ECE 516 Project Description

This project is to design VLSI circuits using Verilog. Three students will be in a group for this project. You should know your team members from my email notice.

As you learn in the class, machine learning algorithms can be implemented in digital hardware circuits for acceleration. You are provided with two Python programs (MNIST_quantized_sim.py) and (MNIST_quantized_weight_save.py). You may use Google Colab (https://colab.research.google.com/) to run these python programs and to understand the machine learning architecture that you will implement.

The first program (MNIST_quantized_sim.py) trains 3-layer fully connected networks for handwriting number recognition using the MNIST dataset. To simplify this project, you can see the bias for each neuron is none by setting "use_bias=False" in lines 52-56. In python, by default, weights for each neural connection are floating type, and its inference accuracy reaches an accuracy of 92-93%. Then, after quantizing the weights into 8-bit integer type, our neural network reaches a similar accuracy. As you know, 8-bit integer type is easy to implement in VLSI circuits, so you will design VLSI circuits to realize 3-layer fully connected networks whose weights are 8-bit integer type.

The second program (MNIST_quantized_weight_save.py) quantizes the weights from floating type to 8-bit integer, and saves well-trained weights into separate CSV files in the folder of "extracted weights from python program". You can see that three CSV files contain floating-type of weights, while the other three CSV files contain 8-bit integer type of weights. You can directly use the generated integer-type weights to design your VLSI circuits.

When you design your VLSI circuits, try your best to speed up the operation and throughput. You may need to consider using pipeline structures for higher clock rate and throughput. You will design an approximate VLSI circuit for the Softmax function. In the python file (MNIST_quantized_sim.py), we extract the first 10 images from the test dataset of MNIST (see code lines 109-135), and use them as inputs to test our neural networks. Therefore, in the Verilog test bench, you should also use these 10 images as test inputs to test your VLSI circuits.

Finally, you will submit a report including your system overview, design considerations, architecture design, Verilog codes, and simulation results.