

# Aly Sultan

978-325-1925 | Reading, MA

[sultan.a@northeastern.edu](mailto:sultan.a@northeastern.edu) | [linkedin.com/in/aly-sultan/](https://www.linkedin.com/in/aly-sultan/) | [github.com/asultan123](https://github.com/asultan123)

## EDUCATION

### Northeastern University

*Ph.D. Computer Engineering*

Boston, MA

2019 – Expected 2025

### American University in Cairo

*B.S. Electronics & Communication Engineering*

Cairo, Egypt

2014 – 2019

## EXPERIENCE

### Graduate Research Assistant

2020 – Present

*Embedded Systems Lab, Northeastern University*

*Boston, MA*

- Developed a SystemC model of a novel Hybrid general matrix multiplication and convolution accelerator (HERO) for accelerating inference in deep neural networks
- Developed a Template Optimization tool (TEMPO) in python that optimizes HERO templates based on available compute resources and a target library of neural networks
- Estimated HERO's latency and energy consumption using a Python-SystemC based simulation environment (HERO-Sim)
- Created a SystemC model of a novel programmable memory primitive called Self Addressable Memory (SAM) used in statically scheduling dataflows in neural network accelerator architectures
- Implemented a SAM program compiler that generates data movement programs from convolution layer descriptions to orchestrate on-chip data movement between SAMs in HERO
- Supervised and mentored a team of three undergraduate students in:
  - \* Implementing SAMs in Verilog
  - \* Integrating HERO into Xilinx's SystemC+QEMU simulation environment

## PROJECTS

### Cache Optimization for CNN Inference | C, Darknet, Intel Pin

2019

- Integrated Intel Pin with the Darknet framework to determine the effect of cache configurations on CNN inference
- Explored effects of cache hierarchy levels, sizing, replacement policy, and prefetching on hit rate
- Cache sizing and associativity was found to have the greatest impact on hit rate

### Darknet Convolution Inference accelerator | C, Darknet, Vivado HLS

2019

- Developed a General Matrix Multiplication (GEMM) accelerator using Vivado HLS
- Integrated accelerator into the Darknet framework
- Accelerated inference time of the tiny darknet network by a factor of 2X over CPU Baseline on a Zynq-7020 soc

### Backend synthesis of a Gaussian Noise Generator | Tcl, Synopsys ICC, Design compiler

2018

- Used TCL scripts to synthesized a Gaussian noise generator with the FreePDK45 Tech node
- Design achieved the following properties:
  - \* A max clock rate of 275 Mhz with clear LVS and DRC and no setup/hold time violations
  - \* Area utilization of 75% with total area equal to 27000  $\mu m^2$

## PUBLICATIONS

### A Compact Low-Power Mitchell-Based Error Tolerant Multiplier | Verilog, Matlab

2018

- Developed a novel approximate multiplier architecture in Verilog and evaluated it's numerical accuracy in JPEG compression using MATLAB
- Design improved power-delay-product by 1.9X with only 20% reduction in peak signal-to-noise ratio in JPEG compression compared to a Xilinx Zynq-7020 DSP at 16 Bit fixed point precision
- Published and presented findings at NGCAS, Malta

## TECHNICAL SKILLS

**Languages:** Python, C/C++, SystemC, Matlab, Verilog, Tcl

**Framework:** Intel Pin, Darknet

**Developer Tools:** Git, Docker, VS Code, Vivado, Vivado HLS, QEMU, Synopsys ICC, Design compiler

**Libraries:** PyTorch, Numpy, Matplotlib