ECE4011 Project Summary

Project Title	Resistive Network System for Neural Network Computation
Team Members (names and majors)	Yunfeng Xin (CompE) Runfeng Chen (CompE) Zheyuan Xu (EE) William Trimmer (CompE) William Scott (CompE) Sho Ko (CompE)
Advisor / Section	Shimeng Yu (YS1)
Semester	2018/Fall Intermediate (ECE4011)
Project Abstract (250-300 words)	Neural networks have demonstrated powerfulness in various tasks such as image recognition and image classification. Standard neural network inference computations consist of matrix addition and multiplication carried out in digital data processors that are not optimized for large-scale matrix operations, resulting in lower throughput and higher power consumption. In such a context, a specialized hardware accelerated solution is more desirable. This project proposes a novel hardware accelerator for computing neural network inference for image classification tasks. The hardware accelerator exploits analog circuit characteristics to avoid costly matrix operations in digital data pipelines that are commonly seen in Central Processing Units (CPUs) or Graphical Processing Units (GPUs). The accelerator consists of a crossbar array of resistors, each with a unique conductance, that takes an input array of analog voltage values representing the input image and outputs analog currents. The weights of the nodes are pre-determined by neural network training algorithms and are mapped to different levels of resistor conductance. The crossbar array forms a natural substrate for the vector-matrix multiplication or dot-product computation in a parallel fashion that is ubiquitous in neural network computation. A microcontroller is also included in the system to perform necessary image preprocessing and convolution for generating immediate inputs for the resistor crossbar array. The microcontroller will also convert grayscale image pixel data into different levels of voltage that is fed into the resistor array and quantize output analog current values back into digital data for final classification. The resulting system will be able to perform hardware acceleration for neural network classification tasks such as handwritten digit recognition.

Project Title	Resistive Network System for Neural Network Computation
List **codes** and **standards** that significantly affect your project. Briefly describe how they influenced your design.	Communication protocols (I2C, SPI, Serial). These would influence the types of chips we obtain for the microcontroller and ADC and potentially amplifier as well. We are planning to use through-hole technology, which would be considered a standard (i.e. not SMD) which will influence the parts available to us. Beyond this, there are few standards and codes we must adhere to, as the project is a proof of concept and much freedom can be taken in the design.
List at least two significant **realistic design constraints** that applied to your project. Briefly describe how they affected your design.	• Size: The design will need to be as small as possible, as the final concept would need to fit on a small chip. After discussion, the team decided to use input data of size 8*8, and 10 output ports, resulting in a total of 640 resistors. • Dataset: The design is targeted toward a specific dataset, so this design may need to be altered on a case-by-case basis with other datasets. Currently, the team decided to use MNIST hand-written digit dataset for image classification.
Briefly explain two **significant trade-offs** considered in your design, including options considered and the solution chosen.	• We considered using an FPGA vs microcontroller for controlling input currents. Ultimately, the team chose to use a microcontroller for the net pinout capability. Our design requires many output pins with on-board calculating capabilities to interface with the resistor network, thus the team chose to use a microcontroller for its advantage in these categories (although it trades cost and physical size for these advantages). • When gauging the number of resistors used for the project, we considered using a smaller number of resistors while doing more network calculations in software. This offloads some of the convolution tasks to the microcontroller, and simplifies the system as it primarily serves as a prototype.
Briefly describe the **computing aspects** of your projects, specifically identifying **hardware-software** tradeoffs, interfaces, and/or interactions. *Complete if applicable; required if team includes CmpE majors.*	The architecture of the finally implementable hardware will significantly affect the speed and performance of the calculation. Under the current stage, multi-layer perceptron would be technically the easiest and the most feasible to build due to the similarity between the structure of the neural network and that of our resistor network. It is possible to complicate the network further and make it into deep-belief as well as multi-dimensional. Voltages for different layers of the network will be deliberately altered to make sure the weights are somewhere between 0 and 1(normalized). As the input voltages are applied at the input layer, the array of input voltages will be produced by different layers, and finally the output. The real challenge lies in how to convert analog, voltage signal to digital signal which can be processed by electronic devices, as a normal signal generator would generate identical-width pulses which result in the nonlinearity in conductance with respect to the pulse frequency. One way would be increasing the widths of the pulses generated and ideally the signal strength would be linear enough to allow for better performance of the network. Another challenge lies in how to reduce the number of sensor data, since it requires resistors between every hidden layer of the

	network to be measured in order to allow for the track of the learning process. It would be better to perform matrix convolutions on the sensor data in order to reduce the number of data that is required to be monitored. Overall, the specific selection of hardware/software strategies will depend largely on performance-economic analysis, as well as the availability of various kinds of resistors.