IT.ENG.6

SOFTWARE HARDWARE CO-DESIGN FOR PROTEIN IDENTIFICATION

S. M. Vidanagamachchi[1], S. D. Dewasurendra[1], R. G. Ragel[1], M. Niranjan[2]

*[1]Department of Computer Engineering,
Faculty of Engineering, University of Peradeniya
[2]School of Electronics and Computer Science, University of Southampton, UK*

In this paper we deal with the problem of protein identification from a known set of peptide sequence data: this problem is characterized by a very rapid accumulation of data which are generated by biologists in the process of identifying unknown organisms/plants or disease related proteins and it challenges currently available processing speeds in either software or hardware. The generic problem is similar for most biological sequence data such as peptides, proteins, DNAs, and genomes. Aho-Corasick algorithm and a few other multiple string matching algorithms have played a major role in developing solutions for these problems. Since we need accelerated methods to deal with the fast growth of data, we have to concentrate on efficient hardware/software design methodologies.

Out of the different architectural choices, we have concentrated on reconfigurable hardware based on the interplay of cost/processing speed/ease of programming and hence, FPGAs. An FPGA consists of a reconfigurable microprocessor core that is highly coupled with a Reconfigurable Functional Unit (RFU). Commercially available reprogrammable processors include the Altera Nios II, Xilinx MicroBlaze and Stretch processors. They support implementing critical parts of an application in hardware using a specialized instruction set and provide a fast and powerful platform for implementing different algorithms with the embedded processor. The instructions from the specialized instruction set are known as custom instructions that involve in implementing a part or entire algorithm in hardware and make it accessible to software through software macros. We have used this feature in a Nios II processor to extend its functionality and therefore to accelerate the protein identification process through Aho Corasick algorithm.

Software hardware co-design for protein identification using peptide sequence data has not been done earlier. Therefore, our objective of this paper is to give a brief introduction to our custom instruction design and to the communication interface with hardware implementation using Nios II Embedded processor for protein identification. Then the performance enhancement of the custom implementation of this algorithm is measured against software-only implementation.

In our design we have used two custom instructions which are coupled with Arithmetic and Logical Unit of Nios II Embedded processor to solve a common exact string matching problem in Computational Biology. Here two custom logics in the hardware side contain the matching algorithm of peptides and in the software side we send the input and receive the matching peptides as the result. According to the results we obtained hardware software co-design system is much faster than software only system. Limitations include the number of maximum states that is used in a finite state machine (FSM) in custom logic and the maximum number of states in a FSM is 64.