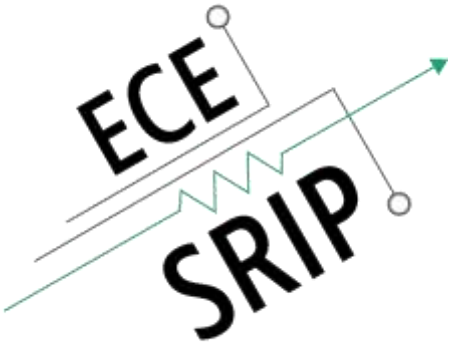


FACULTY MENTOR
Kang, Mingu

PROJECT TITLE
Traffic Sign Recognition on FPGA with XGBoost Algorithm

PROJECT DESCRIPTION
"High-speed / low-power XGBoost algorithm will be implemented on an FPGA platform for traffic sign recognition.

First, students study/understand the training/inference principles of XGBoost algorithm by reading literatures and then running simulations in python.

Second, the algorithm's data flow is mapped on hardware architecture with novel idea to achieve higher energy/delay efficiency.

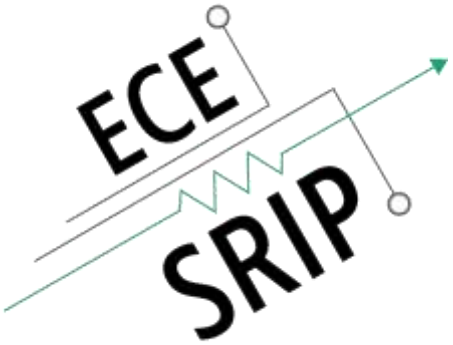
Third, the algorithm is co-optimized considering the hardware requirement and the limitation of FPGA board.

Finally, the hardware is designed in verilog/vhdl codes, and mounted on the FPGA platform to be tested with BelgiumTS Dataset. "

This project can accommodate both remote and in-person students

INTERNS NEEDED
2

PREREQUISITES
FPGA experience welcomed, verilog, python, matlab, and knowledge in machine learning



FACULTY MENTOR
Kang, Mingu

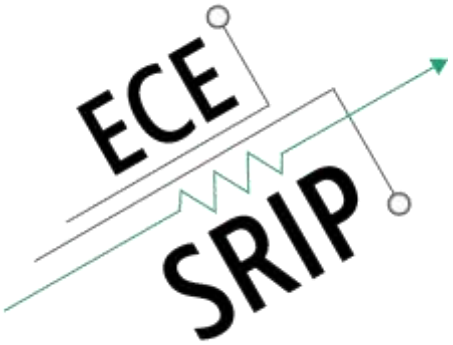
PROJECT TITLE
Optimal bit-quantization and channel skipping in neural networks

PROJECT DESCRIPTION
Bit quantization in the neural network is one of the most efficient methods to achieve significant energy and delay benefits in the hardware implementation. On the other hand, there are many redundant channels in the each layer of neural network models. The goal of this project is to search the most energy efficient point without losing the original accuracy by co-optimizing the bit precision and channel skipping in each layer.

This project can accommodate both remote and in-person students

INTERNS NEEDED
2

PREREQUISITES
Python, pytorch, matlab, deep neural network training/inference experience, and strong mathematic background



FACULTY MENTOR
Kang, Mingu

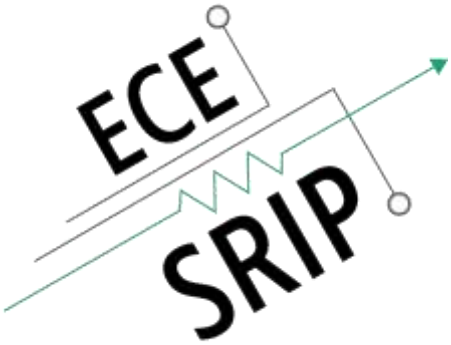
PROJECT TITLE
Deep neural network accelerator with analog switched-capacitors

PROJECT DESCRIPTION
There has been increasing interest in analog-based hardware implementation to achieve aggressive energy efficiency in the neural networks by exploiting the inherent error resiliency in the machine learning algorithms. This project exploits the matrix multiplication with switched-capacitor circuitry in analog domain. Transistor-level full-custom circuit is designed with Cadence Virtuoso, and the statistical error behavior is captured through Monte-Carlo simulations. Finally, this error model is injected in the system simulations with Matlab/Python to observe the impact on application level accuracy.

This project can accommodate both remote and in-person students

INTERNS NEEDED
2

PREREQUISITES
Digital and analog circuit background, transistor-level full custom circuit design, Cadence virtuoso simulation, Hspice, python and knowledge in machine learning (optional)



FACULTY MENTOR
Kang, Mingu

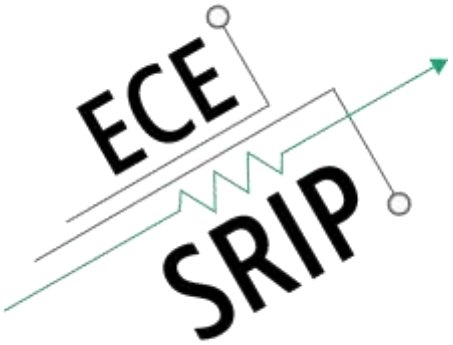
PROJECT TITLE
Dynamic approximate computing for deep neural network with real-time bit precision modulation in FPGA

PROJECT DESCRIPTION
Bit quantization in the neural network is one of the most efficient methods to achieve significant energy and delay benefits in the hardware implementation. On the other hand, edge computing platforms suffer from various time-varying environmental noise, e.g., voltage and temperature fluctuations, limited battery life, and sudden change of input data sparsity. For better error-resiliency in such circumstances, it might be one of the potential solutions to quantize the bit precision dynamically given situation, e.g,. bit precision is reduced when supply voltage drops or battery life is ending. Students are expected to perform system simulations with python (or pytorch) for the bit quantization of neural network models. Then, the algorithm is mapped on FPGA board with a structure, where the bit precision can be modulated flexibly. Finally, the power consumption of the FPGA board is measured in real-time to see the energy-efficiency behavior of the hardware with different bit precision.

This project can accommodate both remote and in-person students

INTERNS NEEDED
2

PREREQUISITES
"FPGA experience welcomed, verilog, python (pytorch), matlab, and knowledge in machine learning"



FACULTY MENTOR
Kang, Mingu

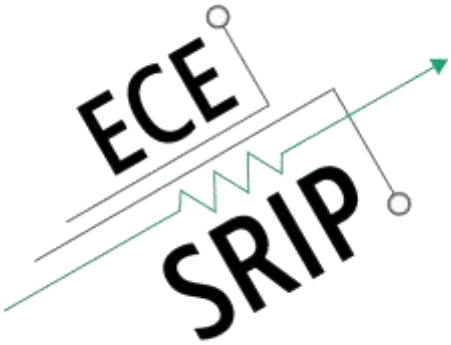
PROJECT TITLE
Data sparsity-aware dynamic frequency modulation for deep neural network acceleration in FPGA

PROJECT DESCRIPTION
It is known that there in high sparsity (zeros in matrices) in the data of neural network algorithms both in activation feature and weights. In many cases, gating-logics are equipped in the arithmetic computing units to skip the computation whenever zero is detected. This is a great opportunity to achieve significant energy savings by skipping the computation with large sparsity. On the other hand, the data sparsity of input is changing over time. So, it is a common practice to assume the worst-case sparsity (dense matrix with no sparsity) in the input, and assign power budget for the worst-case scenario. However, the worst-case data sparsity hardly happens in reality. In this project, students are expected to design a neural network computing hardware in FPGA platform with a data sparsity estimator. Given the detected data sparsity, the frequency is changed dynamically so that the optimal frequency is assigned based on the real-time input data sparsity within the power budget.

This project can accommodate both remote and in-person students

INTERNS NEEDED
2

PREREQUISITES
FPGA experience welcomed, verilog, python (pytorch), matlab, and knowledge in machine learning



FACULTY MENTOR
Kang, Mingu

PROJECT TITLE
Deep neural network training under noisy devices

PROJECT DESCRIPTION
Machine learning training requires expensive hardware resources including a large volume of computing and storage devices. On the other hand, emerging devices (e.g., RRAM and MRAM) or exotic computing paradigms (e.g., in-memory / in-sensor / neuromorphic computing) are gaining lots of interest to provide aggressive energy efficiency. However, those devices tend to suffer from various non-idealities with computing or storage noises. In this project, we study what kind of noise or how much noise can be tolerated during the machine learning training process. Furthermore, we expect to come up with various methods to mitigate the impact from the non-ideality, e.g., re-training, or efficient distributing the training job to both unreliable and reliable devices.

This project can accommodate both remote and in-person students

INTERNS NEEDED
2

PREREQUISITES
Python, pytorch, matlab, deep neural network training/inference experience, and strong mathematic background