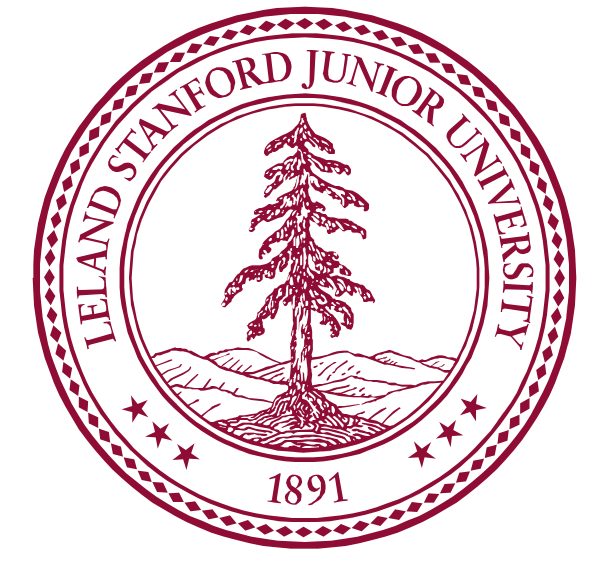


An FPGA Programming System for DNA Sequence Alignment



Albert Ng¹, William Dally^{1,2}, Gill Bejerano^{2,3}

¹Department of Electrical Engineering, ²Department of Computer Science, ³Department of Developmental Biology, Stanford University

Motivation

Sequence alignment is a critical operation in genome research. Commonly used alignment tools, such as BLAST, sacrifice sensitivity in order to run in a feasible amount of time. On the other hand, full dynamic programming alignment algorithms, such as Smith-Waterman and HMMER, are much more sensitive, but require impractically long runtimes.

To bridge this gap for biologists, we are building a FPGA accelerator programming system that will allow non-FPGA experts to easily and flexibly accelerate dynamic programming alignment algorithms. Such a tool will allow biologists to perform analyses that were previously impractical, leading to new discoveries in genome research.

Dynamic Programming Alignment

Example Algorithm

$$W(a_i=b_j) = +2 \quad \text{Match}$$

$$W(a_i \neq b_j) = -1 \quad \text{Mismatch}$$

$$W(a_i, -) = -1 \quad \text{Insert}$$

$$W(-, b_j) = -1 \quad \text{Delete}$$

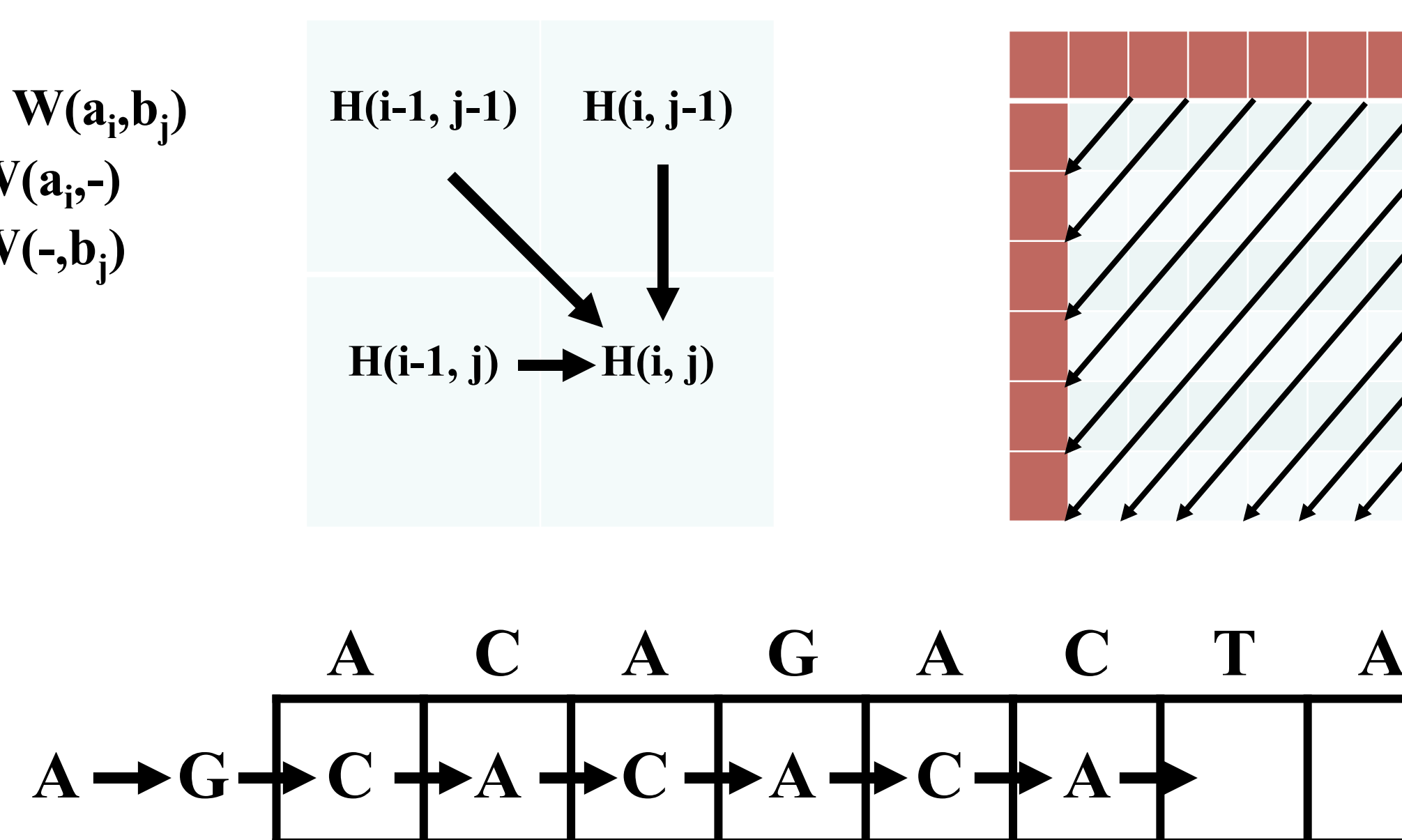
$$H(i,j) = \max \begin{cases} 0 \\ H(i-1, j-1) + W(a_i, b_j) \\ H(i-1, j) + W(a_i, -) \\ H(i, j-1) + W(-, b_j) \end{cases}$$

*	A	C	A	G	A	C	T	A
*	0	0	0	0	0	0	0	0
A	0	2	1	2	1	2	1	0
G	0	1	1	1	4	3	2	1
C	0	0	3	2	3	3	5	4
A	0	2	2	5	4	5	4	6
C	0	1	4	4	4	4	7	6
A	0	2	3	6	5	6	6	8
C	0	1	4	5	5	5	8	7
A	0	2	3	6	5	7	7	9

A-CAGACTA
AGCACAC-A
Score = 9

- Compute 2D score matrix and trace back from high-scoring cells
- $O(n^2)$ computation

Hardware Acceleration



- Due to cell dependencies, the anti-diagonals can be computed in parallel
- Computation can be efficiently implemented using hardware systolic arrays
- Processing elements are small; hundreds of PEs can fit on modern FPGAs

Hardware vs Software

Hardware acceleration with FPGAs has been shown to give significant speedups over software.

However, implementing in hardware requires significantly more time and expertise, prohibiting widespread usage.

We built a preliminary Smith-Waterman accelerator on a FPGA card and a simple software implementation for comparison.

	Hardware	Software
Development Time	3 months	1 hour
Lines of Code	14,000	100
Seconds to Align 100 bp to Human Ref	1	6000

Our Approach

Dynamic Programming Alignment Language

```

alphabet = {A, C, G, T};
signed<10> H[][];
CELL(signed<10> match, signed<10> mismatch,
     signed<10> insert, signed<10> delete) {
  signed<10> substitute;
  if (query_char == ref_char)
    substitute = match;
  else
    substitute = mismatch;

  H[row][col] = MAX(0,
                   H[row-1][col-1] + substitute,
                   H[row-1][col] + insert,
                   H[row][col-1] + delete);
}
    
```

