Md Aamir Raihan

✆ (778) 319 5486
✉ araihan@ece.ubc.ca
🖥 Personal Website
LinkedIn

Computer Engineering Master Student at The University of British Columbia, Vancouver.
Research focus: My research interest spans at the intersection of systems and machine learning.
Areas of Interest: Machine Learning, Computer Architecture, Hardware Accelerators, High-Performance Computing, Machine Learning.

Education

2017–2019 (expected)	**MASc**, *Electrical and Computer Engineering, University of British Columbia*, Vancouver, BC, *GPA – 4.23/4.33*.
2011–2015	**B.Tech.**, *Electronics Engineering*, Indian Institute of Technology (BHU), Varanasi, India. *GPA – 8.54/10.00*

Graduate Research

Sep 2017 – Present **Graduate Research Assistant**, *University of British Columbia*, Vancouver, BC.
Advisor: **Prof. Tor Aamodt**
Funded by the Electrical and Computer Engineering Department

In the past, I have comprehensively investigated NVIDIA's Tensor Core implementation on Volta and Turing architecture and have proposed an architectural model for it. I have implemented the Tensor Core in the GPGPU-Sim, the widely used academic simulator and have also enabled the NVIDIA's Cutlass library. Our implementation achieved 99.6% IPC correlation with real V100 GPU. Following is my current ongoing projects:

– **Accelerating Deep Neural Network Training**.
  · Training DNNs is slow where back-propagation usually dominates the computation time.
  · I am currently investigation the back-propagation algorithm and exploring different ways of approximating gradients such as memoization, sparification to reduce the overall computation time.

Work Experience

July 2016 – July 2017 **Design Engineer (Digital Networking IP Team)**, *NXP Semiconductor*, Noida, Delhi NCR, Manager: Sanjay Jain.

After the rotation, I worked on the verification of DDR4 Controller reporting directly to Marie Sullivan, NXP Austin.
– Developed coverage driven constraint random UVM testbench using the synopsys DDR and AXI VIP.
– Developed constraint random and self checking testcases for different functionality such as transaction size including narrow transfer, DRAM organization, data bus width (x4 and x8), burst size (4 and 8), operation speed (1600 2666,2999,3200 Mhz), ECC error injection and correction, refresh, self refresh, posted refresh and auto-precharge.
– Developed a separate connectivity verification testbench using the cadence formal verifier.

Sep 2015 – July 2016	**Design Engineer *(Engineering Rotation Program)***, *NXP Semiconductor*, Noida, Delhi NCR, Manager: Sharad Kumar.

I was the part of Freescale's(merged with NXP) one year Engineering Rotation Program (ERP), a prestigious program in which only 20 individuals from US, India, and China got selected. The objective of this program is to expose an individual to SOC design cycle.
- *IP Design and Verification team:* Completed migration of AXI VIP from Synopsys to Cadence vendor in the SD host controller testbench.
- *Soc Validation team:* Developed an entire regression framework consisting of around 3000 testcase for exhaustive validation of UART.
- *Soc Application team:* Worked on the USB to SATA bridge using the LS1012A Soc. In this team, I learned about the uboot, yocto project, soc bootup procedure.

May 2014 – July 2014	**Research Associate *(CAD Lab)***, *Indian Institute of Science*, Bangalore, Professor: S.K. Nandy.

Research culminated in the development of a RBF Neural Network on a novel reconfigurable platform called Hyper-Cell.
- Developed an efficient and scalable partitioning and mapping algorithm for implementing any dimension RBFNN on multi-hypercell configuration.
- Developed auxilary communication module for efficient inter-hypercell communication.
- Implemented neural network was found to be around 100x faster than neural network implemented on multi core configuration.
- The internship resulted in 1 conference publication(VLSID 2015).

Selected Publications

1. **Md Aamir Raihan**, Negar Goli, and Tor Aamodt. "Modeling Deep Learning Accelerator Enabled GPUs." , *arXiv preprint arXiv:1811.08309 (2018)*, **(Accepted at ISPASS 2019)**.

2. Mohammadi, Mahnaz, Nitin Satpute, Rohit Ronge, Jayesh Chandiramani, S. K. Nandy, **Aamir Raihan**, Tanmay Verma, Ranjani Narayan, and Sukumar Bhattacharya. "A flexible scalable hardware architecture for radial basis function neural networks.", *IEEE VLSI Design (VLSID), 2015 28th International Conference*, **(VLSID 2015)**.

Technical skills

- **Programming Languages:** *Proficient*: C, C++, CUDA, Python, Verilog, System Verilog, Julia, Matlab.
- **System Simulation:** GPGPU-Sim
- **Deep Learning:** Pytorch, Tensorflow
- **EDA Tools:** Synopsys Design Compiler, Cadence Encounter, Cadence Incisive Enterprise Simulator
- **FPGA Prototyping:** Modelsim, Quartus, Xilinx Virtex 5, Altera DE2.

Graduate Coursework

- **Systems and Architecture**: Computer Architecture (Fall'17), Advanced Computer Architecture (Spring'18), Parallel and Reconfigurable Computing (Fall'16), CAD Algorithms for Integrated Circuits(Spring'18)
- **Machine Learning and Math**: Machine Learning and Data Mining(Fall'17), Advance Machine Learning(Spring'18)
- **Online courses**: Machine Learning (Stanford), Object Oriented Programming (IIT-B), Introduction to Computer Science, Image and Video Processing(Duke University)

Academic Projects

Spring'18   **Generalizability and Training Stability of GANs**, *[CPSC 540]*.
Skills: Tensorflow, Pytorch, Python
- Reviewed several variants of GANs and their problems before summarizing the related work and techniques that have been proposed to address these problems.
- Conducted two experiments: In the first experiment, developed a GAN based semi-supervised architecture and applied it on the dataset provided by the ongoing Kaggle competition: Google Landmark Recognition Challenge and report the classification accuracy. In the second experiment, implemented different GAN variants and compare their learning curves on different datasets.
- Report Link

Spring'18   **Survey of graph partitioning algorithms for VLSI circuit partitioning**, *[EECE 583]*.
Skills: C++, Python, C++
- Surveyed and implemented some of the most commonly used graph partitioning algorithms such as: Tabu Search, Genetic Algorithm, Improved Genetic Algorithm, Simulated Annealing, Spectral Partitioning Algorithm, Modified Spectral Partitioning Algorithm, Parallel Spectral Partitioning Algorithm, Multi-Level Partitioning Algorithm.
- Report Link

Fall'17   **Implementation of Radix-2 DIT FFT algorithm in CUDA and OPENMP** , *[EECE 528]*.
Skills: C, CUDA, OpenMP
- Implemented Radix-2 FFT algorithm in OpenMP and CUDA.
- Achieved an overall speedup of 12x and 7x over baseline in OpenMP and CUDA respectively.
- Report Link

Undergraduate Project   **Implementation of a 32 Bit Pipelined MIPS Processor**.
Skills: Verilog, Modelsim, Xilinx ISE, Xilinx Virtex 5 FPGA
- Designed and implemented a subset of a 32-bit pipelined MIPS processor using Verilog hardware description language.
- The design support forwarding and basic branch prediction logic.
- The design has been tested using a set of assembly language instructions, thereby emulated on Xilinx Virtex 5 FPGA.

Undergraduate Project   **Text Extraction from Business Card**.
Skills: MATLAB
- Implemented Optical Character Recognition using neural network in matlab.
- The preprocessing phase consist of noise removal and character segmentation using text bounding box detection and binarization. Feed forward neural network with one hidden layer was used to recongnize the segmented character.The recognized text is parsed to extract the name.

Undergraduate Project   **Chess Playing Image Processing Robot**.
Skills: MATLAB
- Developed Bot capable of distinguishing between different colors and shapes and is capable of making real time decision, for the selection of the path on the basis of object identified,in a 8*8 grid.
- Secured Second Prize in Optika, event based on Image Processing based Autonomous Bot in AAYAM'13.

Residency Status

Nationality   Indian

Visa   S1 (since August 2017)