TARA CUDA: A HYBRID SIMT IMPLEMENTATION OF ZHANG'S SIX-FRAME ALGORITHM USING CUDA

ARCETA - MENDOZA - TAN

ADVISER: MR. ROGER LUIS UY

BACKGROUND

Zhang's six-frame alignment

- **DNA-protein alignment** is a type of sequence alignment used to identify similarities between DNA and protein sequences.

Zhang's six-frame alignment

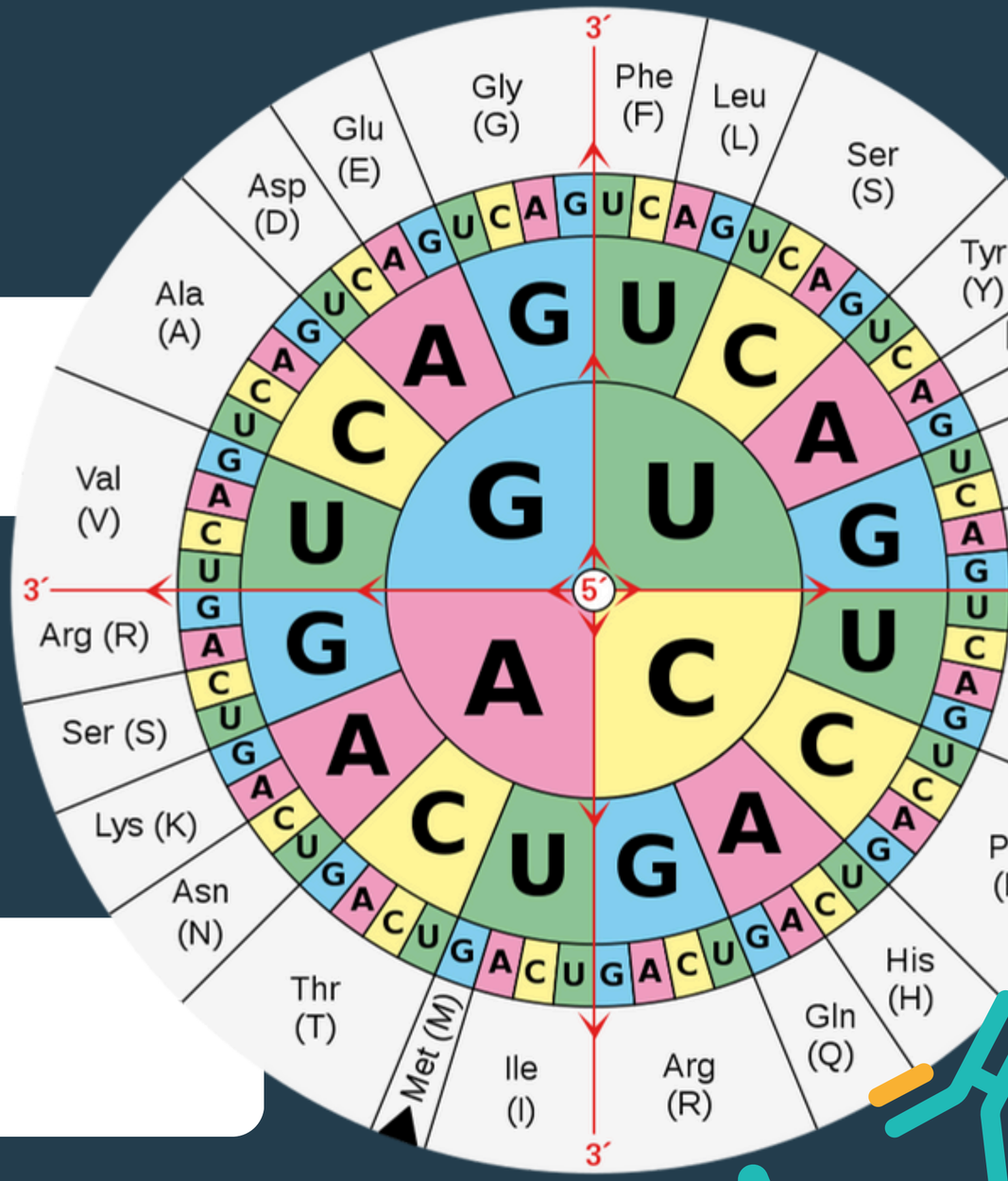
DNA SEQUENCE INPUT

ATGCACTGC...

Zhang's six-frame alignment

ATGCACTGC...

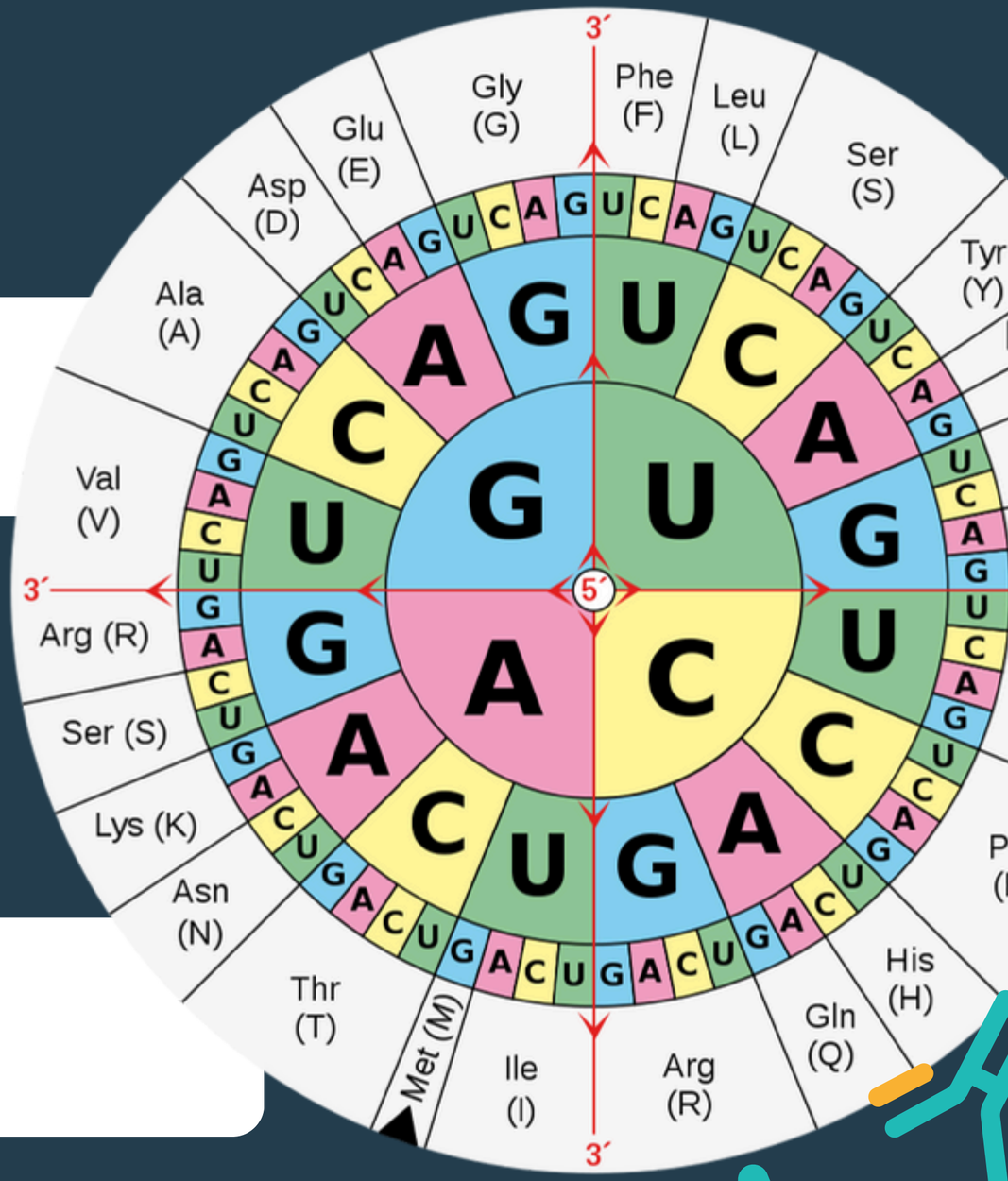
F1	M	H	C ...



Zhang's six-frame alignment

ATGCACTGC...

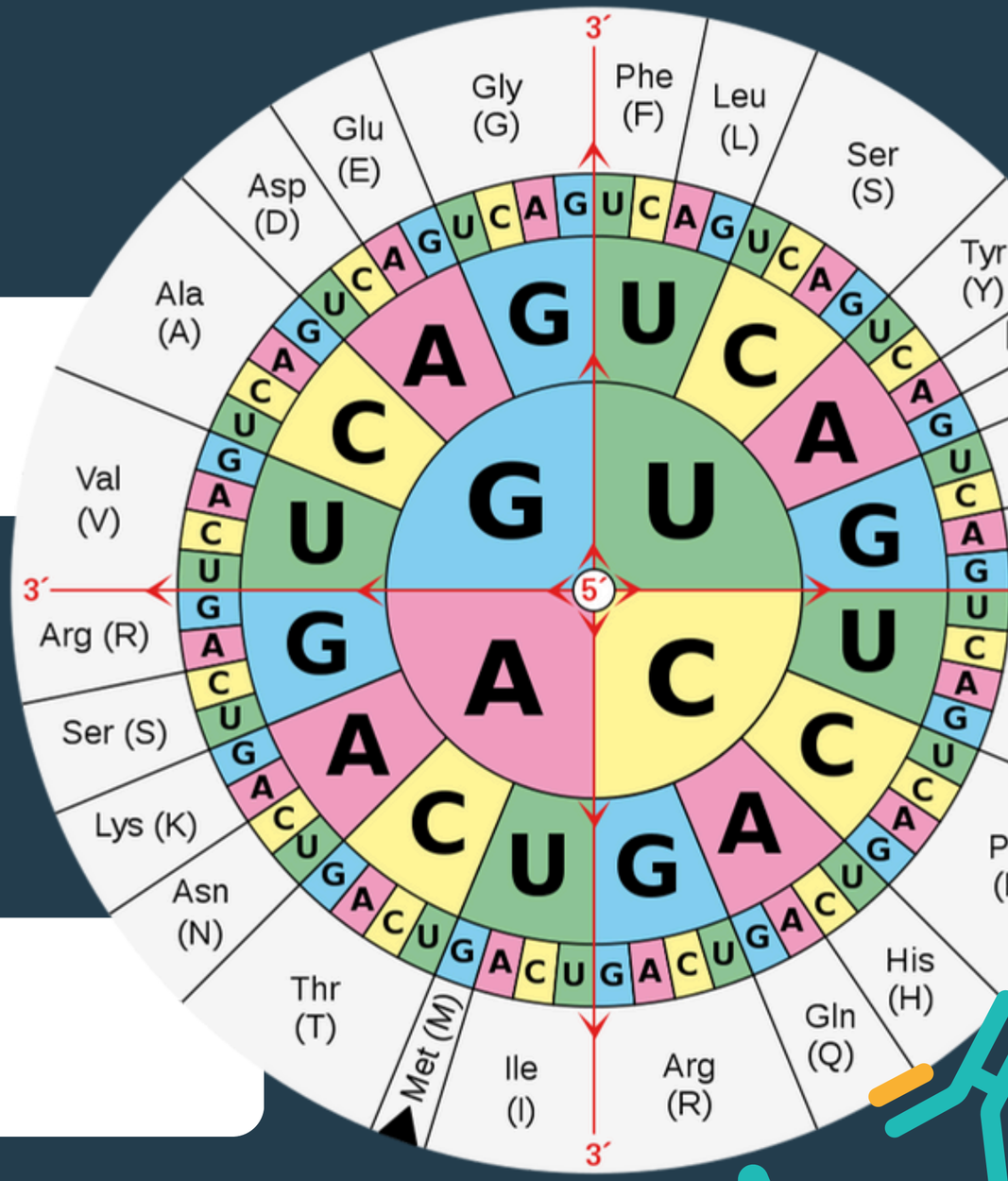
F2 C T ...



Zhang's six-frame alignment

ATGCACTGC...

F3 A L ...



ATGCACTGC...

F1	M	H	C	...
F2	C	T	...	
F3	A	L	...	

Zhang's six-frame alignment

ATGCACTGC...

AGTGTGCAC...

*reverse complement

AGTGTGCAC...

F4	S	V	H	...
F5	V	C	...	
F6	C	A	...	

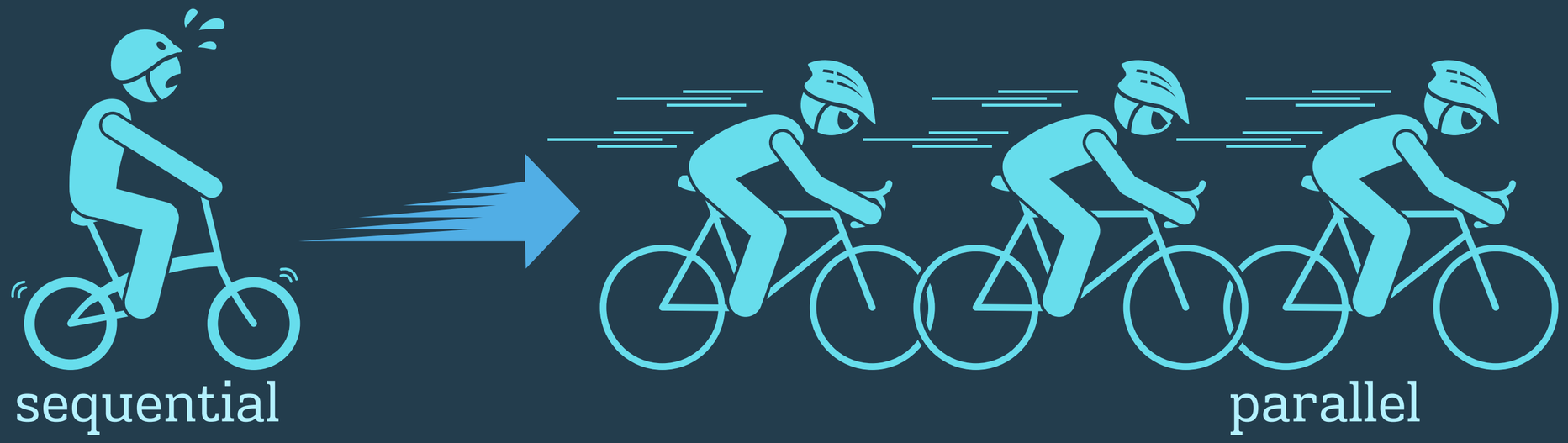
BACKGROUND

Zhang’s six-frame alignment

- **DNA-protein alignment** is a type of sequence alignment used to identify similarities between DNA and protein sequences.

- While **Zhang's six-frame alignment algorithm** is an established method, its sequential implementation can be computationally expensive, particularly when dealing with large datasets.

Hence, to address this issue, **parallel techniques** such as **SIMT (Single Instruction, Multiple Threads)** and SIMD (Single Instruction, Multiple Data) are used to enhance the execution of these algorithms.



GENERAL OBJECTIVE

This research aims to develop an approach to implementing Zhang's Six-Frame Alignment Algorithm in parallel, specifically using the CUDA SIMT execution model.

SPECIFIC OBJECTIVES

- To **develop** a sequential implementation of Zhang's Six-Frame Alignment Algorithm.
- To **create** an implementation scheme for Zhang's Six-Frame Alignment Algorithm to better fit the parallel framework of the SIMT model.
- To **implement** Zhang's Six-Frame Alignment Algorithm with the SIMT execution model using CUDA.
- To **evaluate** the developed parallel implementation by comparing and contrasting its performance to the sequential version and existing SIMD implementation.

SCOPE AND LIMITATIONS

- This research mainly focuses on **DNA to Protein** Sequence Alignment. The algorithm to be implemented is **Zhang's six-frame algorithm.**

- While many parallelization methods exist, the study will focus only on the **SIMT execution model using CUDA.**

- The proposed parallel implementation of this research will mainly focus on the **scoring part** of Zhang's six-frame algorithm.

- The proposed implementation will be evaluated and compared with the **sequential** and **previous parallel implementations**.

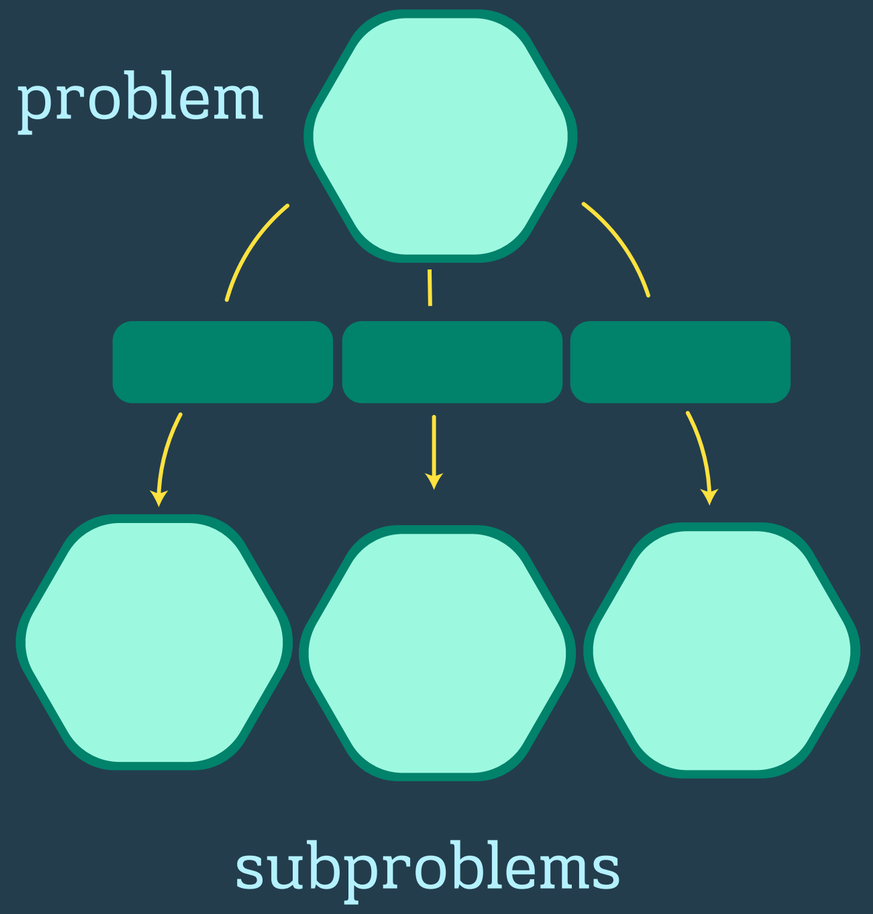
THEORETICAL FRAMEWORK

CUDA & SIMT PARADIGM

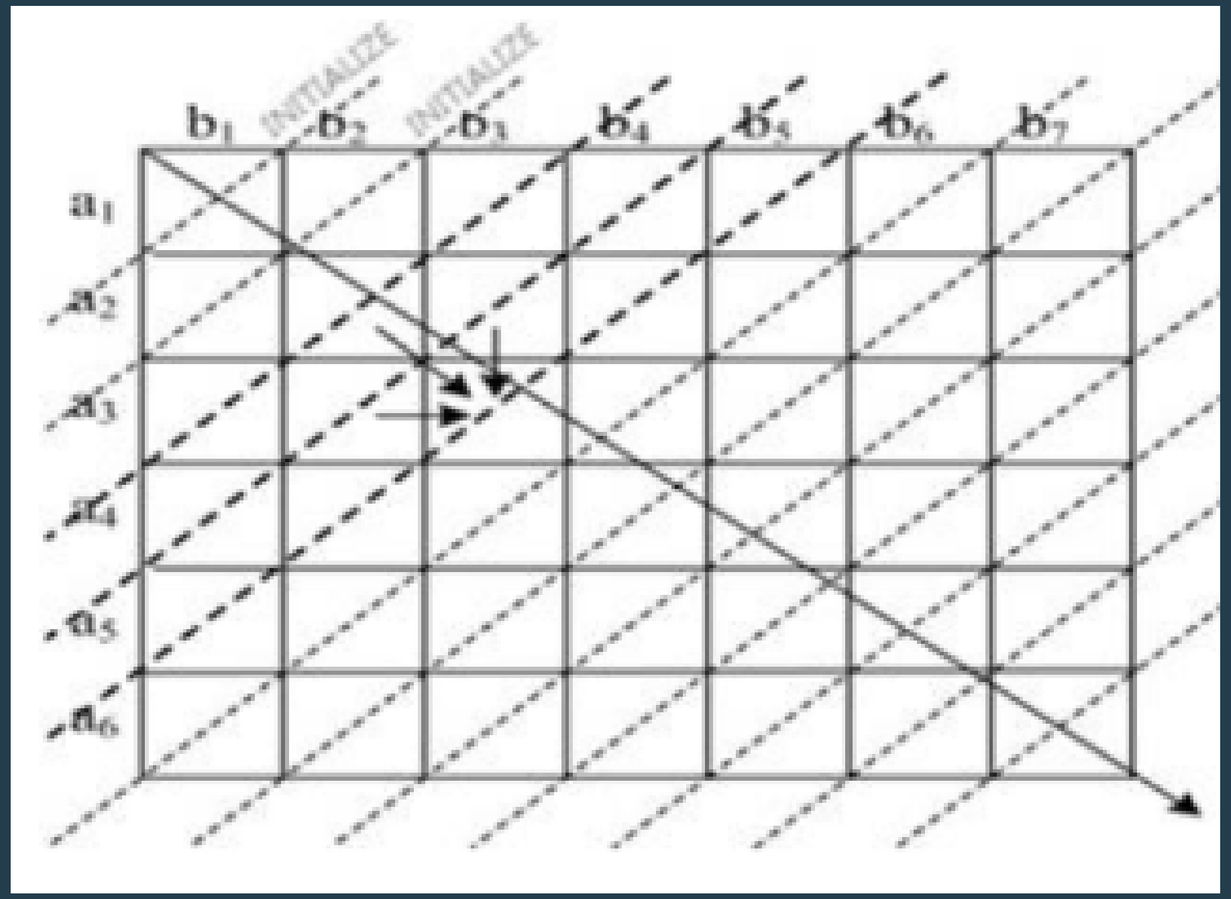
- CUDA's interface **enables high-performance computing tasks** that would otherwise be inefficient in traditional CPU implementations.

- CUDA's design is based on the **SIMT approach**, which allows several threads to execute identical instructions on different data streams, resulting in a massively parallel computation.

CUDA & SIMT PARADIGM



To harness the parallelism capabilities of CUDA, the problem must be broken down into independent tasks that can be performed concurrently. This breakdown makes it possible for many threads to be executed at the same time, which is the main benefit of CUDA.



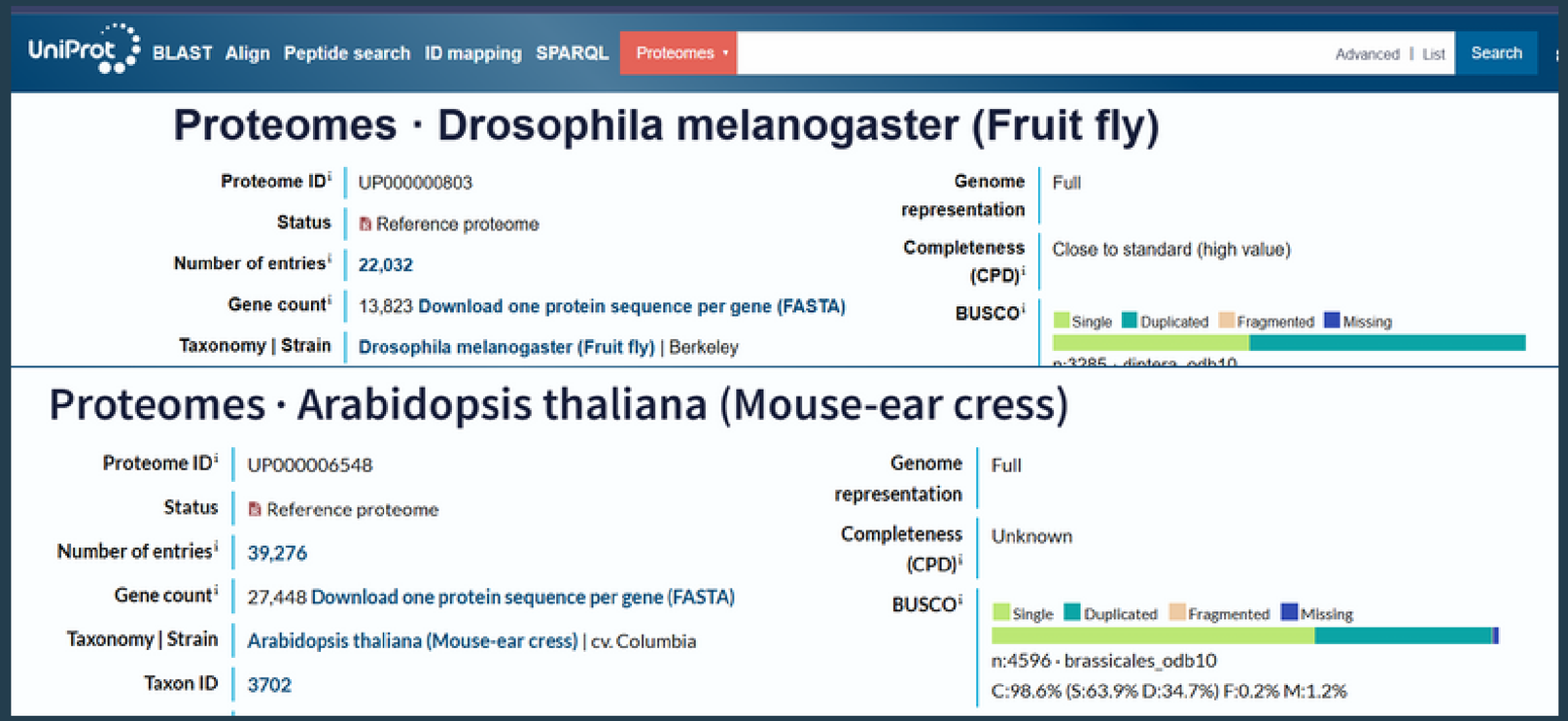
The **Anti-Diagonal Technique** is a technique used for filling in a matrix by processing its elements along anti-diagonals instead of the traditional row-by-row or column-by-column approaches. This enables parallelization for algorithms with top, left, and top-left dependencies.

WAVEFRONT PARADIGM

SYSTEM IMPLEMENTATION

- The datasets utilized in this research comprise protein and DNA sequences for *Drosophila melanogaster* (fruit fly) and. *Arabidopsis thaliana* (mouse-ear cress).

- Scraped via Selenium from UniProt & European Nucleotide Archieve (ENA) website



DATASET

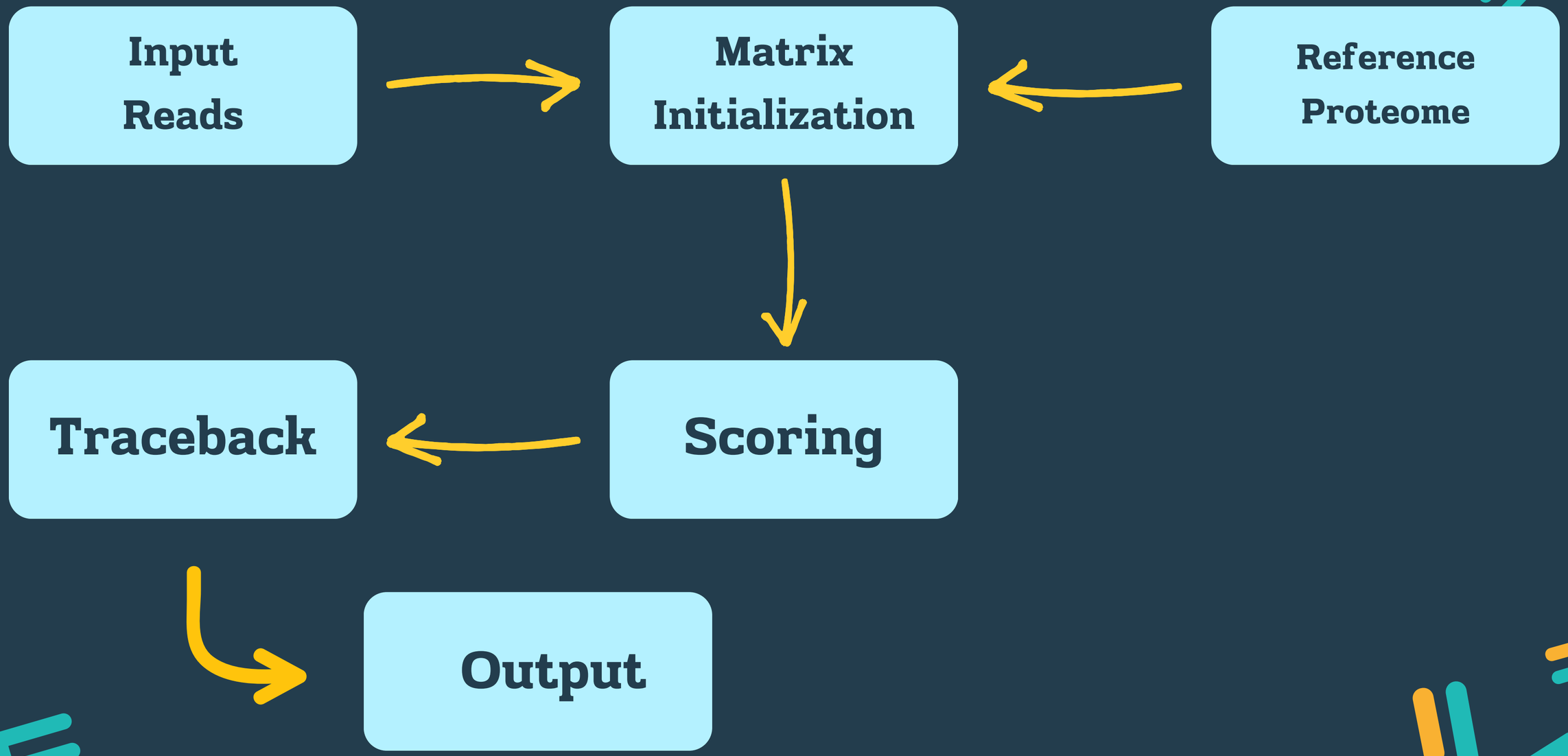
- After preprocessing, the final datasets that will be used consist of valid and fully aligned protein and DNA sequence pairs for *Drosophila melanogaster* (fruit fly) and *Arabidopsis thaliana* (mouse-ear cress).

- *Drosophila melanogaster*
  - Dataset: 21,449 protein sequences (each with DNA translation)
  - Protein lengths: 11 bp (shortest) to 22,949 bp (longest)
  - Average protein length: ~451 bp

- *Arabidopsis thaliana*
  - Dataset: 39,067 protein sequences (higher count)
  - Protein lengths: 5 bp (shortest) to 5,400 bp (longest)
  - Average protein length: ~396 bp

DATASET

TARA CUDA IMPLEMENTATION



TARA CUDA IMPLEMENTATION

Input Reads

GACGCTGTAGGCCATA...

TARA CUDA IMPLEMENTATION

TARA CUDA IMPLEMENTATION

ALL_proteom...000803.fasta | BLOSUM62.txt | seq.cpp

```
36509  TCIGGQTFKILTWLGYSNSAFNPIIYSIFNKEFRDAFKRILTMRNPWCCAQDVGNIHPRN
36510  SDRFITDYAAKNVVVMNSGRSSAELEQVSAI
36511  >sp|P41894|DEI_DROME Helix-loop-helix protein delilah OS=Drosophila melanogaster OX=7227 GN=tx PE=1 SV=3
36512  MKSNTYELHNYADLNDMARATDSKDSRKRKTASARGEKYSLRQKRQKRGSNEDGESANLA
36513  DFQLELDPIAEPASKSRKNAPTKSKTKAPPLSKYRRKTANARERTRMREINTAFETLRHC
36514  VPEAIKGEDAANTNEKLTKITTLRLAMKYITMLTDSIRDPSYESEFIGECLEESANREAR
36515  VDLEANEEAEVELPVPVAKKPAKTKGSGKKSSAASKRQSQKQAKIVPQIPPISSGESCYA
36516  TSSIGSPASSAYASLSSSSNSHSSSSSPGLELDSLVGLNGISALDSLLLDTSDGDSLSCL
36517  SPGYGSLLTGSGESLLPGCRGDLPMSGLEKSDVELSLRLLDQSSKDSFDFASDQQPSACI
36518  SSFSALDGYPFDLLHSDFPDMFLT
36519  >sp|P41964|DMYC_DROME Drosomycin OS=Drosophila melanogaster OX=7227 GN=Drs PE=1 SV=1
36520  MMQIKYLFALFAVLMLVVLGANEADADCLSGRYKGPCAVWDNETCRRVCKEEGRSSGHCS
36521  PSLKCWCEGC
36522  >sp|P42003|MAD_DROME Protein mothers against dpp OS=Drosophila melanogaster OX=7227 GN=Mad PE=1 SV=1
36523  MDTDDVESNTSSAMSTLGSLFSFTSPAVKKLLGWKQGDEEEKWAEKAVDSLVKKLKKRKG
36524  AIEELERALSCPGQPSKCVTIPRSLDGRLQVSHRKGLPHVIYCRVWRWPDLQSHHELKPL
36525  ELCQYPFSAKQKEVCINPYHYKRVESPVLPPVLVPRHSEFAPGHSMLQFNHVAEPSMPHN
36526  VSYSNSGFNSHSLSTSNTSVGSPSSVNSNPNSPYDSLAGTPPPAYSPSEDGNSNNPNDGG
36527  QLLDAQMGDVAQVSYSEPAFWASIAYYELNCRVGEVFHCNNNSVIVDGFTNPSNNSDRCC
36528  LGQLSNVNRNSTIENTRRHIGKGVHLYYVTGEVYAECLSDSAIFVQSRNCNYHHGFHPST
36529  VCKIPPGCSLKIFNNQEFAQLLSQSVNNGFEAVYELTKMCTIRMSFVKGWGAEYHRQDVT
36530  STPCWIEIHLHGPLQWLDKVLTQMGSPHNAISSVS
36531  >sp|P42124|EZ_DROME Histone-lysine N-methyltransferase E(z) OS=Drosophila melanogaster OX=7227 GN=E(z) PE=1 SV=2
36532  MNSTKVPPEWKRRVKSEYIKIRQQKRYKRADEIKEAWIRNWDEHNHNVQDLYCESKVWQA
36533  KPYDPPHVDCVKRAEVTSYNGIPSGPQKVPICVINAVTPIPTMYTWAPTQQNFMVEDETV
36534  LHNIPYMGDEVLDKDGKFIEELIKNYDGKVHGDKDPSFMDDAIFVELVHALMRSYSKELE
36535  EAAPGTATAIKTETLAKSKQGEDDGVVDVDADGESPMKLEKTDSKGDLTEVEKKETEEPL
36536  ETEDADVKPDVEEVKDKLPFPAPIIFQAISANFPDKGTAQELKEKYIELTEHQDPERPQE
36537  CTPNIDGIKAESVSRERTMHSFHTLFCRRCFKYDCFLHRLQGHAGPNLQKRRYPELKPFA
36538  EPCSNSCYMLIDGMKEKLAADSKTPPIDSCNEASSEDSNDSNSQFSNKDFNHENSKDNGL
```

Reference Proteome

TARA CUDA IMPLEMENTATION

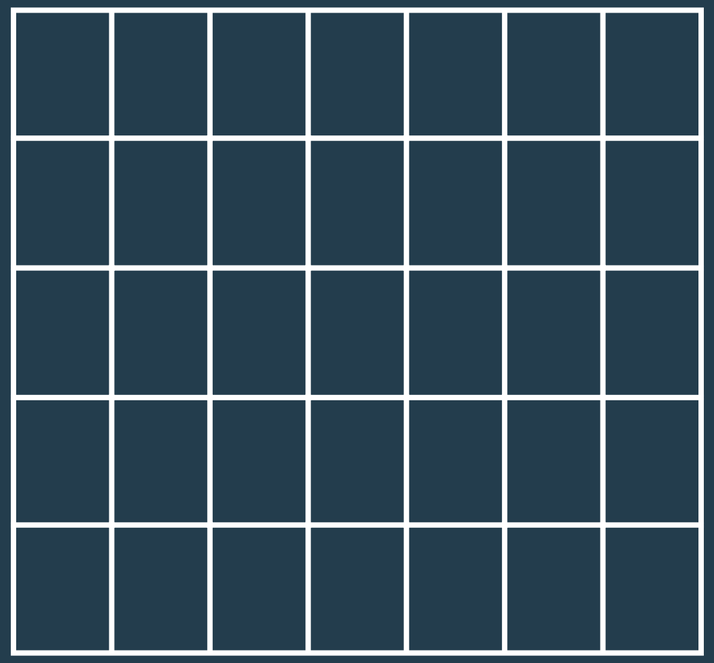
MITDVQLAIFSNVLGV...

Reference Proteome

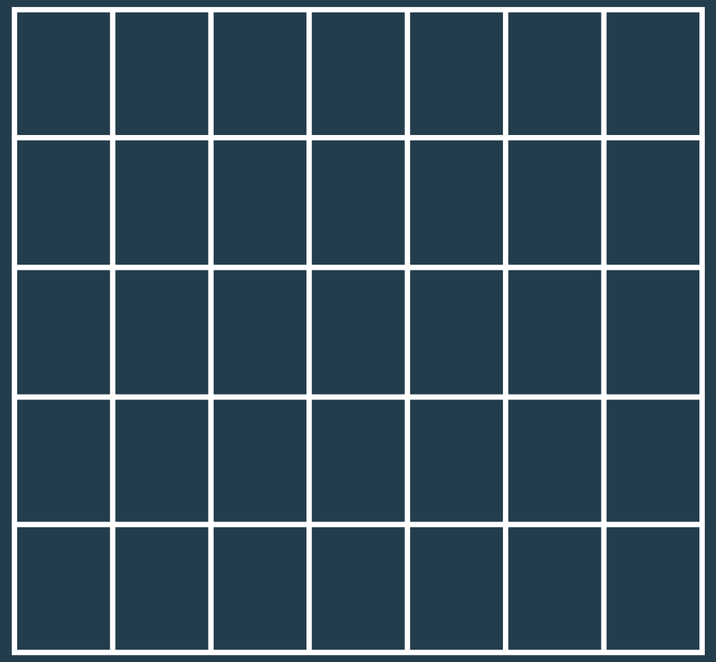
TARA CUDA IMPLEMENTATION

Matrix Initialization

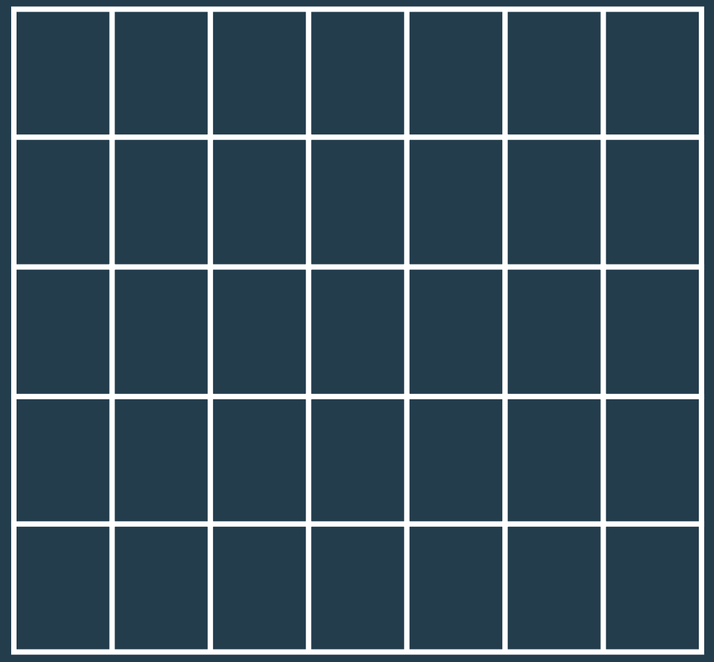
- Initialize the first few columns and rows



Matrix I



Matrix D



Matrix C

TARA CUDA IMPLEMENTATION

Matrix Initialization

- Initialize the first few columns and rows

$$I(j,0) = -\infty$$

$$D(0,j) = D(2,j) = D(3,j) = -\infty$$

$$D(1,j) = C(0,j) - gop - gep$$

$$C(0,0) = 0$$

$$C(0,j) = I(0,j)$$

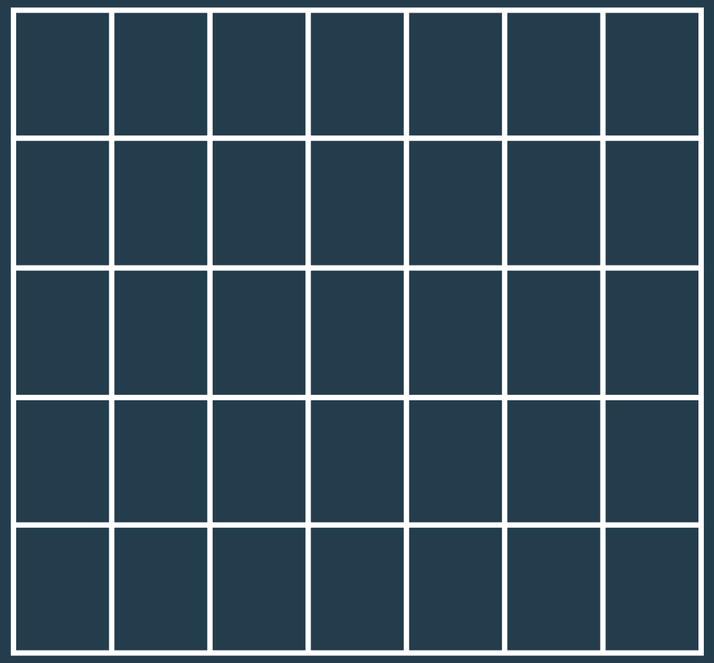
$$C(j,0) = D(j,0)$$

$$C(1,j) = max\begin{cases} I(1,j) \\ D(1,j) \\ C(0,j-1) + S[b_1, a_j] \end{cases}$$

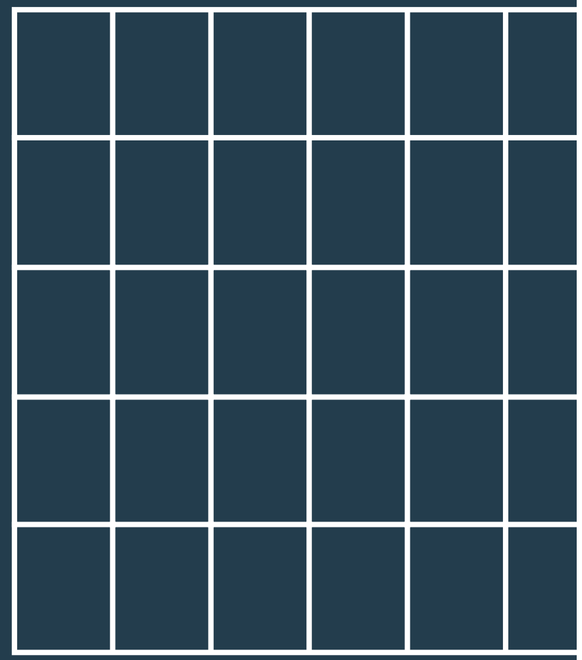
$$C(2,j) = max\begin{cases} I(2,j) \\ C(0,j-1) + S[b_1, a_j] - \pi \end{cases}$$

$$C(3,j) = max\begin{cases} I(3,j) \\ C(1,j-1) + S[b_2, a_j] - \pi \end{cases}$$

$$C(4,j) = max\begin{cases} I(4,j) \\ D(4,j) \\ C(1,j-1) + S[b_4, a_j] \\ C(2,j-1) + S[b_4, a_j] - \pi \end{cases}$$



Matrix I



Mat

x C

TARA CUDA IMPLEMENTATION

Matrix Initialization

- Initialize the first few columns and rows

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞						

Matrix I

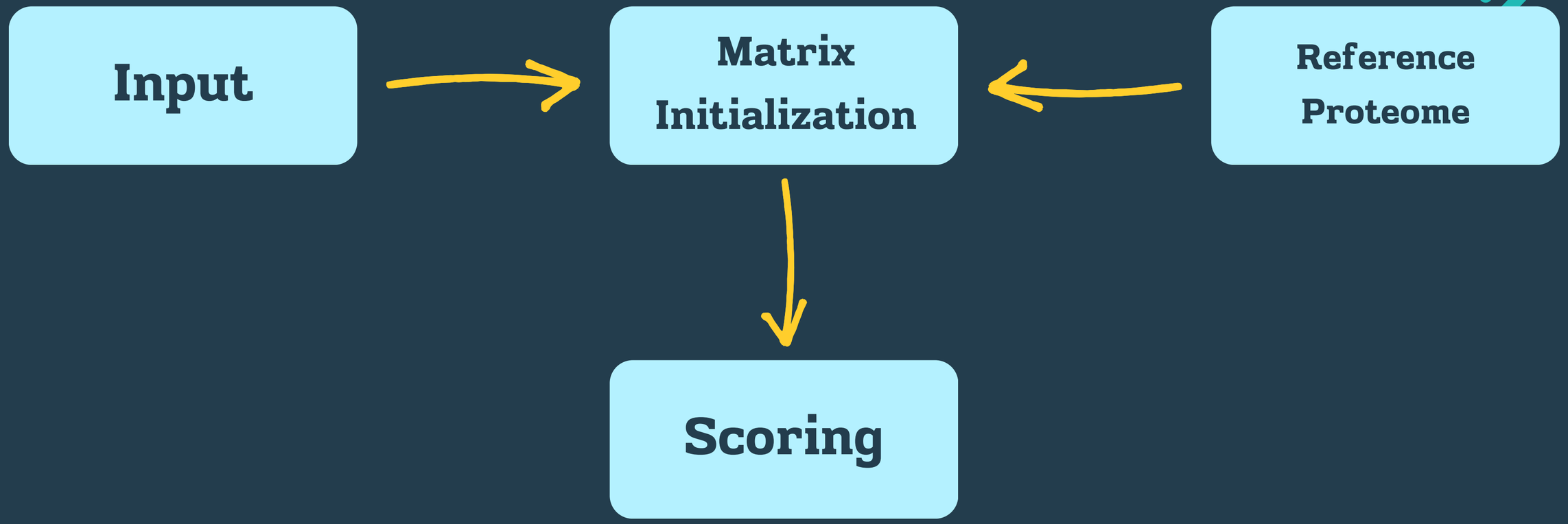
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0						

Matrix C

TARA CUDA IMPLEMENTATION



TARA CUDA IMPLEMENTATION

Scoring

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0						

Matrix C

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞						
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5						
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0						
0						
0						
0						
0						

Matrix C

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞						
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5						
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0						
0						
0						
0						
0						

Matrix C

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞						
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5						
-5						
-5						
-5						
-5						

$$I(i,j) = max\begin{cases} I(i,j-1) - gep \\ C(i,j-1) - gep - gop \end{cases}$$

$$D(i,j) = max\begin{cases} D(i-3,j) - gep \\ C(i-3,j) - gep - gop \end{cases}$$

$$C(i,j) = max\begin{cases} I(i,j) \\ D(i,j) \\ C(i-2,j-1) + S[b_i, a_j] - \pi \\ C(i-3,j-1) + S[b_i, a_j] \\ C(i-4,j-1) + S[b_i, a_j] - \pi \end{cases}$$

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0						
0						
0						
0						
0						

Matrix C

TARA CUDA IMPLEMENTATION

Scoring

This process has 2 options since the program will prompt the user to choose between the **Sequential** or **CUDA (Parallel)** implementation of the selected algorithm.

TARA CUDA IMPLEMENTATION

Scoring

This process has 2 options since the program will prompt the user to choose between the Sequential or CUDA (Parallel) implementation of the selected algorithm.

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5					
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5						
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0						
0						
0						
0						
0						

Matrix C

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	
-∞	-5	-5	-5	-5	-3	-5
-∞	-5					
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5						
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0						
0						
0						
0						
0						

Matrix C

$$I(i,j) = max \begin{cases} I(i,j-1) - gep \\ C(i,j-1) - gep - gop \end{cases}$$

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5					
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0					
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0						
0						
0						
0						
0						

Matrix C

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5					
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0					
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0						
0						
0						
0						

Matrix C

$$D(i,j) = max \begin{cases} D(i-3,j) - gep \\ C(i-3,j) - gep - gop \end{cases}$$

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5					
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0					
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1					
0						
0						
0						
0						

Matrix C

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5					
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0					
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1					
0						
0						

Matrix C

$$C(i,j) = max \begin{cases} I(i,j) \\ D(i,j) \\ C(i-2,j-1) + S[b_i, a_j] - \pi \\ C(i-3,j-1) + S[b_i, a_j] \\ C(i-4,j-1) + S[b_i, a_j] - \pi \end{cases}$$

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4				
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0					
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1					
0						
0						
0						
0						

Matrix C

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4				
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4				
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1					
0						
0						
0						
0						

Matrix C

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4				
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4				
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9				
0						
0						
0						
0						

Matrix C

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4			
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4				
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9				
0						
0						
0						
0						

Matrix C

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4			
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5			
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9				
0						
0						
0						
0						

Matrix C

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4			
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5			
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4			
0						
0						
0						
0						

Matrix C

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2		
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5			
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4			
0						
0						
0						
0						

Matrix C

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2		
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5	-5		
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4			
0						
0						
0						
0						

Matrix C

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2		
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5	-5		
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4	2		
0						
0						
0						
0						

Matrix C

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2	0	
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5	-5		
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4	2		
0						
0						
0						
0						

Matrix C

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2	0	
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5	-5	-4	
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4	2		
0						
0						
0						
0						

Matrix C

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2	0	
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5	-5	-4	
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4	2	3	
0						
0						
0						
0						

Matrix C

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2	0	-2
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5	-5	-4	
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4	2	3	
0						
0						
0						
0						

Matrix C

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2	0	-2
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5	-5	-4	-5
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4	2	3	
0						
0						
0						
0						

Matrix C

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2	0	-2
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5	-5	-4	-5
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4	2	3	0
0						
0						
0						
0						

Matrix C

Sequential Implementation

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2	0	-2
-∞	-5	-5	-5	-4	-5	-5
-∞	-5	-5	-5	-2	-4	-5
-∞	-5	-5	-1	9	7	5
-∞	-5	-3	-3	-1	-3	-4

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5	-5	-4	-5
-5	-5	-5	-5	-5	-5	-5
-5	-5	-5	-5	-3	-5	-5
-5	-2	4	-1	-3	-2	-5
-5	-5	-5	-4	-5	-5	-5

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4	2	3	0
0	0	0	1	0	0	0
0	0	0	3	0	0	0
0	0	4	14	9	7	5
0	2	2	4	0	1	0

Matrix C

TARA CUDA IMPLEMENTATION

Scoring

This process has 2 variants since the program will prompt the user to choose between the Sequential or CUDA (Parallel) implementation of the selected algorithm.

CUDA (PARALLEL) IMPLEMENTATION

- This process includes **configuring the CUDA environment**, as well as **parallelizing the alignment scoring**.

- Due to block synchronization limitations, we developed **a hybrid multi-level anti-diagonal wavefront implementation** using a combination of host-side loops and in-kernel block synchronization.

CUDA (PARALLEL) IMPLEMENTATION

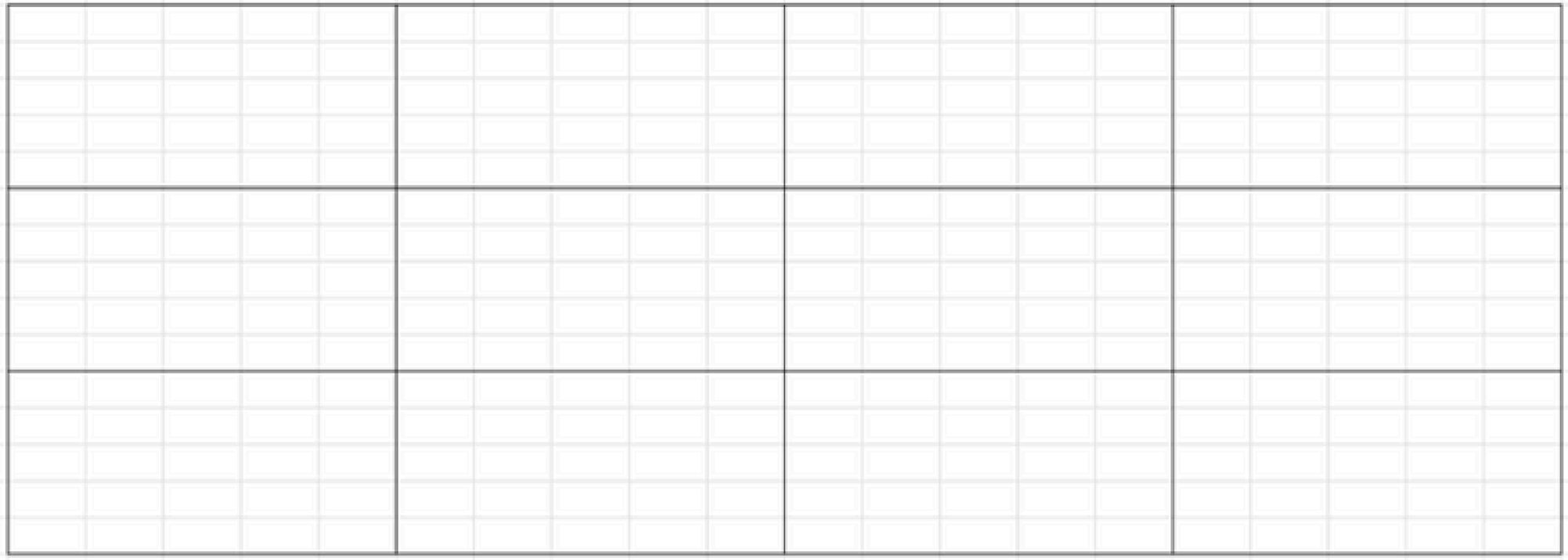


Figure 5-18: Sample Visualization of Submatrix Division

CUDA (PARALLEL) IMPLEMENTATION

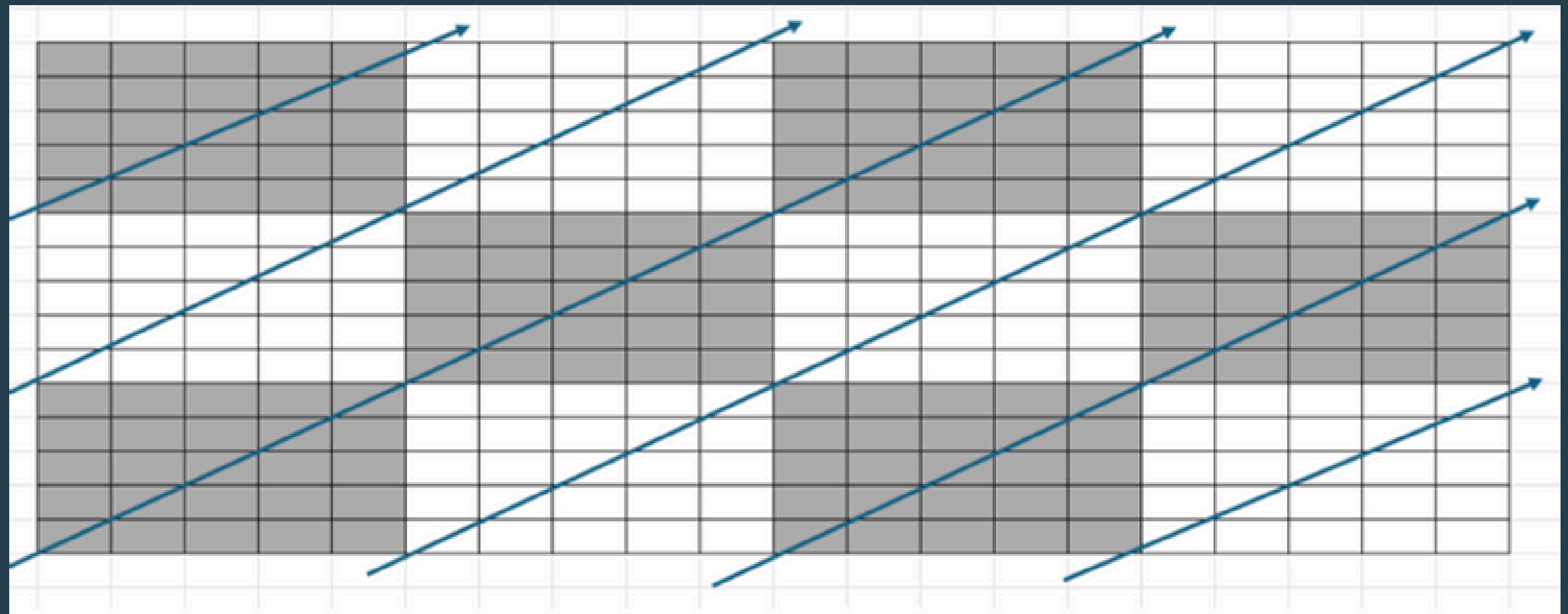


Figure 5-19: Sample Visualization of Macro-Level Matrix Anti-Diagonals

CUDA (PARALLEL) IMPLEMENTATION

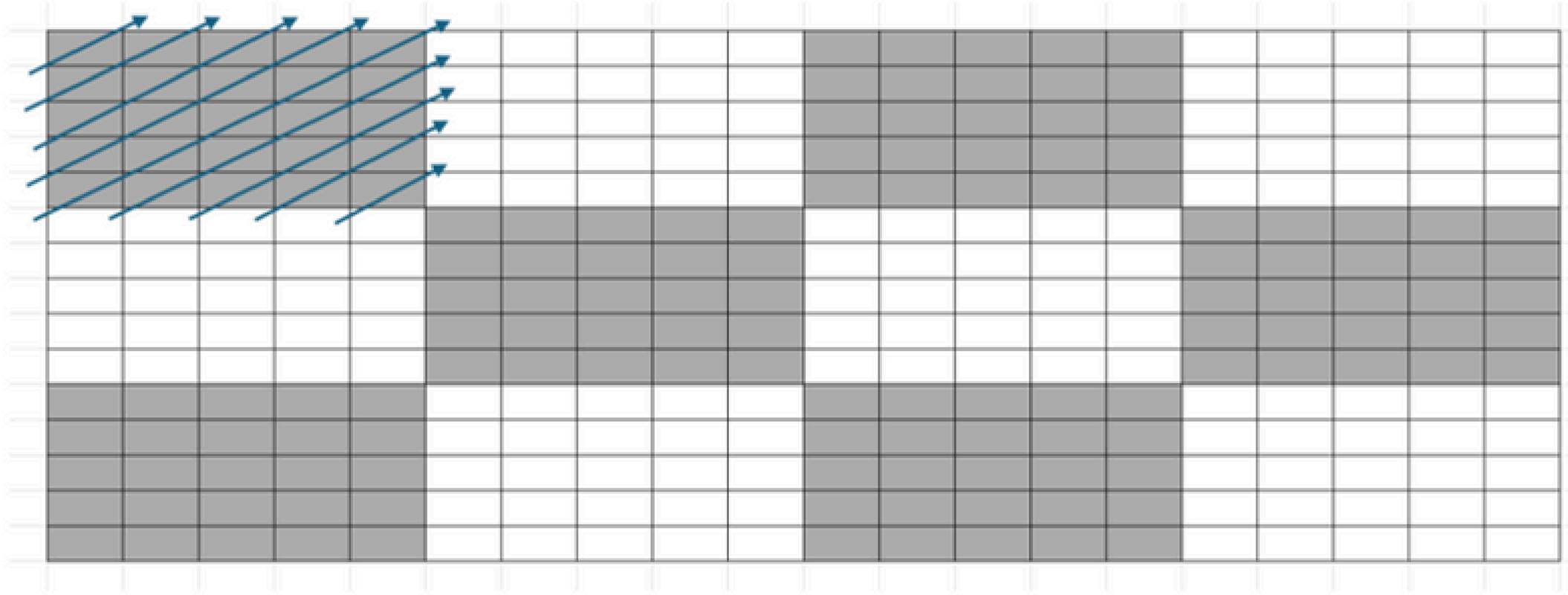


Figure 5-20: Sample Visualization of Submatrix-Level Anti-Diagonals

CUDA (PARALLEL) IMPLEMENTATION

- **Host Side (CPU):** responsible for setting up data, grid and block configuration, memory management, and launching the CUDA kernels.
  - Block dim is set to (32, 32), Grid dim is adjusted to the current anti-diagonal.
  - Handles memory allocation and data transfers (via *cudaMalloc and cudaMemcpy*).
  - Manages kernel execution flow, and synchronizes the kernel launches along each anti-diagonal to wait before the next anti-diagonal is launched.

CUDA (PARALLEL) IMPLEMENTATION

- **Device Side (GPU)**: responsible for the actual parallel computation of the scoring process using **submatrix anti-diagonal parallelization**.
  - A Block is configured **to handle a submatrix** and **strides across all submatrices** in the current anti-diagonal.
  - Threads within a block compute **all** of the **same index locations** i, j of the IDC matrices.
  - Controlled via an **anti-diagonal loop** and **a conditional** for when the thread is in the current wavefront.

CUDA (PARALLEL) IMPLEMENTATION

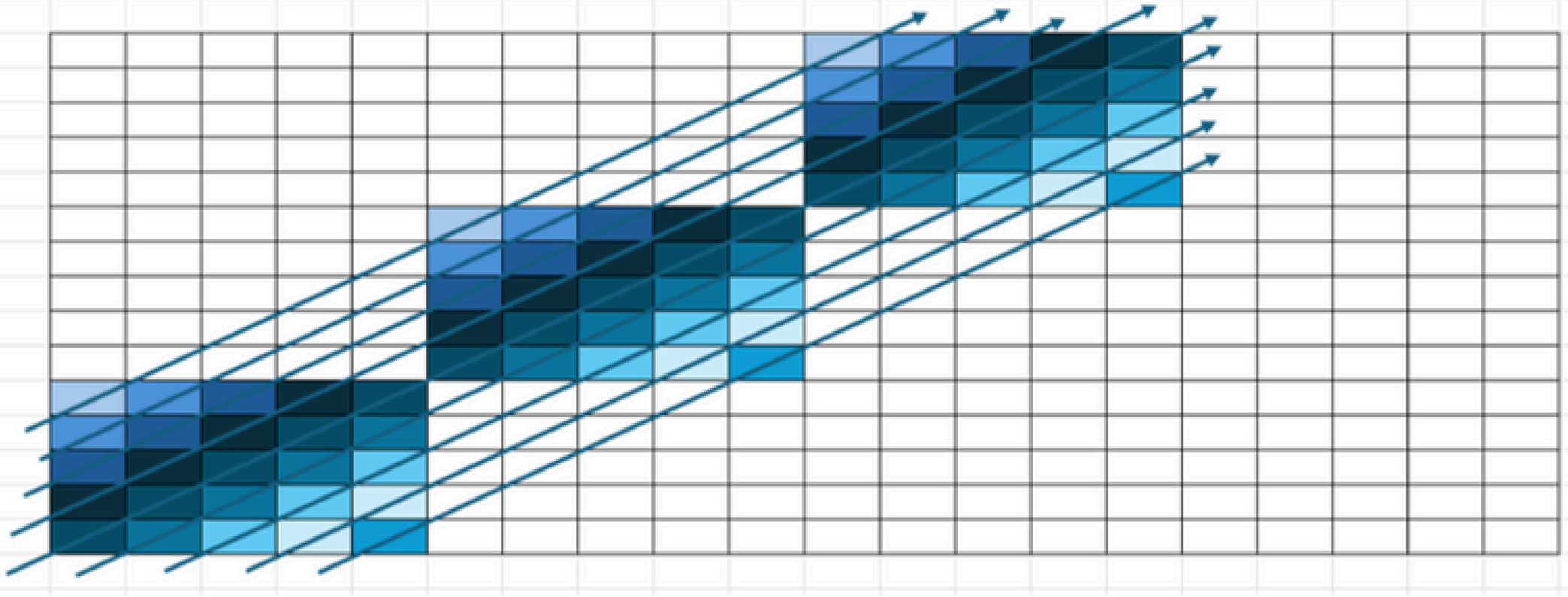


Figure 5-22: Sample Visualization of Grid-Stride Anti-Diagonal Computation in CUDA Kernel

Anti Diagonal 1

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5					
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5						
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0						
0						
0						
0						
0						

Matrix C

Anti Diagonal 1

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	
-∞	-5	-5	-5	-5	-3	-5
-∞	-5					
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5						
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0						
0						
0						
0						
0						

Matrix C

$$I(i,j) = max \begin{cases} I(i,j-1) - gep \\ C(i,j-1) - gep - gop \end{cases}$$

Anti Diagonal 1

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5					
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0					
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0						
0						
0						
0						
0						

Matrix C

Anti Diagonal 1

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5					
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0					
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0						
0						
0						

Matrix C

$$D(i,j) = max \begin{cases} D(i-3,j) - gep \\ C(i-3,j) - gep - gop \end{cases}$$

Anti Diagonal 1

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5					
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0					
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1					
0						
0						
0						
0						

Matrix C

Anti Diagonal 1

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5					
-∞						
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0					
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1					
0						
0						

Matrix C

$$C(i,j) = max \begin{cases} I(i,j) \\ D(i,j) \\ C(i-2,j-1) + S[b_i, a_j] - \pi \\ C(i-3,j-1) + S[b_i, a_j] \\ C(i-4,j-1) + S[b_i, a_j] - \pi \end{cases}$$

Anti Diagonal 2

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-5				
-∞	-5					
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0					
-5						
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1					
0						
0						
0						
0						

Matrix C

Anti Diagonal 2

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-5				
-∞	-5					
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4				
-5	-5					
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1					
0						
0						
0						
0						

Matrix C

Anti Diagonal 2

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-5				
-∞	-5					
-∞						
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4				
-5	-5					
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9				
0	0					
0						
0						
0						

Matrix C

Anti Diagonal 3

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4			
-∞	-5	-5				
-∞	-5					
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4				
-5	-5					
-5						
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9				
0	0					
0						
0						
0						

Matrix C

Anti Diagonal 3

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4			
-∞	-5	-5				
-∞	-5					
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5			
-5	-5	-5				
-5	-5					
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9				
0	0					
0						
0						
0						

Matrix C

Anti Diagonal 3

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4			
-∞	-5	-5				
-∞	-5					
-∞						
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5			
-5	-5	-5				
-5	-5					
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4			
0	0	0				
0	0					
0						
0						

Matrix C

Anti Diagonal 4

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2		
-∞	-5	-5	-5			
-∞	-5	-5				
-∞	-5					
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5			
-5	-5	-5				
-5	-5					
-5						
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4			
0	0	0				
0	0					
0						
0						

Matrix C

Anti Diagonal 4

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2		
-∞	-5	-5	-5			
-∞	-5	-5				
-∞	-5					
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5	-5		
-5	-5	-5	-5			
-5	-5	-5				
-5	-2					
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4			
0	0	0				
0	0					
0						
0						

Matrix C

Anti Diagonal 4

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2		
-∞	-5	-5	-5			
-∞	-5	-5				
-∞	-5					
-∞						

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5	-5		
-5	-5	-5	-5			
-5	-5	-5				
-5	-2					
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4	2		
0	0	0	1			
0	0	0				
0	0					
0						

Matrix C

Anti Diagonal 5

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2	0	
-∞	-5	-5	-5	-4		
-∞	-5	-5	-5			
-∞	-5	-5				
-∞	-5					

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5	-5		
-5	-5	-5	-5			
-5	-5	-5				
-5	-2					
-5						

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4	2		
0	0	0	1			
0	0	0				
0	0					
0						

Matrix C

Anti Diagonal 5

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2	0	
-∞	-5	-5	-5	-4		
-∞	-5	-5	-5			
-∞	-5	-5				
-∞	-5					

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5	-5	-4	
-5	-5	-5	-5	-5		
-5	-5	-5	-5			
-5	-2	4				
-5	-5					

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4	2		
0	0	0	1			
0	0	0				
0	0					
0						

Matrix C

Anti Diagonal 5

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2	0	
-∞	-5	-5	-5	-4		
-∞	-5	-5	-5			
-∞	-5	-5				
-∞	-5					

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5	-5	-4	
-5	-5	-5	-5	-5		
-5	-5	-5	-5			
-5	-2	4				
-5	-5					

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4	2	3	
0	0	0	1	0		
0	0	0	3			
0	0	4				
0	2					

Matrix C

Anti Diagonal 6

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2	0	-2
-∞	-5	-5	-5	-4	-5	
-∞	-5	-5	-5	-2		
-∞	-5	-5	-1			
-∞	-5	-3				

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5	-5	-4	
-5	-5	-5	-5	-5		
-5	-5	-5	-5			
-5	-2	4				
-5	-5					

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4	2	3	
0	0	0	1	0		
0	0	0	3			
0	0	4				
0	2					

Matrix C

Anti Diagonal 6

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2	0	-2
-∞	-5	-5	-5	-4	-5	
-∞	-5	-5	-5	-2		
-∞	-5	-5	1			
-∞	-5	-3				

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5	-5	-4	-5
-5	-5	-5	-5	-5	-5	
-5	-5	-5	-5	3		
-5	-2	4	1			
-5	-5	-5				

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4	2	3	
0	0	0	1	0		
0	0	0	3			
0	0	4				
0	2					

Matrix C

Anti Diagonal 6

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2	0	-2
-∞	-5	-5	-5	-4	-5	
-∞	-5	-5	-5	-2		
-∞	-5	-5	-1			
-∞	-5	-3				

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5	-5	-4	-5
-5	-5	-5	-5	-5	-5	
-5	-5	-5	-5	-3		
-5	-2	4	-1			
-5	-5	-5				

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4	2	3	0
0	0	0	1	0	0	
0	0	0	3	0		
0	0	4	14			
0	2	2				

Matrix C

Anti Diagonal 10

-∞	-5	-5	-5	-5	-5	-5
-∞	-5	0	-2	-4	-5	-4
-∞	-5	-5	-5	-5	-5	-5
-∞	-5	-5	-5	-5	-3	-5
-∞	-5	-4	4	2	0	-2
-∞	-5	-5	-5	-4	-5	-5
-∞	-5	-5	-5	-2	-4	-5
-∞	-5	-5	-1	9	7	5
-∞	-5	-3	-3	-1	-3	-4

Matrix I

-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	-5	-5	-5	-5	-5	-5
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-∞	-∞	-∞	-∞	-∞	-∞	-∞
-5	0	-4	-5	-5	-4	-5
-5	-5	-5	-5	-5	-5	-5
-5	-5	-5	-5	-3	-5	-5
-5	-2	4	-1	-3	-2	-5
-5	-5	-5	-4	-5	-5	-5

Matrix D

0	0	0	0	0	0	0
0	5	1	0	0	1	0
0	0	0	0	0	0	0
0	0	0	0	2	0	0
0	1	9	4	2	3	0
0	0	0	1	0	0	0
0	0	0	3	0	0	0
0	0	4	14	9	7	5
0	2	2	4	0	1	0

Matrix C

TARA CUDA IMPLEMENTATION

Scoring

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ
1	PROTEIN HEADER [FF30]: >sp|A1Z7U3|OST4_DROME Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 4 OS=Drosophila melanogaster OX=7227 GN=CG33774 PE=3 SV=1																																										
2	REF PROTEIN: MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK																																										
3	DNA INPUT: ATGATCACTACCGACGTCAATTGGCATTTTCTCCAACGTTCTGGGCGATATTTCTGCTTCCTGCTGGTCGGTCTATACATTATATTAATGCCAATACTCGGAAAGCCCAGTAAGGCAAAATGA																																										
4				M	I	T	D	V	Q	L	A	I	F	S	N	V	L	G	V	F	L	F	L	L	V	V	A	Y	H	Y	I	N	A	N	T	G	K	P	S	A	K	A	K
5		Uj	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
6	A	0	-999	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
7	T	1	-999	-5	0	-2	-4	-5	-4	-5	-3	-5	-4	-5	-5	-5	-4	-3	-5	-4	-5	-3	-5	-3	-3	-4	-4	-5	-5	-5	-5	-4	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
8	G	2	-999	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
9	A	3	-999	-5	-5	-5	-5	-3	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
10	T	4	-999	-5	-4	4	2	0	-2	-4	-3	-4	-1	-3	-5	-5	-2	-2	-4	-2	-4	-3	-3	-3	-1	0	-1	-3	-5	-5	-5	-1	-3	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
11	C	5	-999	-5	-5	-5	-4	-5	-5	-5	-5	-4	-5	-5	-1	-3	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-4	-5	-5	-5	-5	-4	-4	-4	-4	-5	-5	-5	-1	-3	-5	-4
12	A	6	-999	-5	-5	-5	-2	-4	-5	-5	-5	-5	-5	-5	-5	-4	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-3	3	1	-1	-3	-5	-4	-5	-5	-5	-5	-5	-5	-5	-5
13	C	7	-999	-5	-5	-1	9	7	5	3	1	-1	-3	-3	-3	-5	-5	-3	-4	-5	-4	-5	-5	-4	-4	-1	0	-1	-3	-5	-5	-5	-1	-3	-5	0	-2	-4	-5	-4	-5	-5	-5
14	T	8	-999	-5	-3	-3	-1	-3	-4	-5	-1	-3	-2	-4	-5	-4	-3	-1	-3	-4	-5	-1	-3	-1	-1	-3	-3	-5	-5	-5	-2	-3	-5	-5	-5	-5	-5	-5	-5	-5	-2	-4	-5
15	A	9	-999	-5	-5	-5	-5	2	0	-2	-4	-5	-5	-2	-4	-5	-5	-5	-5	-5	-2	-4	-2	-4	-5	-5	-5	-5	2	0	10	8	6	4	2	0	-2	-4	-5	-5	-5	-5	-5
16	C	10	-999	-5	-5	-3	4	8	6	4	2	0	-2	-4	-2	-3	-5	-5	-5	-4	-5	-5	-5	-5	-5	-4	-1	0	-2	-4	-3	-5	-5	-1	-3	0	-2	-4	-5	-4	-4	-5	-5
17	C	11	-999	-5	-5	-5	-4	4	2	0	-2	-2	-4	-5	-5	-5	-5	-5	-3	-5	-5	-5	-5	-5	-4	-3	-5	-4	-5	-4	-5	3	1	-1	-3	-5	-5	-5	2	0	-2	-3	-5
18	G	12	-999	-5	-5	-5	-5	-2	1	1	-1	-3	-5	-5	-3	-5	-5	-5	-5	-5	-5	-4	-5	-4	-5	-5	-5	-5	-3	2	5	7	5	3	1	-1	-3	-3	-5	-5	-5	-3	-5
19	A	13	-999	-5	-5	-5	2	10	8	6	4	2	0	-2	-4	-1	-3	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-3	-3	-3	-5	3	2	0	0	-2	-1	-3	-5	-2	-4	-5	-5
20	C	14	-999	-5	-5	-5	0	-1	4	2	0	-2	-3	-5	-4	-5	-5	-5	-5	-3	-5	-5	-5	-5	-5	-4	-3	-5	-5	-5	-4	0	3	1	-1	2	0	-2	-3	3	1	-1	-3
21	G	15	-999	-5	-5	-5	-5	-4	3	2	0	-2	-4	-5	-5	-3	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-3	3	2	7	5	3	1	-1	-3	-5	-3	-5	-3	-4
22	T	16	-999	-5	-4	-2	0	5	14	12	10	8	6	4	2	0	3	1	-1	-1	-3	-4	-5	-4	-4	-1	-1	-3	-4	-5	-3	4	2	2	0	0	-2	-3	-5	-5	-1	-3	-5
23	C	17	-999	-5	-5	-5	-4	0	4	4	2	0	-2	-4	-1	-3	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	-4	-5	-5	-3	1	4	2	0	2	0	-2	1	4	2	0
24	A	18	-999	-5	-5	-5	-5	-4	-1	15	13	11	9	7	5	3	1	-1	-3	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	1	0	2	6	4	2	0	-2	-4	-5	-2	-4	-4
25	A	19	-999	-5	-5	-5	-2	3	9	14	12	10	8	6	4	5	3	1	-1	-3	-4	-5	-5	-5	-5	-5	-4	-3	-5	-3	-5	-1	10	8	8	6	4	2	0	-2	-4	0	-2

Matrix I

TARA CUDA IMPLEMENTATION

Scoring

PROTEIN HEADER [FF30]: >sp|A1Z7U3|OST4_DROME Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 4 OS=Drosophila melanogaster OX=7227 GN=CG33774 PE=3 SV=1

REF PROTEIN: MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK

DNA INPUT: ATGATCACTACCGACGTCAATTGGCATTTTCTCCAACGTTCTGGGCGATATTTCTGCTTCCTGCTGGTCGGTCTATACATTATATTAATGCCAATACTCGGAAAGCCCAGTAAGGCAAAATGA

			M	I	T	D	V	Q	L	A	I	F	S	N	V	L	G	V	F	L	F	L	L	V	V	A	Y	H	Y	I	N	A	N	T	G	K	P	S	A	K	A	K
	U	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
A	0	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
T	1	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
G	2	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
A	3	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999	-999
T	4	-5	0	-4	-5	-5	-4	-5	-3	-5	-4	-5	-5	-5	-4	-3	-5	-4	-5	-3	-5	-3	-3	-4	-4	-5	-5	-5	-5	-4	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
C	5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
A	6	-5	-5	-5	-5	-3	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
C	7	-5	-2	4	-1	-3	-2	-5	-3	-4	-1	-4	-5	-5	-2	-2	-5	-2	-4	-3	-3	-3	-1	0	-1	-5	-5	-5	-5	-1	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
T	8	-5	-5	-5	-4	-5	-5	-5	-5	-4	-5	-5	-1	-4	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-4	-5	-5	-5	-5	-4	-4	-4	-4	-5	-5	-5	-1	-4	-5	-4	-5
A	9	-5	-5	-5	-2	-5	-5	-5	-5	-5	-5	-5	-5	-4	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-3	3	-2	-4	-4	-5	-4	-5	-5	-5	-5	-5	-5	-5	-5	-5
C	10	-5	-4	2	9	4	2	0	-2	-3	-3	-3	-3	-5	-4	-3	-4	-4	-4	-5	-5	-4	-3	-1	0	-1	-5	-5	-5	-3	-1	-5	-5	0	-5	-5	-5	-4	-5	-5	-5	-5
C	11	-5	-3	-3	-1	-5	-4	-5	-1	-5	-2	-5	-3	-4	-3	-1	-5	-4	-5	-1	-5	-1	-1	-4	-3	-5	-5	-5	-2	-3	-5	-5	-5	-5	-5	-5	-5	-3	-2	-5	-5	-5
G	12	-5	-5	-5	-4	2	-1	-3	-5	-5	-5	-2	-5	-5	-5	-5	-5	-5	-2	-5	-2	-5	-5	-5	-5	-5	2	1	10	5	3	1	-1	-3	-5	-5	-5	-5	-5	-5	-5	-5
A	13	-5	-5	0	7	8	4	1	-1	-2	-4	-5	-2	-3	-5	-5	-5	-4	-5	-5	-5	-5	-5	-3	-1	0	-3	-5	-3	-5	-3	-1	-5	0	-2	-5	-5	-4	-4	-5	-5	-5
C	14	-5	-5	-5	-3	4	-1	-3	-3	-2	-4	-5	-5	-5	-5	-3	-3	-5	-5	-3	-5	-3	-3	-3	-5	-4	-5	-4	-4	3	-1	-2	-5	-5	-5	-5	2	-3	-4	-3	-5	-5
G	15	-5	-5	-5	-5	0	1	1	-4	-5	-5	-4	-3	-5	-5	-5	-5	-5	-4	-4	-4	-4	-5	-5	-5	-5	0	2	8	7	5	2	1	-2	-5	-3	-5	-5	-5	-3	-5	-3
T	16	-5	-5	-2	5	10	5	4	1	-1	-3	-5	-4	-1	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-3	-2	-3	-3	-5	3	2	-3	0	-2	-1	-3	-5	-2	-5	-5	-5	-5
C	17	-5	-5	-5	0	2	4	-1	-3	-4	-3	-5	-4	-5	-5	-5	-5	-3	-5	-5	-5	-5	-5	-4	-3	-5	-5	-5	-4	1	3	1	-2	2	-3	-5	0	3	-2	-4	-3	-5
A	18	-5	-5	-5	-5	-2	3	2	-1	-4	-5	-5	-5	-3	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	0	6	5	7	4	2	0	-4	-3	-5	-3	-5	-3	-4	-3
A	19	-5	-4	-2	3	8	14	9	7	5	3	1	-1	-3	3	-2	-4	-1	-5	-4	-5	-4	-4	-1	-1	-4	-4	-5	-3	4	0	2	-2	0	-3	-3	-5	-4	-1	-5	-5	-5
T	20	-5	-5	-5	-2	0	4	4	-1	-2	-5	-5	-1	-3	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	-5	-5	-5	-1	1	4	2	0	2	-3	-2	1	4	-1	-3	-3
T	21	-5	-5	-5	-5	-4	1	15	10	8	6	4	2	0	-2	-3	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-4	-2	4	3	5	6	4	1	-2	-3	-4	-5	-2	-4	-4	-3
G	22	-5	-5	-4	1	6	12	14	9	7	5	3	2	5	1	0	-2	-3	-4	-5	-5	-5	-5	-3	-3	-3	-5	-3	-5	2	10	5	8	3	1	-1	-3	-4	-3	0	-5	-5
G	23	-5	-4	-1	-4	-2	4	7	13	8	8	4	2	0	0	1	-4	-2	-5	-3	-5	-3	-3	-2	-2	-4	-3	-5	-5	1	-1	2	1	1	0	-1	-4	-1	2	1	-2	-5

Matrix D

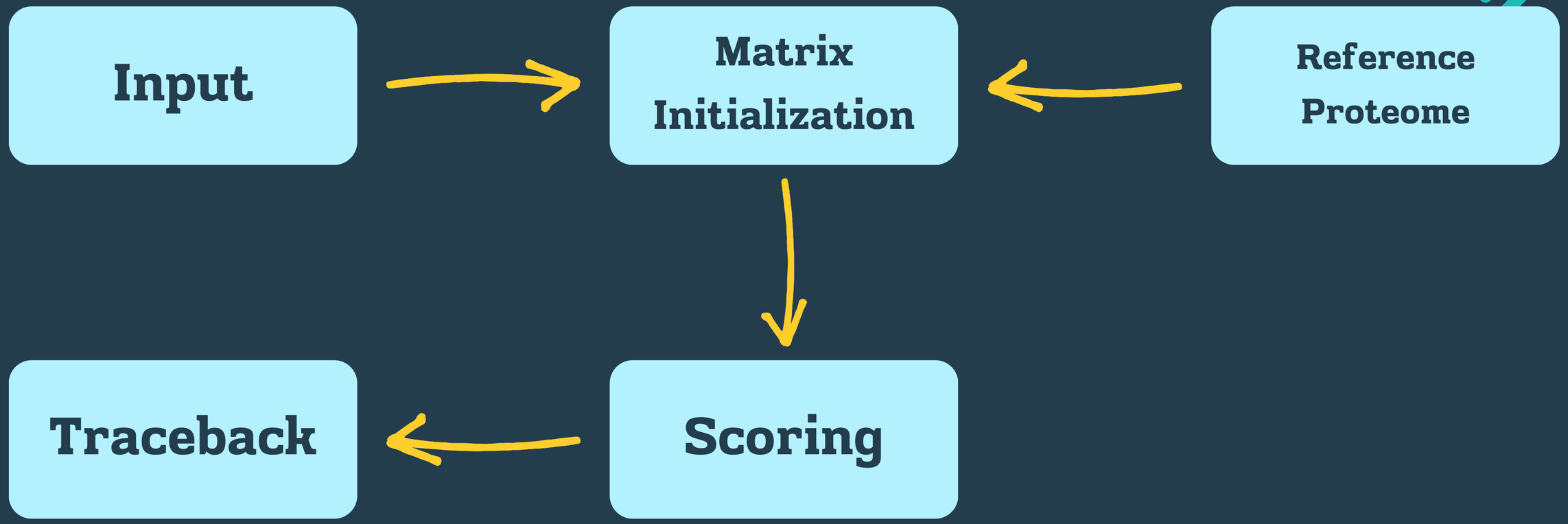
TARA CUDA IMPLEMENTATION

Scoring

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ
1	PROTEIN HEADER [FF30]: >sp|A1Z7U3|OST4_DROME Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 4 OS=Drosophila melanogaster OX=7227 GN=CG33774 PE=3 SV=1																																										
2	REF PROTEIN: MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK																																										
3	DNA INPUT: ATGATCACTACCGACGTCAATTGGCATTTTCTCCAACGTTCTGGGCGATATTTCTGCTTCCTGCTGGTCGGTCTATACATTATATTAATGCCAATACTCGGAAAGCCCAGTAAGGCAAAATGA																																										
4				M	I	T	D	V	Q	L	A	I	F	S	N	V	L	G	V	F	L	F	L	L	V	V	A	Y	H	Y	I	N	A	N	T	G	K	P	S	A	K	A	K
5		ij	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
6	A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	T	1	0	5	1	0	0	1	0	2	0	1	0	0	0	1	2	0	1	0	2	0	2	2	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
8	G	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	A	3	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	T	4	0	1	9	4	2	3	0	2	1	4	1	0	0	3	3	0	3	1	2	2	2	4	5	4	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0
11	C	5	0	0	0	1	0	0	0	0	1	0	0	4	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	1	1	0	0	0	4	1	0	1	0
12	A	6	0	0	0	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2	8	3	1	1	0	1	0	0	0	0	0	0	0	0	0
13	C	7	0	0	4	14	9	7	5	3	2	0	2	2	0	0	2	1	0	1	0	0	1	1	4	5	4	0	0	0	0	4	0	0	5	0	0	0	1	0	0	0	0
14	T	8	0	2	2	4	0	1	0	4	0	3	0	0	1	2	4	0	1	0	4	0	4	4	1	2	0	0	0	3	2	0	0	0	0	0	0	0	0	3	0	0	0
15	A	9	0	0	0	0	7	4	2	0	0	0	3	0	0	0	0	0	0	3	0	3	0	0	0	0	0	7	4	15	10	8	6	4	2	0	0	0	0	0	0	0	0
16	C	10	0	0	2	9	13	9	6	4	3	1	0	3	2	0	0	0	1	0	0	0	0	0	1	4	5	2	0	2	0	0	4	0	5	3	0	0	1	1	0	0	0
17	C	11	0	0	0	1	9	4	2	0	3	0	0	0	0	0	0	2	0	0	0	0	0	1	2	0	1	0	1	0	8	4	3	0	0	0	0	7	2	0	2	0	0
18	G	12	0	0	0	0	3	6	6	1	0	0	0	2	0	0	0	0	0	0	1	0	1	0	0	0	0	2	7	10	12	10	7	6	3	0	2	0	0	0	2	0	2
19	A	13	0	0	0	7	15	10	9	6	4	2	0	0	4	0	0	0	0	0	0	0	0	0	0	0	2	2	2	0	8	7	2	5	0	4	2	0	3	0	0	0	0
20	C	14	0	0	0	5	4	9	4	2	0	2	0	1	0	0	0	0	2	0	0	0	0	0	1	2	0	0	0	1	5	8	6	3	7	2	0	2	8	3	1	2	0
21	G	15	0	0	0	0	1	8	7	4	1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2	8	7	12	9	7	5	1	2	0	2	0	2	1	2
22	T	16	0	1	3	5	10	19	14	12	10	8	6	4	2	8	3	1	4	0	1	0	1	1	4	4	0	1	0	2	9	5	7	2	5	0	2	0	0	4	0	0	0
23	C	17	0	0	0	1	5	9	9	4	3	0	0	4	2	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	2	6	9	7	4	7	2	0	6	9	4	2	2
24	A	18	0	0	0	0	1	4	20	15	13	11	9	7	5	3	2	0	0	0	0	0	0	0	0	0	0	0	0	6	5	7	11	9	6	3	2	1	0	3	1	1	2
25	A	19	0	0	0	3	8	14	19	14	12	10	8	7	10	5	5	3	0	1	0	0	0	0	0	1	2	0	2	0	4	15	10	13	8	6	4	2	1	0	5	0	0
26	T	20	0	1	4	0	0	9	12	18	13	13	9	7	5	5	6	1	3	0	2	0	2	2	3	3	0	2	0	0	6	2	5	6	6	2	4	0	1	5	6	3	0
27	T	21	0	2	2	0	0	5	15	24	19	17	15	13	11	9	7	5	3	1	4	0	4	4	1	1	0	0	0	4	8	5	10	8	8	3	1	0	0	1	3	0	0
28	G	22	0	0	0	1	6	12	14	17	12	10	11	6	5	7	3	3	0	1	0	1	0	0	0	0	0	4	0	4	2	10	12	8	11	6	4	2	0	0	0	2	0

Matrix C

TARA CUDA IMPLEMENTATION



Traceback

-1	-1	-1	-2
-1	-1	-1	3
-1	-1	-1	3
-1	-1	-1	4
-1	-1	-1	-1
3	-1	-1	4

$$TC(i,j) = \begin{cases} -2, & TC(i,j-1) \\ -1, & TC(i-3,j) \\ 1, & TC(i-1,j-1) \\ 2, & TC(i-2,j-1) \\ 3, & TC(i-3,j-1) \\ 4, & TC(i-4,j-1) \end{cases}$$

Traceback

-1	-1	-1	-2
-1	-1	-1	3
-1	-1	-1	3
-1	-1	-1	4
-1	-1	-1	-1
3	-1	-1	4

$$TC(i,j) = \begin{cases} -2, & TC(i,j-1) \\ -1, & TC(i-3,j) \\ 1, & TC(i-1,j-1) \\ 2, & TC(i-2,j-1) \\ 3, & TC(i-3,j-1) \\ 4, & TC(i-4,j-1) \end{cases}$$

Traceback

-1	-1	-1	-2
-1	-1	-1	3
-1	-1	-1	3
-1	-1	-1	4
-1	-1	-1	-1
3	-1	-1	4

$$TC(i,j) = \begin{cases} -2, & TC(i,j-1) \\ -1, & TC(i-3,j) \\ 1, & TC(i-1,j-1) \\ 2, & TC(i-2,j-1) \\ 3, & TC(i-3,j-1) \\ 4, & TC(i-4,j-1) \end{cases}$$

Traceback

-1	-1	-1	-2
-1	-1	-1	3
-1	-1	-1	3
-1	-1	-1	4
-1	-1	-1	-1
3	-1	-1	4

$$TC(i,j) = \begin{cases} -2, & TC(i,j-1) \\ -1, & TC(i-3,j) \\ 1, & TC(i-1,j-1) \\ 2, & TC(i-2,j-1) \\ 3, & TC(i-3,j-1) \\ 4, & TC(i-4,j-1) \end{cases}$$

Traceback

-1	-1	-1	-2
-1	-1	-1	3
-1	-1	-1	3
-1	-1	1	4
-1	-1	2	-1
3	-1	-1	4

$$TC(i,j) = \begin{cases} -2, & TC(i,j-1) \\ -1, & TC(i-3,j) \\ 1, & TC(i-1,j-1) \\ 2, & TC(i-2,j-1) \\ 3, & TC(i-3,j-1) \\ 4, & TC(i-4,j-1) \end{cases}$$

Traceback

0	-1	-1	-2
0	0	-1	3
0	0	-1	3
-1	-1	1	4
-1	-1	2	-1
3	-1	-1	4

$$TC(i,j) = \begin{cases} -2, & TC(i,j-1) \\ -1, & TC(i-3,j) \\ 1, & TC(i-1,j-1) \\ 2, & TC(i-2,j-1) \\ 3, & TC(i-3,j-1) \\ 4, & TC(i-4,j-1) \end{cases}$$

Traceback

-1	-1	-1	-2
-1	-1	-1	3
-1	-1	-1	3
-1	-1	-1	4
-1	-1	2	-1
3	-1	-1	4

$$TC(i,j) = \begin{cases} -2, & TC(i,j-1) \\ -1, & TC(i-3,j) \\ 1, & TC(i-1,j-1) \\ 2, & TC(i-2,j-1) \\ 3, & TC(i-3,j-1) \\ 4, & TC(i-4,j-1) \end{cases}$$

Traceback

-1	-1	-1	-2
-1	-1	-1	3
-1	-1	-1	3
-1	-1	-1	4
-1	-1	2	-1
3	-1	-1	4

$$TC(i,j) = \begin{cases} -2, & TC(i,j-1) \\ -1, & TC(i-3,j) \\ 1, & TC(i-1,j-1) \\ 2, & TC(i-2,j-1) \\ 3, & TC(i-3,j-1) \\ 4, & TC(i-4,j-1) \end{cases}$$

Traceback

-1	-1	-1	-2
-1	-1	-1	3
-1	-1	-1	3
-1	3	-1	4
-1	-1	3	-1
3	-1	-1	4

$$TC(i,j) = \begin{cases} -2, & TC(i,j-1) \\ -1, & TC(i-3,j) \\ 1, & TC(i-1,j-1) \\ 2, & TC(i-2,j-1) \\ 3, & TC(i-3,j-1) \\ 4, & TC(i-4,j-1) \end{cases}$$

Traceback

-1	-1	-1	-2
-1	-1	-1	3
-1	-1	-1	3
-1	3	-1	4
-1	-1	3	-1
3	-1	-1	4

$$TC(i,j) = \begin{cases} -2, & TC(i,j-1) \\ -1, & TC(i-3,j) \\ 1, & TC(i-1,j-1) \\ 2, & TC(i-2,j-1) \\ 3, & TC(i-3,j-1) \\ 4, & TC(i-4,j-1) \end{cases}$$

Traceback

-1	-1	-1	-2
-1	-1	-1	3
-1	-1	-1	3
-1	-1	-1	4
-1	-1	-1	-1
3	-1	-1	4

$$TC(i,j) = \begin{cases} -2, & TC(i,j-1) \\ -1, & TC(i-3,j) \\ 1, & TC(i-1,j-1) \\ 2, & TC(i-2,j-1) \\ 3, & TC(i-3,j-1) \\ 4, & TC(i-4,j-1) \end{cases}$$

Traceback

-1	-1	-1	-2
-1	-1	-1	3
-1	-1	-1	3
-1	-1	-1	4
-1	-1	-1	-1
3	-1	-1	4

$$TC(i,j) = \begin{cases} -2, & TC(i,j-1) \\ -1, & TC(i-3,j) \\ 1, & TC(i-1,j-1) \\ 2, & TC(i-2,j-1) \\ 3, & TC(i-3,j-1) \\ 4, & TC(i-4,j-1) \end{cases}$$

TARA CUDA IMPLEMENTATION

Traceback

- Get the index of the Max Score

-1	-1	-1	-2	-2	-2	-2	-2
-1	-1	-1	3	-2	-2	-2	-2
-1	-1	-1	3	2	-2	-2	-2
-1	-1	-1	4	2	-2	-2	-2
-1	-1	-1	-1	3	3	-2	3
3	-1	-1	3	4	2	-2	4
-1	-1	-1	-1	-1	3	2	-2
-1	-1	-1	-1	3	3	3	-2
-1	2	-1	3	-1	3	4	2
-1	3	-1	-1	-1	2	-2	2
-1	-1	-1	-1	-1	3	-2	3
-1	-1	-1	3	-1	-1	2	4

Traceback

Traceback

-1	-1	-1	-2	-2	-2	-2	-2
-1	-1	-1	3	-2	-2	-2	-2
-1	-1	-1	3	2	-2	-2	-2
-1	-1	-1	4	2	-2	-2	-2
-1	-1	-1	-1	3	3	-2	3
3	-1	-1	3	4	2	-2	4
-1	-1	-1	-1	-1	3	2	-2
-1	-1	-1	-1	3	3	3	-2
-1	2	-1	3	-1	3	4	2
-1	3	-1	-1	-1	2	-2	2
-1	-1	-1	-1	-1	3	-2	3
-1	-1	-1	3	-1	-1	2	4

Index of the Max Score

Traceback

-1	-1	-1	-2	-2	-2	-2	-2
-1	-1	-1	3	-2	-2	-2	-2
-1	-1	-1	3	2	-2	-2	-2
-1	-1	-1	4	2	-2	-2	-2
-1	-1	-1	-1	3	3	-2	3
3	-1	-1	3	4	2	-2	4
-1	-1	-1	-1	-1	3	2	-2
-1	-1	-1	-1	3	3	3	-2
-1	2	-1	3	-1	3	4	2
-1	3	-1	-1	-1	2	-2	2
-1	-1	-1	-1	-1	3	-2	3
-1	-1	-1	3	-1	-1	2	4

K

TARA CUDA IMPLEMENTATION

Traceback

- Get the index of the Max Score
- From here, the traceback will work backward through the matrices.

Traceback

	-1	-1	-1	-2	-2	-2	-2	-2
	-1	-1	-1	3	-2	-2	-2	-2
	-1	-1	-1	3	2	-2	-2	-2
	-1	-1	-1	4	2	-2	-2	-2
	-1	-1	-1	-1	3	3	-2	3
	3	-1	-1	3	4	2	-2	4
	-1	-1	-1	-1	-1	3	2	-2
A	-1	-1	-1	-1	3	3	3	-2
	-1	2	-1	3	-1	3	4	2
	-1	3	-1	-1	-1	2	-2	2
K	-1	-1	-1	-1	-1	3	-2	3
	-1	-1	-1	3	-1	-1	2	4

$$TC(i,j) = \begin{cases} -2, & TC(i,j-1) \\ -1, & TC(i-3,j) \\ 1, & TC(i-1,j-1) \\ 2, & TC(i-2,j-1) \\ 3, & TC(i-3,j-1) \\ 4, & TC(i-4,j-1) \end{cases}$$

Traceback

	-1	-1	-1	-2	-2	-2	-2	-2
	-1	-1	-1	3	-2	-2	-2	-2
	-1	-1	-1	3	2	-2	-2	-2
	-1	-1	-1	4	2	-2	-2	-2
K	-1	-1	-1	-1	3	3	-2	3
	3	-1	-1	3	4	2	-2	4
	-1	-1	-1	-1	-1	3	2	-2
A	-1	-1	-1	-1	3	3	3	-2
	-1	2	-1	3	-1	3	4	2
	-1	3	-1	-1	-1	2	-2	2
K	-1	-1	-1	-1	-1	3	-2	3
	-1	-1	-1	3	-1	-1	2	4

$$TC(i,j) = \begin{cases} -2, & TC(i,j-1) \\ -1, & TC(i-3,j) \\ 1, & TC(i-1,j-1) \\ 2, & TC(i-2,j-1) \\ 3, & TC(i-3,j-1) \\ 4, & TC(i-4,j-1) \end{cases}$$

Traceback

	-1	-1	-1	-2	-2	-2	-2	-2
A	-1	-1	-1	3	-2	-2	-2	-2
	-1	-1	-1	3	2	-2	-2	-2
	-1	-1	-1	4	2	-2	-2	-2
K	-1	-1	-1	-1	3	3	-2	3
	3	-1	-1	3	4	2	-2	4
	-1	-1	-1	-1	-1	3	2	-2
A	-1	-1	-1	-1	3	3	3	-2
	-1	2	-1	3	-1	3	4	2
	-1	3	-1	-1	-1	2	-2	2
K	-1	-1	-1	-1	-1	3	-2	3
	-1	-1	-1	3	-1	-1	2	4

$$TC(i,j) = \begin{cases} -2, & TC(i,j-1) \\ -1, & TC(i-3,j) \\ 1, & TC(i-1,j-1) \\ 2, & TC(i-2,j-1) \\ 3, & TC(i-3,j-1) \\ 4, & TC(i-4,j-1) \end{cases}$$

Traceback

-1	-1	-1	-2	-2	-2	-2	-2	
-1	-1	-1	3	-2	-2	-2	-2	SA
-1	-1	-1	3	2	-2	-2	-2	
-1	-1	-1	4	2	-2	-2	-2	
-1	-1	-1	-1	3	3	-2	3	K
3	-1	-1	3	4	2	-2	4	
-1	-1	-1	-1	-1	3	2	-2	
-1	-1	-1	-1	3	3	3	-2	A
-1	2	-1	3	-1	3	4	2	
-1	3	-1	-1	-1	2	-2	2	
-1	-1	-1	-1	-1	3	-2	3	K
-1	-1	-1	3	-1	-1	2	4	

$$TC(i,j) = \begin{cases} -2, & TC(i,j-1) \\ -1, & TC(i-3,j) \\ 1, & TC(i-1,j-1) \\ 2, & TC(i-2,j-1) \\ 3, & TC(i-3,j-1) \\ 4, & TC(i-4,j-1) \end{cases}$$

Traceback

-1	-1	-1	-2	-2	-2	-2	-2	
-1	-1	-1	3	-2	-2	-2	-2	SA
-1	-1	-1	3	2	-2	-2	-2	
-1	-1	-1	4	2	-2	-2	-2	
-1	-1	-1	-1	3	3	-2	3	K
3	-1	-1	3	4	2	-2	4	
-1	-1	-1	-1	-1	3	2	-2	
-1	-1	-1	-1	3	3	3	-2	A
-1	2	-1	3	-1	3	4	2	
-1	3	-1	-1	-1	2	-2	2	
-1	-1	-1	-1	-1	3	-2	3	K
-1	-1	-1	3	-1	-1	2	4	

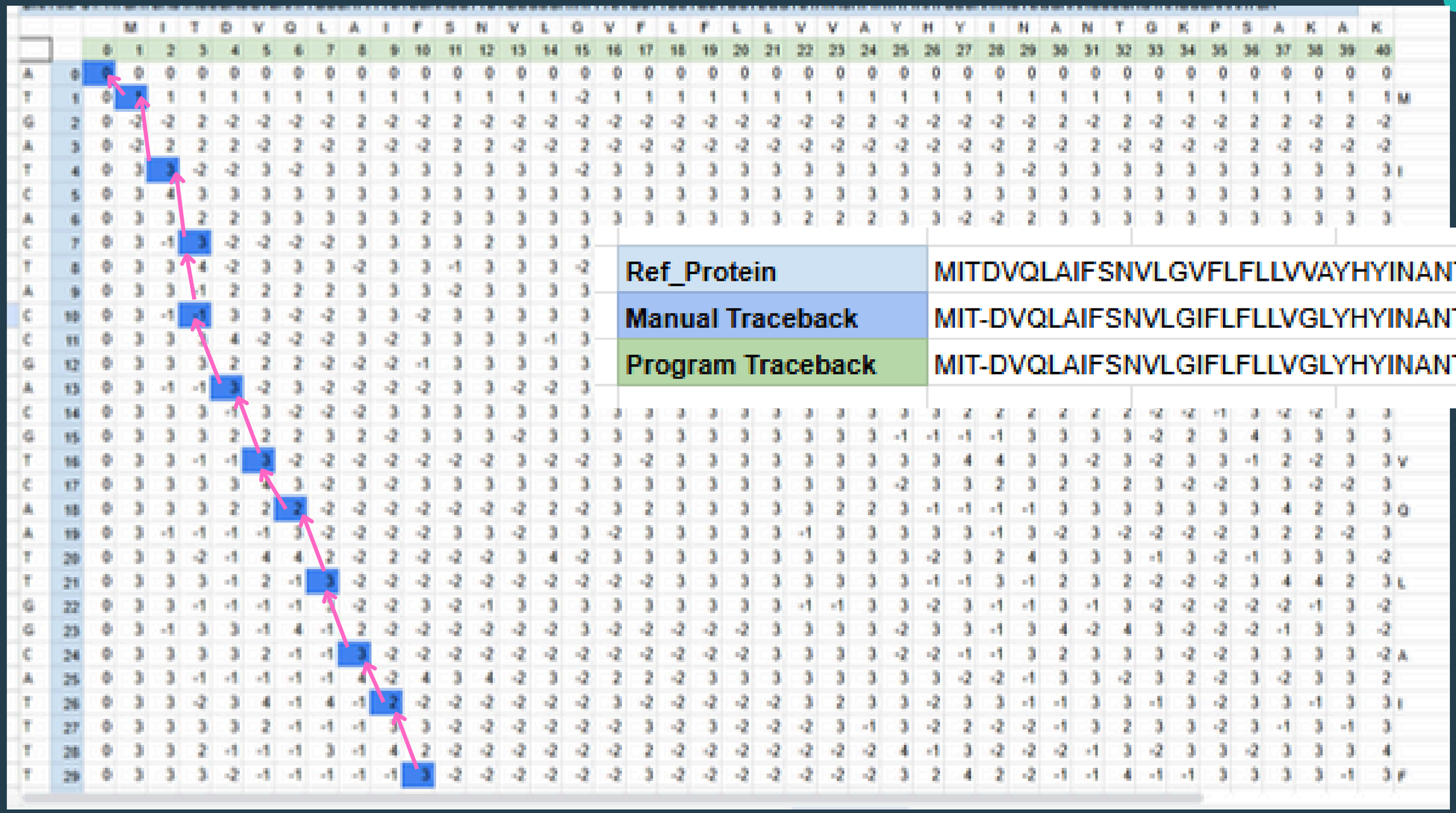
$$TC(i,j) = \begin{cases} -2, & TC(i,j-1) \\ -1, & TC(i-3,j) \\ 1, & TC(i-1,j-1) \\ 2, & TC(i-2,j-1) \\ 3, & TC(i-3,j-1) \\ 4, & TC(i-4,j-1) \end{cases}$$

TARA CUDA IMPLEMENTATION

Traceback

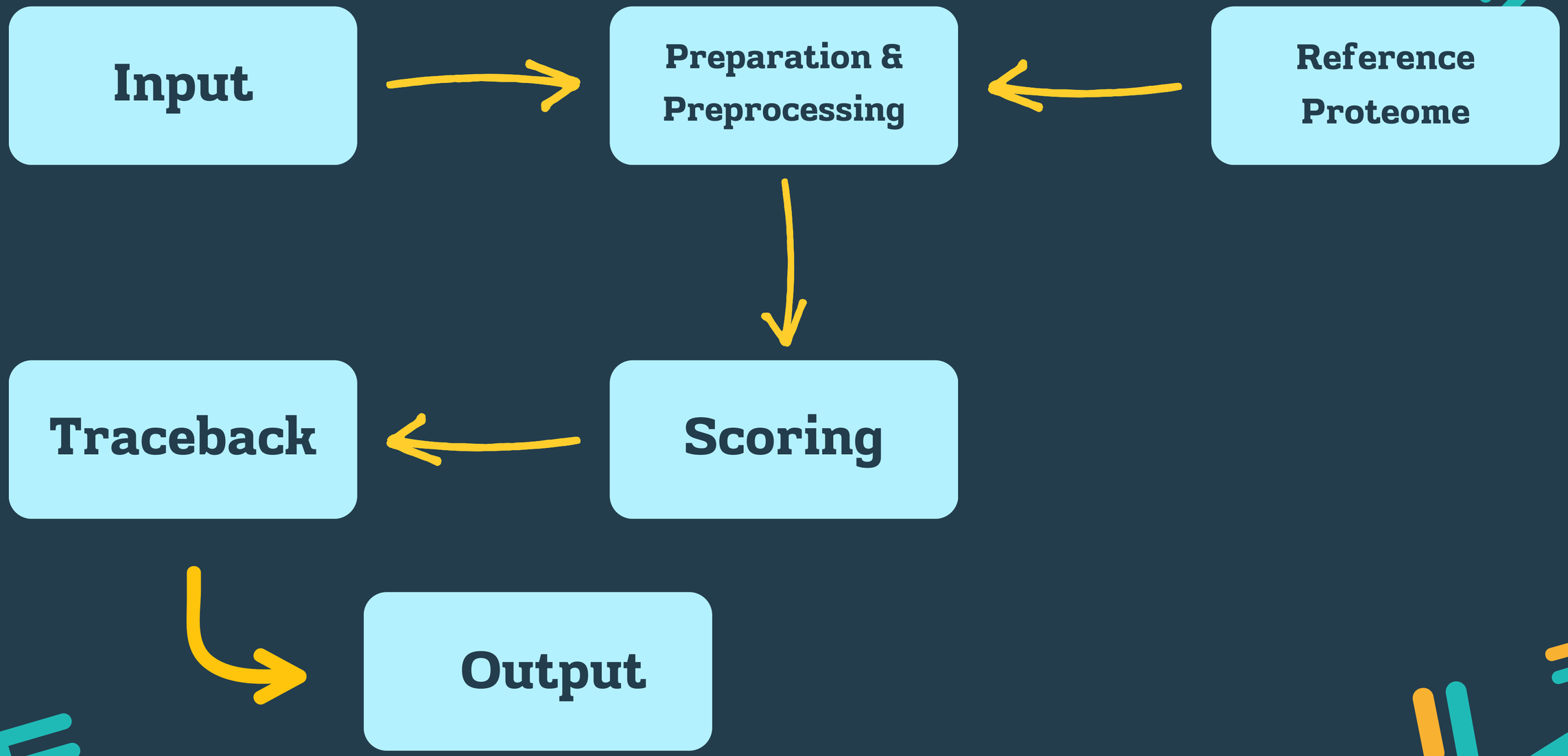
- Get the index of the Max Score
- From here, the traceback will work backward through the matrices.
- The traceback will continue until it encounters a cell with a score of zero

Traceback



Ref_Protein	MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK
Manual Traceback	MIT-DVQLAIFSNVLGIFLFLLVGLYHYINANTGKPSAKAK
Program Traceback	MIT-DVQLAIFSNVLGIFLFLLVGLYHYINANTGKPSAKAK

TARA CUDA IMPLEMENTATION



TARA CUDA IMPLEMENTATION

REF P. MITDVQLAIFSNVLGVFL...

O. DNA MITTDVQLAIFSNVLGIFL...

O. PROT MIT-DVQLAIFSNVLGIFL...

Output

TARA CUDA IMPLEMENTATION

```
frame 1:        M I T T D V N W H F L Q R S G R Y F C F L L V G L Y I I L M P I L G K P S K A K X
frame 2:         X S L P T S I G I F S N V L G D I S A S C W S V Y T L Y X C Q Y S E S P V R Q N
frame 3:          D H Y R R Q L A F S P T F W A I F L L P A G R S I H Y I N A N T R K A Q X G K M
Output DNA:     M I T T D V Q L A I F S N V L G I F L F L L V G L Y H Y I N A N T G K P S - K A K
Frameshift:                   *     *           *     *           *           *
Output Prot:    M I T - D V Q L A I F S N V L G V F L F L L V V A Y H Y I N A N T G K P S A K A K
Reference Seq: MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK
Time in ms: 332.675
```

Program Output

RESULTS & ANALYSIS

RESULTS & ANALYSIS

Upon completion of development, the implementation will be subjected to a series of tests. These evaluations will focus on ensuring **correctness**, assessing **runtime performance**, verifying **scalability**, and **analyzing cross-platform performance.**

1. Correctness Test
2. Runtime Test
3. Scalability Test
4. Cross-Platform Test

CORRECTNESS TEST

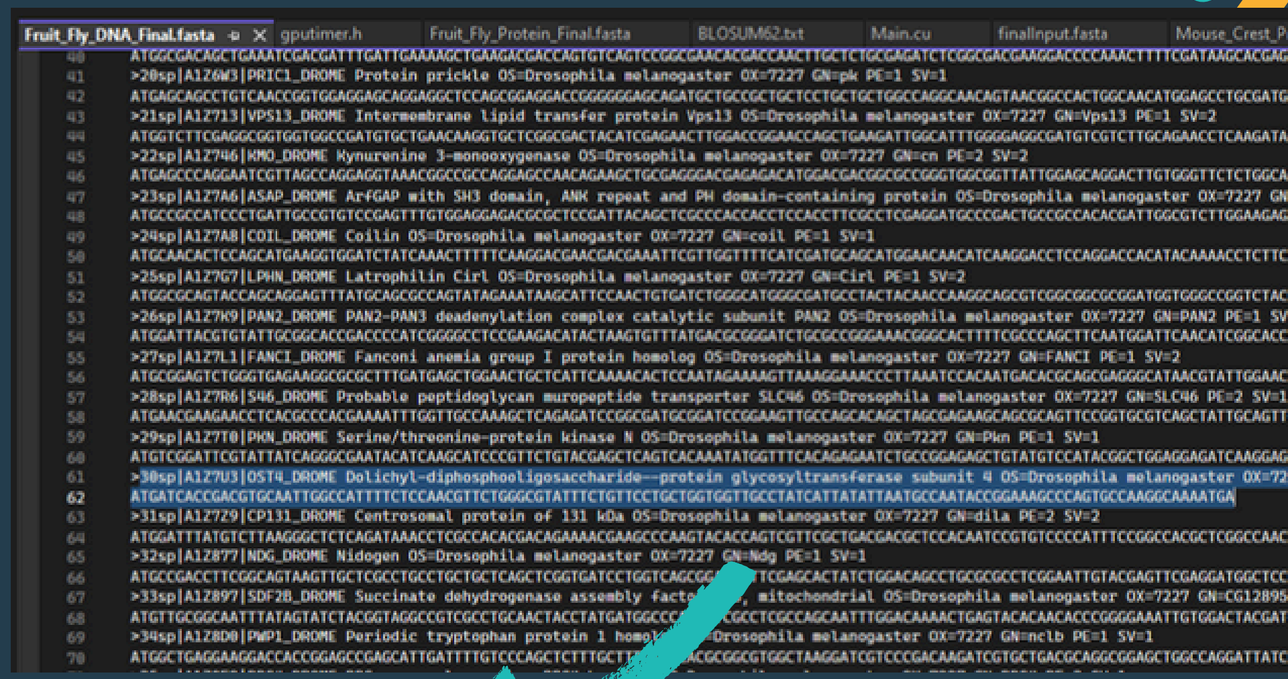
Evaluated the **algorithm's correctness** using aligned DNA and Protein sequence datasets, where DNA[i] aligns directly with Protein[j] via Manual Checking and External Cross-Checking.

CORRECTNESS TEST

- Evaluated the algorithm's correctness using aligned DNA and Protein sequence datasets, where DNA[i] aligns directly with Protein[j] via Manual Checking and External Cross-Checking.
  - Perfect Alignment

CORRECTNESS TEST

- **Perfect Alignment**




```
>0FFread30_match
ATGATCACCGACGTGCAATTGGCCATTTTCTCCAACGTTCTGGGCGTATTTCTGTTCCTGCT
GGTGGTTGCCTATCATTATATTAATGCCAATACCGGAAAGCCCAGTGCCAAGGCAAAATGA
```

Figure 6-1: DNA Substring (read30) - Match Case

read30 from the fruit fly dataset*

CORRECTNESS TEST

- Evaluated the algorithm's correctness using aligned DNA and Protein sequence datasets, where DNA[i] aligns directly with Protein[j] via Manual Checking and External Cross-Checking.
  - Perfect Alignment
  - With Insertion

CORRECTNESS TEST

- With Insertion

Dna Input v1
>1FFread30ins1
ATGATCACTCGACGTGCAATTGGCCATTTTCTCCAACGTTCTGAGGCGTATTTCTGTTCCTG
CTGGTGGTTGCCTATCATATATATTAATGCCAATACCGGAACAGCCCAGTGCCAAGGCAATG
AATGA

Dna Input v2
>2FFread30ins2
ATGATCACCGACGTGCAATTTAGGGCCATTTTCTCCAACGTTCTGGGCGTATTTCTGTTCCT
GCTGGTGGTTGCCTATCATTATATTAATGCCAATACCGGAAATAAGCCCAGTGCCAAGGCAAA
ATGA

Figure 6-2: DNA Substring - With Insertion

CORRECTNESS TEST

- Evaluated the **algorithm's correctness** using aligned DNA and Protein sequence datasets, where DNA[i] aligns directly with Protein[j] via Manual Checking and External Cross-Checking.
  - Perfect Alignment
  - With Insertion
  - With Deletion

CORRECTNESS TEST

- With Deletion

Dna [30]v1:
ATGATCACCGACGTGCAATTGGCCATTTTCTCCAACGTTCTGGGCGTATTTCTGTTCCTGCT
GGTGGTTGCCTATCATTATATTAATGCCAATACCGGAAAGCCCAGTGCCAAGGCAAAATGA

Input: 108
>3FFread30del1
ATGATCACCGACGTGCAATTGGCCATTTTCTCCAACGTTCTGGTATTTTTCCTGCTGGTGGT
TGCCTATCATTATATTAATGCCAATACCGGAAAGCCCAGTGCCAAGGCAAAATGA

Dna [30]v2:
ATGATCACCGACGTGCAATTGGCCATTTTCTCCAACGTTCTGGGCGTATTTCTGTTCCTGCT
GGTGGTTGCCTATCATTATATTAATGCCAATACCGGAAAGCCCAGTGCCAAGGCAAAATGA

Input: 117
>4FFread30del2
ATGATACCGACGTGCAATTGGCCATTTTCTCCAACTTCTGGGCGTATTTCTGTTCTGCTGGT
GGTTGCCTATCATTATATTAATGCCAATACCGGAAAGCCCAGTGCCAGGCAAAATGA

Figure 6-3: DNA Substring - With Deletion

CORRECTNESS TEST

- Evaluated the algorithm's correctness using aligned DNA and Protein sequence datasets, where DNA[i] aligns directly with Protein[j] via Manual Checking and External Cross-Checking.
  - Perfect Alignment
  - With Insertion
  - With Deletion
  - With Insertion and Deletion

CORRECTNESS TEST

Dna [30]
ATGATCACTACCGACGTGCAATTGGCCATTTTCTCCAACGTTCTGGGCGATATTTCTGCTTC
CTGCTGGTGGTTGCCTATCATTATATTAATGCCAATACTCGGAAAGCCCAGTGCCAAGGCAA
AATGA

Input - 123
ATGATCACTACCGACGTCAATTGGCATTTTCTCCAACGTTCTGGGCGATATTTCTGCTTCCTG
CTGGTCGGTCTATACATTATATTAATGCCAATACTCGGAAAGCCCAGTAAGGCAAAATGA

>read30_indelv2
ATGATTGTCACCGACGTGCAATTGGCCATTTTCTCCAACGTTCTGGGCGTATTTCTGTTCCT
GCTGGTGGTTGCCTATCATTATATTAATGCCAATACCGGAAAGCCCAGTGCCAAGGCAAAAT
GA

>read30_indelv2
ATGATTGTCACCGACGTGCAATTGGCCATTTTCTCCAACGTTCTGGGCGTATTTCTGTTCCT
GCTGGTGGTTGCCTATCATTATATTAATGCCAATACCGGAAAGCCCGCCAAGGCAAAATGA

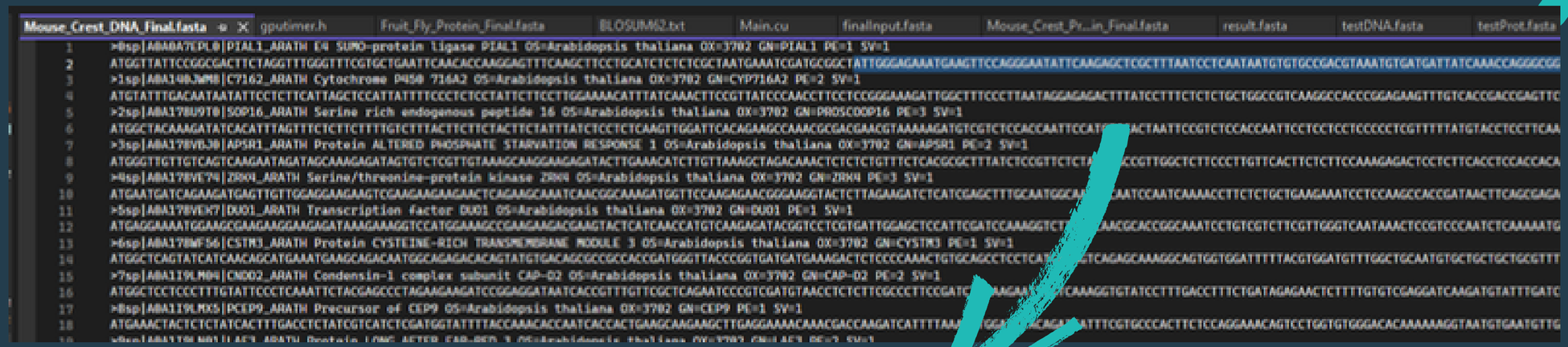
Figure 6-4: DNA Substring - With Insertion and Deletion

- With Insertion and Deletion

CORRECTNESS TEST

- Evaluated the algorithm's correctness using aligned DNA and Protein sequence datasets, where DNA[i] aligns directly with Protein[j] via Manual Checking and External Cross-Checking.
  - Perfect Alignment
  - With Insertion
  - With Deletion
  - With Insertion and Deletion
  - Complete Mismatch

CORRECTNESS TEST



>MCread0
TGTGCCGACGTAAATGTGATGATTATCAAACCAGGGCGGTGGTAATGGCGCTGATGATCTCA
GTCAAGAGTGCTTGTCAGCTTGGATGGTTCCCGGAGAGAGAGACTCAAGAATTGTTGGCTA
TCATAGACTTGATGTGGAATGGTTTCA

read0 from the mouse ear cress dataset*

Complete Mismatch

CORRECTNESS TEST

- Evaluated the algorithm's correctness using aligned DNA and Protein sequence datasets, where DNA[i] aligns directly with Protein[j] via **Manual Checking** and External Cross-Checking.
  - Perfect Alignment
  - With Insertion
  - With Deletion
  - With Insertion and Deletion
  - Complete Mismatch

MANUAL CHECKING

PROTEIN HEADER [FF30]: >sp|A1Z7U3|OST4_DROME Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 4 OS=Drosophila melanogaster OX=7227 GN=CG33774 PE=3 SV=1

REF PROTEIN: MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK

DNA INPUT: ATGATCACTACCGACGTCAATTGGCATTTTCTCCAACGTTCTGGGCGATATTTCTGCTTCCTGCTGGTCGGTCTATACATTATATTAATGCCAATACTCGGAAAGCCCAGTAAGGCAAAATGA

			M	I	T	D	V	Q	L	A	I	F	S	N	V	L	G	V	F	L	F	L	L	V	V	A	
	i,j	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	0	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
T	1	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
G	2	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
A	3	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
T	4	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
C	5	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
A	6	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
C	7	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
T	8	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
A	9	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
C	10	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
C	11	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
G	12	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
A	13	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
C	14	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

Compared the 3 scoring matrix (*c_mat, i_mat, d_mat*) from the program output to the manual excel

MANUAL CHECKING

Perfect Alignment

```
Score: 198

frame 1:        M  I  T  D  V  Q  L  A  I  F  S  N  V  L  G  V  F  L  F  L  L  V  V  A  Y  H  Y  I  N  A  N  T  G  K  P  S  A  K  A  K  X
frame 2:         X  S  P  T  C  N  W  P  F  S  P  T  F  W  A  Y  F  C  S  C  W  W  L  P  I  I  I  L  M  P  I  P  E  S  P  V  P  R  Q  N
frame 3:          D  H  R  R  A  I  G  H  F  L  Q  R  S  G  R  I  S  V  P  A  G  G  C  L  S  L  Y  X  C  Q  Y  R  K  A  Q  C  Q  G  K  M
Output DNA:     M  I  T  D  V  Q  L  A  I  F  S  N  V  L  G  V  F  L  F  L  L  V  V  A  Y  H  Y  I  N  A  N  T  G  K  P  S  A  K  A  K
Frameshift:
Output Prot:    M  I  T  D  V  Q  L  A  I  F  S  N  V  L  G  V  F  L  F  L  L  V  V  A  Y  H  Y  I  N  A  N  T  G  K  P  S  A  K  A  K
Reference Seq: MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK
Protein Indexes: 1 to 40

Time in ms: 203.027
```

Ref_Protein	MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK
Manual Traceback	MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK
Program Traceback	MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK

Score: 198

MANUAL CHECKING

- Perfect Alignment

Protein	Length	Score	s_index	e_index	Time(in sec)	
24	634	44	300	336	453.5861	
25	1697	50	692	722	1161.257	
26	1241	47	562	586	851.089	**Top Score**
27	1481	42	281	321	1020.578	198
28	519	47	249	287	357.5385	**Found in**
29	1190	57	1024	1065	816.8593	30
30	40	198	1	40	27.8574	
31	1134	34	224	252	787.6624	
32	1350	52	803	836	925.3988	
33	156	39	18	57	107.6762	
34	459	39	328	359	313.4625	
35	486	41	307	353	332.5456	
36	545	44	180	231	383.0587	

MANUAL CHECKING

With Insertion

```
Score: 184

frame 1:        M  I  T  D  V  Q  F  R  A  I  F  S  N  V  L  G  V  F  L  F  L  L  V  V  A  Y  H  Y  I  N  A  N  T  G  N  K  P  S  A  K  A  K  X
frame 2:         X  S  P  T  C  N  L  G  P  F  S  P  T  F  W  A  Y  F  C  S  C  W  W  L  P  I  I  I  L  M  P  I  P  E  I  S  P  V  P  R  Q  N
frame 3:          D  H  R  R  A  I  X  G  H  F  L  Q  R  S  G  R  I  S  V  P  A  G  G  C  L  S  L  Y  X  C  Q  Y  R  K  X  A  Q  C  Q  G  K  M
Output DNA:     M  I  T  D  V  Q  F  R  A  I  F  S  N  V  L  G  V  F  L  F  L  L  V  V  A  Y  H  Y  I  N  A  N  T  G  N  K  P  S  A  K  A  K
Frameshift:
Output Prot:    M  I  T  D  V  Q  L  -  A  I  F  S  N  V  L  G  V  F  L  F  L  L  V  V  A  Y  H  Y  I  N  A  N  T  G  -  K  P  S  A  K  A  K
Reference Seq: MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK
Protein Indexes: 1 to 40

Time in ms: 212.524
```

Ref_Protein	MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK
Manual Traceba	MITDVQFRAIFSNVLGVFLFLLVVAYHYINANTGNKPSAKAK
Program Traceb	MITDVQFRAIFSNVLGVFLFLLVVAYHYINANTGNKPSAKAK

Score: 184

MANUAL CHECKING

With Insertion

Protein	Length	Score	s_Index	e_index	Time (in ms)
29	1190	48	1024	1065	877.3087
30	40	184	1	40	31.1185
31	1134	36	198	252	830.4659
32	1350	49	896	928	1007.453
33	156	32	21	51	126.094
34	459	35	192	227	337.4611
35	486	44	313	353	356.6825
36	545	39	201	228	394.7275
37	2346	40	1646	1691	1712.253
38	351	48	310	350	254.8981
39	1977	46	1151	1195	1453.055
40	163	35	79	113	119.6926

Top Score
184
Found in
30

MANUAL CHECKING

With Deletion

```
Score: 149

frame 1:        M  I  N  V  Q  L  A  I  F  S  N  V  L  V  F  F  L  L  V  V  A  Y  H  Y  I  N  N  T  G  K  P  S  A  K  A  X
frame 2:         X  S  T  C  N  W  P  F  S  P  T  F  W  Y  F  S  C  W  W  L  P  I  I  I  L  I  I  P  E  S  P  V  P  R  H
frame 3:          D  Q  R  A  I  G  H  F  L  Q  R  S  G  I  F  P  A  G  G  C  L  S  L  Y  X  X  Y  R  K  A  Q  C  Q  G  M
Output DNA:     M  I  -  N  V  Q  L  A  I  F  S  N  V  L  -  V  F  -  F  L  L  V  V  A  Y  H  Y  I  N  -  N  T  G  K  P  S  A  K  A
Frameshift:
Output Prot:    M  I  T  D  V  Q  L  A  I  F  S  N  V  L  G  V  F  L  F  L  L  V  V  A  Y  H  Y  I  N  A  N  T  G  K  P  S  A  K  A
Reference Seq: MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK
Protein Indexes: 1 to 39

Time in ms: 178.679
```

Ref_Protein	MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK
Manual Traceba	MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKA
Program Tracel	MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKA

MANUAL CHECKING

With Deletion

Protein	Length	Score	s_index	e_index	Time (in ms)		
27	1481	41	598	631	950.1122		
28	519	46	172	204	338.4439		Top Score
29	1190	54	1024	1065	809.6146		149
30	40	149	1	39	26.4733		Found in
31	1134	39	563	601	713.4803		30
32	1350	38	1300	1333	860.4029		
33	156	30	96	127	98.7649		
34	459	37	327	357	289.1428		
35	486	41	9	54	300.9241		
36	545	52	339	372	348.116		
37	2346	31	1792	1826	1504.074		
38	351	44	185	216	225.2745		
39	1977	47	598	632	1290.232		
40	163	30	79	114	103.3844		

MANUAL CHECKING

With Insertion and Deletion

```
Score: 147

frame 1:        M  I  T  T  D  V  N  W  H  F  L  Q  R  S  G  R  Y  F  C  F  L  L  V  G  L  Y  I  I  L  M  P  I  L  G  K  P  S  K  A  K  X
frame 2:         X  S  L  P  T  S  I  G  I  F  S  N  V  L  G  D  I  S  A  S  C  W  S  V  Y  T  L  Y  X  C  Q  Y  S  E  S  P  V  R  Q  N
frame 3:          D  H  Y  R  R  Q  L  A  F  S  P  T  F  W  A  I  F  L  L  P  A  G  R  S  I  H  Y  I  N  A  N  T  R  K  A  Q  X  G  K  M
Output DNA:     M  I  T  T  D  V  Q  L  A  I  F  S  N  V  L  G  I  F  L  F  L  L  V  G  L  Y  H  Y  I  N  A  N  T  G  K  P  S  -  K  A  K
Frameshift:                       *        *                 *        *                 *                 *
Output Prot:    M  I  T  -  D  V  Q  L  A  I  F  S  N  V  L  G  V  F  L  F  L  L  V  V  A  Y  H  Y  I  N  A  N  T  G  K  P  S  A  K  A  K
Reference Seq: MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK
Protein Indexes: 1 to 40

Time in ms: 203.175
```

Ref_Protein	MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK
Manual Traceback	MIT-DVQLAIFSNVLGIFLFLLVGLYHYINANTGKPSAKAK
Program Traceback	MIT-DVQLAIFSNVLGIFLFLLVGLYHYINANTGKPSAKAK

MANUAL CHECKING

With Insertion and Deletion

Protein	Length	Score	s_index	e_index	Time (in ms)		
29	1190	46	955	990	811.477		Top Score
30	40	147	1	40	27.6793		147
31	1134	45	36	61	780.2066		Found in
32	1350	47	28	66	938.0438		30
33	156	33	103	117	106.9801		
34	459	42	150	191	313.0785		
35	486	41	305	336	335.1678		
36	545	44	344	387	372.5739		
37	2346	38	486	523	1678.204		
38	351	47	171	221	242.6604		
39	1978	50	1261	1292	1366.69		
40	163	31	72	109	114.8696		
41	652	49	327	357	474.3161		
42	950	40	296	328	676.9619		
43	883	45	51	83	615.9441		

MANUAL CHECKING

Complete Mismatch

```
Score: 28

frame 1:        D A V G H N M I A T C Y S R L N P P D V T E A L Q H Y Q R S I Q I D P R Q S E V V I D A C E L L V K
frame 2:         T L X A I T X L P L A T V G X T R P M X Q R P C S T T N V L F K L I R V S Q R L S L T P A N F W X
frame 3:          R C R P X H D C H L L Q X A E P A R C D R G P A A L P T F Y S N X S A S V R G C H X R L R T S G E
Output DNA:     M V T E A - L A A L T N V L - - F K L I R V V R G C H X T P A N S G E
Frameshift:       *           *     *                 *           *       *
Output Prot:    M I T D V Q L A I F S N V L G V F L F L L V V - A Y H Y I N A N T G K
Reference Seq: MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK
Protein Indexes: 1 to 34
```

Ref_Protein	MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK
Manual Traceback	MVTEALAALTNVLFKLIRVV-GCHXTPANSGK
Program Traceback	MVTEALAALTNVLFKLIRVV-GCHXTPANSGK

MANUAL CHECKING

Complete Mismatch

Protein	Length	Score	s_index	e_index	Time(in sec)	Time (in ms)	
19	635	46	147	191	535.6914		
20	1299	54	698	745	1087.269		Top Score
21	3321	60	1235	1276	2809.251		60
22	465	49	74	113	405.6597		Found in
23	1155	58	544	586	1022.003		21
24	634	44	344	389	555.4611		
25	1697	54	657	709	1464.229		
26	1241	49	879	933	1070.537		
27	1481	49	115	163	1280.757		
28	519	48	384	427	448.088		
29	1190	57	439	484	1028.707		
30	40	30	4	40	34.1434		
31	1134	52	449	489	977.2827		

MANUAL CHECKING

The Same Score

Protein	Length	Score	s_Index	e_index	Time (in ms)	
6519	151	46	4	42	165.1742	
6520	623	246	131	180	592.6589	**Top Score**
6521	130	68	29	86	123.718	246
6522	646	246	122	171	588.7805	**Found in**
6523	702	246	122	171	667.0111	6520
6524	891	49	389	433	898.3162	**6522**
6525	633	50	184	225	566.7814	6523
6526	524	51	238	280	475.9851	
6527	555	51	269	311	495.0546	
6528	604	56	235	278	544.8025	
6529	1526	56	729	772	1392.644	
6530	1375	56	578	621	1269.324	
6531	185	36	137	169	170.9544	
6532	137	36	13	52	161.6379	
6533	182	51	48	77	170.9909	

EXTERNAL CROSS-CHECKING: BLASTX

To further ensure correctness, the results of the implementation were compared with outputs from **BLASTX** and Zhang's Six Frame Algorithm in SIMD.

EXTERNAL CROSS-CHECKING: BLASTX

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Select columns Show 100

select all 1 sequences selected GenPept Graphics

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc. Len	Accession
RecName: Full=Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 4 [Drosophila melanogaster]	Drosophila melan...	57.4	57.4	98%	2e-14	100.00%	40	A1Z7U3.1

Figure 6-26: BLASTX Results for Perfect Match

Download GenPept Graphics Next Previous Descriptions

RecName: Full=Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 4 [Drosophila melanogaster]

Sequence ID: A1Z7U3.1 Length: 40 Number of Matches: 1

Range 1: 1 to 40 GenPept Graphics Next Match Previous Match

Score	Expect	Identities	Positives	Gaps	Frame
57.4 bits(137)	2e-14	40/40(100%)	40/40(100%)	0/40(0%)	+1

```
Query  1    MITDVQLAIFSNvlgvflfllvvAYHYINANTGKPSAKAK  120
            MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK
Sbjct  1    MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK  40
```

Related Information

Gene - associated gene details

AlphaFold Structure - 3D structure displays

Identical Proteins - Identical proteins to A1Z7U3.1

Figure 6-27: BLASTX Results for Alignment (Perfect Match)

EXTERNAL CROSS-CHECKING: BLASTX

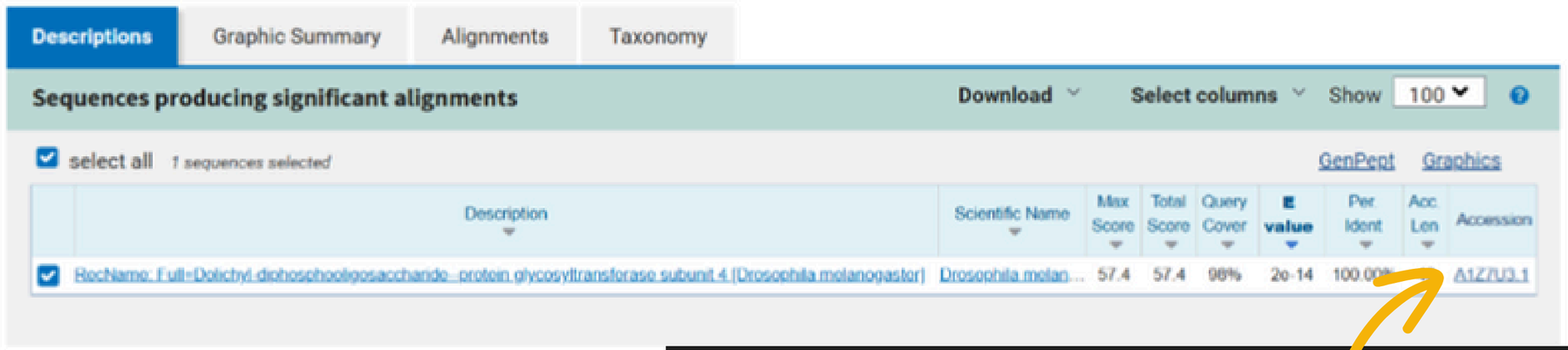
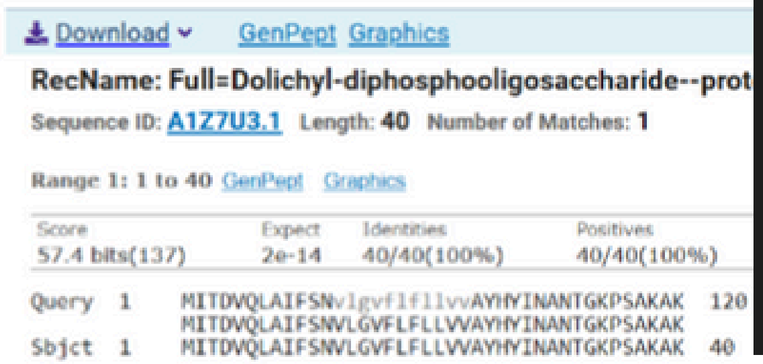


Figure 6-26: BLAS



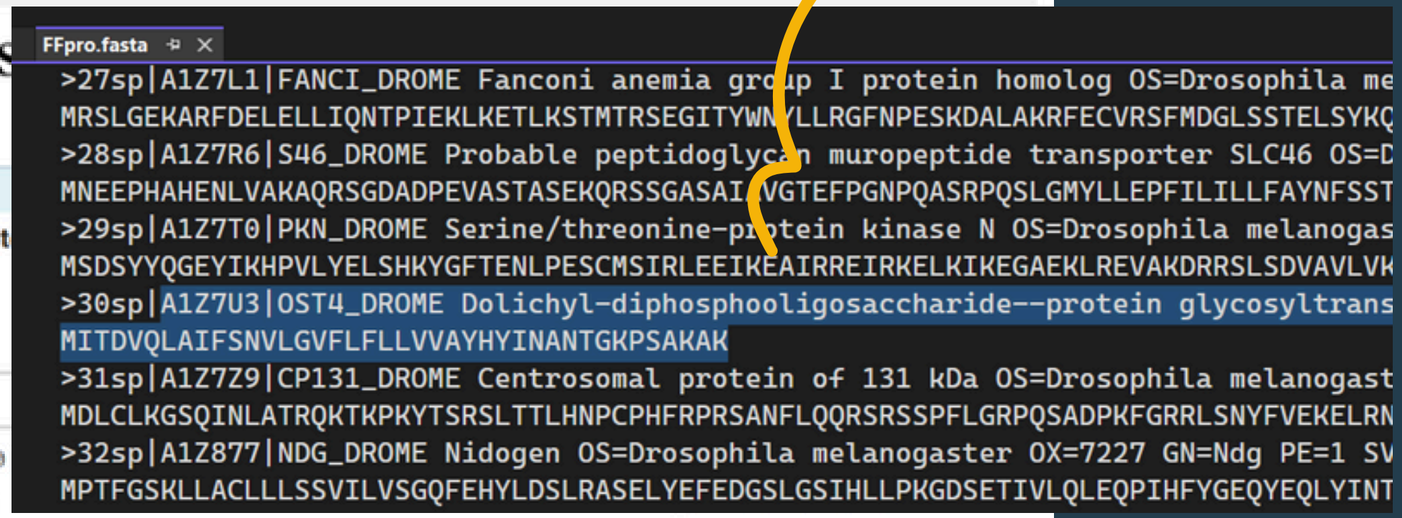
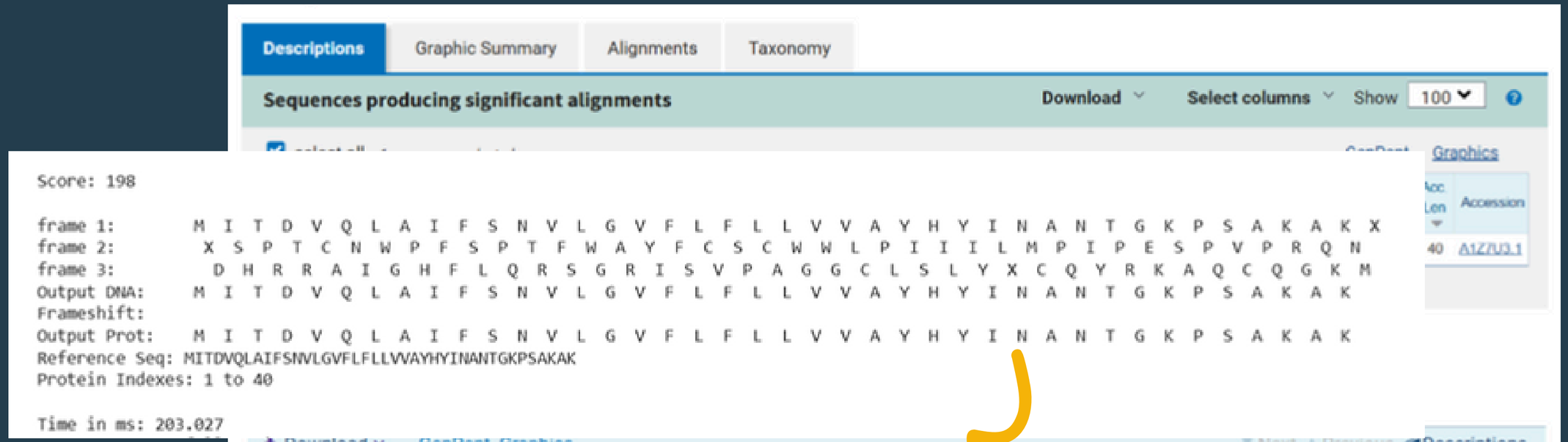


Figure 6-27: BLASTX Results for Alignment (Perfect Match)

EXTERNAL CROSS-CHECKING: BLASTX



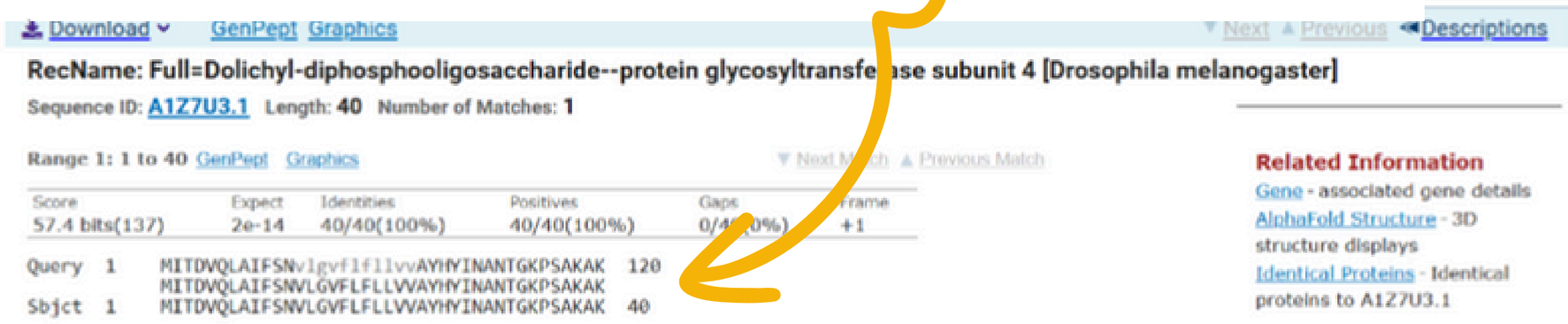


Figure 6-27: BLASTX Results for Alignment (Perfect Match)

EXTERNAL CROSS-CHECKING: BLASTX

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments — Download — Select columns — Show 100

select all 1 sequences selected — GenPept Graphics

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
RecName: Full=Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 4 [Drosophila melanogaster]	Drosophila melan...	46.6	46.6	98%	4e-10	92.86%	40	A1Z7U3.1

Figure 6-28: BLASTX Results for With Insertion

Download — GenPept Graphics — Next Previous Descriptions

RecName: Full=Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 4 [Drosophila melanogaster]

Sequence ID: A1Z7U3.1 Length: 40 Number of Matches: 1

Range 1: 1 to 40 GenPept Graphics — Next Match Previous Match

Score	Expect	Identities	Positives	Gaps	Frame
46.6 bits(109)	4e-10	39/42(93%)	39/42(92%)	2/42(4%)	+1

```
Query  1    MITDVQFRAIFSNvlgvflfllvvAYHYINANTGNKPSAKAK  126
            MITDVQ  AIFSNVLGVFLFLLVVAYHYINANTG KPSAKAK
Sbjct  1    MITDVQL-AIFSNVLGVFLFLLVVAYHYINANTG-KPSAKAK  40
```

Related Information

Gene - associated gene details

AlphaFold Structure - 3D structure displays

Identical Proteins - Identical proteins to A1Z7U3.1

Figure 6-29: BLASTX Results for Alignment (With Insertion)

EXTERNAL CROSS-CHECKING: BLASTX

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Select columns Show 100

select all *1 sequences selected* GenPept Graphics

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
RecName: Full=Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 4 [Drosophila melanogaster]	Drosophila melan...	53.1	53.1	97%	8e-13	82.50%	40	A1Z7U3.1

Figure 6-30: BLASTX Results for With Deletion

Download GenPept Graphics Next Previous Descriptions

RecName: Full=Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 4 [Drosophila melanogaster]

Sequence ID: A1Z7U3.1 Length: 40 Number of Matches: 1

Range 1: 1 to 40 GenPept Graphics Next Match Previous Match

Score	Expect	Identities	Positives	Gaps	Frame
53.1 bits(126)	8e-13	33/40(83%)	35/40(87%)	2/40(5%)	+1

```
Query  1    MITDVQLAIFSNvlvffllvv--AYHYINANTGKPSAKAK  114
            MITDVQLAIFSNVL  FL ++  AYHYINANTGKPSAKAK
Sbjct  1    MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK  40
```

Related Information

Gene - associated gene details

AlphaFold Structure - 3D structure displays

Identical Proteins - Identical proteins to A1Z7U3.1

Figure 6-31: BLASTX Results for Alignment (With Deletion)

EXTERNAL CROSS-CHECKING: BLASTX

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Select columns Show 100

select all *1 sequences selected* GenPept Graphics

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
RecName: Full=Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 4 [Drosophila melanogaster]	Drosophila melan...	49.3	49.3	95%	3e-11	92.50%	40	A1Z7U3.1

Figure 6-32: BLASTX Results for With Insertion and Deletion

Download GenPept Graphics Next Previous Descriptions

RecName: Full=Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 4 [Drosophila melanogaster]

Sequence ID: A1Z7U3.1 Length: 40 Number of Matches: 1

Range 1: 1 to 40 GenPept Graphics Next Match Previous Match

Score	Expect	Identities	Positives	Gaps	Frame
49.3 bits(116)	3e-11	37/40(93%)	39/40(97%)	1/40(2%)	+1

```
Query  4    IVTDVQLAIFSNvlgvflfllvvAYHYINANTGKP-AKAK  120
            ++TDVQLAIFSNVLGVFLFLLVVAYHYINANTGKP AKAK
Sbjct  1    MITDVQLAIFSNVLGVFLFLLVVAYHYINANTGKPSAKAK  40
```

Related Information

Gene - associated gene details

AlphaFold Structure - 3D structure displays

Identical Proteins - Identical proteins to A1Z7U3.1

Figure 6-33: BLASTX Results for Alignment (With Insertion and Deletion)

EXTERNAL CROSS-CHECKING: BLASTX

Figure 6-34: BLASTX Results for Complete Mismatch

EXTERNAL CROSS-CHECKING: SIMD

To further ensure correctness, the results of the implementation were compared with outputs from BLASTX and Zhang's Six Frame Algorithm in SIMD.

EXTERNAL CROSS-CHECKING: SIMD

Protein	Length	Score (SIMD)	Score (CUDA)		
0	774	58	58		TRUE
1	88	37	37		TRUE
2	2718	67	67		TRUE
3	1866	57	57		TRUE
4	965	61	61		TRUE
5	780	67	67		TRUE
6	343	47	47		TRUE
7	834	58	58		TRUE
8	2036	74	74		TRUE
9	400	57	57		TRUE
10	2176	245	245		TRUE
11	446	62	62		TRUE
12	668	55	55		TRUE
13	279	46	46		TRUE
14	1876	61	61		TRUE

Figure 6-35: Excel Snapshot of SIMD and CUDA Score Results Comparison

CORRECTNESS TEST

	Manual Checking	BLASTX	Zhang's Six Frame in SIMD
Zhang's Six Frame in SIMT	✓	✓	✓

RESULTS & DISCUSSION

Upon completion of development, the implementation will be subjected to a series of tests. These evaluations will focus on ensuring **correctness**, assessing **runtime performance**, verifying **scalability**, and **analyzing cross-platform performance.**

1. Correctness Test
2. Runtime Test
3. Scalability Test
4. Cross-Platform Test

RUNTIME TEST

- This test aimed to **assess the runtime performance of the CUDA implementation in comparison** to the sequential and SIMD implementations. This was done by recording the time before and after the execution of the kernel and computing the elapsed time.

- There would be two variants of this test: **Sequential vs SIMT** and **SIMD vs SIMT.**

RUNTIME TEST: SEQUENTIAL VS SIMT

PC uses

- CPU: Intel i7-12700F (12-core, 20-thread)
- GPU: NVIDIA RTX 4070 Super
- Memory: 32GB DDR4

Visual Studio is configured to

- CUDA Runtime: shared CUDA Runtime Library
- Optimization: Full Optimization (/Ox)
- C/C++ Runtime: Multi-Threaded (/MT)

RUNTIME TEST: SEQUENTIAL VS SIMT

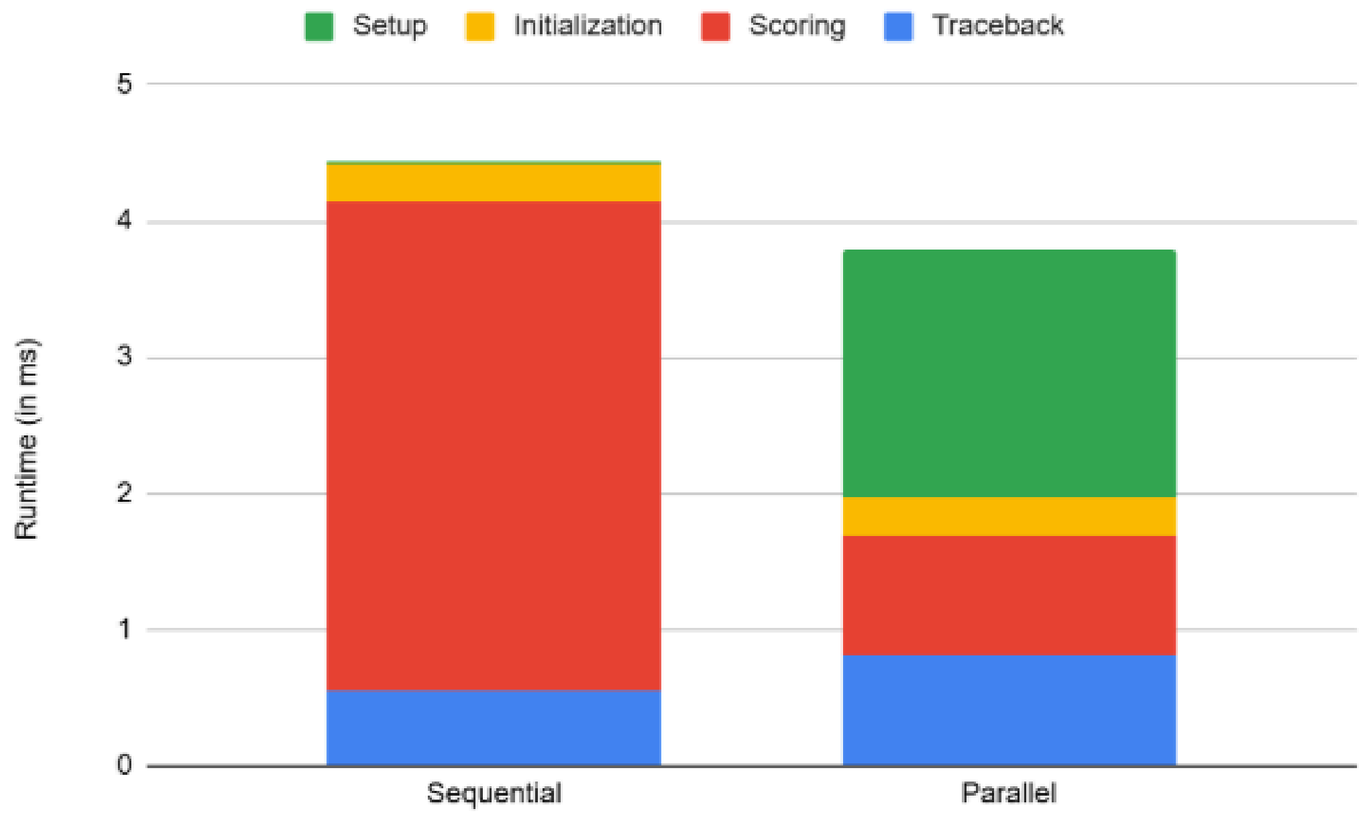


Figure 6-45: Visualization of Average Runtime Contribution of Each Stage For Single Read

- Scoring is a significant portion of the program

Tested on desktop with an NVIDIA RTX 4070 Super GPU and an Intel Core i7-12700F (Ox)

RUNTIME TEST: SEQUENTIAL VS SIMT

Stage	Sequential (in ms)	Fraction of the program	CUDA (in ms)	Fraction of the program
Program Setup	0.0117	0.27%	1.8238	48.09%
Matrix Initialization	0.2768	6.24%	0.2748	7.25%
Matrix Scoring	3.5892	80.98%	0.8726	23.01%
Traceback	0.5544	12.51%	0.8216	21.66%
Total	4.4321	100%	3.7928	100%

Table 6-5: Average Runtime Contribution of Each Phase

Tested on desktop with an NVIDIA RTX 4070 Super GPU and an Intel Core i7-12700F (Ox)

RUNTIME TEST: SEQUENTIAL VS SIMT

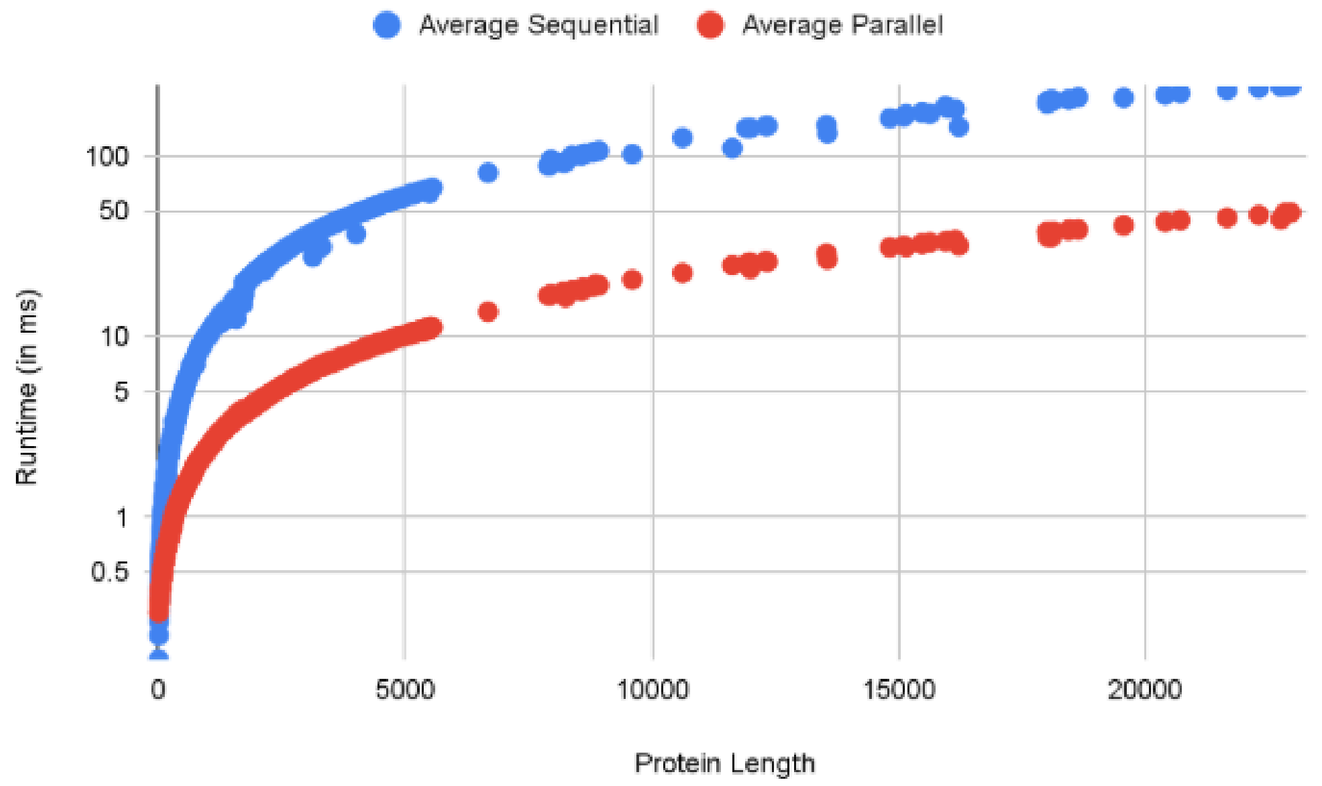


Figure 6-46: Runtime vs. Sequence Length for Sequential and Parallel Implementations (for Fruit Fly Dataset)

- CUDA performs better than seq, especially for longer reads

Tested on desktop with an NVIDIA RTX 4070 Super GPU and an Intel Core i7-12700F (Ox)

RUNTIME TEST: SEQUENTIAL VS SIMT

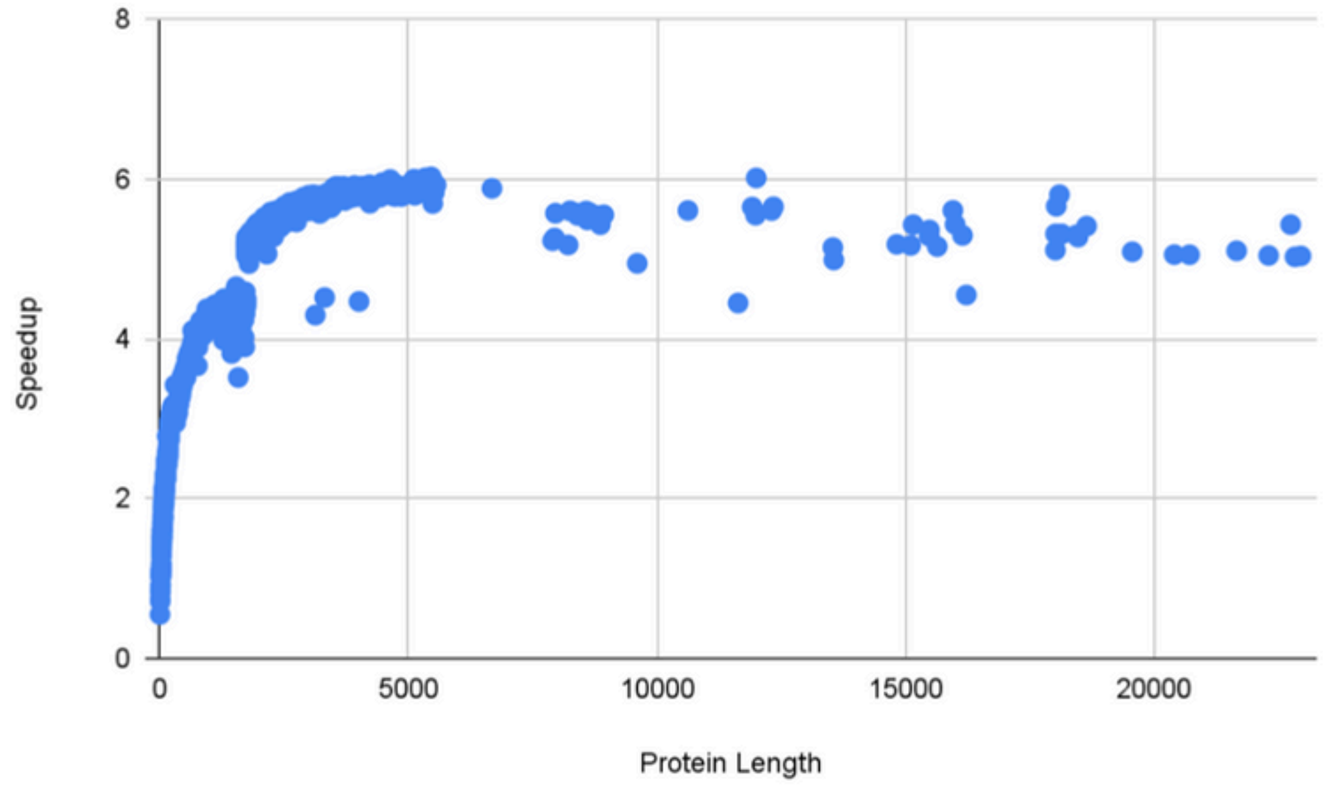


Figure 6-47: Speedup vs. Sequence Length for Sequential and Parallel Implementations (for Fruit Fly Dataset)

- Speedup generally increases as sequence length increases
- CUDA averages about a 3.51x speedup over Sequential

Tested on desktop with an NVIDIA RTX 4070 Super GPU and an Intel Core i7-12700F (Ox)

RUNTIME TEST: SEQUENTIAL VS SIMT

	Sequential (in ms)	SIMT (in ms)	Speedup
Average	7.1402	1.7300	3.5104x
Lowest	0.1606	0.2903	0.5468x
Highest	245.6083	48.7752	6.0252x

Table 6-6: Average Runtime and Speedup for Sequential and SIMT Implementation (for Fruit Fly Dataset)

Sequential (in s)	SIMT (in s)	Speedup
153.1505	37.1068	4.1273x

Table 6-7: Total Runtime and Speedup for Sequential and SIMT Implementation (for Fruit Fly Dataset)

- Slowdowns for shorter sequences
- Max Speedup of 6.03x
- Total Speedup of 4.13x

Tested on desktop with an NVIDIA RTX 4070 Super GPU and an Intel Core i7-12700F (Ox)

RUNTIME TEST: SEQUENTIAL VS SIMT

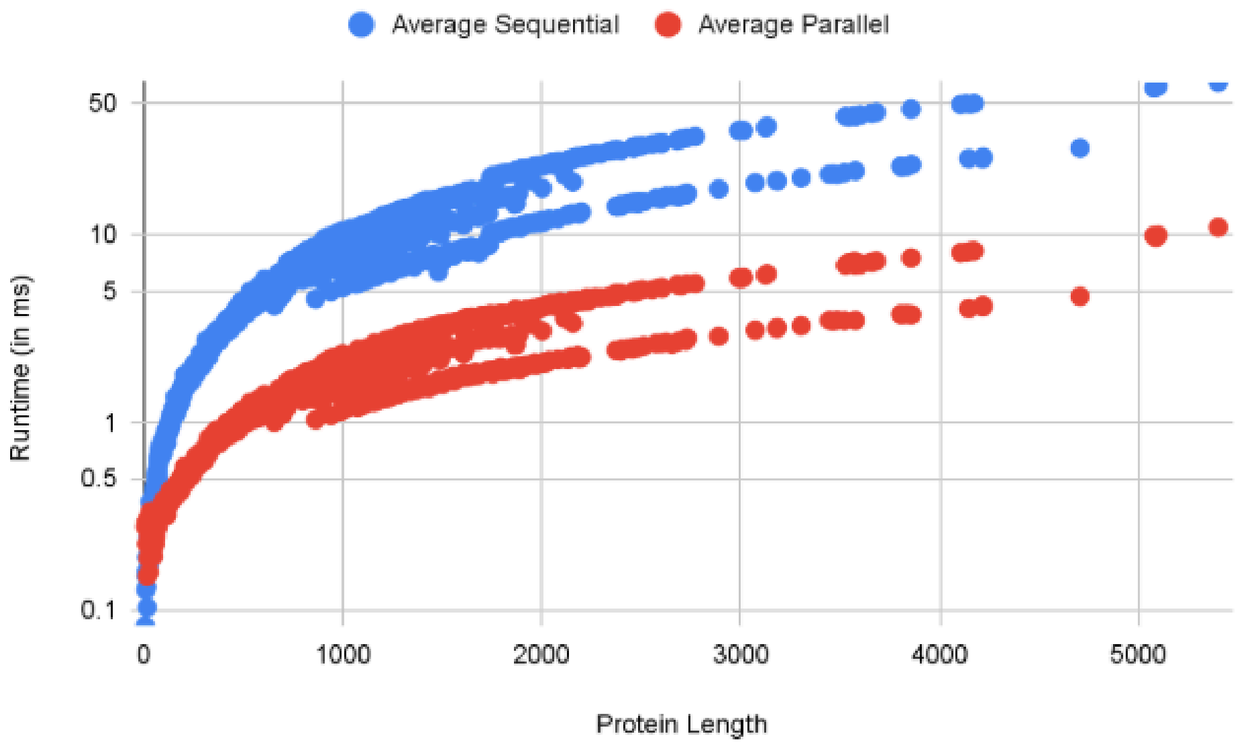
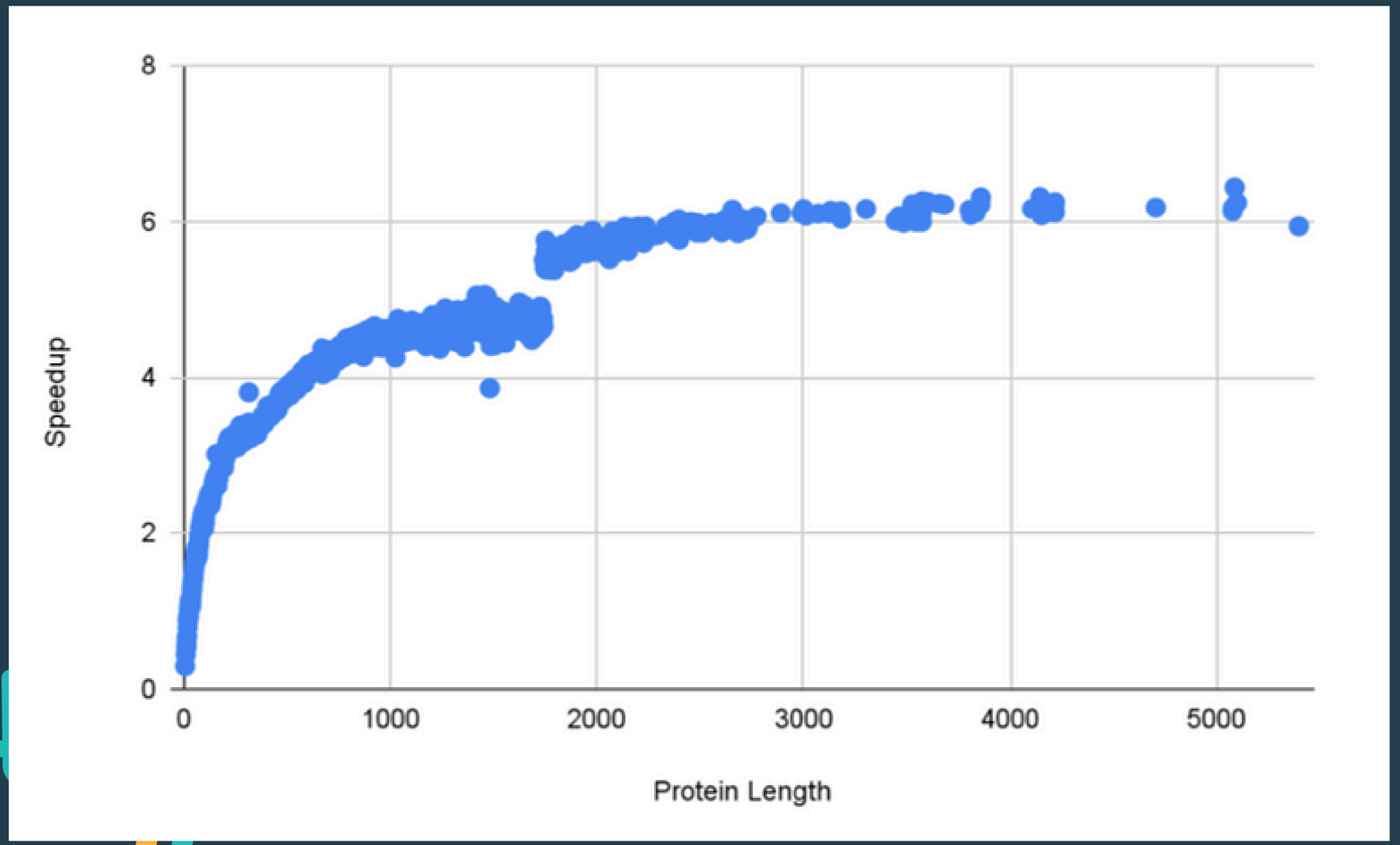


Figure 6-48: Runtime vs. Sequence Length for Sequential and Parallel Implementations (for Mouse Ear Cress Dataset)

- Behavior is consistent across datasets

Tested on desktop with an NVIDIA RTX 4070 Super GPU and an Intel Core i7-12700F (Ox)

RUNTIME TEST: SEQUENTIAL VS SIMT



- Behavior is consistent across datasets
- CUDA averages about a 3.44x speedup over Sequential

Tested on desktop with an NVIDIA RTX 4070 Super GPU and an Intel Core i7-12700F (Ox)

RUNTIME TEST: SEQUENTIAL VS SIMT

	Sequential (in ms)	SIMT (in ms)	Speedup
Average	3.4489	0.9023	3.4412x
Lowest	0.0822	0.1450	0.2954x
Highest	65.0062	10.9524	6.4371x

Table 6-8: Runtime and Speedup for SIMT and Sequential Implementation (for Mouse Ear Cress Dataset)

Sequential (in s)	SIMT (in s)	Speedup
134.9088	35.2936	3.8224x

Table 6-9: Total Runtime and Speedup for Sequential and SIMT Implementation (for Mouse Ear Cress Dataset)

- Behavior is consistent across datasets
- Max Speedup of 6.43x
- Total Speedup of 3.82x

Tested on desktop with an NVIDIA RTX 4070 Super GPU and an Intel Core i7-12700F (Ox)

RUNTIME TEST: SIMD VS SIMT

- Laptop utilizes the following specs:
  - CPU: **Ryzen 7 5800H** (8-core, 16-thread)
  - GPU: **NVIDIA RTX 3050 Ti Mobile**
  - Memory: **16GB** DDR4

- **Compatibility issues** with the SIMD exe (Would not run with windows 11 systems).

RUNTIME TEST: SIMD VS SIMT

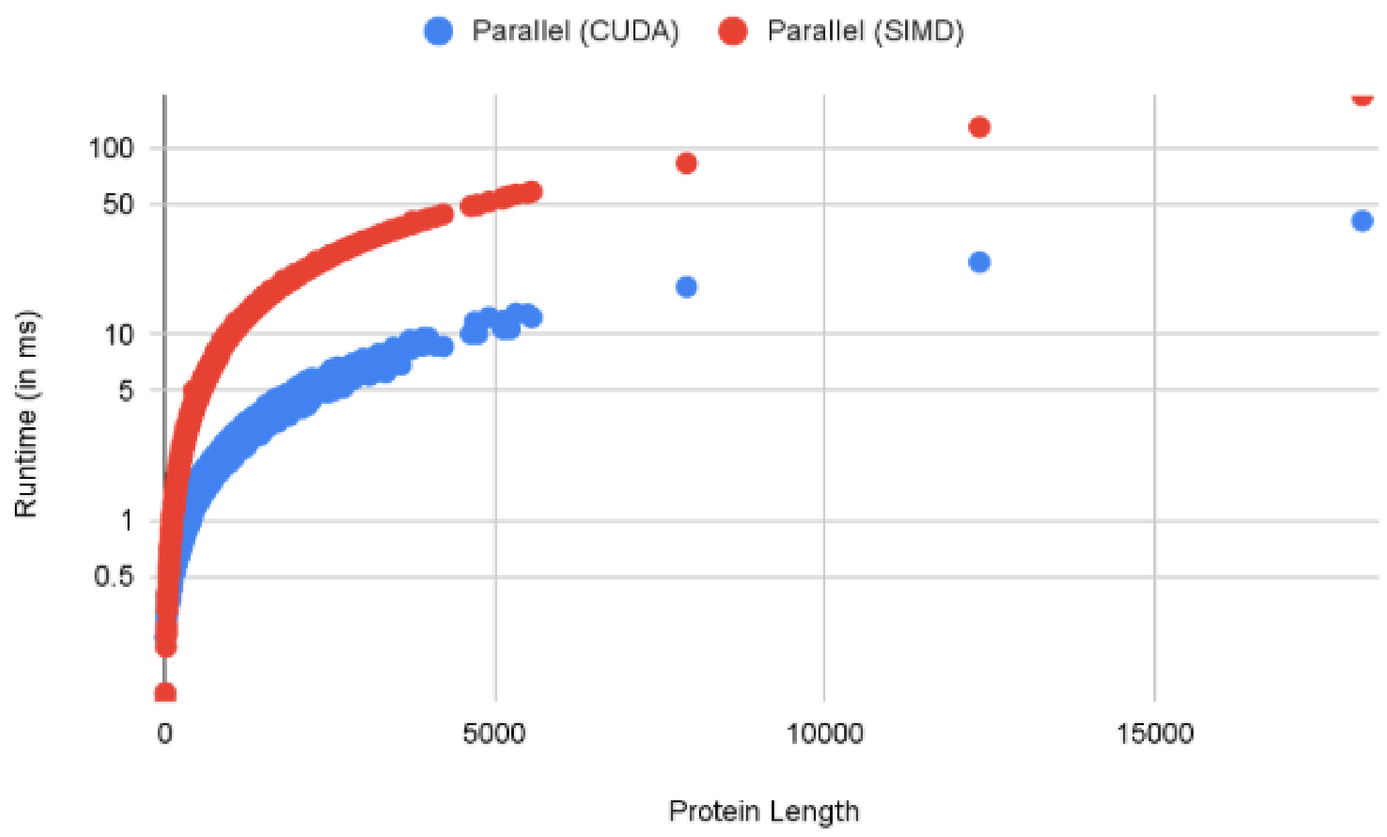


Figure 6-52: TARA CUDA Parallel Implementation vs FICS Parallel (SIMD) Implementation

- CUDA is faster than AVX2 (SIMD) by 3.14x

Tested on laptop with an NVIDIA RTX 3050 Ti Laptop GPU and Ryzen 7 5800H processor (Ox)

RUNTIME TEST: SIMD VS SIMT

	SIMD (in ms)	SIMT (in ms)	Speedup
Average	6.4093	1.7712	3.1435x
Lowest	0.1064	0.2353	0.4518x
Highest	192.3529	40.8517	5.6077x

Table 6-10: Average Runtime and Speedup for SIMD and SIMT Implementation

- **Slowdowns for shorter sequences**
- **Max Speedup of 5.61x**

Tested on laptop with an NVIDIA RTX 3050 Ti Laptop GPU and Ryzen 7 5800H processor (Ox)

RESULTS & DISCUSSION

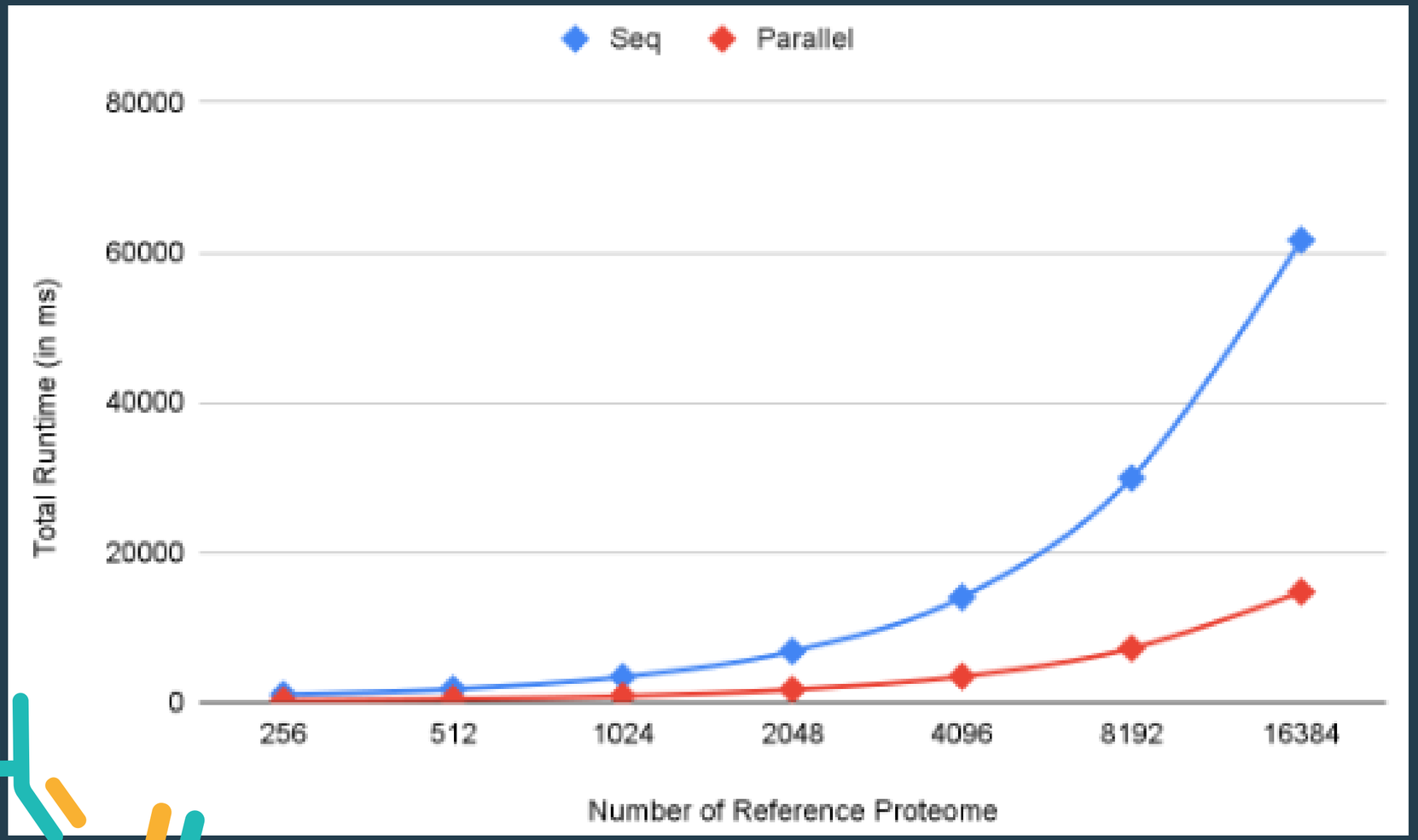
Upon completion of development, the implementation will be subjected to a series of tests. These evaluations will focus on ensuring **correctness**, assessing **runtime performance**, verifying **scalability**, and **analyzing cross-platform performance.**

1. Correctness Test
2. Runtime Test
3. Scalability Test
4. Cross-Platform Test

SCALABILITY TEST

Two scalability tests were conducted: input scalability and compiler optimization. **Input scalability** examines how the runtime grows with *increasing proteome input volume*. While **compiler optimization** measures the *impact of different compiler optimization levels* on execution speed.

SCALABILITY TEST: INPUT



- Polynomial growth
- Parallel implementation exhibits better scalability

Tested on desktop with an NVIDIA RTX 4070 Super GPU and an Intel Core i7-12700F (Ox)

SCALABILITY TEST: INPUT

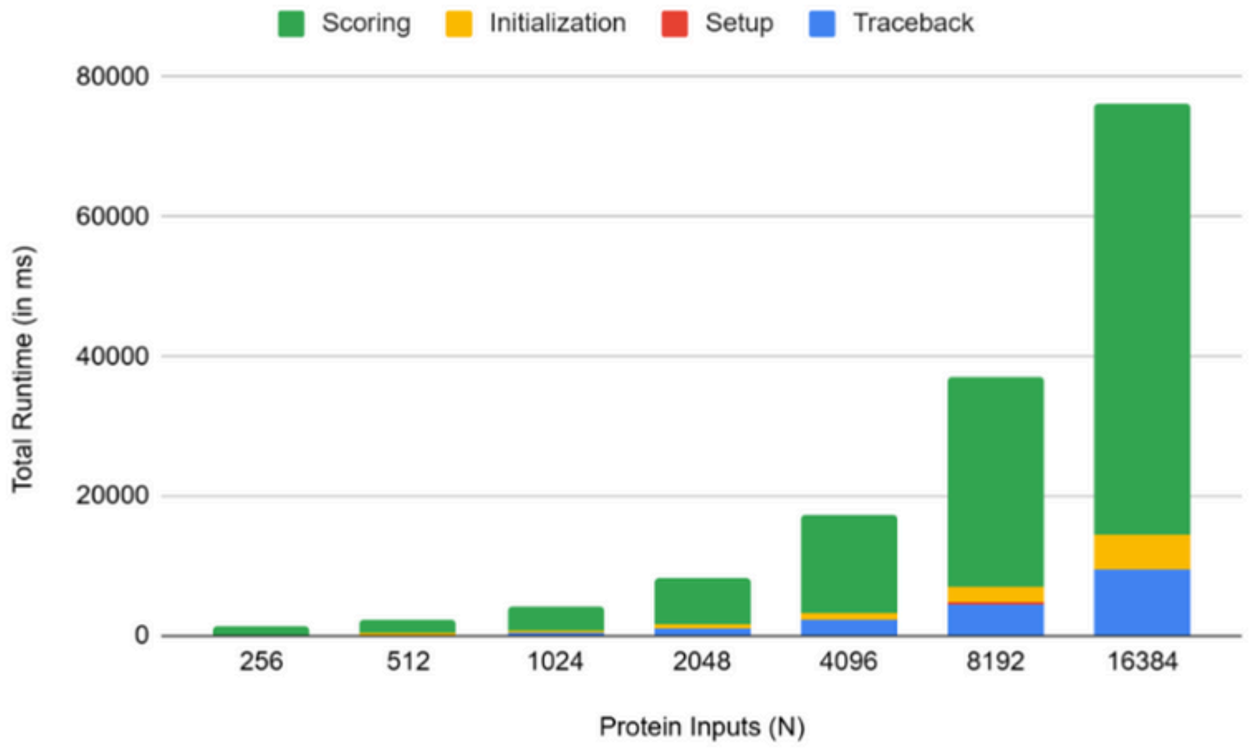


Figure 6-57: Scalability of the Runtime Contribution of Each Phase (Sequential)

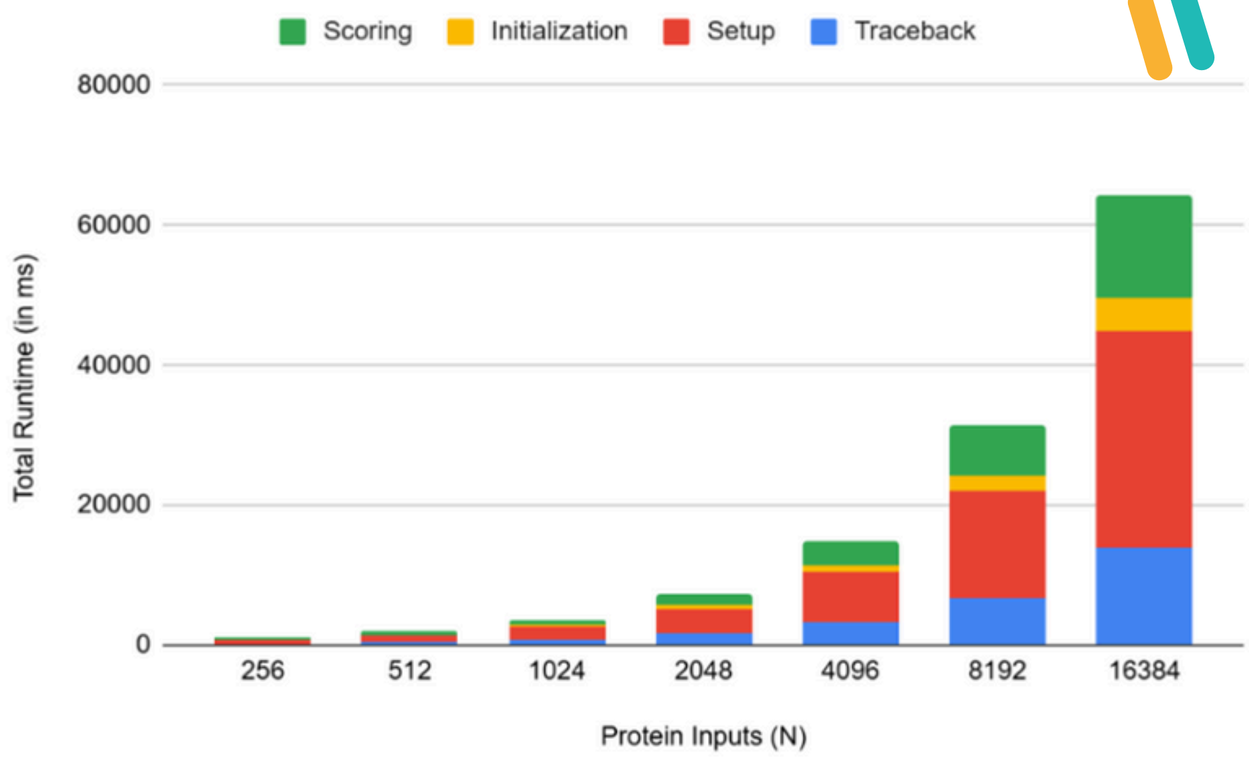


Figure 6-58: Scalability of the Runtime Contribution of Each Phase (Parallel)

Tested on desktop with an NVIDIA RTX 4070 Super GPU and an Intel Core i7-12700F (Ox)

SCALABILITY TEST: INPUT

	Setup (in ms)	Initialization (in ms)	Scoring (in ms)	Traceback (in ms)	Total
256	3.8860 (0.30%)	92.3057 (7.12%)	1041.9393 (80.42%)	157.4784 (12.15%)	1295.6095
512	6.5797 (0.30%)	148.7886 (6.84%)	1750.4340 (80.44%)	270.2063 (12.42%)	2176.0086
1024	12.1077 (0.29%)	266.9724 (6.37%)	3384.3033 (80.74%)	528.3830 (12.61%)	4191.7664
2048	23.4261 (0.28%)	513.5938 (6.11%)	6803.7110 (80.94%)	1065.6262 (12.68%)	8406.3572
4096	48.4944 (0.28%)	1082.7895 (6.26%)	13995.0016 (80.88%)	2177.4801 (12.58%)	17303.7656
8192	102.3685 (0.28%)	2372.0643 (6.41%)	29925.7021 (80.85%)	4614.8657 (12.47%)	37015.0007
16384	201.0066 (0.26%)	4836.5212 (6.34%)	61761.0711 (80.96%)	9486.8561 (12.44%)	76285.4550

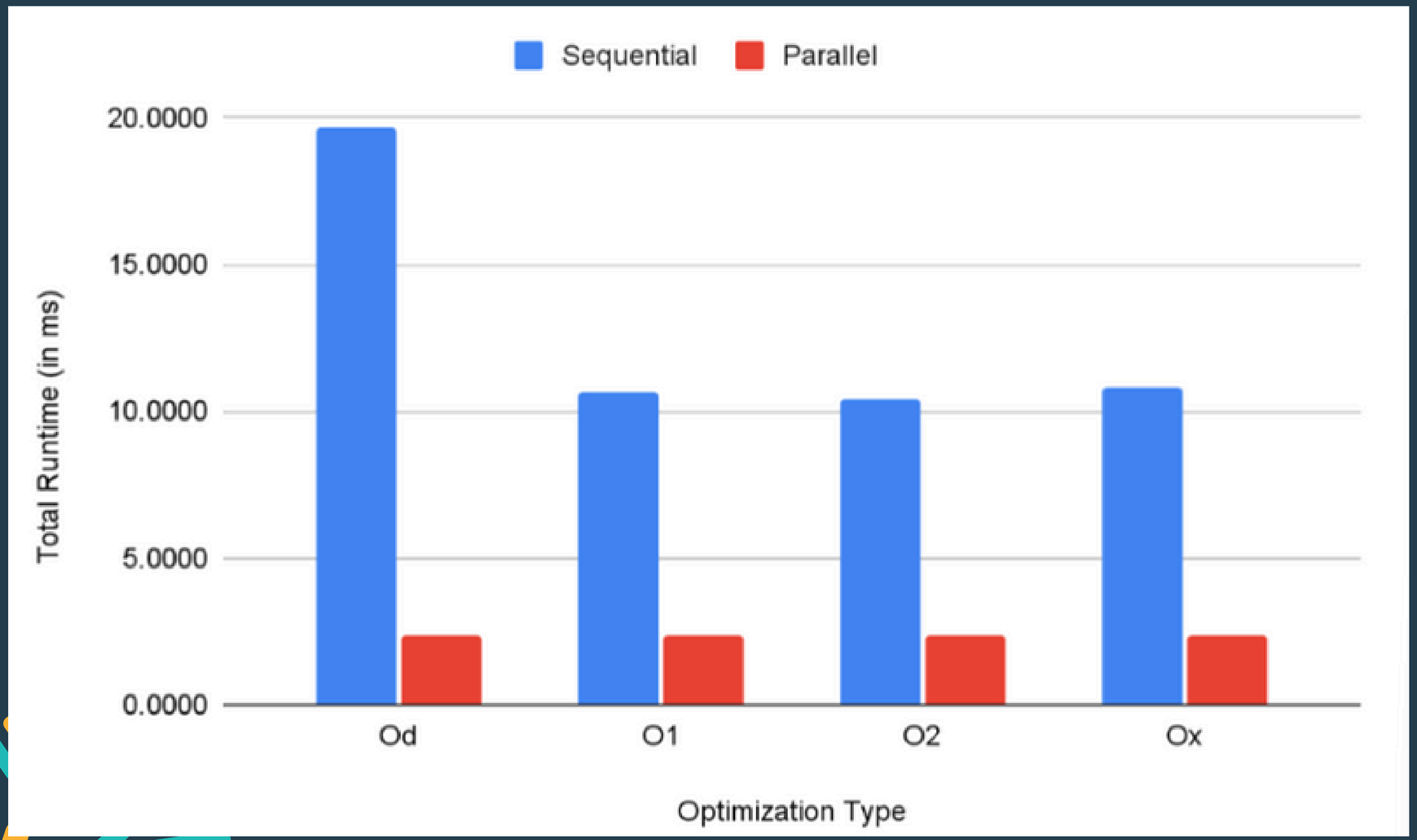
Table 6-12: Summary of Total Runtime Contribution of Each Phase (Sequential)

	Setup (in ms)	Initialization (in ms)	Scoring (in ms)	Traceback (in ms)	Total
256	508.4128 (47.34%)	92.8346 (8.64%)	242.4676 (22.58%)	230.1847 (21.43%)	1073.8998
512	910.3084 (48.11%)	150.0760 (7.93%)	428.8005 (22.66%)	402.8938 (21.29%)	1892.0788
1024	1771.5835 (48.19%)	268.3095 (7.30%)	844.4160 (22.97%)	791.5481 (21.53%)	3675.8571
2048	3541.3975 (48.20%)	509.8643 (6.94%)	1701.1427 (23.15%)	1595.3716 (21.71%)	7347.7762
4096	7211.7521 (48.09%)	1074.5773 (7.17%)	3457.5802 (23.06%)	3251.9264 (21.69%)	14995.8360
8192	15059.1698 (47.79%)	2364.3335 (7.50%)	7243.2962 (22.99%)	6844.1329 (21.72%)	31510.9325
16384	30850.5456 (47.86%)	4804.9875 (7.45%)	14812.6448 (22.98%)	13989.0137 (21.70%)	64457.1915

Table 6-13: Summary of Total Runtime Contribution of Each Phase (Parallel)

- Proportions of the program segments are consistent as dataset size grows

SCALABILITY TEST: COMPILER OPT



- Minimal impact on the CUDA implementation

Tested on desktop with an NVIDIA RTX 4070 Super GPU and an Intel Core i7-12700F

RESULTS & DISCUSSION

Upon completion of development, the implementation will be subjected to a series of tests. These evaluations will focus on ensuring **correctness**, assessing **runtime performance**, verifying **scalability**, and analyzing **cross-platform performance.**

1. Correctness Test
2. Runtime Test
3. Scalability Test
4. Cross-Platform Test

CROSS-PLATFORM PERFORMANCE TEST

- This test provides insights into how the implemented program performs **across different hardware platforms.**

- Three platforms were examined in this section, namely, the **desktop** with an Intel i7-12700F and NVIDIA RTX 4070 Super, the **laptop** with a Ryzen 7 5800H and NVIDIA RTX 3050 Ti Mobile, and the **CCSCloud** Jupyter Environment with an Intel Xeon Silver 4210 and NVIDIA Tesla V100 PCIe 32 GB.

CROSS-PLATFORM PERFORMANCE TEST

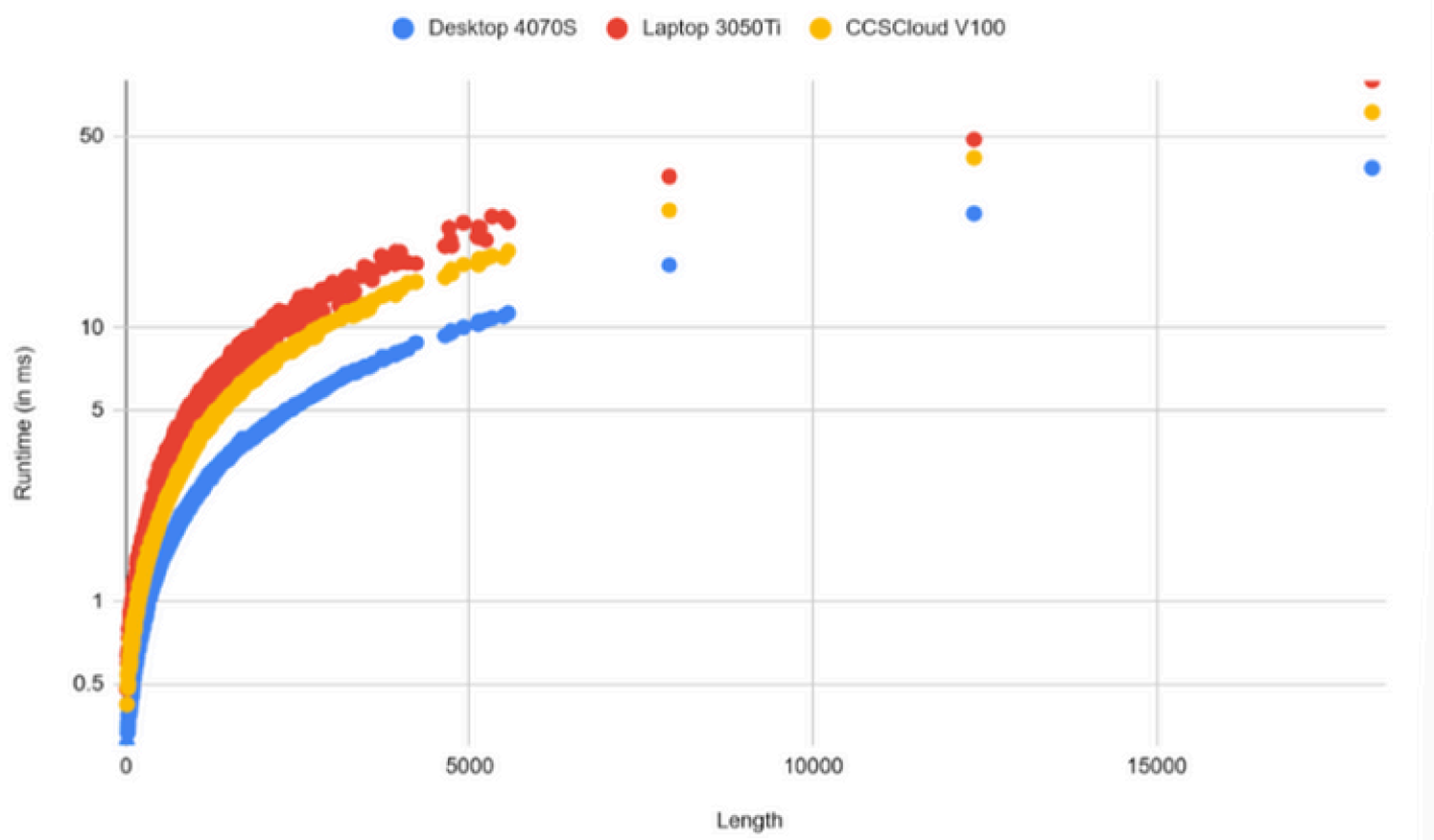


Figure 6-60: Parallel Implementation Runtime Across GPU Platforms

- Desktop provides best performance, followed by CCSCloud and then Laptop

CROSS-PLATFORM PERFORMANCE TEST

	Laptop 3050Ti	CCSCloud V100	Desktop 4070 Super
Average	1.7544x	8.5373x	3.5212x
Min	0.3225x	1.1129x	0.5468x
Max	3.2502x	12.4351x	5.9915x

	Laptop 3050Ti (in ms)		CCSCloud V100 (in ms)		Desktop 4070 Super (in ms)	
	Sequential	SIMT	Sequential	SIMT	Sequential	SIMT
Average	11.1005	5.0417	37.4361	3.7952	10.8321	2.3915
Min	0.1619	0.4788	0.4767	0.421	0.1638	0.2983
Max	209.7535	80.1586	744.9806	61.4128	204.0286	38.4254

Table 6-16: Runtime Comparison Across GPU Platforms

CROSS-PLATFORM PERFORMANCE TEST

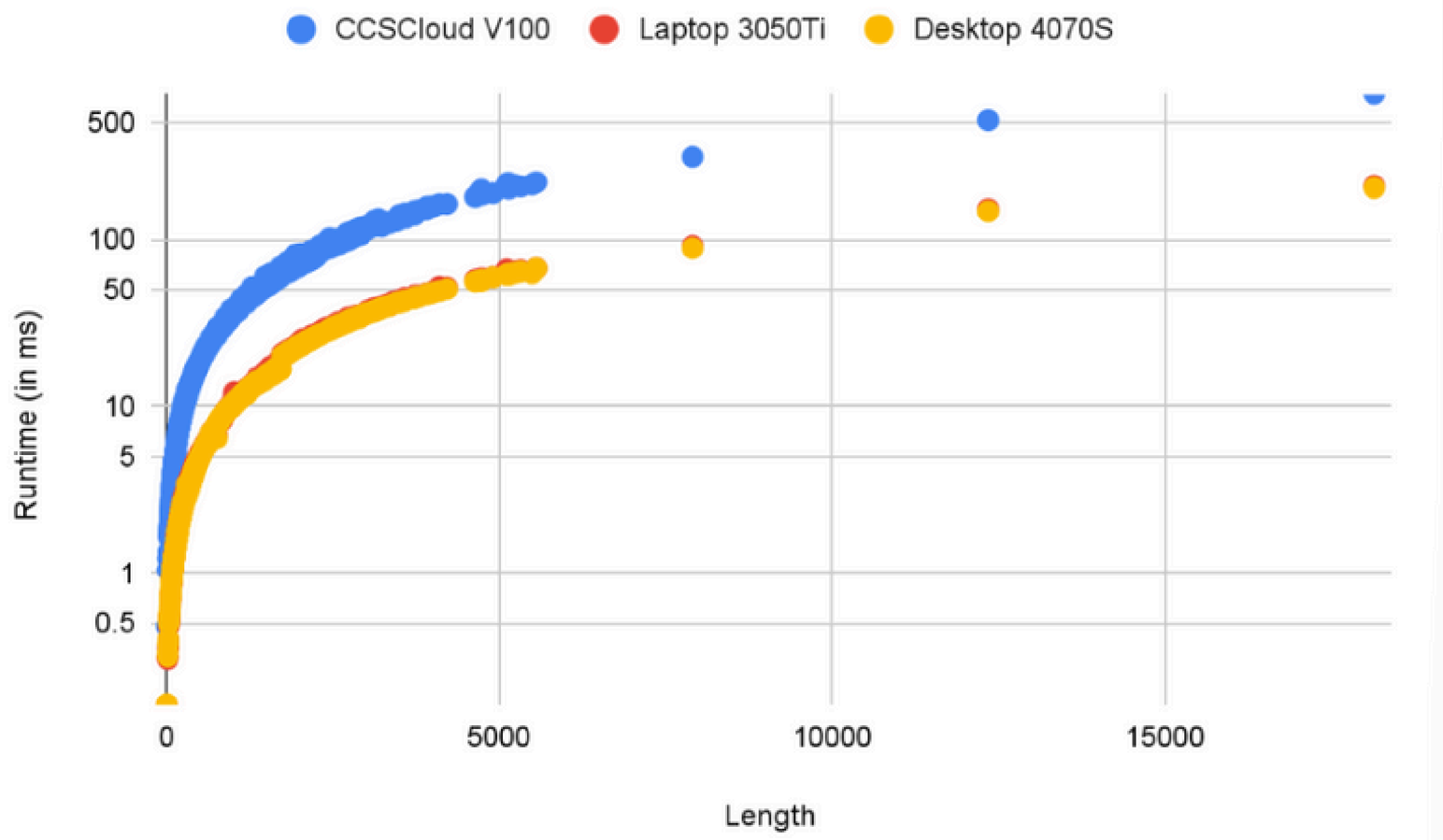


Figure 6-61: Sequential Implementation Runtime Across GPU Platforms

- 4070 Super < Tesla V100 < 3050ti Mobile

SPECIFIC OBJECTIVES

- **Developed** a sequential implementation of Zhang's Six-Frame Alignment Algorithm.
- **Created** an implementation scheme for Zhang's Six-Frame Alignment Algorithm to better fit the parallel framework of the SIMT model.
- **Implemented** Zhang's Six-Frame Alignment Algorithm with the SIMT execution model using CUDA.
- **Evaluated** the developed parallel implementation by comparing and contrasting its performance to the sequential version and existing SIMD implementation.

CONCLUSION & RECOMMENDATIONS

CONCLUSION

- **Demonstrated SIMT Acceleration:** Utilized the CUDA SIMT execution model to reduce computational costs and accelerate Zhang's Six-Frame Algorithm.
- **Parallelization:** Implemented a hybrid multi-level anti-diagonal wavefront technique, adapted for the SIMT model, proving the viability of this paradigm in CUDA.
- **Significant Speedups:** Achieved ~3.5x speedup over Sequential implementation. Also achieved ~3.1x speedup with SIMD (AVX2).

CONCLUSION

- **Hardware Impact:** Showcased the impact of GPU hardware quality on CUDA code performance.
- **Compiler Impact:** Compiler optimizations had minimal effect on CUDA performance (due to GPU-specific compilation and architecture factors), but showed considerable improvement for the Sequential CPU implementation.
- **Proof of Concept:** Validates the ability of GPUs for complex bioinformatics algorithms, enabling faster data processing and downstream genetic analysis.

RECOMMENDATIONS

- **Implementation Optimization:** Use of a purely device implementation scheme can potentially enable better performance.
- **Holistic Optimization:** Exploring speedups in the other phases of the program, including performing initialization using device code, as well as in the setup phase, particularly with regard to data transfers between the device and the host.

RECOMMENDATIONS

- **Comprehensive Hardware Testing:** Some improvements to the testing methodology can also be made, particularly with regard to testing performance across multiple devices.

- **Utilize CUDA C/C++ Libraries:** To explore the use of libraries of CUDA C/C++ that can lessen the overhead/computation cost.

DEMO

TARA CUDA: A HYBRID SIMT IMPLEMENTATION OF ZHANG'S SIX-FRAME ALGORITHM USING CUDA

ARCETA - MENDOZA - TAN

ADVISER: MR. ROGER LUIS UY