



PipeBSW: A Two-Stage Pipeline Structure for Banded Smith-Waterman Algorithm on FPGA

Luyi Li, Jun Lin, and Zhongfeng Wang

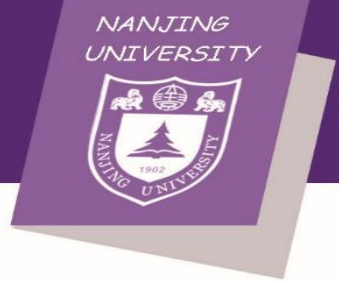
Integrated Circuits and Intelligent Systems (ICAIS) Lab

Nanjing University, China

ISVLSI 2021

Outline

- Background
- Hardware Architecture
- Experimental Validation
- Conclusion



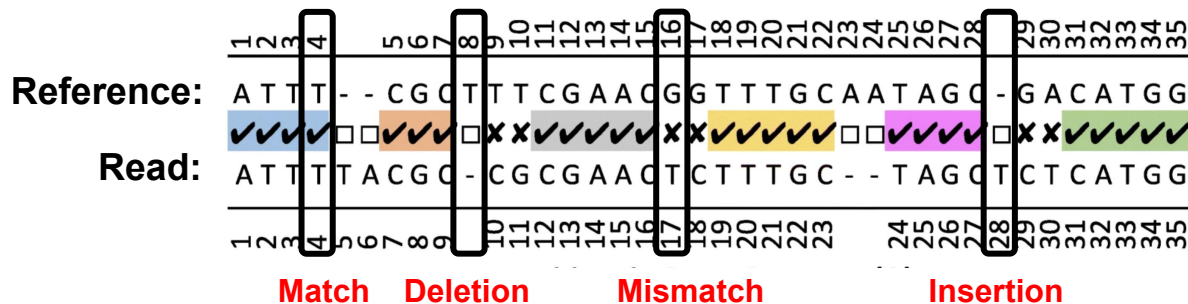
- Background
- Hardware Architecture
- Experimental Validation
- Conclusion

Background



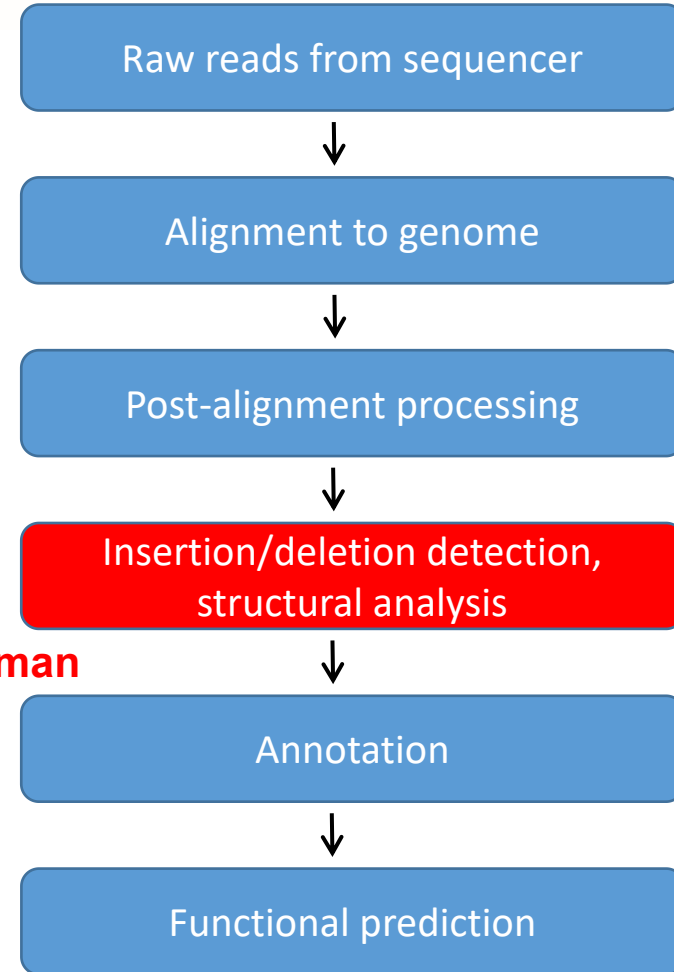
● DNA Sequencing

- ◆ It has a wide range of application scenarios, such as **early cancer detection**, **gene editing** and **virus vaccine research**.
- ◆ DNA sequencing is the process of determining the nucleic acid sequence.
- ◆ Sequence alignment is to **align the sequences to a known reference genome** that may reveal relationships between the sequences.



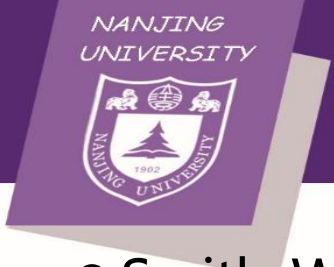
PS: Picture quoted from Google

Smith-Waterman



A common sequencing procedure

Background



● Smith-Waterman (S-W) Algorithm[1]

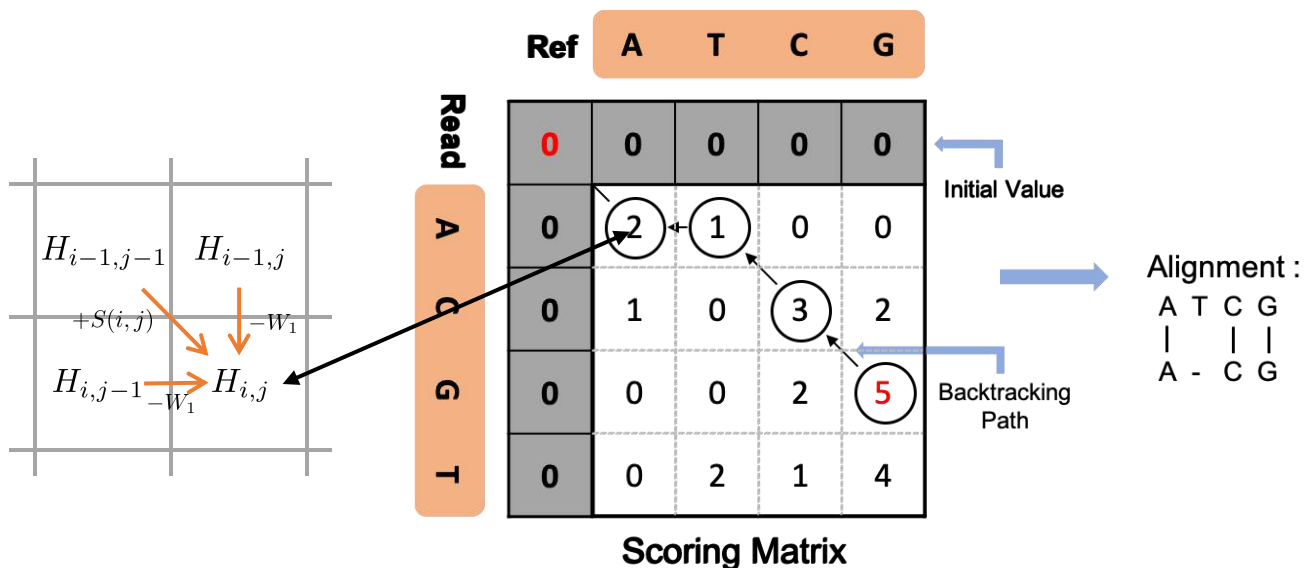
◆ Scoring Step

- Record a scoring matrix
- Most time-consuming, but suitable for parallel acceleration

◆ Bracktracking Step

- Generate an alignment path
- Commonly on CPU, need to read the scoring matrix from memory (memory bottleneck)

$$H_{i,j} = \max \begin{cases} H_{i-1,j-1} + S(i, j) & : \text{(Mis)Match} \\ H_{i-1,j} - W_1 & : \text{Insertion} \\ H_{i,j-1} - W_1 & : \text{Deletion} \\ 0 & : \text{Baseline} \end{cases}$$



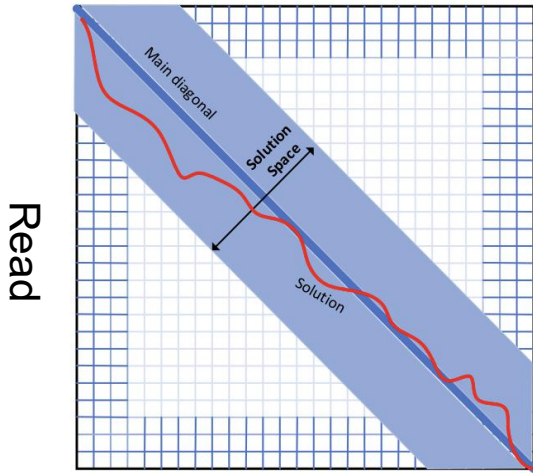
[1] T. F. Smith, M. S. Waterman, et al., "Identification of common molecular subsequences," Journal of molecular biology, vol. 147, no. 1, pp. 195–197, 1981.

Background



● Banded S-W[2]

Reference



- ◆ Aligning sequences with a limited number of mismatches, deletion and insertion.
- ◆ If only a 10% mismatch is acceptable, the solution can be only 10% upper or lower than the main diagonal.



Relatively low resource consumption to record the matrix, so we can store it in registers (can mitigate the memory bottleneck problem)

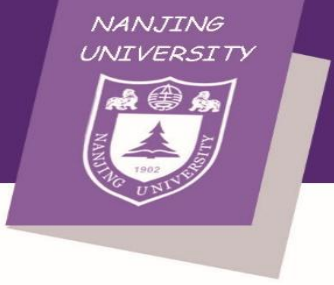
● Direction Matrix[3]

↖	↖	←	0	0	↖	←	0
↑	↑	↖	0	↖	↑	↖	↖
0	0	↖	←	↑	↖	0	↑
0	0	↑	↖	↖	←	←	↖
↖	↖	↑	↑	↑	↖	←	←
↖	↖	←	←	↑	↑	↖	←
↑	↑	↖	↖	←	↑	↖	←
0	↑	↑	↑	↖	←	↑	↖

- ◆ Match (the score inherits from ↖), mismatch (from ↖), insertion (from ↑), and deletion (from ←), can be represented within 2 bits.
- ◆ The direction matrix still keeps 2 bits per element as the score is increasing.

[2] Z. Nawaz, M. Nadeem, H. van Someren, and K. Bertels, "A parallel fpga design of the smith-waterman traceback," in 2010 International Conference on Field-Programmable Technology, pp. 454–459, IEEE, 2010.

[3] K.-M. Chao, W. R. Pearson, and W. Miller, "Aligning two sequences within a specified diagonal band," Bioinformatics, vol. 8, no. 5, pp. 481–487, 1992.



● Motivation

◆ Improve calculation parallelism and performance

- Optimize the lookahead calculation cell
- Design a hardware backtracking module to avoid communication with the memory

◆ Mitigate resource consumption and memory bottleneck

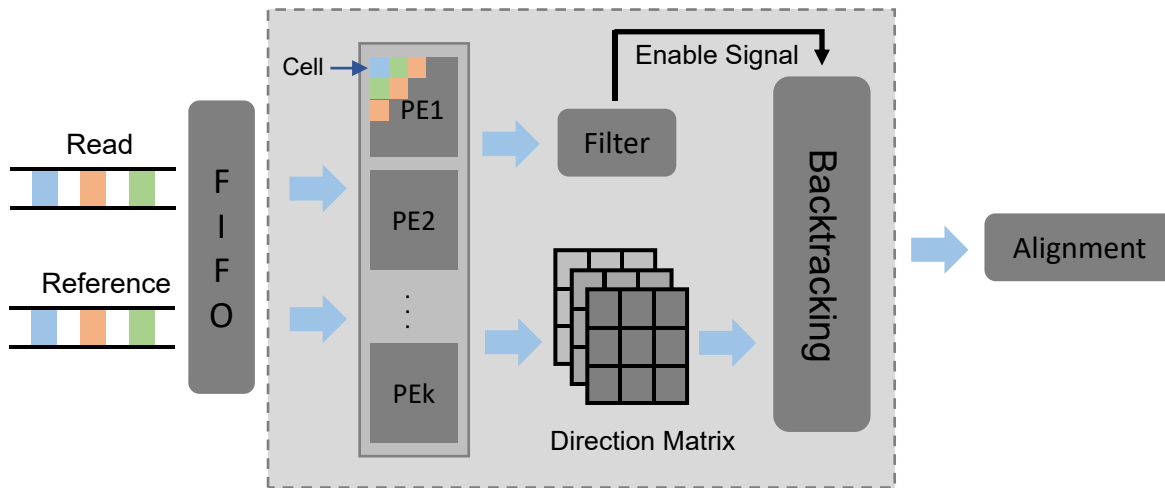
- Combine the banded S-W and the direction matrix
- Design a two-stage pipeline structure to increase the reuse rate of modules

The whole process of the algorithm is implemented on FPGA.



- Background
- Hardware Architecture
- Experimental Validation
- Conclusion

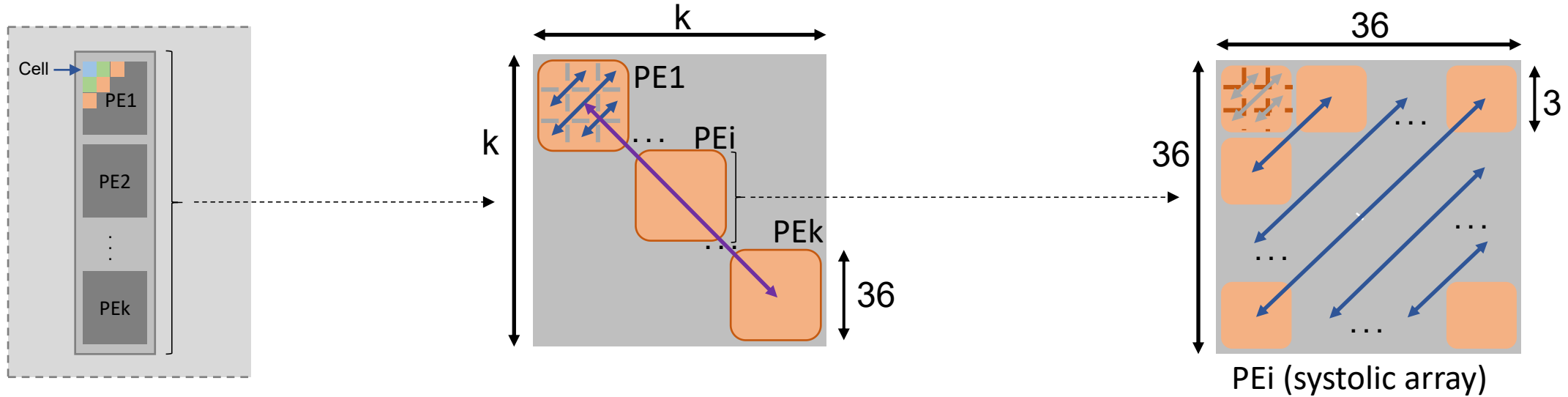
- Overall Workflow



- ◆ Read and reference are input through FIFO, sliced into segments, and then sent to relative PEs to do the scoring step.
- ◆ One PE consists of several calculation cells, and outputs a direction matrix for the backtracking step.
- ◆ The filter module estimates the number of errors along with scoring, in order to identify bad candidate reads early in the procedure.

Hardware Architecture

● Scoring Calculation Structure

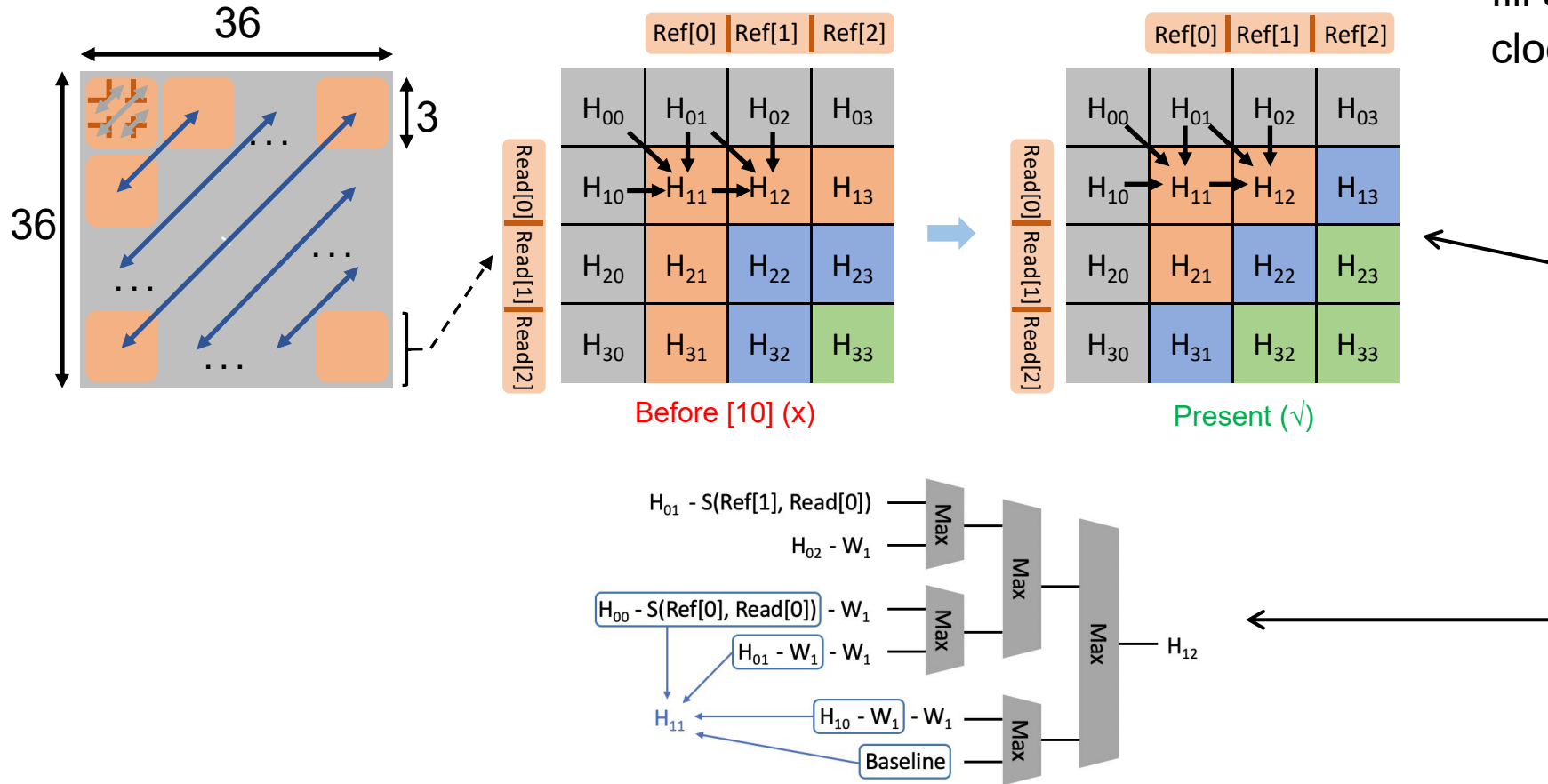


- ◆ A structure for global alignment in the banded S-W.
- ◆ Used among PEs to construct the main diagonal band.

- ◆ A systolic array structure consists of small calculation cells.
- ◆ Positions in the diagonal direction can do parallel calculation.

Hardware Architecture

● Calculation Cell



◆ Lookahead calculation technique, fill a 3×3 scoring matrix within three clock cycles.

■ More balanced critical path.

□ Six candidate results of $H_{1,2}$ are directly compared, without knowing $H_{1,1}$.

□ Within the same 3 cycles, only 3 positions ($H_{1,2}$, $H_{2,1}$, $H_{3,3}$) need lookahead technique after modification.

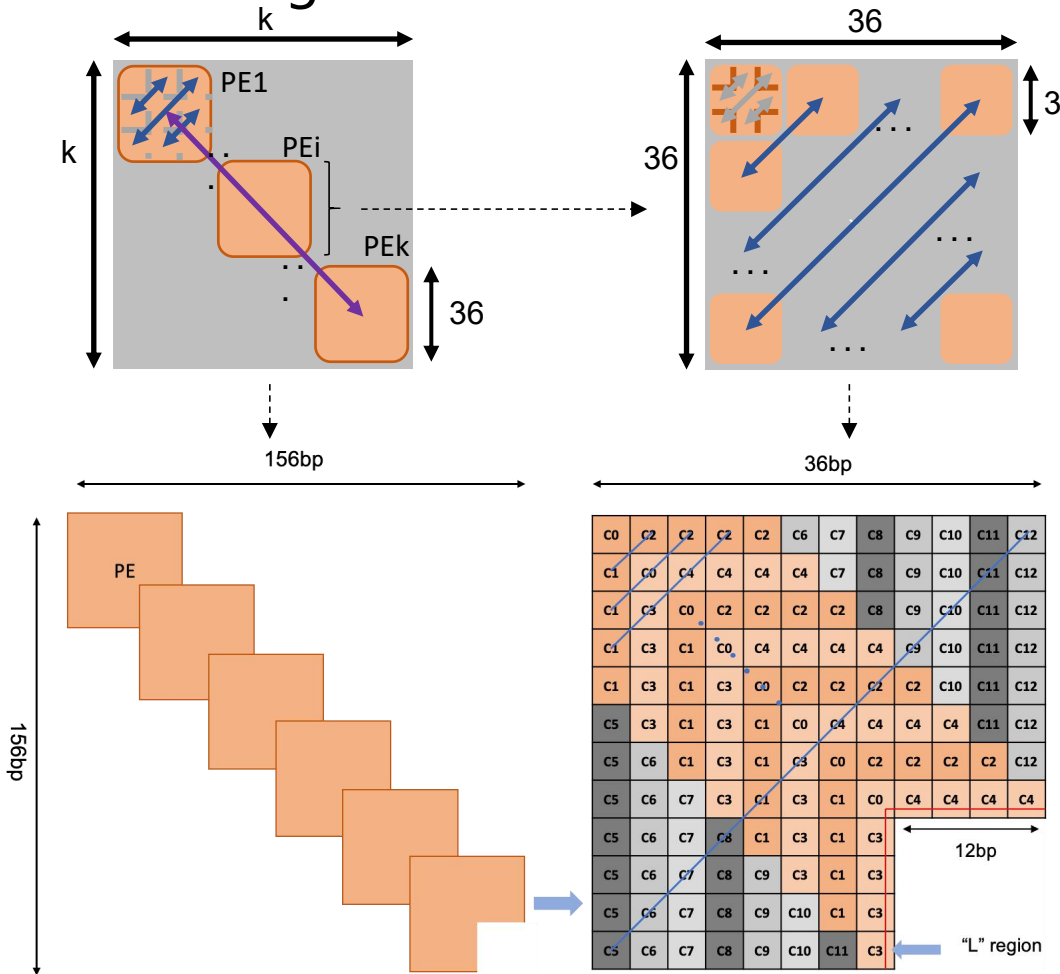
■ More parallel and reusable comparison.

□ Parallel comparison.

□ Parts of the intermediate results can be shared.

Hardware Architecture

● Main Diagonal Band



◆ Systolic array structure

- Thirteen 3×3 cells are sufficient to constitute a 36×36 processing element (PE).
- A 12×12 overlap between each two PEs.

◆ Diagonal Band

- 25 positions at the end of the PE compose a “L” region.
- The “L” region is expanded from the end to the beginning of the PE.
- All the cells along “L” regions need to record the direction information to generate a direction matrix.

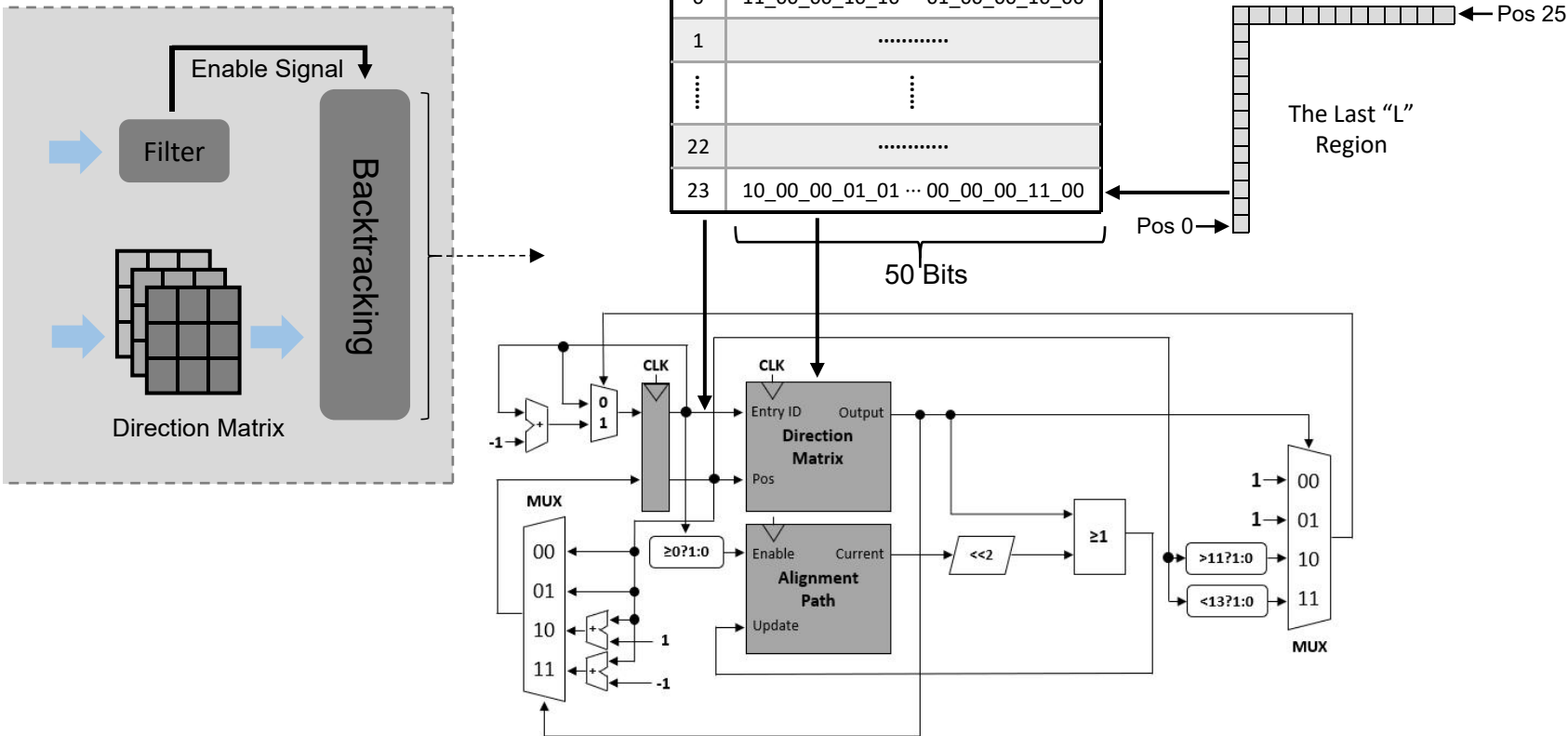
Why only cells along the “L” region? The filter module ensures that the backtracking path is bounded in the band.



Hardware Architecture

● Backtracking Module

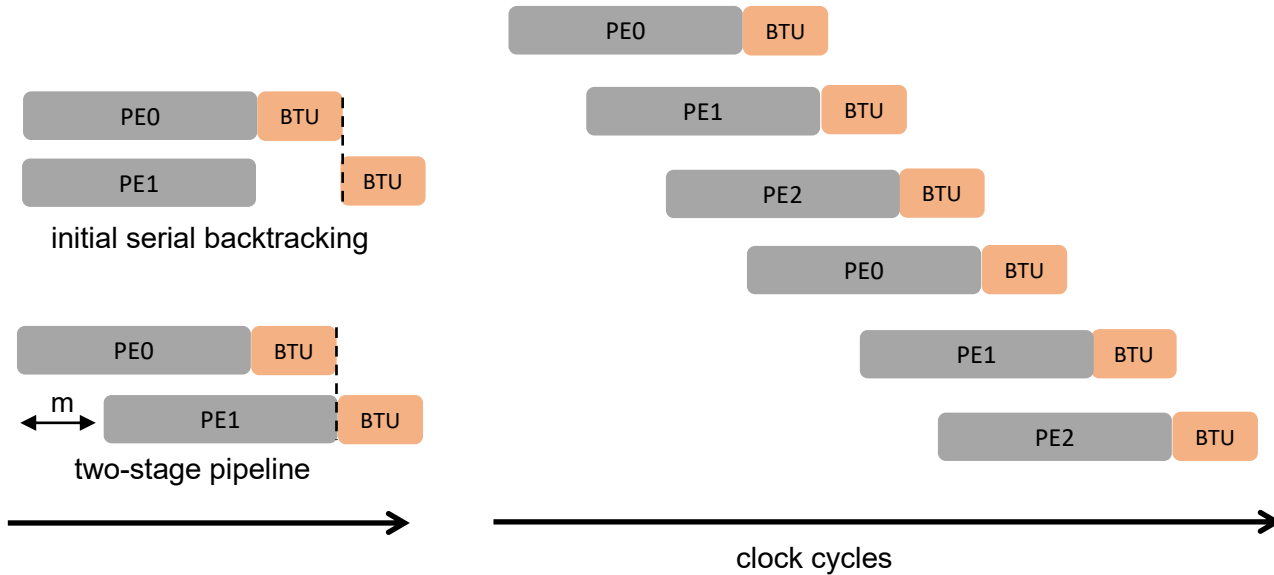
- ◆ A register buffer is designed to store the direction matrix and a hardware module is designed to do the backtracking step.
- ◆ The module can directly read information from the buffer, instead of sending the matrix to a general-purpose processor.
- ◆ The entry id pointer and the position pointer are updated in each cycle based on the information that is read.
- ◆ Avoiding communication with the memory, the entire process **only spends 24-36 cycles.**





Hardware Architecture

● Two-Stage Pipeline Structure



- ◆ Parallel scoring vs. Serial scoring
 - **Parallel:** fast but resource-consuming
 - **Serial:** accurate, resource-saving, but slow
- ◆ Parallel backtracking vs. Serial backtracking
 - **Parallel:** inaccurate (x)
 - **Serial:** accurate (✓)
- ◆ v1: “Parallel scoring + Serial backtracking”
 - Some PEs are in the idle state when waiting for BTU.
 - They can be reused for next-round calculation.
- ◆ v2: “Two-stage pipeline”
 - To be more re-configurable, the start time of PE1 can be delay by m cycles after PE0.
 - The **reuse rate of PE modules can be significantly improved** and **the resources saved are approaching 50%**.

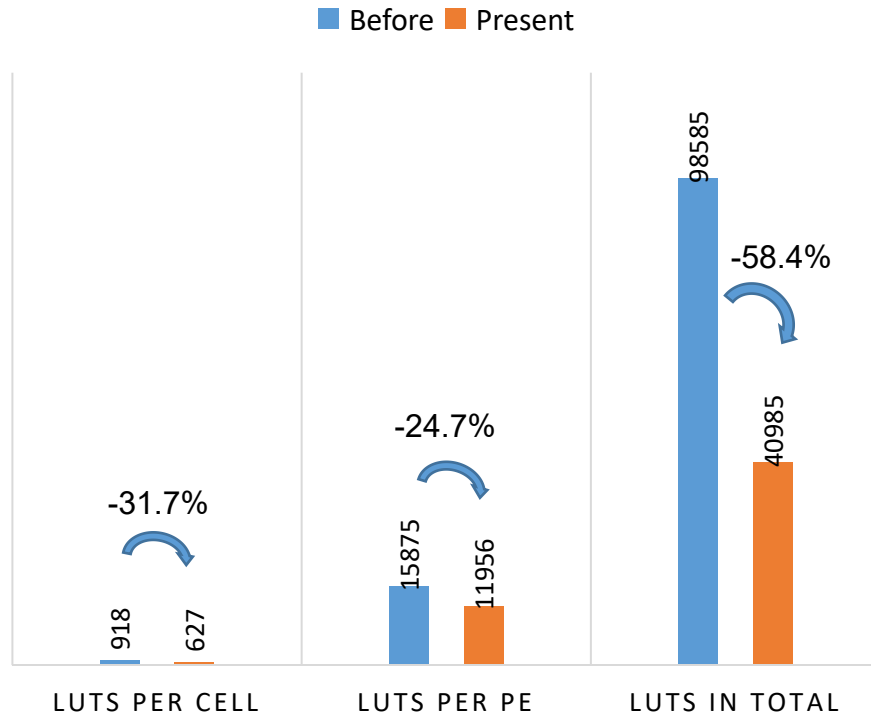
Outline

- Background
- Hardware Architecture
- **Experimental Validation**
- Conclusion



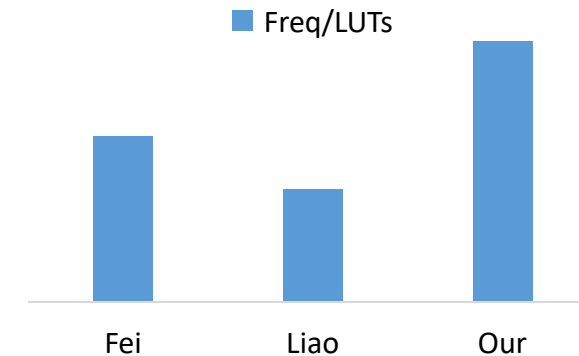
Experimental Validation

- Resource consumption (6 PEs)



- Comparison

Work	Backtrack	Freq(MHz)	LUTs	Arch
Nawaz[5]	Yes	79.3	-	FPGA
Fei[6]	Yes	150.0	57870	FPGA
Liao[7]	Yes	125.0	70839	FPGA
Our	Yes	166.7	40985	FPGA



[5] Z. Nawaz, M. Nadeem, H. van Someren, and K. Bertels, "A parallel fpga design of the smith-waterman traceback," in 2010 International Conference on Field-Programmable Technology, pp. 454–459, IEEE, 2010.

[6] X. Fei, Z. Dan, L. Lina, M. Xin, and Z. Chunlei, "Fpgasw: Accelerating large-scale smith–waterman sequence alignment application with backtracking on fpga linear systolic array," Interdisciplinary Sciences: Computational Life Sciences, vol. 10, no. 1, pp. 176–188, 2018.

[7] Y.-L. Liao, Y.-C. Li, N.-C. Chen, and Y.-C. Lu, "Adaptively banded smith-waterman algorithm for long reads and its hardware accelerator," in 2018 IEEE 29th International Conference on Application-specific Systems, Architectures and Processors (ASAP), pp. 1–9, IEEE, 2018.

Outline

- Background
- Hardware Architecture
- Experimental Validation
- Conclusion

Conclusion



- PipeBSW is proposed, a re-configurable system **implementing the whole Smith-Waterman algorithm on FPGA.**
- The lookahead calculation technique improves parallelism of cells, **reducing the time** it takes to complete the scoring.
- The hardware backtracking module directly generates the alignment path from the direction matrix buffer, **mitigating potential memory bandwidth issue.**
- The pipeline structure adjusts the calculation timing between modules, **reducing the resource consumption by 58.4%.**
- Compared with previous works, PipeBSW achieves both **high frequency (166.7MHz)** and **low resource consumption (40985LUTs).**



Thanks for Listening

If You Have Any Question, Please Contact Us at
luyli@smail.nju.edu.com

ICAIS Lab, Nanjing University, China