Santosh Pandey

Personal webpage | Github | LinkedIn | santosh.pandey@rutgers.edu

Research Interest

My research interests lie at the intersection of systems and ML, focusing on full circle innovation–using ML to design better systems (more current focus) and building optimized systems to advance ML/HPC applications. Notably, I have devised novel techniques for creating accurate and reusable ML-based microarchitecture simulators, while also accelerating and scaling these simulators for performance modeling with hardware software co-design, demonstrating the synergy between system optimization and ML.

Education

Ph.D. in Computer Engineering — Spring, 2023 – Present
Rutgers University, NJ — *Advisor: Dr. Hang Liu*

Ph.D. in Computer Engineering — Aug, 2019 – Dec 2022
Stevens Institute of Engineering, NJ [Master's Degree Awarded] — *Advisor: Dr. Hang Liu*

Bachelor's in Computer Engineering — 2017
Tribhuvan University, Nepal — *Advisor: Dr. Subarna Shakya*

Professional Experiences

Student Researcher @ System Research Group | Google — May 2024 – August 2024
Mentor/Host: Victor Lee, Amir Yazdanbakhsh, Mohammad Alizadeh

- Worked on accurate/explainable/generalizable ML techniques for microarchitecture simulation to support design space exploration.

Research Intern | Brookhaven National Lab — May 2020 – August 2022
Mentor/Host: Lingda Li, Adolfy Hoisie

- Worked on developing Deep Learning-based architecture simulation for CPU and GPU architectures.
- Accelerated Deep Learning-based microarchitecture simulators with workflow optimizations on GPU, distributed and parallel simulation providing >1000× speedup.

Research Intern | Lawerence Berkeley National Lab — May 2019 – August 2019
Mentor/Host: Xiaoye Sherry Li

- Designed scalable GPU-accelerated graph triangle counting achieving unprecedented throughput beyond 100 billion TEPS rate; Won 2019 MIT/DARPA Graph Challenge Award for the work.

Selected Publications

- **S. Pandey**, A. Yazdanbakhsh, H. Liu. *TAO: Re-Thinking DL-based Microarchitecture Simulation.* Proceedings of the ACM on Measurement and Analysis of Computing Systems (**SIGMETRICS**). ACM, 2024.
- **S. Pandey**, L. Li, T. Flynn, A. Hoisie, H. Liu. *Scaling Deep Learning-based Microarchitecture Simulation on GPUs.* In International Conference for HPC, Networking, Storage and Analysis (**SC**). ACM, 2022.
- **S. Pandey***, Z. Wang* and others. *TRUST: Triangle Counting on GPUs.* In the Transactions on Parallel and Distributed Systems (**TPDS**). IEEE, 2021.
- **S. Pandey**, L. Li, A. Hoisie, X. Li and H. Liu. *C-SAW: A Framework for Graph Sampling and Random Walk on GPUs.* In International Conference for HPC, Networking, Storage and Analysis (**SC**). ACM, 2020.
- Other publications: Google scholar

Patent

- **S. Pandey**, H. Liu. "Accelerating Microarchitecture Simulation with Machine Learning" (RU Docket 2024-101). U.S. Provisional Application 63/539,950 on April, 2024.

Programming Skills

Languages: Python, C / C++, CUDA **Infrastructure tools:** Docker, Git
Machine Learning Frameworks: LibTorch, TensorRT, PyTorch, JAX

Awards and Professional Activities

Awards: 2022 IEEE TCHPC student travel award, 2019 MIT Graph Challenge Competition
Reviewer: TPDS'24, ISCA MLArchSys'24, OSDI'24 AE, ATC'24 AE, Micro'24 AE, PPoPP'22 AE
Workshop Program Committee: IEEE BigData'23 GTA^3,BigData'22 GTA^3