

Snehil Verma

MASTER'S STUDENT · DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

2501 Speedway, EER 5.860, Austin, TX 78712, United States

☎ (+1) 737-217-5056 | ✉ snehilv@utexas.edu | ✉ snehilverma41@gmail.com | 🏠 snehilverma41.github.io



Education

The University of Texas at Austin

4.0*/4.0

M.S. IN ELECTRICAL AND COMPUTER ENGINEERING

Fall 2018 – Spring 2020

TRACK: ARCHITECTURE, COMPUTER SYSTEMS, AND EMBEDDED SYSTEMS (ACSES)

Indian Institute of Technology, Kanpur

8.9/10

B.TECH IN ELECTRICAL ENGINEERING (WITH DISTINCTION)

Fall 2014 – Spring 2018

MINOR IN COMPUTER SYSTEMS, COMPUTER SCIENCE AND ENGINEERING

* Calculated at the end of Spring'19

Publications

- **S. Verma**, Q. Wu, B. Hanindhito, G. Jha, E. John, R. Radhakrishnan, and L. John, "Metrics for Machine Learning Workload Benchmarking," International Workshop on Performance Analysis of Machine Learning Systems (FastPath), In conjunction with *ISPASS*, March 2019. [Publication] [Presentation]
- R. Radhakrishnan, **S. Verma**, Q. Wu, B. Hanindhito, G. Jha, E. John, and L. John, "Demystifying Hardware Infrastructure Choices for Deep Learning Using MLPerf," *NVIDIA GPU Technology Conference (GTC)*, March 2019. [Presentation]
- **S. Verma**, N. Deshmukh, P. Agrawal, B. Panda, and M. Chaudhuri, "DFCM++: Augmenting DFCM with Early Update and Data Dependence-driven Value Estimation," 1st Championship Value Prediction (CVP-1), In conjunction with *45th International Symposium on Computer Architecture (ISCA 2018)*, June 2018. [Publication] [Presentation] [Code]

Experience and Projects

GPU Hardware Intern at Samsung SARC | ACL

Summer'19 – Present

- **PPA (Power, Performance, and Area)** mentored by *Raghavan R. Srinivasa*: Executed power and performance flows on Cadence's Palladium Z1 enterprise **SoC emulation** platform to identify the performance bottlenecks and the blocks using high power; developed **microbenchmarks** in OpenCL and OpenGL targeted at specific architectural and compiler features; initiated the research on **power prediction** based on machine learning
- **Architecture and Modeling** mentored by *Sushant Kondguli*: Delved into the design and working of **Texture Cache**, analyzed its performance, studied state-of-the-art **Texture Compression** techniques, and explored various trade-offs
- **ML Strategic Planning** mentored by *Rama Harihara*: Studied existing papers on **model quantization** for inference to understand the area trade-off between integer and floating-point units on the GPU
- **Workload characterization and analysis** mentored by *Brent Kelley*: Aimed to identify and characterize hot-spots on Compute/ML workloads like AI Benchmark, and investigate opportunities in workload tracing

Qualitative and Quantitative analysis of the MLPerf benchmark suite

UT Austin

GRADUATE RESEARCH ASSISTANT AT LAB FOR COMPUTER ARCHITECTURE UNDER PROF. LIZY K. JOHN

Fall'18 – Present

- **FastPath, ISPASS'19**: Proposed a **new metric for benchmarking ML workloads** that consider time and accuracy from the perspective of comparing the hardware used for training. Showed that merely taking into account the time for training to multiple thresholds makes the metric less sensitive to the specific threshold chosen and the seed values
- **NVIDIA GTC'19**: An extensive study on the **impact of hardware infrastructure choices on deep learning performance**. Presented quantitative analysis on various configurations of Dell systems with NVIDIA GPUs using MLPerf [v0.5]
- **arXiv e-print**: Analyzed and characterized the MLPerf [v0.5] benchmark suite exposing various **system-level trends**

Improving Data Locality by Kernel Fusion in DNNs [PRESENTATION]

UT Austin

COURSE PROJECT FOR COMPARCH: PARALLELISM AND LOCALITY UNDER PROF. MATTAN EREZ

Spring'19

- Studied *Convolutional Sequence to Sequence Learning* model for translation, a part of **Facebook AI Research Sequence-to-Sequence Toolkit (FairSeq)** implemented using **PyTorch** (lacks support for explicit memory management)
- Explored various methods of performing kernel fusion involving libraries like **CUTLASS**, **cuBLAS**, and **cuDNN**
- Integrated our C++/CUDA extensions with PyTorch and showed **~2x reduction** w.r.t the global memory/L2\$/DRAM writes

Graph Placement Optimization on a HMS [REPORT] [PRESENTATION]

UT Austin

COURSE PROJECT FOR COMPARCH: USER SYSTEM INTERPLAY UNDER PROF. MATTAN EREZ

Fall'18

- Proposed a novel optimization technique that **statically** makes **fine-grain placement** decisions based on the natural properties of a graph: the number of incoming/outgoing edges, topology, frontier composition
- Modified a light-weight shared memory graph processing framework (**Ligra**) to incorporate the proposed method
- Evaluated the same, demonstrating its good adaptability and up to **2x performance improvement**

Value Prediction: DFCM++ [PUBLICATION] [PRESENTATION] [POSTER] [CODE]

IIT Kanpur

UNDERGRADUATE RESEARCHER UNDER PROF. B. PANDA AND PROF. M. CHAUDHURI

Spring'18

- **CVP-1, ISCA'18**: Proposed a series of enhancements on top of existing DFCM predictor: **Early Update, Value Estimator, PC Blacklister**, and **Dynamic Context Length**. The design achieved an IPC improvement of **28.1%** with respect to the baseline, i.e, without any value predictor, and **40.2%** in comparison to the base DFCM

Perceptron Learning driven Cache Replacement policy [REPORT]

Texas A&M University

VISITING RESEARCH SCHOLAR AT HIGH PERFORMANCE COMPUTING LAB UNDER PROF. EUN J. KIM

Summer'17

- Proposed and modeled **Coherence-Aware Reuse Prediction** on **ZSim** that achieved a geometric mean speedup of **20%** over LRU and resulted in a **40%** drop in average MPKI with respect to LRU, when evaluated on PARSEC for 4 MB LLC

Emerging Non-Volatile Memory [PRESENTATION] [TERM PAPER]

IIT Kanpur

COURSE PROJECT UNDER PROF. YOGESH S. CHAUHAN AND PROF. BAQUER MAZHARI

Spring'18

- Studied various emerging **flexible** NVMs like **ReRAM, FeRAM, PCRAM**, and **Flash**, including the approaches for making them, their operating principles, and some common architectures. Additionally, performed a literature survey on **binary metal-oxide resistive switching RAM** encompassing its switching mechanism, design, and electrical characteristics

Other Projects

IIT Kanpur

- Coded an **assembler** and a **cycle-accurate simulator** for LC-3b RISC ISA capable of handling virtual memory translation
- Implemented **best-offset hardware prefetcher** highlighting its IPC and MPKI characteristics against other prefetchers
- Designed a low-power **PLL** [REPORT], a 2.4GHz **inductorless LNA**, and a two-stage folded cascode **OTA** [REPORT]
- Implemented a BSIM4-like model on **Verilog-A** and extracted parameters using **IC-CAP** simulation software [REPORT]
- Built an **all-terrain vehicle** capable of autonomous navigation using Embedded Systems and Google Maps API [PPT]
- Selected among the **top 5 best ideas** for a game developed using Unity3D Game Engine for **Microsoft Code.Fun.Do**

Technical Skills

Programming languages

C, C++, CUDA, OpenCL, OpenGL, Regent, Python, Bash, Verilog(-A), HSPICE perf, NVProf, NVVP, CACTI, PAPI, SimPoints, PINTool, Docker, Git, ~~TeX~~, Cadence Virtuoso, Synopsys, Silvaco (Athena and Atlas), PSPICE, Microcap,

Tools / Platforms

Mentor Graphics, Ardupilot, Arduino, Processing, MATLAB, GNU Octave

Selected Coursework

UT AUSTIN

Computer Architecture*

Comp Arch: User-System Interplay*

Comp Arch: Parallelism and Locality*

Operating Systems

Superscalar Microprocessor Architecture*

High-Speed Computer Arithmetic*

IIT KANPUR

Computer Architecture

Microelectronics-I (Circuits), II (Devices)

Data Structures and Algorithms

Modern Memory Systems*

Digital Electronics

Probability and Statistics

Principles of Data Base Systems

Analog/Digital VLSI Circuits*

Compact Modeling*

* indicates Graduate Level Courses

Scholastic Achievements

- **Professional Development Award, UT Austin** - research presentation at **FastPath, ISPASS'19** 2019
- **Second position** in unlimited track, **1st Championship Value Prediction, ISCA'18** 2018
- **Microsoft Research India Travel Grant** - research presentation at **CVP-1, ISCA'18** 2018
- **ISCA 2018 Student Travel Grant Award** and **Departmental (E.E.) Travel Grant Award**, IIT Kanpur 2018
- **TAMU-IITK summer undergraduate research scholarship** - awarded to two students per department 2017
- **Academic Excellence Award** - awarded to top 7% students in the institute 2015, 2017
- **JEE Advanced 2014, All India Rank 387** amongst 120,000 candidates 2014
- **KVPY National Fellowship**, Department of Science and Technology, Government of India 2014
- **Certificate of Merit** at National Level, HBCSE - **International Chemistry Olympiad 2013-14** 2014

Teaching Experience

Academic Mentor

IIT Kanpur

INSTITUTE COUNSELLING SERVICE

Fall'15 - Spring'16

- Tutored students having difficulties in **Engineering Design and Graphics** by conducting institute level remedial classes and doubt-clearing sessions. Personally mentored academically weaker students to cope with their academic load