Shreya

+1 (801) 859-8332 | 296shreya@gmail.com | github.com/Shreya

Research interests

HPC, AI Compilers, Program analysis and optimization

Research Experience

TritonComm May 2024 – Present

Project Advisor: Angélica Moreira, Roshan Dathathri, Ponnuswamy Sadayappan, Atanas Rountev

- Collective communication support in Triton using MSCCLPP

Optimizing Sparse Tensor Computation using MLIR March 2023 – Present

Advisor: Professor P. (Saday) Sadayappan

- Creating custom abstraction for optimization of sparse tensor in MLIR

Precise and efficient point-to analysis in LLVM IR Mar 2021 – June 2022

Advisor: Professor Uday Khedker and Professor Swati Jaiswal

- Worked on making points-to analysis for live variables more efficient in handling function calls
- Modelled LLVM instructions for retaining precision in points-to analysis
- Abstracted LLVM IR based on pointer suitable modeling
- bi-directional value-context for context-sensitive variant

Education

University of Utah Salt Lake city, US

PhD CSE, advised by Professor P. (Saday) Sadayappan *August 2022 – Present*

ACM India Summer School 2021 Hyderabad, India

on Programming Language Analysis and Optimization *July 2021*

Rajkiya Engineering College Sonbhadra Sonbhadra, India

B.Tech ECE *August 2016 – July 2020*

Publications

PEAK: Generating High-Performance Schedules in MLIR, Amir Mohammad Tavakkoli, Sameeran Joshi, **Shreya Singh**, Yufan Xu, P. Sadayappan, and Mary Hall. In Proceedings of the 36th International Workshop on Languages and Compilers for Parallel Computing (LCPC23), Oct 2023

Award

Graduate Research Fellow, University of Utah *2022-2023*

Experience

Teaching Mentor Aug 2024 – Dec 2024

University of Utah *Salt Lake City, Utah*

- High Performance Computing course

Microsoft Research Intern May 2024 – Aug 2024

Microsoft Research *Redmond, WA*

- Mentors: Angélica Moreira, Roshan Dathathri
- Worked on supporting collective communication in Triton using MSCCLPP

Eklavya Summer Intern May 2018 – July 2018

IIT Bombay *Mumbai, India*

- Advisor: Professor D.B. Phatak
- Secured 21st rank nationwide in the eklavya software quota contest
- Developed a system using the open source tool DSpace for archiving high loads of generated data from collaborative community services
- Deployed the system on docker containers and contributed to the middle-ware

Projects

DSL for abstracting transform dialect in MLIR Jan 2023 – April 2023

- Advisor: Professor P. (Saday) Sadayappan
- Created an embedded DSL in C++ for abstracting the complexities of transform dialect in MLIR.
- developed various optimizations for Matrix Multiplication in proposed DSL

Optimizing Matrix Multiplication: Parallelization Strategies with OpenMP and CUDA Oct 2022 – Dec 2022

- Advisor: Professor P. (Saday) Sadayappan
- Worked on optimizing matrix multiplication via OpenMP and CUDA, using various strategies like loop unrolling, tiling with different tile sizes, and shared memory optimization. Used Nsight for performance analysis.

Modeling Implementation and Optimization in LLVM Aug 2021 – April 2022

- Advisor: Professor Uday Khedker
- Automatically generates abstract representation of LLVM IR
- Uses textual description of the modeling as input
- Eliminates the need of finding manual patterns in the IR by conservatively replacing uses with defs for the LLVM generated temporaries
- Optimizes the representation to match source level constructs

Live Variable Analysis in LLVM March 2021

- Implemented intra-procedural live variable analysis in LLVM
- Upgraded to an inter-procedural version

Strong Live Variable Analysis in LLVM March 2021

- Implemented intra-procedural and inter-procedural strong live variable analysis in LLVM

Modelling LLVM IR instructions Feb 2021 – Mar 2021

- Modelled LLVM IR instructions into a small grammar close to C
- Handled memory instructions, Load, Store, Alloca, BitCast, and GetElementPointer
- Binary arithmetic operations, Add, Mul, Sub, and UDiv

Early Flood Detection and Avoidance System Aug 2019 – Oct 2019

- Capable of sending early alert messages through social media
- We used NodeMCU platform

Skills

C, C++, Python, LLVM, MLIR, Racket, CUDA, OpenMP, MPI

Courses

Parallel Programming, Advance Parallel Programming Many-Core, Operating Systems, High Performace Machine learning Seminar, Programming Languages, Graduate Algorithms