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 | 𝒮 snehilverma41@gmail.com | ⌂ snehilverma41.github.io

Education

Cockrell School of Engineering, The University of Texas at Austin   *4.0*/ 4.0*

M.S. IN ELECTRICAL AND COMPUTER ENGINEERING

TRACK: ARCHITECTURE, COMPUTER SYSTEMS, AND EMBEDDED SYSTEMS (ACSES)   *Expected 2020*

Indian Institute of Technology, Kanpur   *8.9/ 10*

B.TECH IN ELECTRICAL ENGINEERING (WITH DISTINCTION)

MINOR IN COMPUTER SYSTEMS, COMPUTER SCIENCE AND ENGINEERING   *2018*

Modi Public School, Kota   *90.8 %*

ALL INDIA SENIOR SCHOOL CERTIFICATE EXAMINATION, CBSE   *2014*

Delhi Public School, Jamshedpur   *10/ 10*

ALL INDIA SECONDARY SCHOOL EXAMINATION, CBSE   *2012*

* Calculated at the end of Spring'19

Research Interests

Computer Architecture, Memory Systems, Performance Evaluation

Publications

- **Snehil Verma**, Qinzhe Wu, Bagus Hanindhito, Gunjan Jha, Eugene John, Ramesh Radhakrishnan, and Lizy Kurian John, "Demystifying the MLPerf Benchmark Suite," arXiv:1908.09207 [cs.LG], 2019. [arXiv e-print]

- **Snehil Verma**, Qinzhe Wu, Bagus Hanindhito, Gunjan Jha, Eugene John, Ramesh Radhakrishnan, and Lizy Kurian John, "Metrics for Machine Learning Workload Benchmarking," International Workshop on Performance Analysis of Machine Learning Systems (FastPath), In conjunction with *International Symposium on Performance Analysis of Systems and Software (ISPASS 2019)*, Madison, USA, 2019. [Publication] [Presentation]

- Ramesh Radhakrishnan, **Snehil Verma**, Qinzhe Wu, Bagus Hanindhito, Gunjan Jha, Eugene John, and Lizy Kurian John, "Demystifying Hardware Infrastructure Choices for Deep Learning Using MLPerf," GPU Technology Conference (GTC), *NVIDIA's Deep Learning & AI Conference 2019*, Silicon Valley, USA, 2019. [Presentation]

- **Snehil Verma**, Nayan Deshmukh, Prakhar Agrawal, Biswabandan Panda, and Mainak 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)*, Los Angeles, USA, 2018. [Publication] [Presentation] [Code]

Industrial Experience

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

Research Experience

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 ~$2\times$ **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*

- Explored trade-offs offered by **Heterogeneous Memory System** in the domain of **graph analytics**
- 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 $2\times$ **performance improvement**

DFCM++ Value Predictor [PUBLICATION] [PRESENTATION] [POSTER] [CODE]
IIT Kanpur

COURSE PROJECT FOR COMPUTER ARCHITECTURE UNDER PROF. B. PANDA AND PROF. M. CHAUDHURI *Spring'18*

- Reviewed the literature on **computational** and **context-based** value predictors
- Implemented multiple value predictors like last-value, stride, (D)FCM, and (D)VTAGE (state-of-the-art) predictors
- 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
- Showed the effectiveness of our enhancements on some of the state-of-the-art value predictors such as (D)VTAGE
- Presented at 1st **Championship Value Prediction (CVP-1)**, **ISCA'18** and secured **second** position in the unlimited track
- Showcased the work at a poster presentation session in **Graduate and Industry Networking (GAIN) 2019**, UT Austin

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** non-volatile memory technologies like **ReRAM**, **FeRAM**, **PCRAM**, and **Flash**
- Prof. B. Mazhari guided the research as a part of the course Introduction to Flexible Electronics. The work comprised of the approaches for making flexible NVMs, their operating principles, and some common architectures
- Performed a literature survey on **binary metal-oxide resistive switching RAM**. The study, supervised by Prof. Chauhan, encompasses the switching mechanism, design, and electrical characteristics of various binary metal-oxide ReRAMs

Perceptron Learning Driven Coherence-Aware Reuse Prediction for LLC [REPORT]
Texas A&M University

VISITING RESEARCH SCHOLAR AT HIGH PERFORMANCE COMPUTING LAB UNDER PROF. EUN J. KIM *Summer'17*

- Performed extensive literature survey on replacement policies, and inclusive, non-inclusive and exclusive caches
- Familiarized myself with various cache performance improvement techniques such as **Reuse Prediction**, **Inclusive Cache Management** and **Sharing Awareness Cache Management**
- Used an execution-driven simulator **ZSim** to model detailed micro-architectural behaviors
- Employed 8 multi-threaded applications and kernels from the **PARSEC benchmark** suite for evaluation
- Proposed **Coherence-Aware Reuse Prediction** that achieved a **geometric mean speedup of 20%** over LRU and resulted in a **40% drop in average MPKI** with respect to LRU, for 4 MB LLC
- Extended the work under the supervision of Prof. B. Panda, Department of Computer Science and Engineering, IIT Kanpur. Studied the correlation between the shared status of a cache block and its chances of being reused **[PPT]**

Phase Locked Loop (Design and Implementation) [REPORT]
IIT Kanpur

UNDERGRADUATE PROJECT UNDER PROF. SHAFI QURESHI *Spring'17*

- Studied PLL and its various blocks i.e **Phase-Frequency Detector**, **Charge Pump** and **Voltage Controlled Oscillator**
- Realized the whole circuit on **Cadence Virtuoso** using SCL's 180nm CMOS technology library
- Designed a **low power linear Current Starved VCO** which consumed a maximum power of 182µW
- Performed **stability analysis** on the whole circuit and the Low Pass Filter was modified to attain enhanced stability
- Pre-layout simulation results: Settling time of the PLL (±5%) came out to be around 33µs and Lock time around 125µs

Design of 2.4 GHz Inductorless Low-Noise Amplifier (LNA)

IIT Kanpur

RESEARCH PROJECT UNDER PROF. YOGESH S. CHAUHAN

Summer'16

- Studied the **noise cancellation techniques** of inductorless LNAs, and effectively applied them to design a better circuit
- Designed the schematic of the circuit on **Cadence Virtuoso Analog Design Environment** (IC 616)
- Extracted the netlist and modified the same in **SPICE3** in accordance with the commercial **Tower Semiconductor/SCL** 180 nm CMOS technology library for the simulation
- Coded for **S-parameter and Linearity Analyses** and simulated the same on **Synopsys HSPICE (RF)**
- Validated that LNA designed without on-chip inductors achieves **performance comparable** to inductor-based designs

Selected Projects

Verilog-A implementation and parameter extraction for BSIM4 like model [REPORT]

IIT Kanpur

COURSE PROJECT FOR COMPACT MODELLING UNDER PROF. YOGESH S. CHAUHAN

Spring'18

- Implemented a threshold voltage based model taking second-order effects, such as **mobility degradation** with vertical field, **velocity saturation**, channel length modulation (**CLM**), and drain induced barrier lowering (**DIBL**), into account
- Extracted the parameters using IC-CAP simulation software which were then tuned to match the measured TCAD data
- Examined the model for **Gummel Symmetry Test**, **Derivative Test**, and **Inverter Characteristics**

Mini Railway Inquiry System

IIT Kanpur

COURSE PROJECT FOR PRINCIPLES OF DATA BASE SYSTEMS UNDER PROF. MEDHA ATRE

Spring'18

- Designed a website implementing a miniature version of the railway inquiry system to handle standard queries such as trains between stations, fetch train route, and all reachable stations
- Optimized SQL queries by creating indexes on the most frequently used queries and creating a plan tree

Two-Stage Folded Cascode OTA Suitable for Large Capacitive Loads [REPORT]

IIT Kanpur

COURSE PROJECT FOR ANALOG/DIGITAL VLSI CIRCUITS UNDER PROF. SHAFI QURESHI

Fall'17

- Modified the circuit design mentioned in the paper titled *Enhanced Single-Stage Folded Cascode OTA Suitable for Large Capacitive Loads* [PAPER] and optimized the same for low power, better output voltage swing and slew rate
- Employed **Adaptive biasing** and **current folding stage**, that provide class AB stage with **dynamic current boosting**
- Simulated the schematic and **layout** design, in **Mentor Graphics**, using TSMC's 180 nm CMOS technology library

Advances in MIMO : System Model and Potentials [REPORT]

IIT Kanpur

TERM PAPER FOR PRINCIPLES OF COMMUNICATION UNDER PROF. ADITYA K. JAGANNATHAM

Fall'16

- Conducted a literature survey on Multiple-Input Multiple-Output (MIMO) systems and its potential in **4G** and **5G**
- Explained the **mathematical modeling** of MIMO systems along with their advantages and drawbacks
- Provided a detailed review on latest development in MIMO domain such as **Multi-user MIMO**, **Massive MIMO** and **MIMO-OFDM** techniques, and further emphasizing their importance in cellular communication systems

Evaluating "Reducing Risk In Type 1 Diabetes Using H_∞ Control" [REPORT]

IIT Kanpur

COURSE PROJECT FOR ROBUST CONTROL SYSTEMS UNDER PROF. RAMPRASAD POTLURI

Fall'16

- Evaluated and reproduced the results of the paper titled *Reducing Risk In Type 1 Diabetes Using H_∞ Control* [PAPER]
- Designed a H_∞ **controller** and tuned it for desired performances in order to make the system as **robust** as possible
- Designed the **Iodine Feedback Loop** and conceptualised the role of **safety mechanism** used in the research paper

Semi-Autonomous Surveillance and Transportation Robot (SASTR) [PPT]

IIT Kanpur

SUMMER PROJECT UNDER ROBOTICS CLUB, STUDENTS GYMKHANA

Summer'15

- Designed an all-terrain vehicle that could travel autonomously from one place to another
- Programmed **Arduino** microcontroller to receive the data from different sensors like **GPS** and **IMU**, and transmit the required data to the base (a computer) via wireless module **Telemetry**
- Implemented **PID controller** in order to minimize the deviation of the robot from the actual path
- Implemented the **Direction Cosine Matrix (DCM) Algorithm** to extract roll, pitch and yaw from the **IMU** of the robot
- Developed a **GUI** application, using **C#** language, to input destination from the user and find the shortest path to the destination. Application ensured smooth transmission of data between user and the robot via **serial communication**

Minor Projects

UT Austin

- Developed a simulator to evaluate **PC-indexed branch prediction** schemes like one- and two-level branch predictors
- Programmed a parallel renderer in **CUDA** capable of drawing colored circles, and a parallel version of PageRank in Regent
- **Analytically modeled** the locality for cache-aware and cache-oblivious dense matrix-matrix multiplication algorithms
- Evaluated (Static and Dynamic) RRIP **cache replacement** policy on SimPoints generated by PinPoints for SPEC 2006
- Coded an **assembler** and a **cycle-accurate simulator** for LC-3b RISC ISA capable of handling virtual memory translation

Minor Projects (continued)

- Analyzed the performance of **SHiP++** and **Hawkeye** on SPEC 2006 CPU suite under single and multi-core configuration
- Implemented **best-offset hardware prefetcher** highlighting its IPC and MPKI characteristics against other prefetchers
- **Instrumented** SPEC 2006 binaries compiled for the IA-32 ISA with **PIN** and investigated the properties of x86 instructions
- Enhanced the **gshare branch predictor** under a maximum hardware budget of 64KB by considering branch biasness
- Modified default **DRAM scheduling** policy on ChampSim, a trace-based simulator, to improve the row-buffer hit rates

Scholastic Achievements

- Recipient of the **Professional Development Award** by **UT Austin** for presenting research at **FastPath, ISPASS'19**
- Secured **second position** in the unlimited track of 1st **Championship Value Prediction, ISCA'18**
- Awarded with a **travel grant** of $1500 by **Microsoft Research Labs, India** for attending and presenting at **CVP-1, ISCA'18**
- Recipient of the **ISCA 2018 Student Travel Grant Award** and the **Departmental (E.E.) Travel Grant Award**, IIT Kanpur
- Recipient of **TAMU-IITK summer undergraduate research scholarship 2017** (awarded to two students per branch)
- Received **Academic Excellence Award** for outstanding academic performance (awarded to top 7% students in the institute) for the academic years 2014-15 and 2016-17
- Received **A* grade** in 3 courses, including Electrical Engineering Lab I (awarded to **top 1-2%** students in a course)
- Secured **All India Rank 387** in **JEE Advanced 2014**, amongst 120 thousand successful candidates selected from over 1.4 million aspirants who appeared for JEE Mains 2014
- Selected for the **Kishore Vaigyanik Protsahan Yojana (KVPY)** Scholarship in the year 2014, funded by the Department of Science and Technology, Government of India, and secured **All India Rank 236** in the national test
- Awarded **Certificate of Merit** by HBCSE in **International Chemistry Olympiad 2013-14** at the **National Level**

Technical Skills

Programming languages	C, C++, C#, CUDA, OpenCL, OpenGL, Regent, Java, Python, Bash, Perl, Verilog, Verilog-A, HSPICE, MySQL, HTML
Tools / Platforms	perf, NVProf, NVVP, CACTI, PAPI, SimPoints, PINTool, Docker, Cadence Virtuoso, Synopsys, Silvaco (Athena and Atlas), PSPICE, Microcap, Mentor Graphics, Ardupilot, Arduino, Processing, MATLAB, GNU Octave, CodeVisionAVR, MS Visual Studio, Git, LaTeX, Unity, AutoCAD, SolidWorks
Operating Systems	Linux, Windows

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)	Introduction To Flexible Electronics*
Modern Memory Systems*	Digital Electronics	S/C Optical Communication Devices*
Principles of Data Base Systems	Analog/Digital VLSI Circuits*	Power Electronics
Data Structures and Algorithms	Compact Modeling*	Robust Control Systems*

* indicates Graduate Level Courses

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 as well as Hall level remedial classes and doubt-clearing sessions
- Personally mentored academically weaker students to cope with their academic load

Extra-Curricular Activities

- Selected among the **top 5 best ideas** for a game developed using Unity3D Game Engine for **Microsoft Code.Fun.Do**
- Designed an LED Matrix and coded **ATmega32** in order to simulate a game "Space Invaders" for the event Electromania in Techkriti'15, inter-college technical festival of IIT Kanpur
- Designed a **hand gesture controlled robot** using **flex sensors** in Takneek'15, inter-hostel technical competition of IITK
- Worked as a member of **brakes designing team**, a part of **Society of Automotive Engineers (SAE)** IIT Kanpur team
- Fabricated a **remote controlled aeroplane model** for the event Aviator in Takneek'14 and won **3rd prize** for the same

Miscellaneous

- Talks given:
  - Prediction Fusion [PPT] *Spring'19*
  - Contrasting LRU with (S/D)RRIP [PPT] *Fall'18*
  - Plasticine: A Reconfigurable Architecture For Parallel Patterns [PPT] *Spring'18*
  - Flexible Non-volatile Memory [PPT] *Spring'18*
  - BeBoP: A Cost Effective Predictor Infrastructure for Superscalar Value Prediction *Spring'18*
  - Phase Change Memory [PPT] *Fall'17*
  - Memory Power Management via Dynamic Voltage/Frequency Scaling *Fall'17*
  - Perceptron Learning Driven Coherence-Aware Reuse Prediction for LLC [PPT-TAMU] [PPT-IITK] *Summer'17, Fall'17*
- Member of **SIGARCH ACM** and **TCCA**
- Blogs:
  - Talk attended on Qualcomm Datacenter Technologies & Centriq 2400 Processor by Dr. Niket Choudhary
  - My ISCA-2018 experience