Sana Damani

Website: www.sanadamani.com LinkedIn: www.linkedin.com/in/sanadamani

SUMMARY

A PhD student with engineering experience seeking a research role at the intersection of compiler and architecture design where I can make an impact on the future of high performance computing in a collaborative and fast-paced work environment.

EDUCATION

PhD, Computer Science Aug 2017-Present
Georgia Institute of Technology
Advisor: Dr. Vivek Sarkar
Thesis: Instruction and Work Scheduling for Thread-Parallel Architectures

Bachelor of Engineering May 2012
Computer Science
University of Pune

SKILLS

- Compiler optimizations
- GPU architecture
- Performance analysis
- LLVM, MLIR
- Programming in C, C++

PUBLICATIONS

Memory Access Scheduling to Reduce Thread Migrations
S.Damani, P.Barua, and V.Sarkar.
Under submission. Copy available on request.

GPU Subwarp Interleaving
S.Damani, M.Stephenson, R.Rangan, D.R.Johnson, R.Kulkarni, and S.W.Keckler.
To appear in the 28th IEEE International Symposium on High-Performance Computer Architecture (HPCA 2022).

Speculative reconvergence for improved SIMT efficiency
S.Damani, D.R.Johnson, M.Stephenson, S.W.Keckler, E.Yan, M.McKeown, O.Giroux.
In Proceedings of the 18th ACM/IEEE International Symposium on Code Generation and Optimization (CGO 2020).

Common Subexpression Convergence
S.Damani and V.Sarkar.
32nd Workshop on Languages and Compilers for Parallel Computing (LCPC 2019).

OTHER PROJECTS

- Run-time branch divergence detection for optimal per-warp branch handling in the LLVM-based HARP compiler for the SIMD HARMONICA architecture, advised by Dr. Hyesoon Kim.
- Compiler support for a novel Sparse Matrix Engine to accelerate Sparse DNNs on CPUs, in collaboration with the Synergy lab led by Dr. Tushar Krishna.
- MLIR tutorial, Workshop on MLIR for HPC, 2019.

PATENTS

Convergence among concurrently executing threads
D. Johnson, J. Choquette, O. Giroux, M McKeown, M.W. Stephenson, S. Damani
Publication Date: 03/12/2020

TEACHING EXPERIENCE

Graduate Teaching Assistant Fall 2020
Georgia Institute of Technology
- Co-instructor for a graduate level course on Parallelizing Compilers (CS 6245) with Dr. Vivek Sarkar.

Graduate Teaching Assistant Spring 2018
Georgia Institute of Technology
- Teaching assistant for a graduate level course on GPU Architecture (ECE 8823) with Dr. Sudhakar Yalamanchili.

WORK EXPERIENCE

Graduate Research Assistant 2017-2022
Georgia Institute of Technology
- Research on instruction and work scheduling for parallel architectures.

Graduate Intern Summer 2021
Architecture Research Group, Nvidia
- Developed a dynamic register allocation technique to increase warp occupancy.

Graduate Intern Summer 2020
Architecture Research Group, Nvidia
- Designed and implemented subwarp interleaving, a hardware feature to improve warp latency of divergent programs on Nvidia GPUs.

Graduate AI Intern Summer 2019
TensorFlow Team, Intel
- Enabled quantization-aware training on object detection programs in the TensorFlow toolchain to improve the accuracy of quantized inference.
- Developed a prototype MLIR dialect for quantization optimizations and designed a toolchain for integration of Intel's TensorFlow tools into the MLIR compiler.

Graduate Intern Summer 2018
GPU Compiler Team, Nvidia
- Designed and implemented speculative reconvergence, a compiler optimization that improved SIMT efficiency and run-time for programs with divergent execution on Volta GPUs. Collaborated with the engineering team to help with productization.

Senior System Software Engineer 2012-2017
GPU Compiler Team, Nvidia
- Designed compiler back-end analyses and transformations, including dead store elimination, redundant type conversion elimination, memory dependence analysis and early exit, to improve the performance of graphics and compute workloads on mobile and desktop GPUs.

Undergraduate Intern 2011-2012
GPU Compiler Team, Nvidia
- Implemented register allocation and register promotion passes in LLVM.