Raveesh Garg

Curriculum Vitae

266 Ferst Drive NW
Klaus Adv Computing Bldg Rm 3305
Atlanta, GA, 30332
raveesh.g@gatech.edu
sites.gatech.edu/raveesh/

Research Interests

Computer Architecture

Programmable Spatial Accelerators

Accelerators for Artificial Intelligence and Scientific Computing

Education

2021-Present **Georgia Institute of Technology**, *PhD in Electrical and Computer Engineering*, Atlanta, GA, USA
- Advisor: *Dr. Tushar Krishna*
- Broad Research Area: *Hardware Accelerators for AI and HPC.*
- GPA - *4/4*

2019-2021 **Georgia Institute of Technology**, *Master of Science in Electrical and Computer Engineering*, Atlanta, GA, USA
- Advisor: *Dr. Tushar Krishna*
- Master's Thesis: *Understanding the Design Space of Dataflows for Graph Neural Network Accelerators.*
- GPA - *4/4*

2015-2019 **Birla Institute of Technology and Science, Pilani**, *Bachelor of Engineering in Electronics & Instrumentation Engineering*, Pilani, Rajasthan, India
- GPA - *9.29/10*

Skills

Programming Verilog, C/C++, Assembly Language, Python, MATLAB/Octave

Simulators and EDA Tools gem5 garnet on-chip network simulator, Structural Simulation Toolkit (SST), SESC SuperScalar simulator, Xilinx ISE and Vivado, ModelSim, Icarus iverilog, Cadence Encounter RTL Compiler, SPICE, Cadence Virtuoso, Synopsys Design Vision, Cadence Innovus.

Publications and Pre-prints

IPDPS 2022 **(Best Paper Nominee)** *Raveesh Garg*, Eric Qin, Francisco Muñoz-Martínez, Robert Guirado, Akshay Jain, Sergi Abadal, José L Abellán, Manuel E Acacio, Eduard Alarcón, Sivasankaran Rajamanickam, and Tushar Krishna. "Understanding the Design-Space of Sparse/Dense Multiphase GNN dataflows on Spatial Accelerators", 36th IEEE International Parallel & Distributed Processing Symposium (IPDPS 2022)

arXiv 2022 Eric Qin, *Raveesh Garg*, Abhimanyu Bambhaniya, Michael Pellauer, Angshuman Parashar, Sivasankaran Rajamanickam, Cong Hao, and Tushar Krishna. "Enabling Flexibility for Sparse Tensor Acceleration via Heterogeneity." arXiv preprint arXiv:2201.08916 (2022).

INDICON 2018 *Raveesh Garg* and Karri Babu Ravi Teja, "A High-Speed Pipelined Architecture for Block Motion Estimation Using Hexagon- Based Search Algorithm," 2018 15th IEEE India Council International Conference (INDICON), Coimbatore, India, 2018

Research Projects

May 2022 - Present **Mapping Exploration for AR/VR workloads**
- Mapping Optimization and Cost Modelling for AR/VR workloads on DNN Accelerators.

Nov 2021 - Present **Hardware Accelerator for High Performance Computing Applications**
- Currently pursuing research on microarchitecture of an accelerator for High-Performance Computing.

Sept 2020 - Oct 2021 **Dataflow Design-Space Exploration for GNN Accelerators**
- Proposed a taxonomy for description of dataflows capturing the dataflows of individual phases SpMM and DenseGEMM and pipelined parallelism between the two phases and encoded it into a simulation framework OMEGA.
- OMEGA uses STONNE simulator to model GEMM and SpMM individually and an analytical model to compute pipelined statistics from individual kernel statistics.

Feb 2020 - Apr 2020 **Inter-layer parallelism for DNN Acceleration**
- Analytically modelled the impact of mapping parallel branches of layers on a reconfigurable accelerator (MAERI) on the amount of data reuse and bandwidth requirement of the accelerator.

Internship Experience

Aug 2022 - Nov 2022 **Meta Reality Labs**, *Part-time Student Researcher*, (Remote) Atlanta, GA, USA
- Research Project: Mapping Optimization and Cost Modelling for AR/VR workloads on DNN Accelerators.

May 2022 - Aug 2022 **Meta Reality Labs**, *Research Scientist Intern*, Sunnyvale, California, USA
- Research Project: Mapping Optimization and Cost Modelling for AR/VR workloads on DNN Accelerators.

May 2018 - July 2018 **Defense Research and Development Organization**, *Research and Development Intern*, Hyderabad, Telangana, India
- Project: Implemented a Fault Tolerant NOR Flash Memory controller and a branch predictor for an indigenous processor in RTL.

Workshop, Tutorials and Talks

ModSim 2022 **Workshop on Modeling & Simulation of Systems and Applications 2022**
SST-STONNE: Enabling cycle-level simulation of flexible spatial accelerators for DNNs and GNNs with a detailed memory hierarchy.

ASPLOS 2022 **Young Architect Workshop 2022**
A Communication-Centric Dataflow Accelerator for High-Performance Conjugate Gradient.

ASPLOS 2022 **Tutorial: STONNE+OMEGA: Cycle-level Simulation of Dense/Sparse DNN and GNN Accelerators**, *(Organizer and Presenter)*
Delivered a talk about dataflow design-space exploration of Graph Neural Network mappings and demo and hands-on exercise on the OMEGA framework that models the metrics for GNN dataflows.

SIAM PP22 **Minisymposium: Co-Design of Data Flow Accelerators for Scientific Simulations and Machine Learning**, *(Presenter)*
Discussed the design-space of dataflows for multiphase kernels with sparse and dense computations like GNNs in a minisymposium at SIAM PP22.

Honors and Awards

IPDPS 2022 Best Paper Award Nomination (First authored). Top 5/474 submissions.

Service

HPCA 2022 Artifact Evaluation PC Reviewer