Smith Waterman Algorithm for Sequence Alignment

Karthika Jayaprakash

CSE 633 Parallel Algorithms (Dr. Russ Miller)
April 21, 2020

University at Buffalo
Department of Computer Science and Engineering
School of Engineering and Applied Sciences

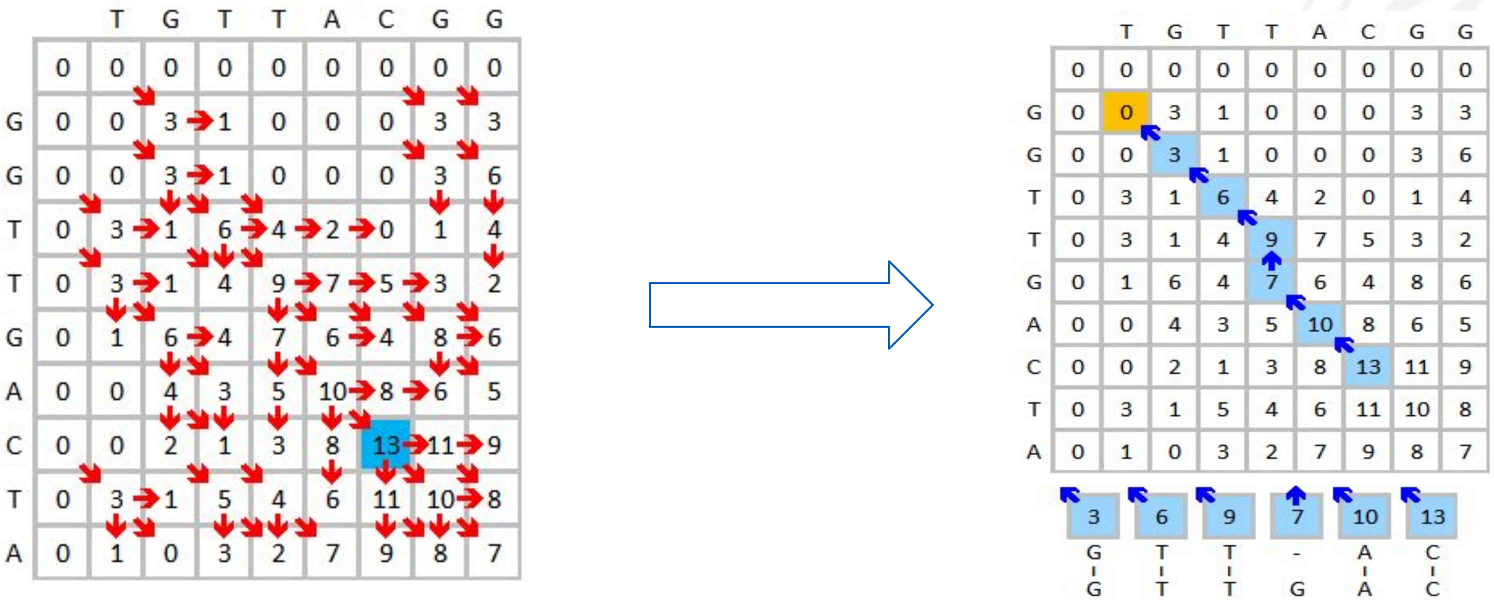
Smith Waterman Overview

Biological sequence alignment is a frequently performed task in bioinformatics.

The Smith-Waterman algorithm, based on dynamic programming, is one of the most fundamental algorithms used in local sequence alignment.

Smith Waterman Overview

Given two sequences find the best local alignment.

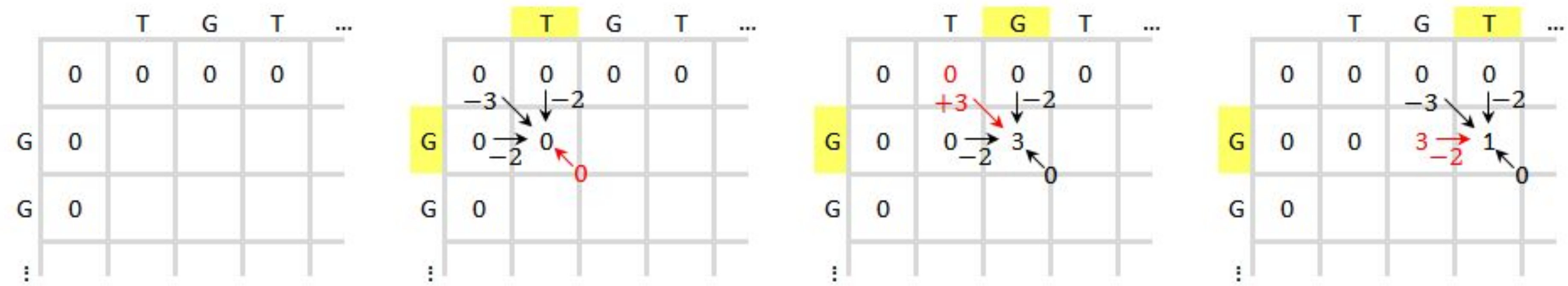


		T	G	T	T	A	C	G	G
	0	0	0	0	0	0	0	0	0
G	0	0	3	1	0	0	0	3	3
G	0	0	3	1	0	0	0	3	6
T	0	3	1	6	4	2	0	1	4
T	0	3	1	4	9	7	5	3	2
G	0	1	6	4	7	6	4	8	6
A	0	0	4	3	5	10	8	6	5
C	0	0	2	1	3	8	13	11	9
T	0	3	1	5	4	6	11	10	8
A	0	1	0	3	2	7	9	8	7

		T	G	T	T	A	C	G	G
	0	0	0	0	0	0	0	0	0
G	0	0	3	1	0	0	0	3	3
G	0	0	3	1	0	0	0	3	6
T	0	3	1	6	4	2	0	1	4
T	0	3	1	4	9	7	5	3	2
G	0	1	6	4	7	6	4	8	6
A	0	0	4	3	5	10	8	6	5
C	0	0	2	1	3	8	13	11	9
T	0	3	1	5	4	6	11	10	8
A	0	1	0	3	2	7	9	8	7

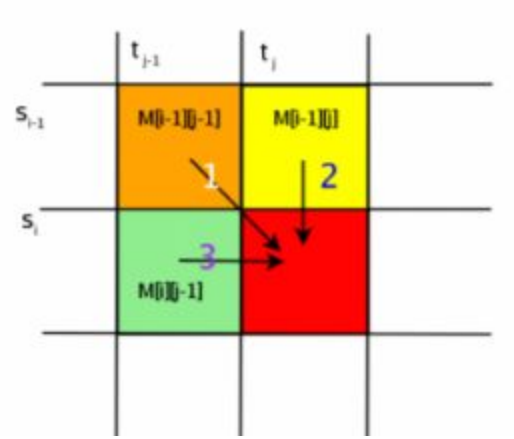
3	6	9	7	10	13
G	T	T	-	A	C
G	T	T	G	A	C

Smith Waterman Overview



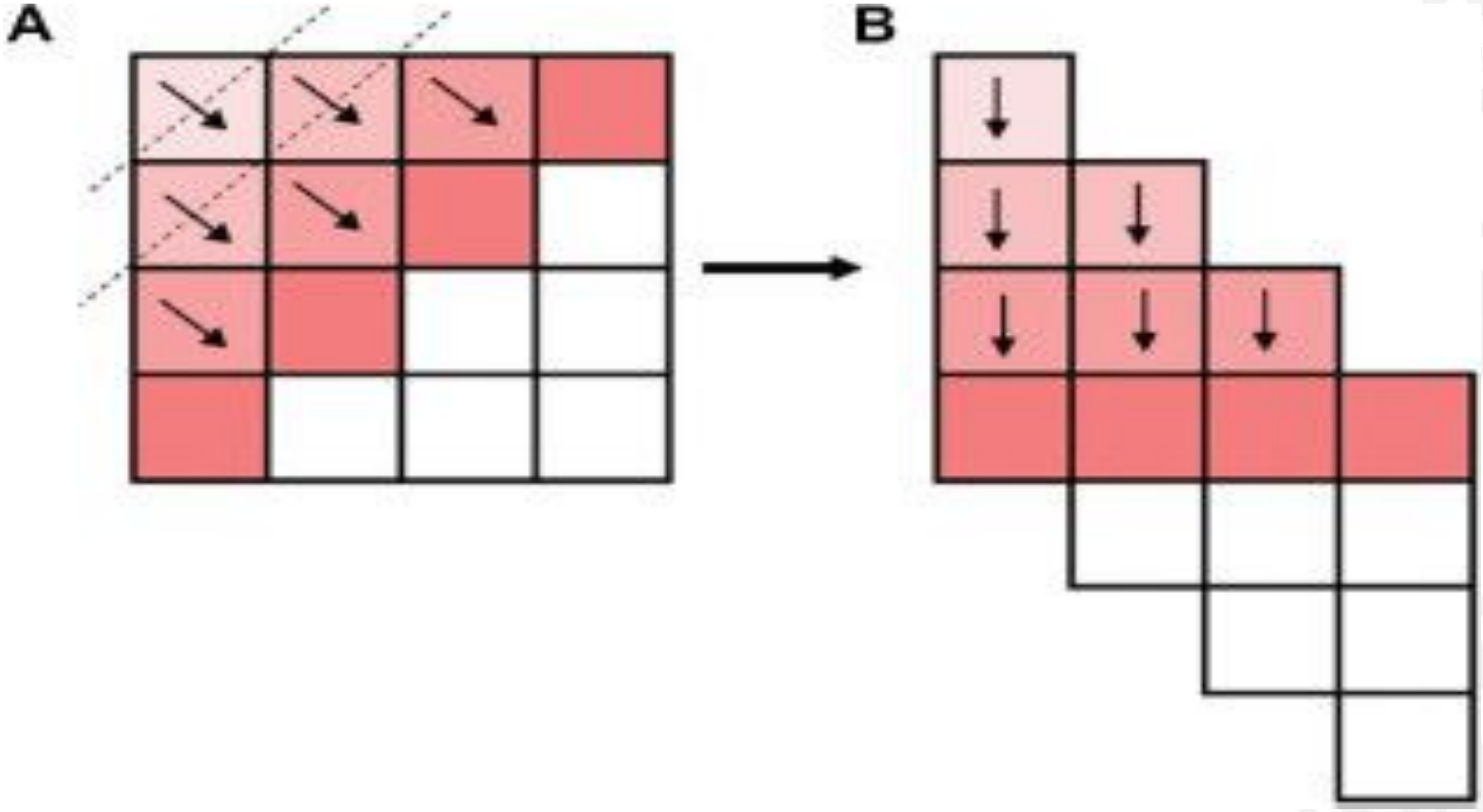
Smith Waterman - Parallel Approach

- **Diagonal Calculations are independent of each other.**
- **Hence Diagonal Computations are parallelized.**

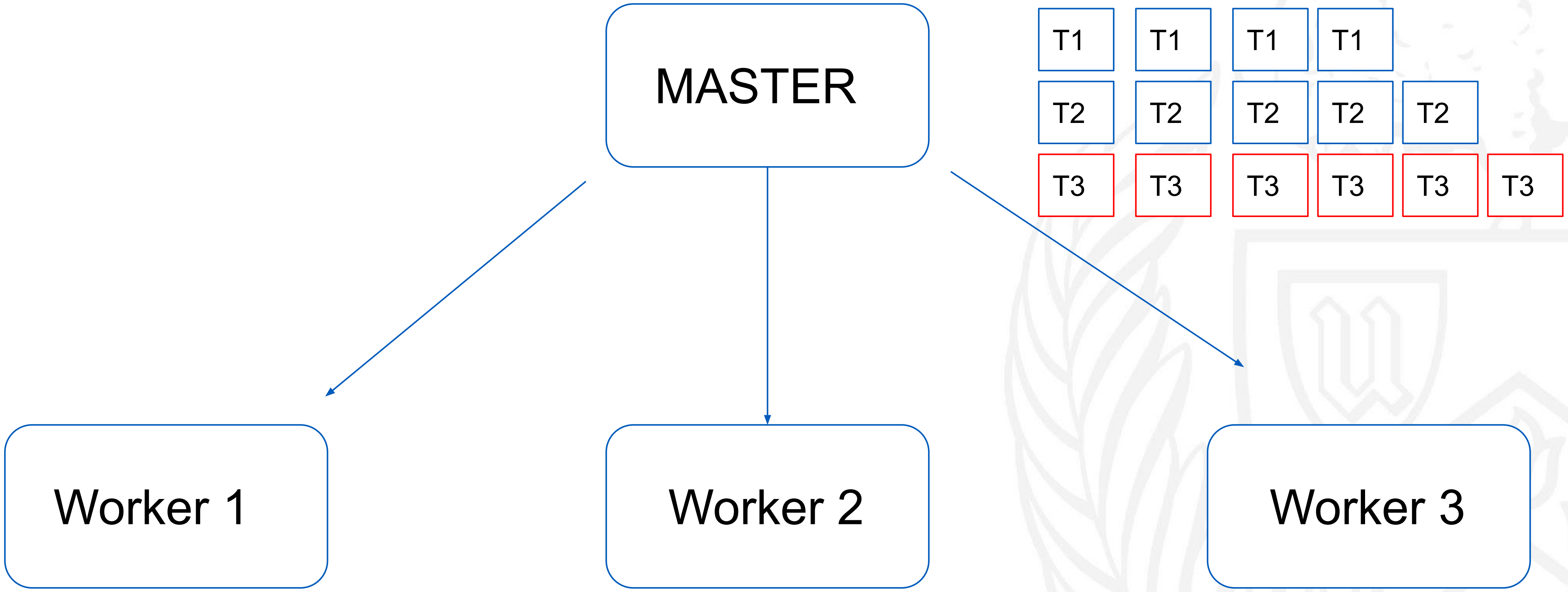


	-	C	G	G	G	T	A	T	C
-	0	0	0	0	0	0	0	0	0
C	0	T1	T2	T3	T4	T5	T6	T7	T8
C	0	T2	T3	T4	T5	T6	T7	T8	T9
C	0	T3	T4	T5	T6	T7	T8	T9	T10
T	0	T4	T5	T6	T7	T8	T9	T10	T11
A	0	T5	T6	T7	T8	T9	T10	T11	T12
G	0	T6	T7	T8	T9	T10	T11	T12	T13
G	0	T7	T8	T9	T10	T11	T12	T13	T14
T	0	T8	T9	T10	T11	T12	T13	T14	T15

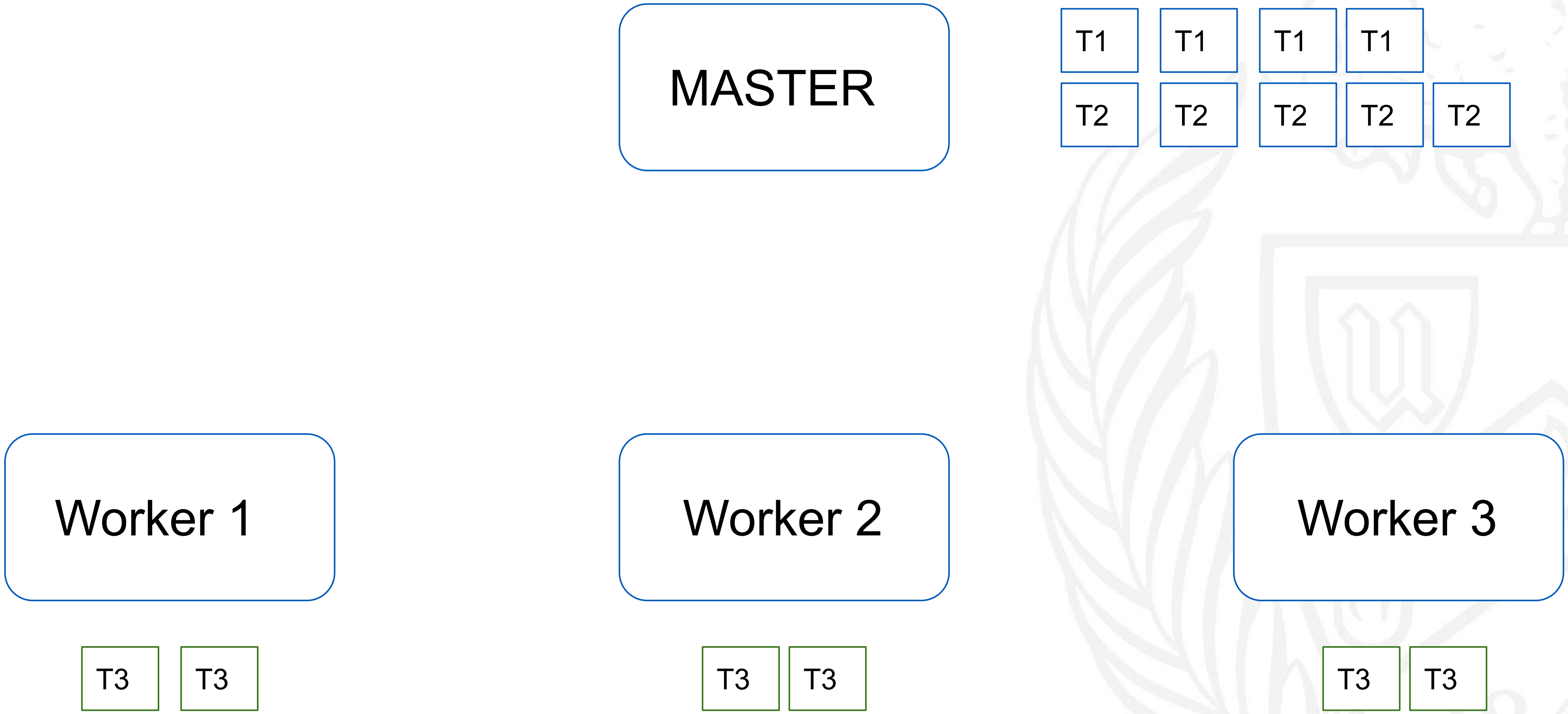
Smith Waterman - Parallel Approach



Smith Waterman - Parallel Approach



Smith Waterman - Parallel Approach



MASTER

T1 T1 T1 T1

T2 T2 T2 T2 T2

Worker 1

T3 T3

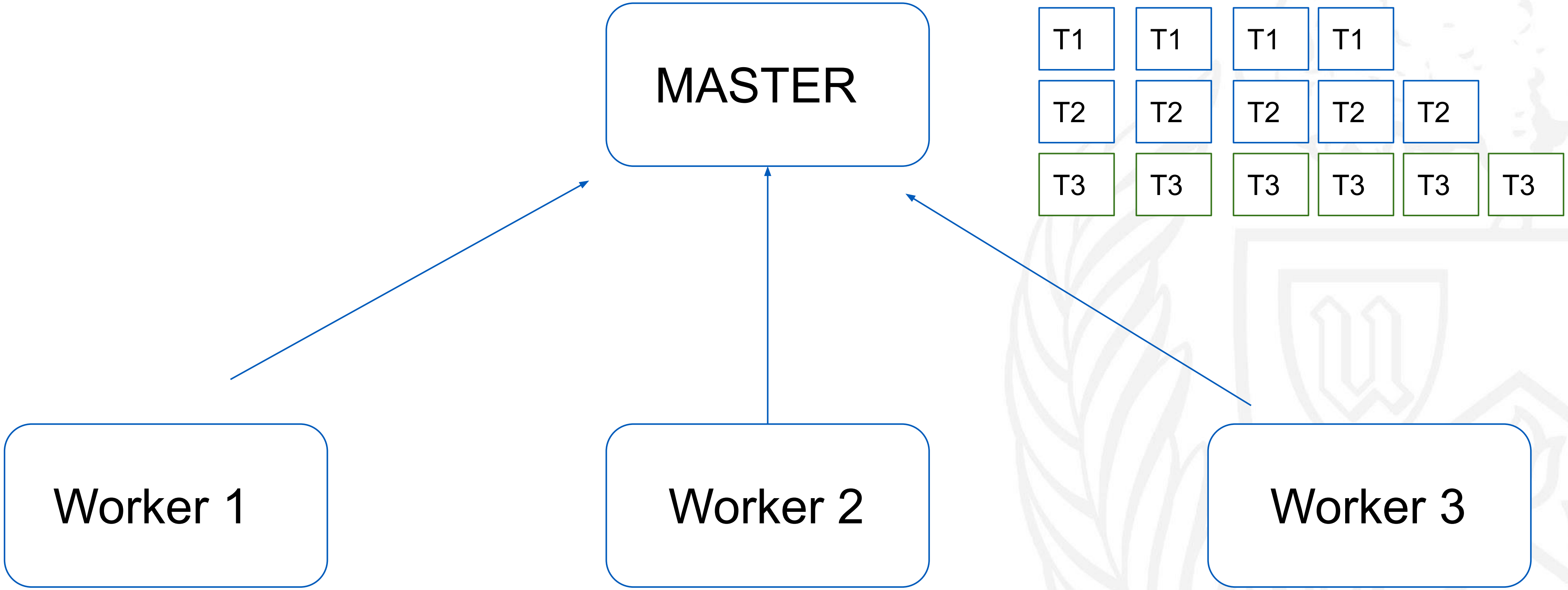
Worker 2

T3 T3

Worker 3

T3 T3

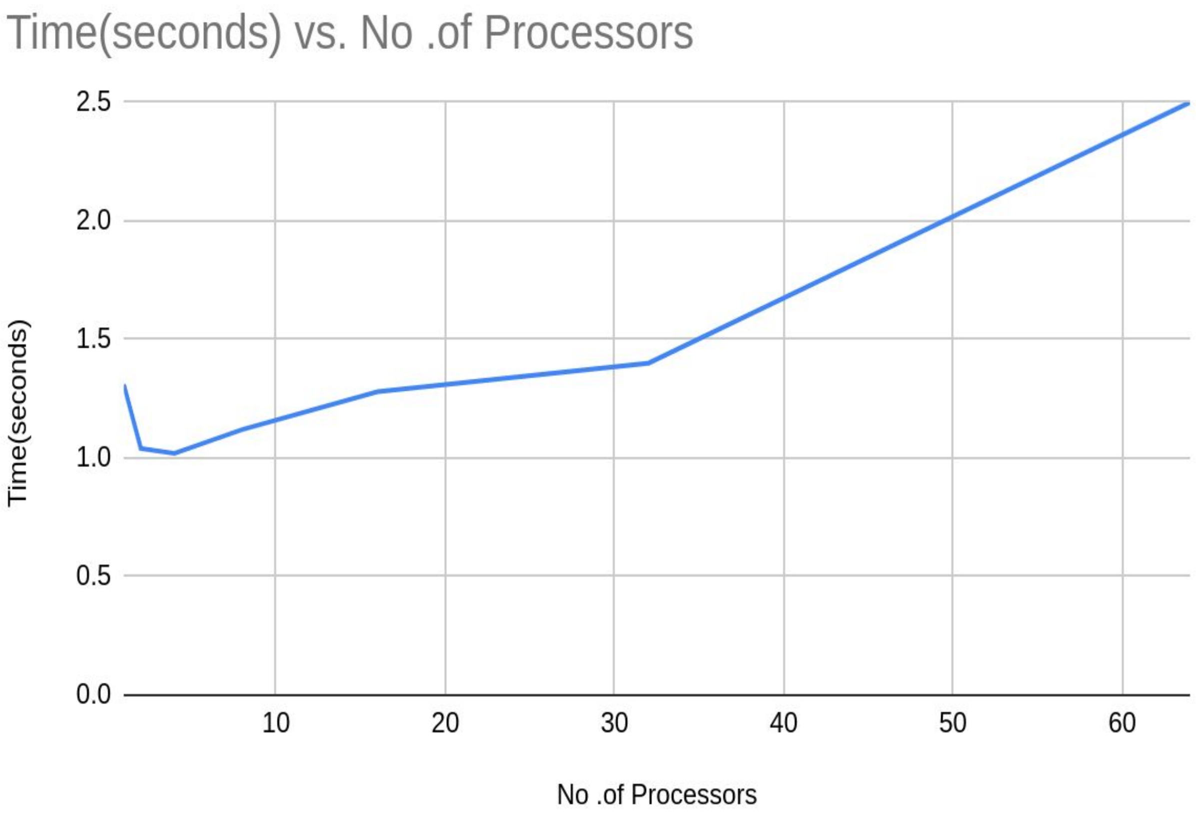
Smith Waterman - Parallel Approach



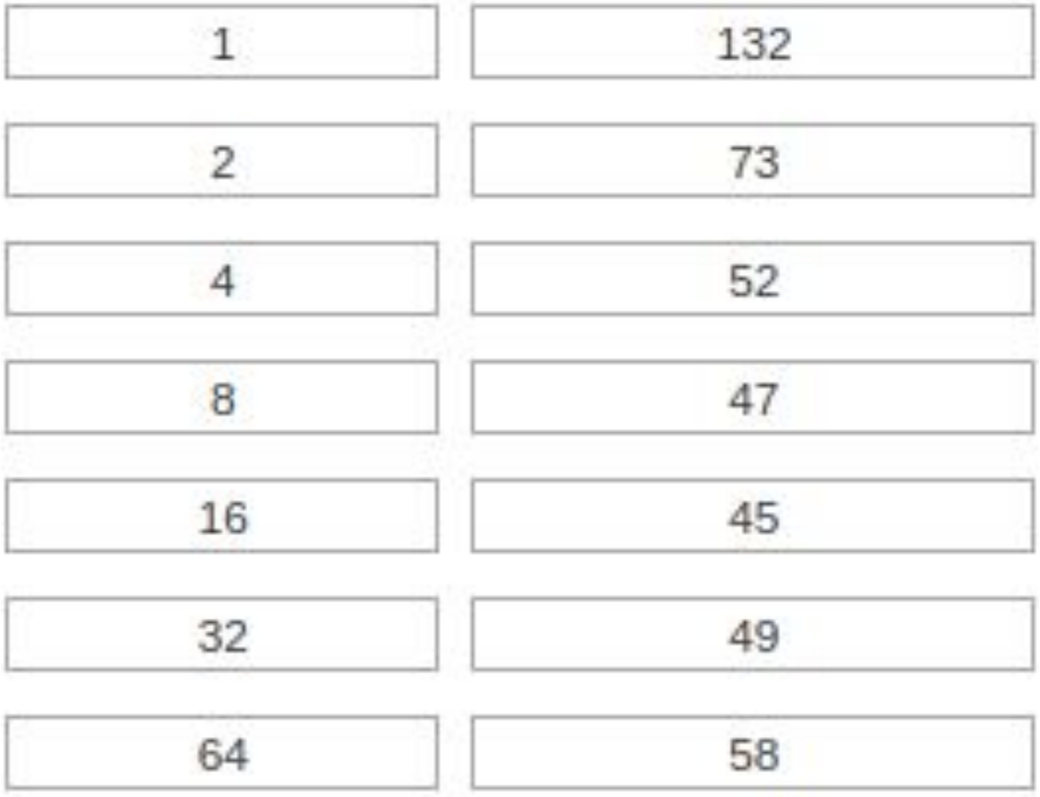
Runtime Comparison

1000 x 1000 matrix

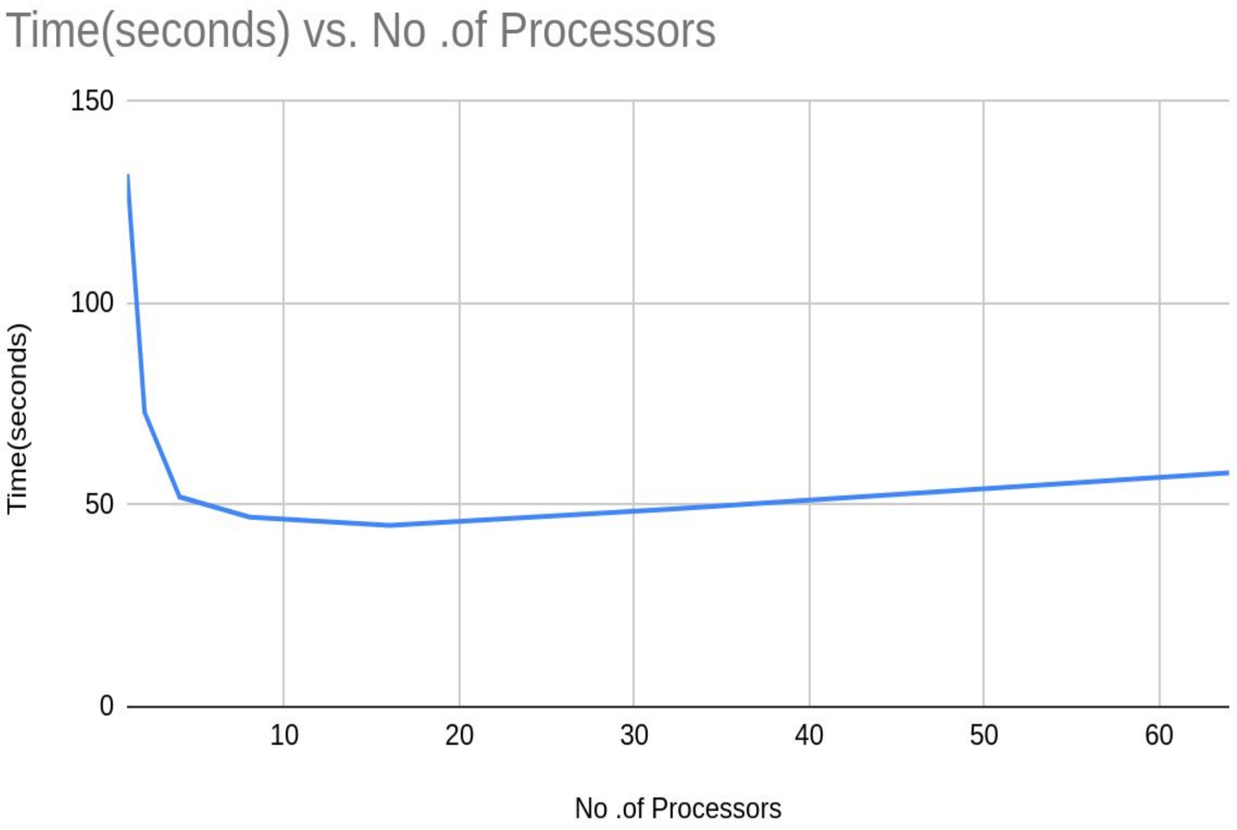
1	1.31
2	1.04
4	1.02
8	1.12
16	1.28
32	1.4
64	2.5



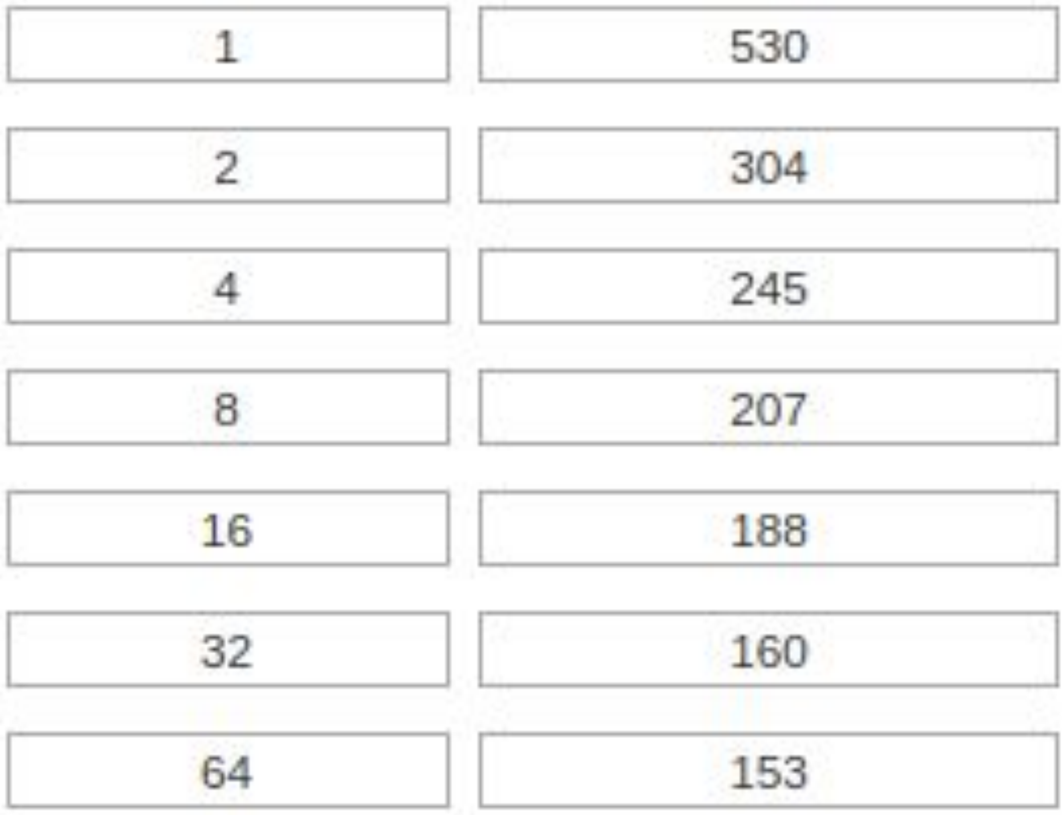
10000 x 10000 matrix



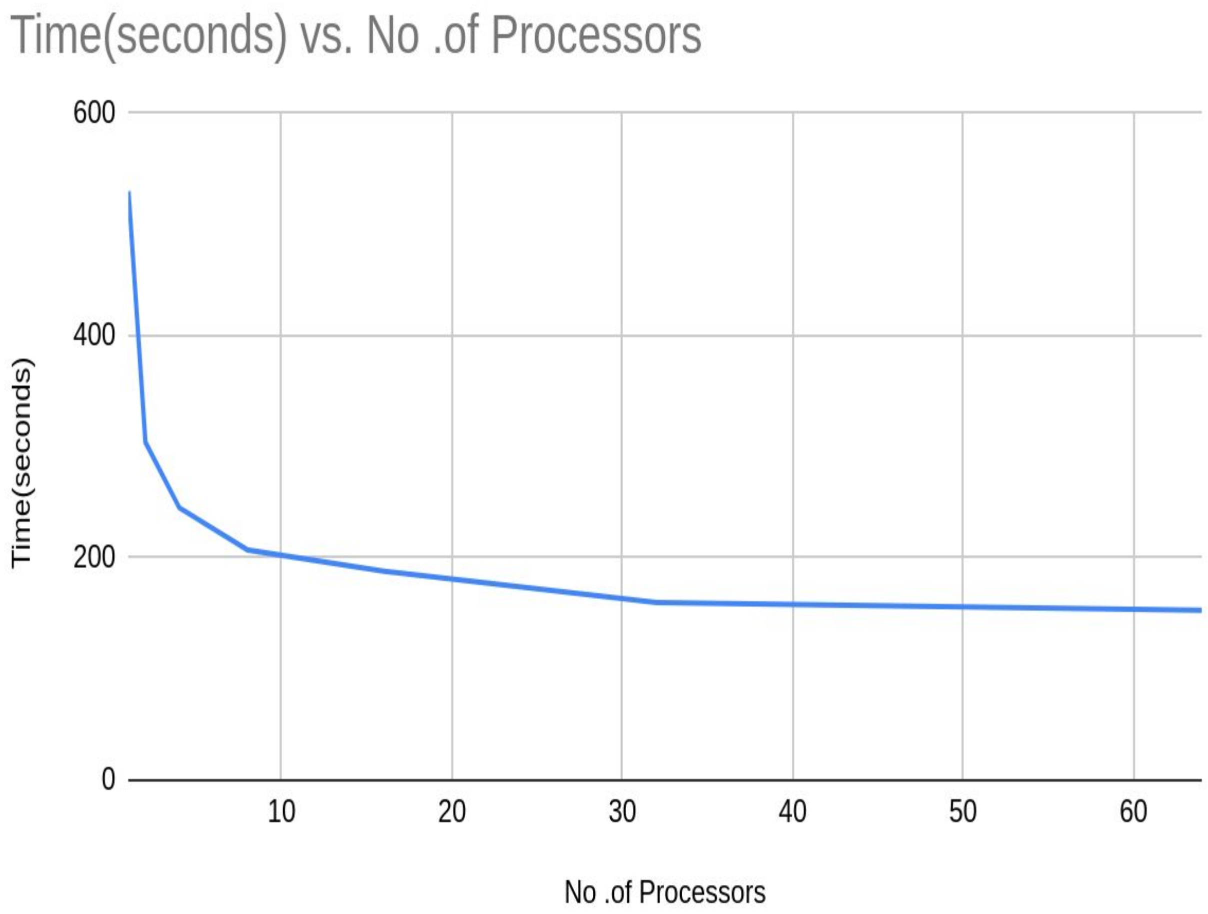
1	132
2	73
4	52
8	47
16	45
32	49
64	58



20000 x 20000 matrix



1	530
2	304
4	245
8	207
16	188
32	160
64	153



Conclusion

The computations which can be executed parallely are identified.

The diagonals are computed parallely at each step.

The data in each diagonal are split among the processes to compute scores.

Parallelization has resulted in speedup of computations.

References

Fa ZHANG1 , Xiang-Zhen QIAO1 , Zhi-Yong LIU “A Parallel Smith-Waterman Algorithm Based on Divide and Conquer” Proceedings of the Fifth International Conference on Algorithms and Architectures for Parallel Processing

MPI4PY documentation - https://mpi4py.readthedocs.io/en/stable/tutorial.html

CCR Resources - https://ubccr.freshdesk.com/support/home

Thank You