving the scalability of HPC systems.

2007-2009 **Reliable and Efficient Checkpointing for Cycle-Sharing Systems**
In this project, we have developed a reliable and efficient distributed checkpointing system for cycle- sharing systems. The contributions include a novel failure model for predicting the availability of volatile grid resources; balancing the load on these machines; transferring data efficiently through the shared network; and a scalable framework that can be transparently deployed in any High-Throughput Computing (HTC) environment. Our solution eliminates the need for dedicated checkpoint servers, reduces contention for shared resources such as network bandwidth and improves the performance of applications up to **44%** on a shared-grid environment. This research resulted in several publications and a best student paper nomination at SC'09 [12, 9, 23].