Peter Ferrarotto – Proposal for Senior Design II Project to Matthew Campisi

Summary:

Implementation of DNA analysis on an FPGA board for finding matches to segments given in input.

Proposal:

In Stem Cell research, a large amount of time is spent on manually finding and identifying strands of DNA in stem cells that are refactored from tissue samples, such as fibroblasts and muscle culture. During my internship at the New York Stem Cell Foundation (NYSCF), I observed researchers combing through pages and pages of DNA to find exact matches for DNA after the use of gel electrophoresis to separate the DNA they need, and to ensure that what was cut by agents such as CRISPR is correct.

This process reduced the time researchers were able to spend in the lab. What could be done is that the input could be passed into software that combs through the DNA and finds matches for the user, but this could be slow and resource-intensive, and lock up computers for an extended period of time. My proposal is to develop this program in VHDL and put it on an FPGA board.

In recent years, the use of FPGA chips to create efficient and fast implementations of algorithms has become increasingly popular. I would like to develop a VHDL program that takes the input of the strand of DNA being searched through, the particular segment that the user wants to find, and any other extraneous variables that may be added during the development of this application.

The reason to put this program on an FPGA chip rather than writing software is that chips that are devoted solely to the execution of one particular function or algorithm can do it much more efficiently, and significantly faster. Rather than run multiple threads of the analysis program on one computer, risking filling the RAM entirely or having a stack overflow or negatively impacting the use of the computer. Then, once the schematic for the chip is completely nailed down, if this project succeeds and I decide to move forward with it, the boards could be produced en masse and sold as modular devices for quickly analyzing data that is passed into it.

This would allow for the researchers in the field to generate the file by other sequencing devices, drop it into a folder or external drive and give it to the array of chips running this procedure. This ultimately will increase efficiency and allow the team to research other projects while they wait for the results.

I am returning to the New York Stem Cell Foundation over the winter vacation and I know the analysis they perform on DNA for finding sequences and their positions is a standardized practice across all research facilities. As such, this would not utilize any proprietary rules or methods, and could be used in facilities outside of NYSCF for the same purpose.

Timeline:

I plan to gather as much information as I can on the rules for this analysis of DNA while I am at NYSCF over the summer, ensuring that I can have a well-developed rule set for the procedure.

After that, I would like to follow the following timeline throughout the semester:

|  |  |
| --- | --- |
| Time Period | Goal |
| Weeks 1 – 4 | Have a program running on an FPGA board that will take a text file as input and spit out output generated using the contents of the file into another file with a name derived from the name of the input file. This is to ensure that the processing of files is possible on FPGA boards.  (I’ve already researched it and determined that text processing is possible in VHDL and on an FPGA board, so this is more to just establish how this is done on this particular model of board.) |
| Weeks 4 – 6 | Have a program running on an FPGA board that will take the input from a file containing the string representing the sequenced DNA. It will then parse all of it into a continuous binary vector. Ideally, each nucleotide will be represented by 4 bits to facilitate hexadecimal evaluation.  Then, by using a bit shifter, each nucleotide in the sequence will be translated back into its ASCII counterpart and written to a new text file to recreate the original DNA sequence. This will lay the foundation for working with the nucleotides in a sequence of DNA. |
| Weeks 6 – 9 | Have a program running on an FPGA board that will take the DNA input file, and then look through it using a passed in sequence of nucleotides that the user is looking for. Using the same method as above and an additional binary vector as a temporary store for how much the sequence matches by, it will comb through the DNA, and once a matching sequence has been found, it will output into a new file the position at which the match is found.  It will continue doing this until the entirety of the passed in DNA segment is complete. |
| Weeks 9 – 15 | Completion of the project, fixing any and all bugs, and the utilization of additional modules on the board (or built-in LEDs/seven segment displays) to inform the user of the progress made so far in the program.  Stretch goal: Connecting the board to a computer in a way that allows for the user to start and stop the program through a UI. This would slightly impact performance, but not enough to warrant not using an FPGA board. |

Budget:

Ideally, this would cost nothing, as I’ve already purchased an FPGA board for my Senior Design 1 class – the Nexys 4 DDR. However, depending on the size of the project moving forward, I may have to purchase a larger board. Also depending on how the Nexys 4 handles text files, I may have to invest in either modules or a new board altogether to implement file manipulation.