An Algorithmic Replacement for BLAST

Craig Steffen, NCSA  Matt Hudson, Crop Sciences

BLAST (Basic Local Alignment Search Tool) is a heuristic search algorithm for genetic matches. Different tools in the BLAST toolset perform DNA/DNA, DNA-protein, and protein algorithms.

BLAST in scalar CPU form modifies its searches in place to achieve very efficient searches. That is, each step depends intimately on the step before. This strategy leans heavily on the speed of the CPU and this data dependent algorithm. This data dependency forces computing pipelines to be short.

Our work is attempting to solve the same sort of problem that BLAST solves (genetic sequence comparison), but doing it in away that leverages the strengths of FPGAs. FPGAs lend themselves to parallel, deeply pipelined, data independent algorithms; the opposite of BLAST.

Our solution to the FPGA part of the problem was to create an exhaustive element-element comparison over all offsets. All possible combinations are compared, requiring an enormous number of comparators. The match score for an offset is the sum of the number of correct matches for that offset.

The sample problem shown here is searching to see if a protein comes from a given genome.

This design has been implemented on an SRC MapStation 6. We are optimizing to increase the number of simultaneous comparators.

Currently we are working on making the output something that will be biologically meaningful.

```c
for(i_protblock=0;i_protblock<protein_blocks;i_protblock++){
    int8_t a0,a1,a2,a3,a4,a5,a6,a7;
    int raw_dna_inword_index;
    // load candidate protein into one side of comparator
    for(i_word=0;i_word<PROTEIN_PIPE_LENGTH;i_word++){
        ...
        #pragma src parallel sections
        {
            #pragma src section
            {
                // for(i_dnaA=DNA_start_offset_forward;i_dnaA<N_dna_bases;i_dnaA++){
                for(i_dnaAA=0;i_dnaAA<(N_dna_amino_acids+(2*PROTEIN_PIPE_LENGTH));i_dnaAA++){
```