Md Shihab Shahriar Khan

Stack Overflow: /u/4553309/
GitHub: /shihab-shahriar
Email: khanmd@msu.edu
Phone: (+1) 517-775-9450

EXPERIENCE

Los Alamos National Laboratory — NM, USA
Graduate Research Intern — June 2024 –Aug. 2024

- Completed the development of an open-source Hydrodynamics and Radiative Diffusion Solver (HARD) within this timeframe, achieving performance comparable to state-of-the-art solutions. For more details, please see the research poster presented at Supercomputing'24. (**link**)
- Led the parallelization and GPU acceleration of the framework as part of a diverse team of astrophysicists and mathematicians, utilizing C++, Kokkos, MPI, CUDA, and Nsight profilers.

Michigan State University — MI, USA
Research Assistant II — Mar. 2022 – Present

- Developing a performance-portable simulation framework for particulate flows, designed for seamless scaling across multi-node systems and on-node with GPU acceleration.
- Exploring how machine learning techniques can accelerate the compute-intensive evaluation of N-body hydrodynamic interactions. Tech stack includes C++, CUDA, OpenMP/MPI, Kokkos, Python, Pytorch.

Samsung R&D Institute Bangladesh — Dhaka, Bangladesh
Machine Learning Engineering Intern, Advanced Research Group — Jan. 2018 – Jun. 2018

- Worked primarily in computer vision, particularly image classification, inference on edge etc.
- Technologies included tensorflow/keras, numpy, pandas etc.

PUBLICATIONS

1. **Khan, Shihab S**, Bryce Palmer, Christopher Edelmaier, and Hasan M Aktulga. "OpenRAND: A performance portable, reproducible random number generation library for parallel computations." SoftwareX 27 (2024): 101773.

2. **Shihab Shahriar Khan**, Nishat Tasnim Niloy, Md. Aquib Azmain and Ahmedul Kabir. "Impact of Label Noise and Efficacy of Noise Filters in Software Defect Prediction". *International Conference on Software Engineering and Knowledge Engineering (SEKE)* KSIR Virtual Conference Center, USA, 2020.

3. Asad, Moumita, Rafed Muhammad Yasir, **Shihab Shahriar Khan**, Nadia Nahar, and Md Nurul Ahad Tawhid. "Analyzing Program Comprehensibility of Go Projects." In SEKE, pp. 255-260. 2021

EDUCATION

- **Michigan State University** — MI, USA
  Ph.D. in Computer Science & Engineering — Aug. 2021 – Present
  - CGPA: 3.92/4.00 (completed 18 credits)

- **Institute of Information Technology, University of Dhaka** — Dhaka, Bangladesh
  M.S. in Software Engineering — Jan. 2019 – Dec. 2020
  - Thesis Title: "Noise Robust Classification Using Instance Level Analysis". CGPA: 3.83/4.00

- **Institute of Information Technology, University of Dhaka** — Dhaka, Bangladesh
  B.S. in Software Engineering — Jan. 2015 – Dec. 2018
  - CGPA: 3.68/4.00

Technical Skills

High Performance Computing
C/C++, CUDA, Kokkos, OpenMP/MPI, Trilinos, CMake/Make, Paraview/VTK, NSight Systems/Compute

Machine Learning and Data Science
Python, Pytorch, Numpy/Pandas, R, Scipy, Scikit-learn, SQL, Tensorflow

Common Skills and Tools
Advanced data structures, Technical writing, Git, Linux, CI/CD, Sphinx/Doxygen, Haskell, Java

Projects

OpenRAND (Link)
Reproducible random number generation for parallel computations (Lead Developer)   Sept. 2023 – Present
- An open-source, statistically robust C++17 library aimed at facilitating reproducible scientific research through the generation of reproducible, parallel random number streams
- Performance-portable, designed to work seamlessly across various software and hardware platforms (including GPUs.) and often faster than native libraries like libstdc++, Nvidia's Curand or rocRAND.
- Merged into HOOMD-blue, slated for inclusion in Mundy (from Flatiron Institute).

HARD (Link)
Hydrodynamics And Radiative Diffusion Solver based on FleCSI framework   June 2024 – Present
- The first publicly available code coupling radiation and hydrodynamics that is highly scalable and portable to heterogeneous HPC architectures.
- GPU acceleration led to around 15 times performance improvement over a 64-core intel CPU on a A100 GPU, resulting in more than twice cell updates per second in single node configuration over best known literature result.

Scikit-clean (Link)
A python ML library for classification in the presence of label noise.   Dec. 2019 – Dec. 2021
- Tools to simulate artificial noise, create complex pipelines and evaluate them
- scikit-learn API compatible- all scikit-learn's building blocks can be seamlessly integrated into workflow
- Equipped with units tests, extensive documentation, CI pipeline and available in pypi.

DeepPaint (Link)
Deep learning based computer vision tool to automatically colorize and stylize paintings   Jul. 2018 – Nov. 2018
- Colorizes a sketch using color hints. Stylizes a colored painting along the style of any given painting. Implemented using python, pytorch, pyqt.

Activities

Open Source
Several code-level pull requests accepted in projects like Kokkos, Spack, Imbalance-learn, and MLAlgorithms. OpenRAND is currently part of HOOMD-blue, slated for inclusion in Mundy (from Flatiron Institute), possibly Stan.

Technical Writing Samples
- **StackOverflow**: Currently over 5k "reputation", mostly in Python, Scikit-learn, Pytorch and Algorithm tags.
- **Article**: "CUDA vs ROCm: A Case Study"
- **Article** in Towards Data Science titled "An Introduction to Classification Using Mislabeled Data".