Fall
2016

DEEPFINNS

FPGA IMPLEMENTATION OF A DEEP NEURAL NETWORK FOR SPEECH RECOGNITION

UNIVERSITY OF CENTRAL FLORIDA

Group 3 Members

LINDSAY DAVIS
Electrical Engineering

ESTELLA GONG
Computer Engineering

MICHAEL LOPEZ-BRAU
Electrical Engineering

CEDRIC ORBAN
Electrical Engineering

Initial Project and Group Identification

1 Introduction

People have the remarkable ability of being able to make vast inferences about the world with very little information, as well as being able to communicate these inferences to others. How are humans able to effectively use language, an unfathomably flexible and informal tool, to clearly exchange ideas and develop new ones? Is it possible to train machines to communicate with us in this way? These questions and many like them have piqued the interest of many computer scientists and engineers, leading to research thrusts in computational linguistics and natural language processing.

Recent advances in hardware have caused a resurgence of development in deep neural network modelling (DNN), now commonly known as *deep learning*. In the past, neural network models with many hidden layers were seen as inferior due to their computational complexity but now many of the most common algorithms can be run on a middle-to-high end laptop. Deep learning approaches have become very popular as of late due to their ability to dominate other machine learning algorithms on almost every benchmark. We have noticed that these algorithms also perform exceptionally well in various linguistic domains, such as speech recognition, a subset of computational linguistics.

In this project, we focus on the design and applications of speech recognition, particularly in the domain of human-computer interfaces. Speech recognition systems generally deal with several important steps, namely:

1. Data acquisition and pre-processing
2. Recognition/decoding
3. Application interface (e.g. moving a prosthetic arm)

Conventional speech recognition systems use hidden Markov models (HMM) in their recognition/decoding step. With the popularity and success of deep learning, we would like to design a system that mirrors current research thrusts by implementing a deep learning framework in our speech recognition system, specifically on an FPGA chip. FPGA chips have the advantage of being significantly cheaper than ASIC variants for small-scale applications. They also have the ability to run a neural network faster than a CPU and with less power consumption than a GPU. We would our speech recognition system to have a small vocabulary of pre-registered words that the user can utilize to get started, with the possibility of adding their own words in the future.

Our motivation for this approach comes from the desire to design a speech recognition system that is affordable, energy-efficient, low-cost, and portable while being able to maintain the computational sophistication needed for keeping the word error rates as low as possible. More generally, our motivation for this project stems from its interdisciplinary nature: speech recognition technologies often employ tools from a myriad of disciplinary areas, such as computer hardware, linear algebra, signal processing, and machine learning to name a few. Though our application is not solidified, ultimately, we aim to create an FPGA alternative to the typical CPU/GPU setups that are often used in machine learning applications.

2 Requirements and Constraints

2.1 Money

We will need to purchase two major components at the least. Each of these components can be very expensive. We plan on applying for multiple project sponsors in order to reach our fundraising goal and be able to purchase all the necessary parts. SoarTech is one of the sponsors we plan to reach out to. We will have to wrestle with using smaller, less expensive components while trying to maintain accurate speech recognition.

2.2 Time

The largest challenge to this project is the hard deadline for completion. We only have 7 months to complete this project. We will need to work a lot on our own and keep up with our own share of the workload in order to finish the project with a promising deliverable. Efficiently scheduling our time and carefully planning project steps will help us succeed.

2.3 FPGA

The size of our DNN is directly proportional to the amount of logic we have available for use on the chip. Using a bigger chip implies a bigger neural net and higher recognition accuracy at the expense of actual currency.

Using an FPGA entails routing many more traces on our PCB than a typical microprocessor would make use of. These traces could belong to programming interfaces such as JTAG and USB and connect to break-out pins to serve as general purpose IO (GPIO) pins. This factor directly affects the number of layers needed to implement our design as well as the width and height of the board itself. We will need to examine our project carefully to determine what connections are actually needed in order to keep costs down.

Additionally, some vendors only provide a free programming environment for some specific models of their FPGAs. For example, Altera only provides synthesis and fitting functionality in their Quartus Prime tool for their lower-end FPGAs (the Cyclone series and some Arria models).

3 Specifications

- Hardware
    - PCB with dedicated memory and power supply
    - PCB shall have inputs for:
        - FPGA
        - USB
        - JTAG
        - Microphone
    - FPGA will need to process a certain amount of voice information
    - Memory will need be able to store the weights of the DNN (or the parameters for the HMM) as well as a small preset vocabulary
    - High-quality microphone

- Software
    - Speech signal data set for training the DNN
    - Command set for when it should start deciphering voice signals
    - Detect and filter out any background noise

4 House of Quality Table

			Marketing Requirements				
			Signal Integrity	Performance	Power Consumption	Accuracy	7) Cost
			+	+	+	+	+
Engineering Requirements	Signal Integrity	+		↑↑	↑	↑↑	↑↑
	FPGA Size	+	↑	↑↑	↑	↑↑	↑↑↑
	Performance	+	↑		↑↑	↑↑	↑
	Power Consumption	+	↑	↑		↑	↑
	PCB Dimensions	+					↑
	Accuracy	+	↑↑	↑	↑		↑
	Cost	-	↓	↓↓		↓	

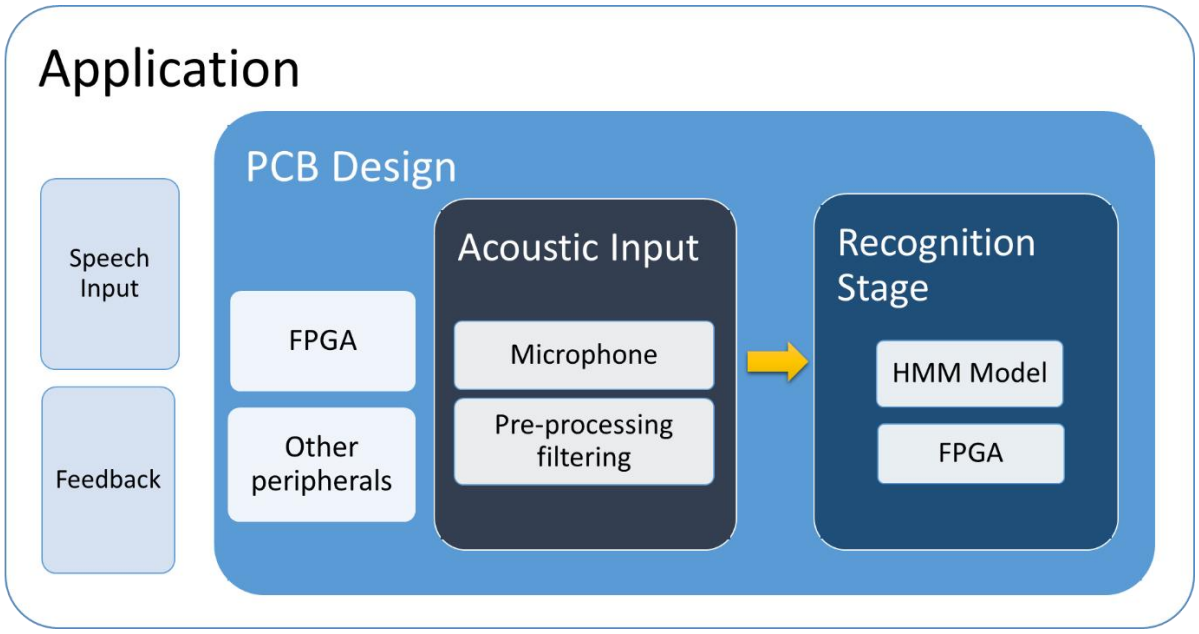
- Signal Integrity - The fidelity and signal-to-noise ratio of the speech signal after signal processing has been performed.
- FPGA Size – The amount of logic available for use on the FPGA. Cost typically goes up as this the number of logic resources increase. This directly correlates to the size of our DNN and whether or not signal processing will be performed on the FPGA.
- Performance – How quickly and how efficiently our device can decode speech signals. This depends on how fast we clock the logic in our FPGA, which in turn depends on the model of FPGA we use and what kind of logic we perform (i.e. whether we use DSP or not and whether we use a DNN or HMM).

- Accuracy – The ability of our device to correctly decode speech signals. We aim to have a low error rate.

5 System Block Diagram

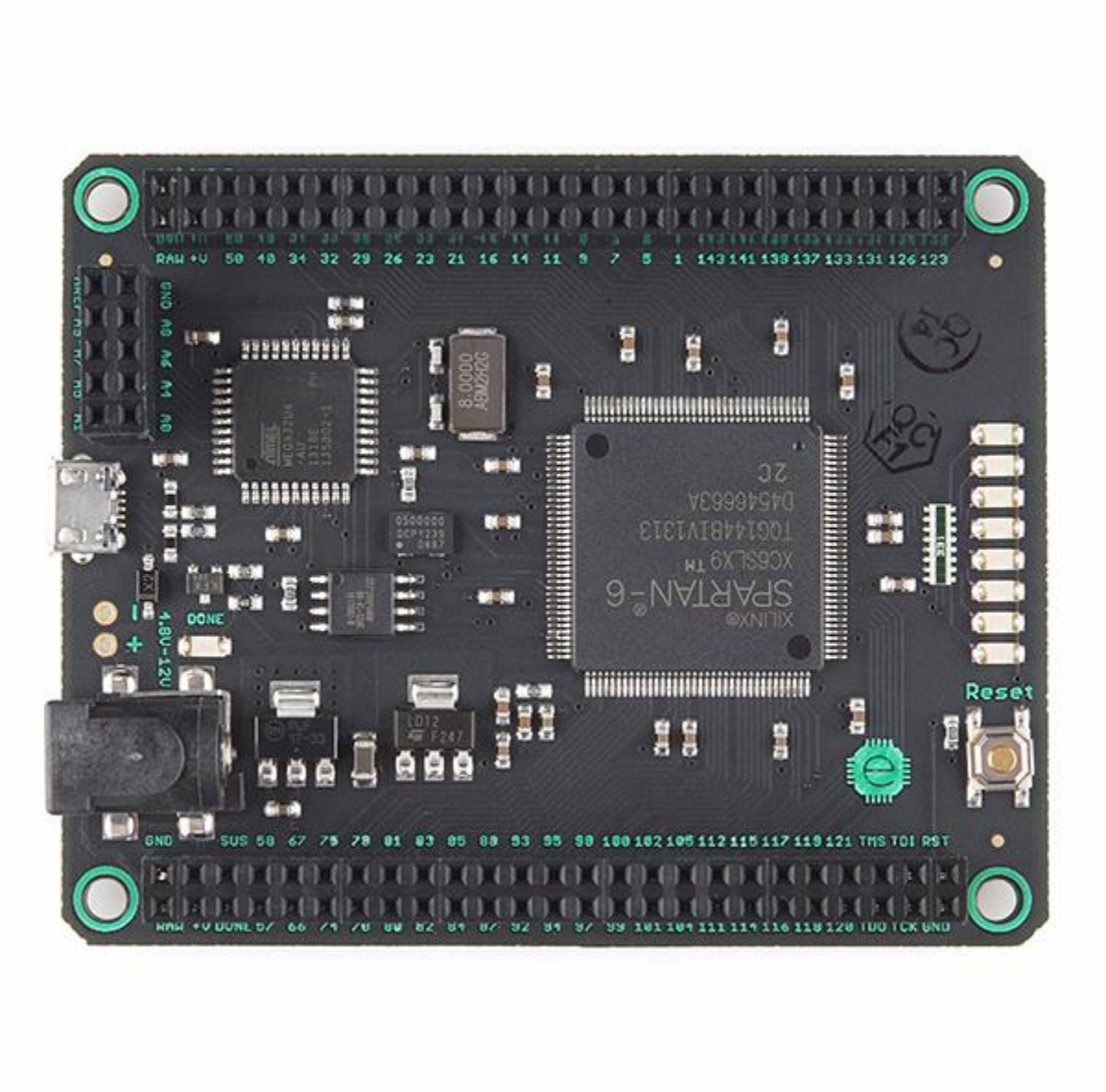
Our project can be broken down into 4 parts:

1. Acoustic input, which involves hardware (mic) and pre-processing (filtering, maybe some DSP).
2. Recognition, or decoding stage. This includes developing the DNN/HMM model and coding it on the FPGA.
3. Designing the PCB for the FPGA, mic, and any other peripherals we decide to use, like USB ports.
4. Application we decide to use the speech input on.



Primary Responsibility:

Our final PCB would look similar to:



6 Financing

6.1 Budget

This is a very rough outline of our projected budget. The final amount we spend will depend on whether we receive corporate sponsorship and the amount the group itself agrees to contribute to the final design.

The breakdown goes as follows:

FPGA Chip	$300 - $400
Onboard DSP	$80 - $90
Onboard Processor	$10 - $20
PCB Printing Service	$60 - $70
PCB Pick and Place Service	$50 - $100
Various Electronic Components (EEPROM, DRAM, ADC… etc)	$30 - $50

These parts and values are vague because we have not yet determined exactly what will be needed to satisfy our goals. For example, the price and model of the FPGA is dependent on whether or not we perform speech recognition using neural nets or Hidden Markov Models. Some of the components listed may not be used in the project at all. After more research we may determine that it is advantageous to perform digital signal processing on the FPGA itself, freeing us from purchasing a discrete DSP integrated circuit. However, it is clear that some purchases are necessary. These include the PCB printing and pick and place services.

6.2 Sponsors

We plan on submitting a proposal to SoarTech, State Farm, Duke Energy, SAIC, and Leidos to see if there is any potential in having our project sponsored by them. If sponsorship is not feasible, our team will be bootstrapping the entire project development.

7 Milestones

Milestones	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.
Research Phase								
Finalize Application Idea								
FPGA								
Machine Learning Methods								
PCB Design								
Embedded Programming								
Speech Recognition								
Design and Documentation Phase								
Application Interface								
PCB interfacing with FPGA								
PCB interfacing with Microphone								
PCB interfacing with other peripherals								
Acoustic Input								
Recognition Stage								
Implementation Phase								
Acoustic Hardware								
Acoustic Processing								
Application Software								
Recognition Stage Software								
Recognition Stage Hardware (PCB's etc.)								
Integration and Testing Phase								
Acoustic Integration								
Acoustic Testing								
PCB Integration								
PCB Testing								
FPGA Integration								
FPGA Testing								
Recognition/Machine Learning Testing								
Recognition Stage Integration								
Entire System Integration								
Entire System Testing								