A Tile-based Parallel Viterbi Algorithm for Biological Sequence Alignment on GPU with CUDA

Zhihui Du[1], Zhaoming Yin[2], David A. Bader[3]

1, Department of Computer Science, Tsinghua University.
2, School of Software and Microelectronics, Peking University.
3, School of Computing, Georgia Institute of Technology
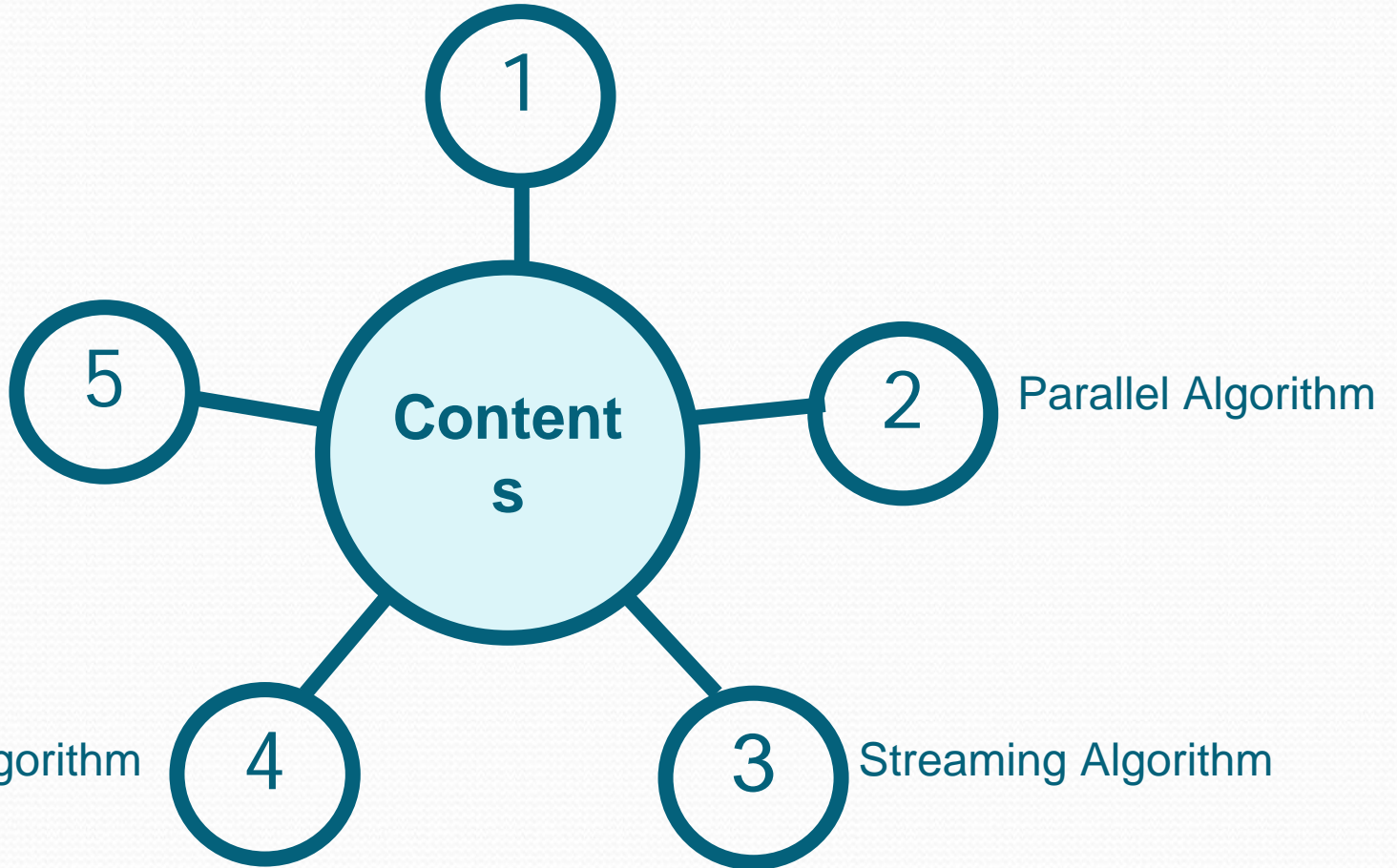
Contents
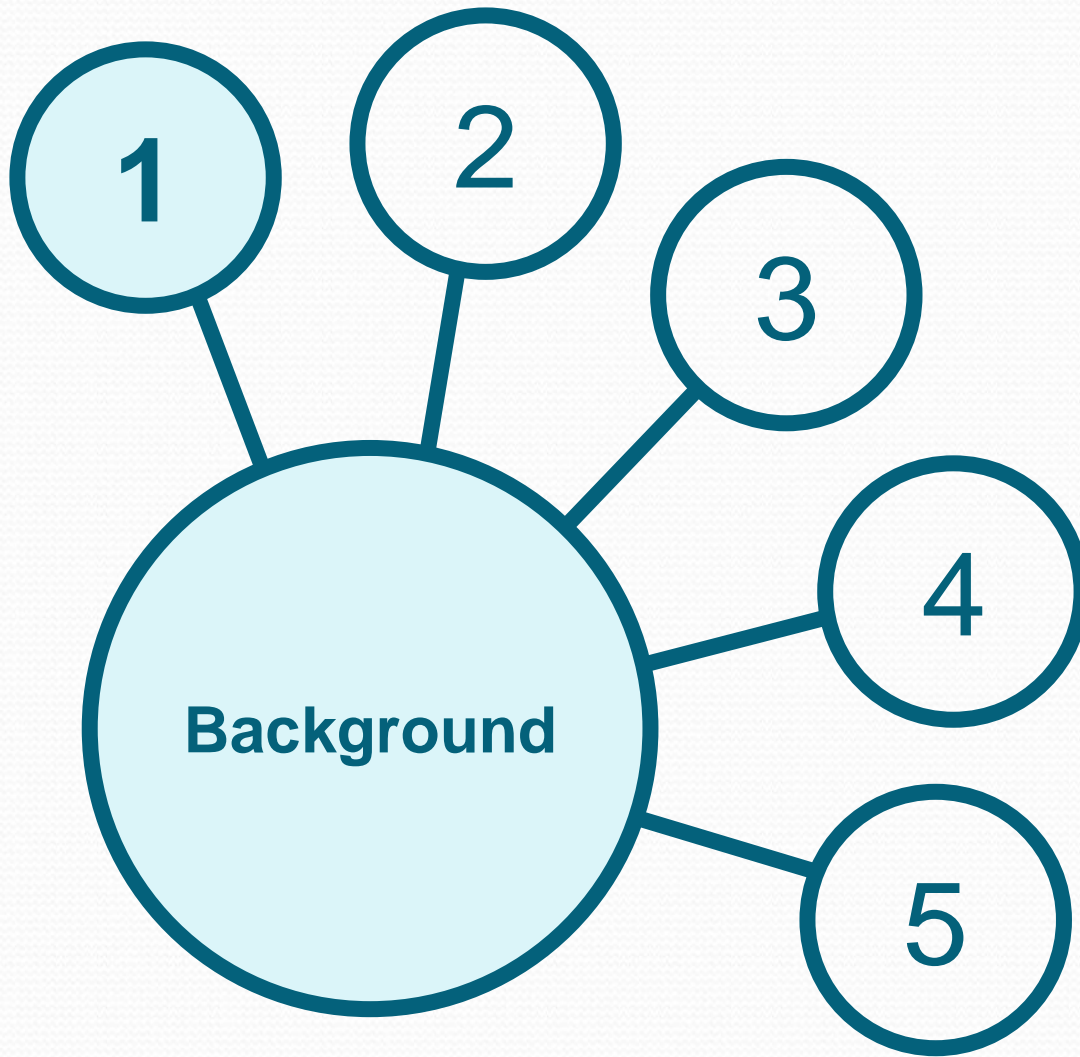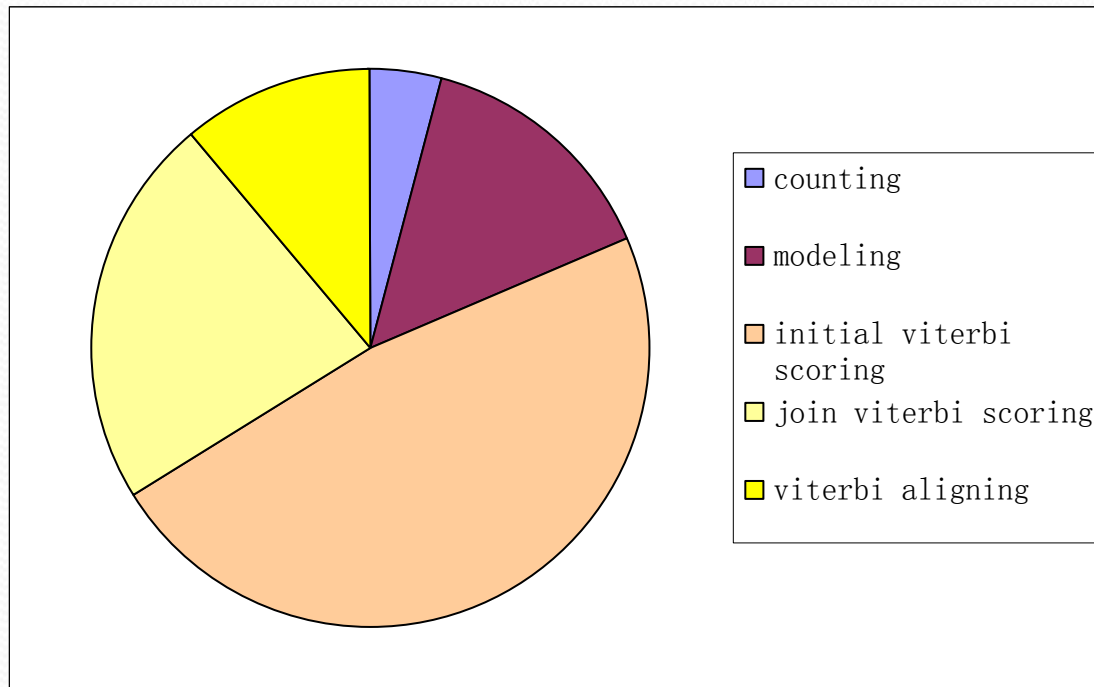
Background

The Importance of Accelerating Viterbi Algorithm

- In the test of SATCHMO, Viterbi algorithm occupy anout 80% compuation time

counting

modeling

initial viterbi scoring

join viterbi scoring

viterbi aligning

Dynamic Programming Matrix

				H+			H-			H-		
	$-\infty$	$-\infty$	$-\infty$	$-\infty$	-1.0	$-\infty$	$-\infty$	-2.0	$-\infty$	$-\infty$	-3.0	$-\infty$
	N/A	N/A	N/A	N/A	л	N/A	N/A	DE	N/A	N/A	DE	N/A
H-	$-\infty$	$-\infty$	-1.3	-0.8	-2.1	-1.8	-1.85	-1.4	-2.6	-2.4	-2.2	-3.8
	N/A	N/A	л	л	IN	DE	DE	MA	DE	DE	DE	DE
H-	$-\infty$	$-\infty$	-2.6	-2.6	-3.9	-1.8	-1.45	-2.6	-2.0			
	N/A	N/A	IN	IN	IN	MA	MI	IN	DE	MA	MA	DE

Emit Probability

Transition Probability

Forward Variable

-2.25 -2.40 -3.0
-2.25 + log(0.5) + log(0.8)
= -2.25
-1.4 + log(0.1) + log(0.5)
= -2.7
-2.6 + log(0.3) + log(0.5)
= -3.4

Trace Back

	H+	H-	H-
	N/A N/A N/A	N/A Л N/A	N/A DE N/A	N/A DE N/A
H-	N/A N/A Л	Л IN DE	DE MA DE	DE DE DE
H-	N/A N/A IN	IN IN MA	MI IN DE	−2.25 −2.45 −3.0 / MA MA DE

0:H-
  H-
1:H-H-
 H- H-
2:H-H- H+
  H- H- -

result

H+H-H-
- H- H-

Related Work

- W. Liu, B. Schmidt, G. Voss, W. Muller Wittig. "Streaming Algorithms for Biological Sequence Alignment on GPUs" IEEE TPDS, Vol. 18, No. 9. (2007), pp. 1270-1281.

- Y. Munekawa, F. Ino, and K. Hagihara. Design and Implementation of the Smith-Waterman Algorithm on the CUDA-Compatible GPU. 8th IEEE International Conference on BioInformatics and BioEngineering, pages 1 C6, Oct .2008

Related Work

- S.A. Manavski, G. Valle. CUDA compatible GPU cards as efficient hardware accelerators for Smith-Waterman sequence alignment. BMC Bioinformatics. 2008 Mar 26;9 Suppl 2:S10

- R. Horn, M. Houston, P. Hanrahan. ClawHMMer: A streaming HMMer –search implementation. Proc. Supercomputing (2005).
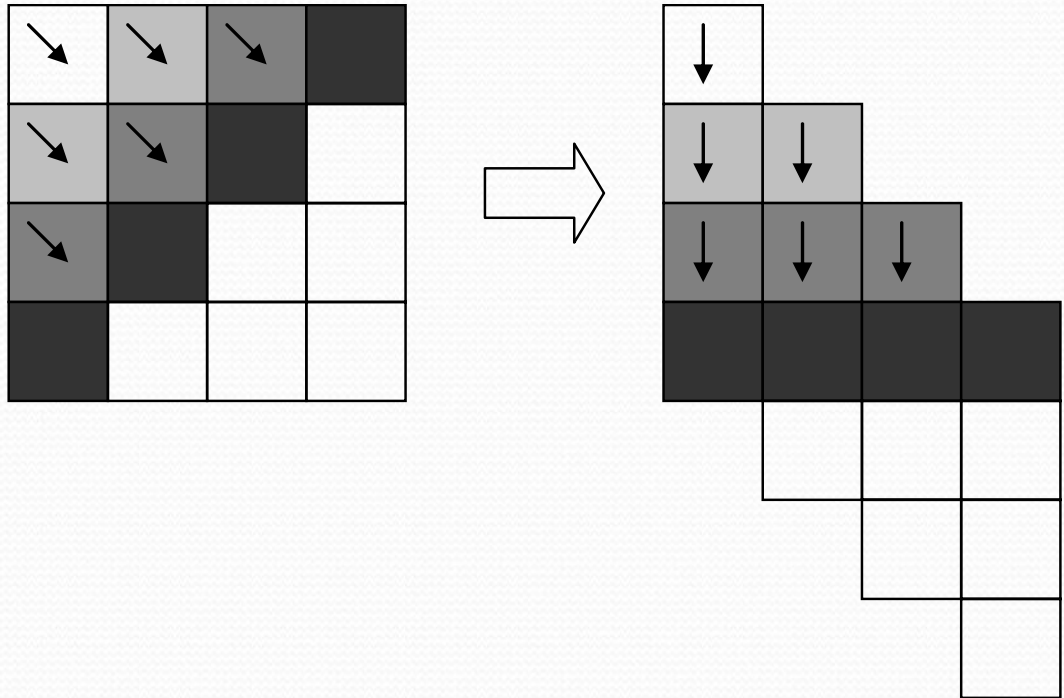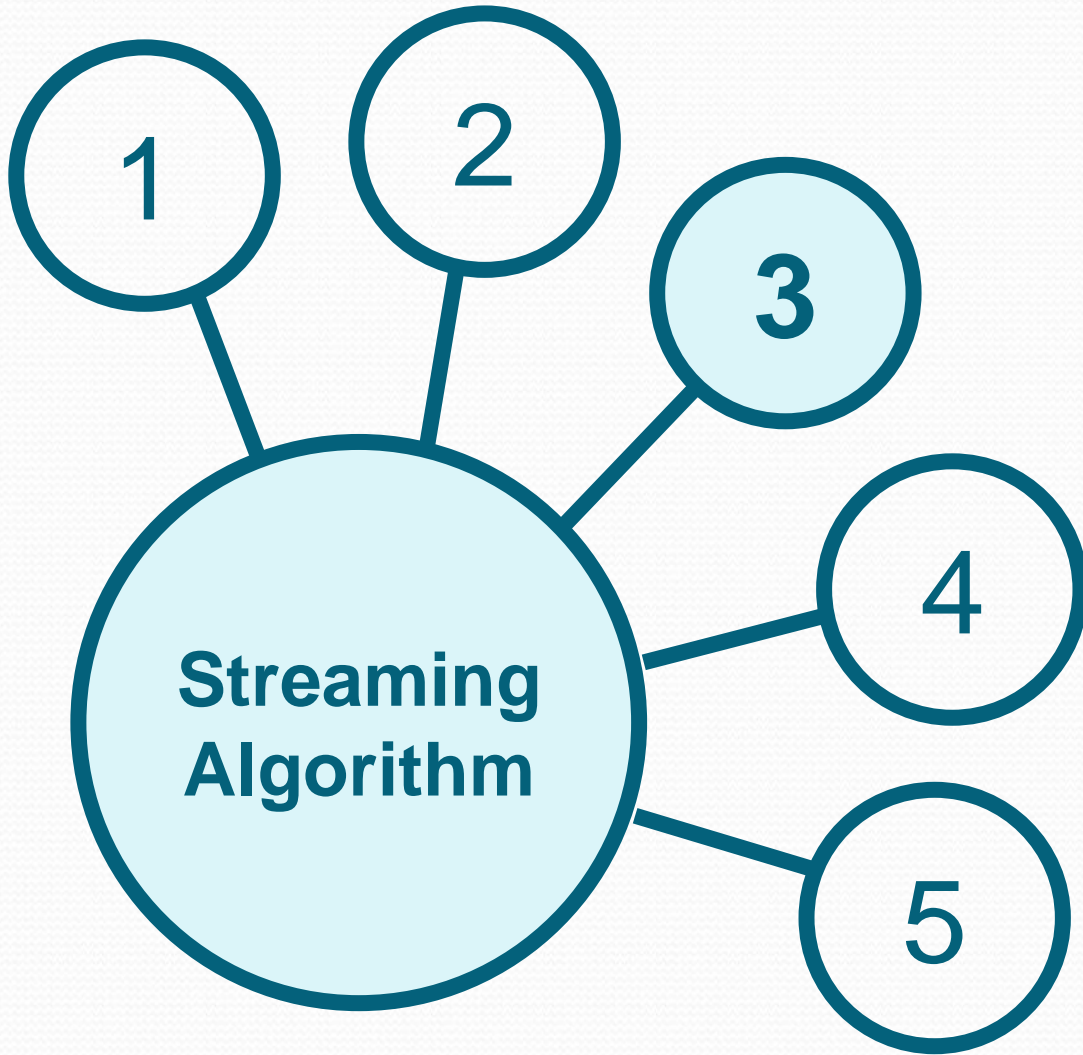
Parallel Algorithm
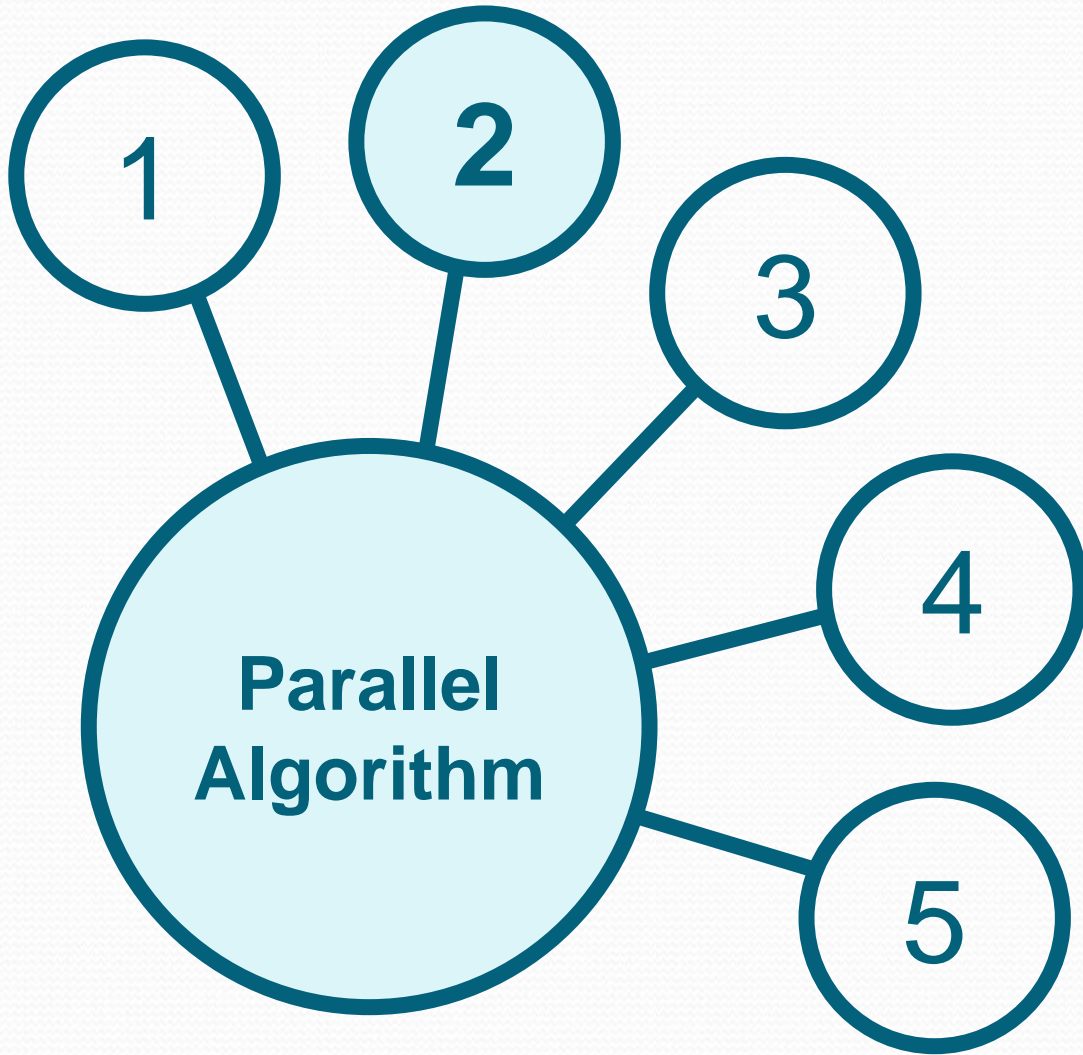
Wave-front Algorithm

wave-front algorithm:
The computing procedure
is similar to a frontier of a
wave to fill a matrix, where
each block's value in the
matrix is calculated based
on the values of the
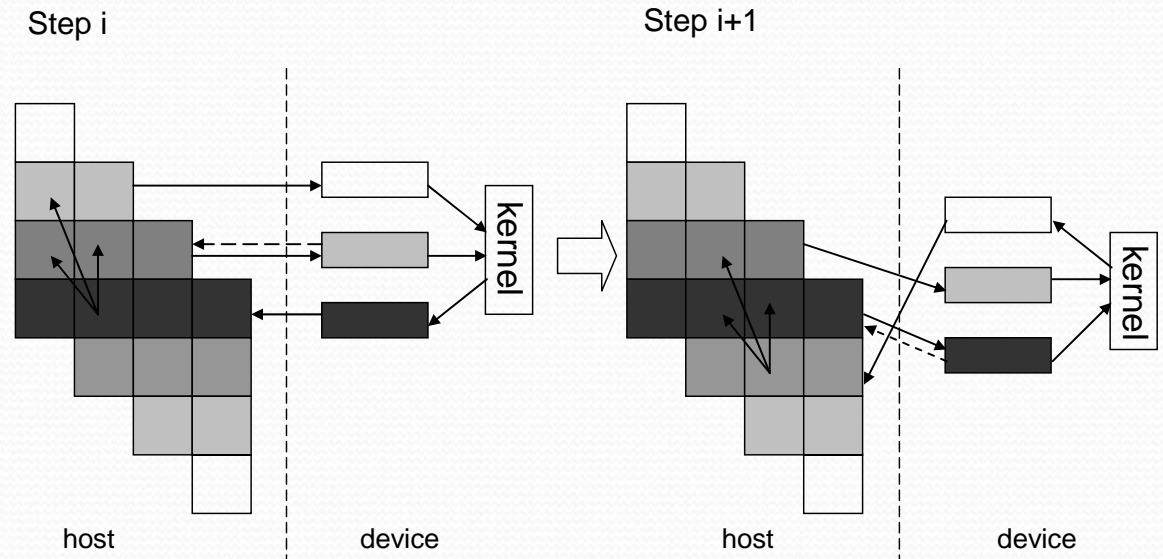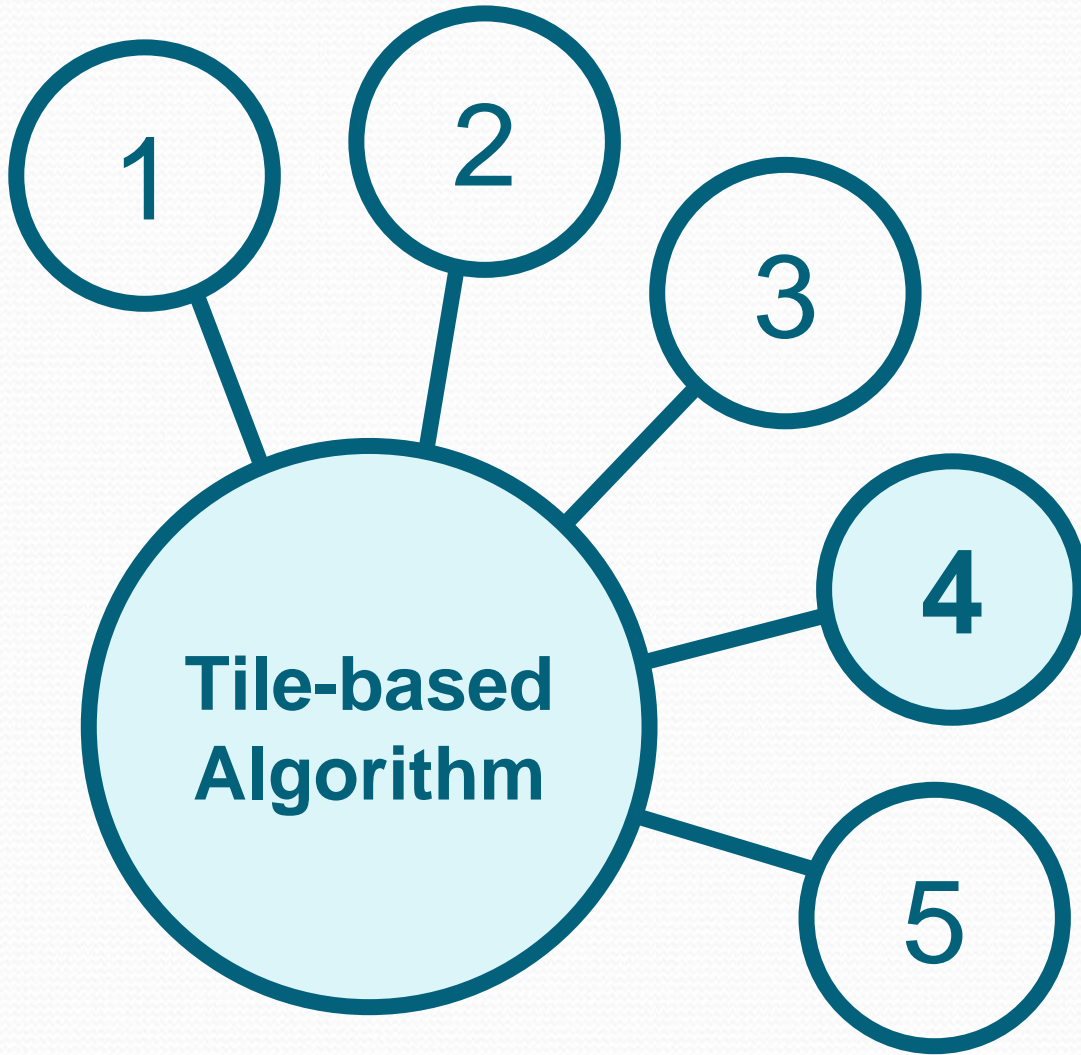previously-calculated blocks.

Streaming Algorithm

If the Sequence Length is Too Long!

Streaming Algorithm:
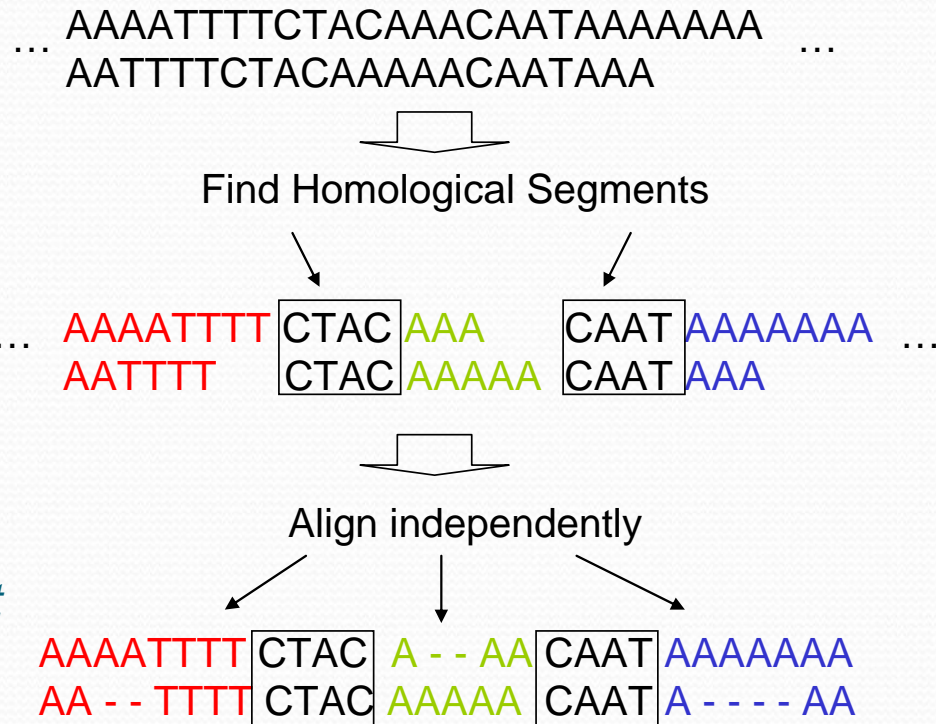Transfering data between Host and Device.

The Tile-based Algorithm

The Tile Based Algorithm

Step1:
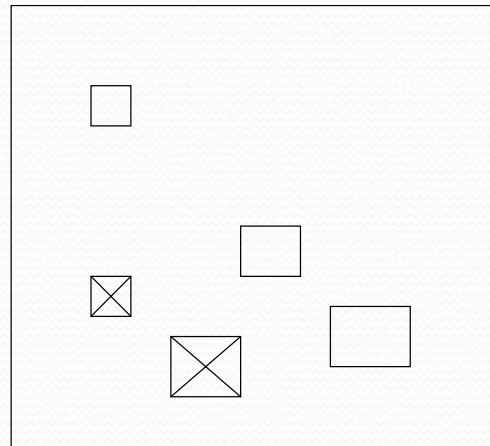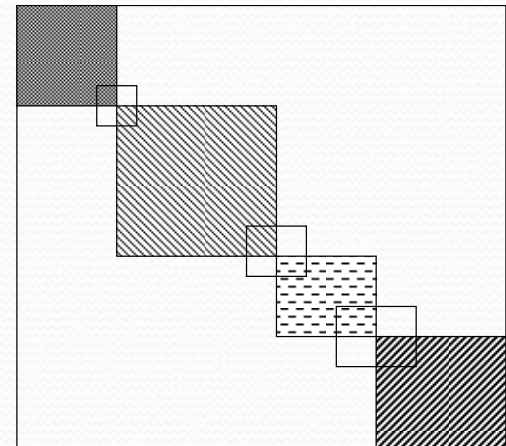Utilizing Homological Segments to divide long sequence

See:
M. Katoh and M. Kuma. "MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform".
In Nucleic Acids Res.
30:3059-3066 2002.

… AAAATTTTCTACAAACAATAAAAAAA …
AATTTTCTACAAAAACAATAAA

Find Homological Segments

… AAAATTTT CTAC AAA CAAT AAAAAAA …
AATTTT CTAC AAAAA CAAT AAA

Align independently

AAAATTTT CTAC A - - AA CAAT AAAAAAA
AA - - TTTT CTAC AAAAA CAAT A - - - - AA

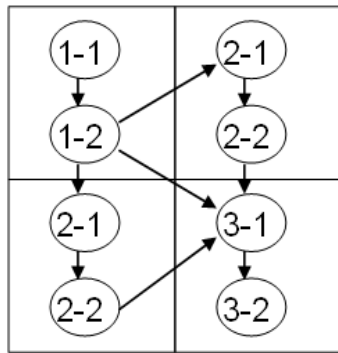# The Tile Based Algorithm

Step2:
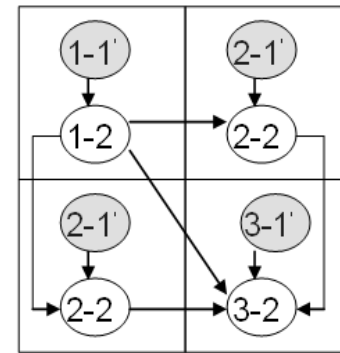Align sub-matrices

Find homological segment pairs

Divide sequence(shaded area)
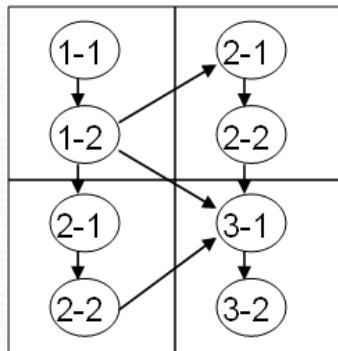
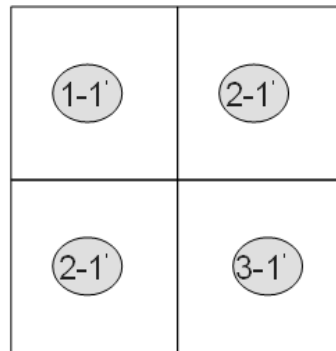# Partition of Different Kind of Computation
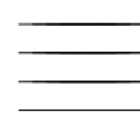

[A * B] * C = A * [B * C]

Formula Transformation
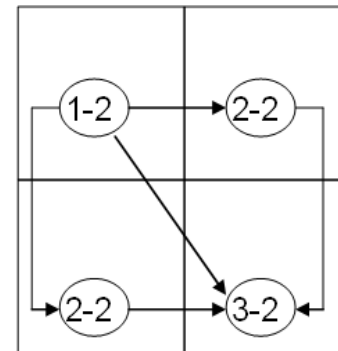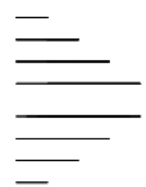
Partition of dependent and independent computations

Thread load

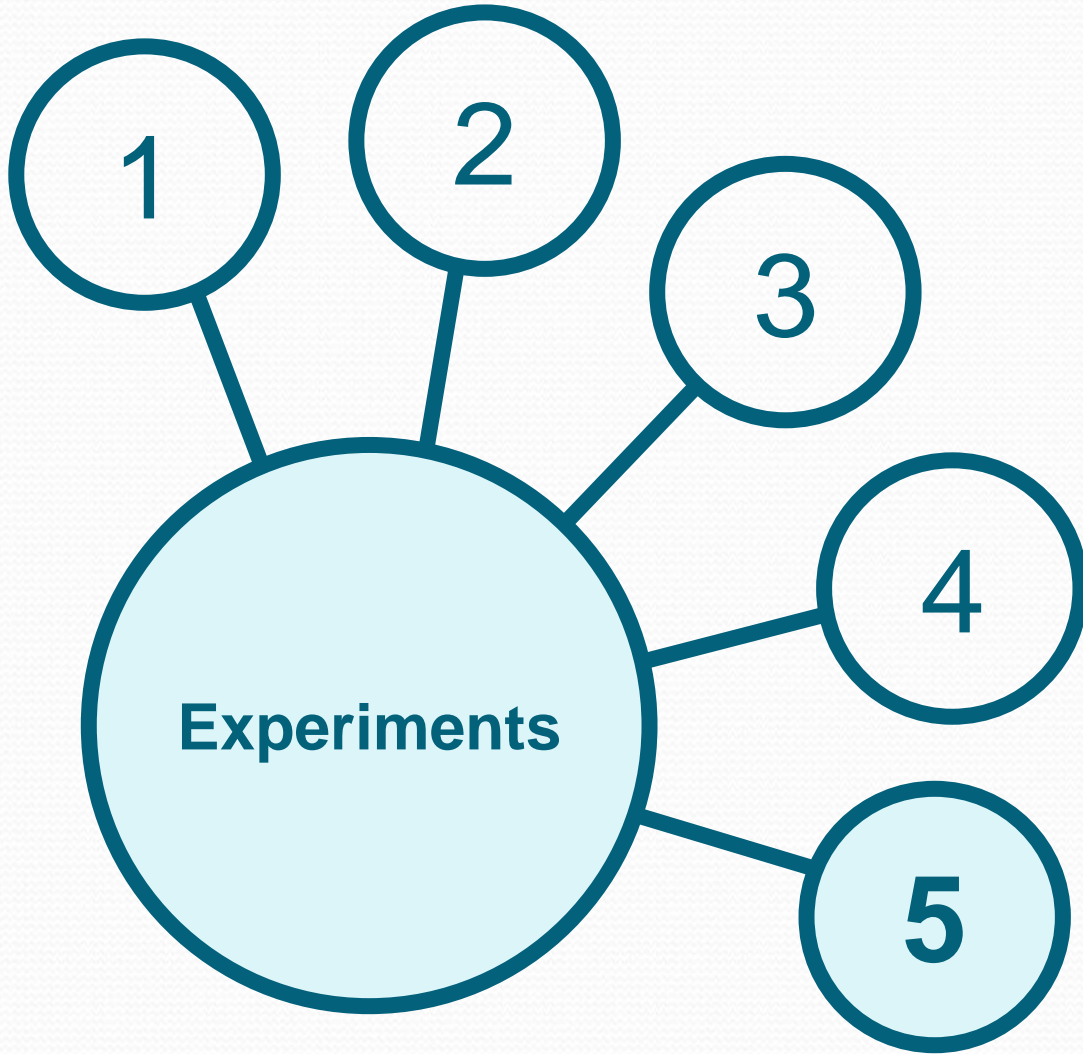Thread load

Experiments

Results

Intel Core 2
NVIDIA GeForce 9800

| Seq-Length		Execution Time (Second)/Speedup						
		serial	*Simple Wave-front*		*Streaming*		*Tile-based*	
100	DW	0.73	0.37	1.97	0.38	1.92	0.28	2.61
	RW	0.017	0.007	2.42	0.02	0.85	0.006	2.83
	DL	0.063	0.008	7.87	0.023	2.74	0.007	9
	RL	0.027	0.007	3.86	0.023	1.17	0.007	3.86
200	DW	2.34	0.39	6	0.44	5.32	0.39	6
	RW	0.05	0.03	1.67	0.061	0.82	0.028	1.79
	DL	0.324	0.035	9.26	0.065	4.98	0.029	11.17
	RL	0.142	0.035	4.06	0.065	2.18	0.029	4.9
300	DW	5.89	0.42	14.02	0.46	12.8	0.43	13.7
	RW	0.12	0.068	1.76	0.1	1.2	0.055	2.18
	DL	0.647	0.07	9.26	0.112	5.78	0.054	11.98
	RL	0.283	0.068	4.16	0.116	2.44	0.054	5.24
400	DW	9.93	0.50	19.86	0.52	19.1	0.45	22.07
	RW	0.21	0.13	1.61	0.159	1.32	0.098	2.14
	DL	1.112	0.12	9.27	0.2	5.56	0.099	11.23
	RL	0.485	0.122	3.98	0.174	2.79	0.097	5
500	DW	15.9	0.54	29.44	0.54	29.44	0.52	30.58
	RW	0.34	0.19	1.78	0.239	1.42	0.174	1.95
	DL	1.783	0.198	9	0.262	6.8	0.155	11.5
	RL	0.783	0.191	4.10	0.251	3.12	0.153	5.12
1000	DW	62.1	0.99	62.73	1.10	56.45	0.86	72.21
	RW	1.34	0.64	2.09	0.686	1.95	0.554	2.42
	DL	6.98	0.64	10.91	0.725	9.63	0.53	13.17
	RL	3.07	0.635	4.83	0.62	4.952	0.512	6.0

Test of Streaming Algorithm
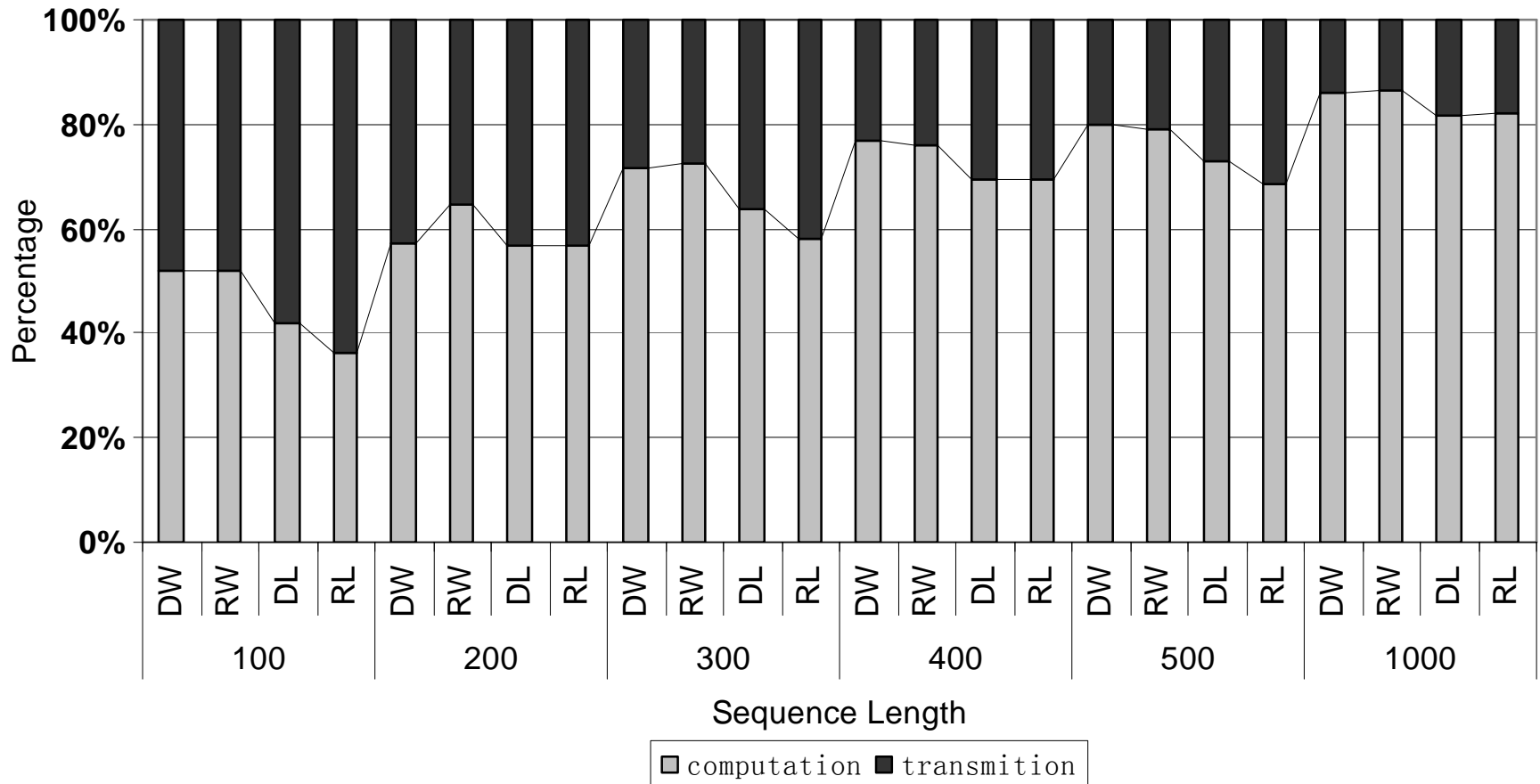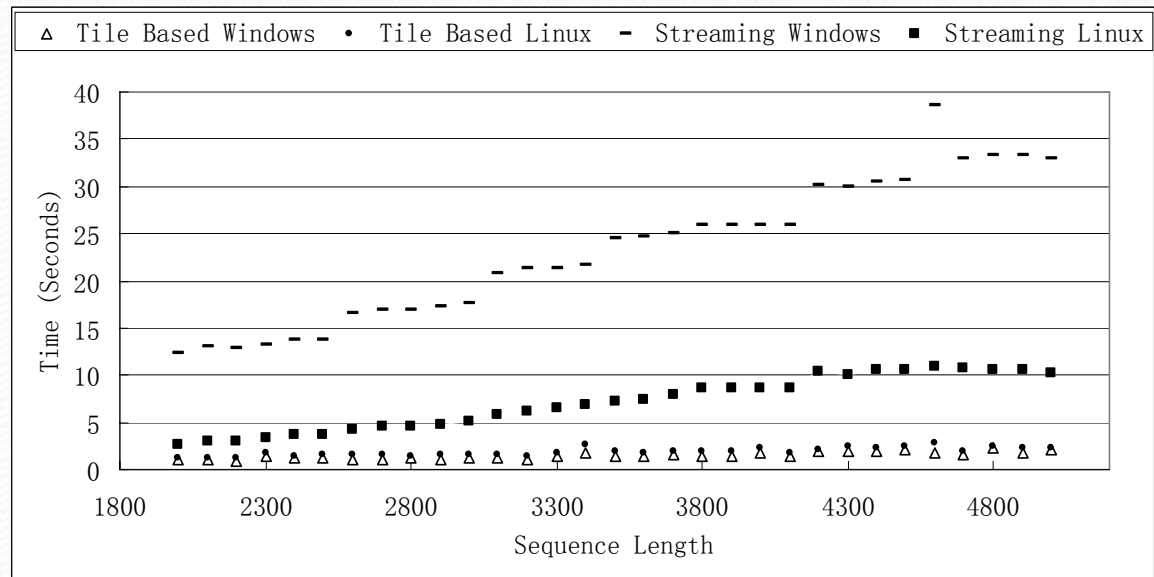
Test of Tile based Algorithm
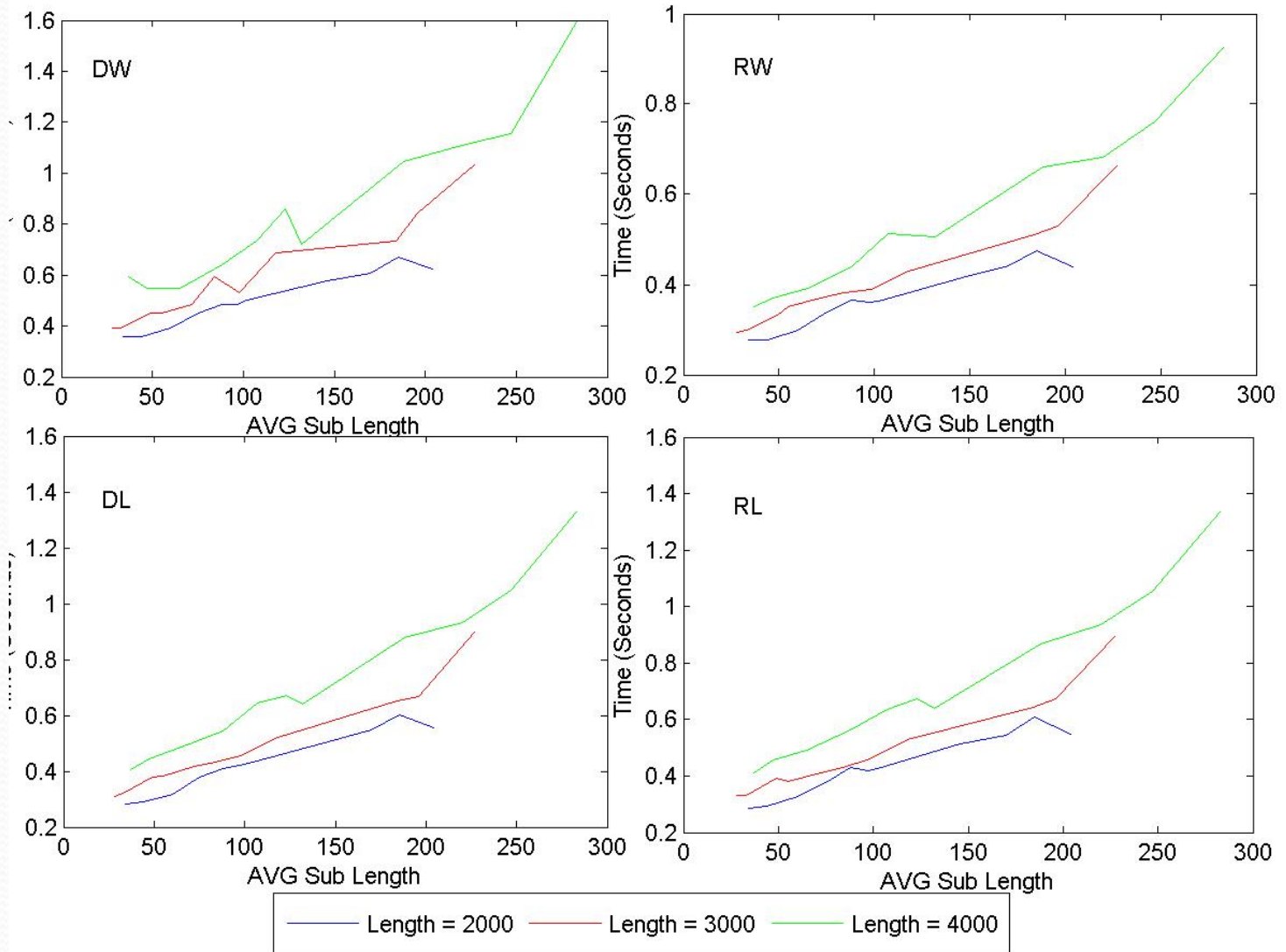
Test of Long Sequences
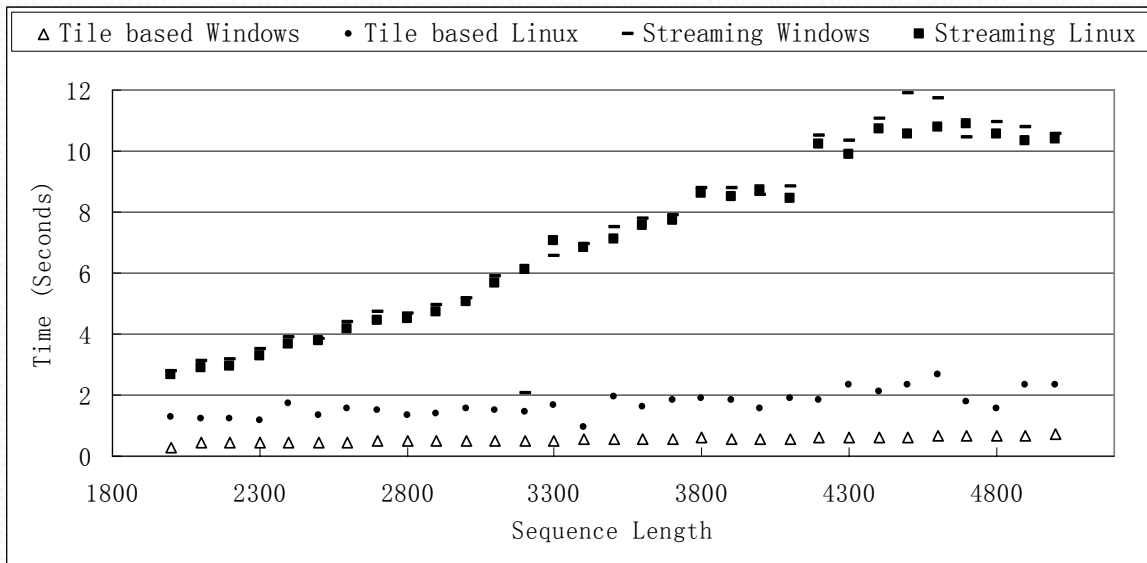
Under Linux System

Windows System

Future work

- 1: Experiments on multiple GPUs
- 2: New Architectures such as FERMI

Q&A?