Comparison between Double Stranded DNA with Restriction Enzymes and Single Stranded DNA with Primers for Solving Boolean Matrix Multiplication

Nordiana Rajaee[1], Awang Ahmad Sallehin Awang Hussaini[2] and Azham Zulkharnain[2]
[1]*Faculty of Engineering, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia*.
[2]*Faculty of Resource Science and Technology, 94300 Kota Samarahan, Sarawak, Malaysia*.
rnordiana@unimas.my

***Abstract*— Boolean matrix multiplication is the basis for most computing algorithms and is widely used in many fields. In this paper, we compare and discuss two methods to solve Boolean matrix multiplication with DNA computing. The first method utilizes double stranded DNA sequences with Restriction Enzymes meanwhile the second method utilizes single stranded DNA sequences with primers. We prove that while both methods are able to solve the Boolean matrix multiplication problem, these two methods differ in their performance and output results. We compare the advantages of the latter method in terms of easier sequence designs and more efficient analysis of results.**

***Index Terms*—DNA Computing; Bio-Inspired Computing; Boolean Matrix Multiplication; Graph Problem.**

I. INTRODUCTION

Matrix multiplication problem is a basic problem which is heavily utilized in areas such as signal processing, graph theory, parallel computing and digital control. In matrix multiplication problems, the zero-one data in matrices are widely applied in network fault identification, social networks, data mining, knowledge management and clustering analysis [1]. In computing, a number of approaches are proposed to implement matrix multiplication in parallel systems for higher performance. However, the current sequential computers have their limits in providing high performance computing due to the bottlenecking factor of Moore's Law. As an alternative to sequential computers, deoxyribonucleic acid (DNA) computing has emerged as a new computation medium. DNA computing attracts researches from various fields due to its massive parallel processing capabilities. The DNA computer is estimated to possess 104 faster computing prowess than the speed of supercomputers in a mix of mix of 1018 DNA strands [2]. Although the idea of using nature for computation is not exactly new, physical implementation of DNA computation was only proven by L. M Adleman in 1994 when he solved a seven-node Hamiltonian Path Problem. The computation was carried out by encoding the problem in double stranded DNA molecules and using bio-molecular tools to execute the computation [3]. Based on Adleman's architecture, many other proposals to compute Boolean operations using DNA have been published including a proposal to compute Boolean matrix multiplication by John S. Oliver in 1998, which also proposed using double stranded DNA oligonucleotides and restriction enzymes for executing the DNA computation [4]. In this paper, we compare two approaches to solve Boolean matrix multiplication with DNA computing. The first method utilizes double stranded DNA sequences with Restriction Enzymes and the second method utilizes single stranded DNA sequences and primers. We aim to compare these two methods and study the differences in the extraction outputs when using Restriction Enzymes as cutting reaction in DNA computing and primers as indicators for row and column in matrices.

II. BOOLEAN MATRIX MULTIPLICATION

Two matrices X and Y and their product matrix can be translated into a directed graph whereby the first column indicators in the first matrix X are denoted as initial vertices, the second column indicators in the matrix X and first column indicators in the matrix Y are the intermediate vertices, while the second column indicators in the matrix Y denoted as the terminal vertices. In such a way, the product matrix column indicators are consists of initial and terminal vertices and for each value of 1 indicators, an edge is created to link a vertex to the corresponding vertex, while no edge is created for the value 0. The values of the elements in the product matrix are later determined by the existence of path from an initial vertex to a terminal vertex by the connection of edges via the intermediate vertices. An example of such Boolean matrix multiplication is shown as in Figure 1 consisting of two matrices X and Y and their product matrix Z in (a) and its directed graph representation of initial vertices, intermediate vertices, terminal vertices and corresponding edges in (b) [4].

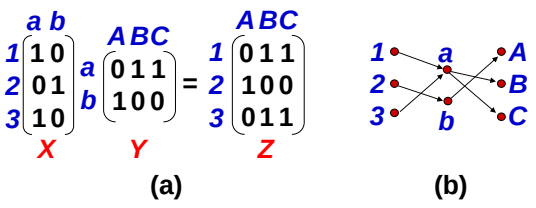


Figure 1: Boolean matrix multiplication and its directed graph representation

III. DNA COMPUTING

Deoxyribonucleic acid (DNA) computing requires the biomolecular tools to execute and extract the results of DNA computation. Basic biomolecular tools such as hybridization-ligation, polymerase chain reaction, parallel overlap assembly and gel electrophoresis process are explained briefly in the following section.

Hybridization is a process where single stranded DNA sequences anneal to its reverse complementary strands in order to form a double stranded DNA. In the formation of initial paths, this step is most critical to ensure correct annealing of sequences for the extraction phase in gel electrophoresis.

Ligation is a process where ligase is used to seal or glue covalent bonds in DNA sequences after the single stranded DNA annealed to each other.

Denaturation is a melting process in vitro. By using heat to separate the double stranded DNA sequences into two single stranded DNA sequences, the sequence strands can be preserved due to much weaker hydrogen bonds between the molecules.

Cutting process is carried out by using Restriction Enzymes. Restriction Enzymes recognize a specific sequence of DNA known as a restriction site. Any DNA that contains the restriction site within its sequence is cut by the Restriction Enzyme at that point.

Parallel Overlap Assembly (POA) is a process where DNA strands corresponding to the same position string are overlapped during annealing step in the assembly process while the remaining parts of the DNA strands are extended by polymerase. During each cycle in POA, the DNA strands self-assemble and extend/elongate as the denaturation and annealing processes are repeated causing the number of target strands decreasing while the lengths of the newly formed strands increasing.

Polymerase Chain Reaction (PCR) is process whereby a specific sequence of DNA strand is amplified exponentially. DNA sequences known as primers are utilized to signal start and end points of amplification.

Gel electrophoresis is a technique used for separation of DNA strands according to their sizes. The size of the DNA strands refers to the weight of the DNA strands which is proportional to the lengths of their sequences. The results of gel electrophoresis process can be viewed by staining gel with fluorescent dye and photographed under UV light [5].

IV. DOUBLE STRANDED DNA AND RESTRICTION ENZYMES (METHOD I)

We consider computing the Boolean matrix multiplication problem in Figure 1 where the initial and terminal vertices are embedded with Restriction Enzymes sites. Table 1 shows "cutting" sites for each Restriction Enzyme type and Table 2 shows the DNA sequences for the encoded Initial/ Terminal Vertices and their complements to form double strands.

Table 1
Restriction Enzymes Sites encoded in Initial and Terminal Vertex Sequences

Vertex	Restriction Enzymes	Cutting Site
V1	EcoRV	GAT¦ATC CTA¦TAG
V2	EcoRI	G¦AATTC CTTAA¦G
V3	SmaI	CCC¦GGG GGG¦CCC
VA	BamHI	G¦GATCC CCTAG¦G
VB	DraI	TTT¦AAA AAA¦TTT
VC	HindIII	A¦AGCTT TTCGA¦A

Table 2
DNA Sequences for Initial and Terminal Vertices and their reverse complements

Vertex	DNA Sequence (5' – 3')	Length (mer)
V1 [V1]	GAT↓ATCtagcacacgaaccc gggttcgtgtgctaGAT↓ATC	20
V2 [V2]	G↓AATTCgtgagggaggagtg cactcctccctcacG↓AATTC	20
V3 [V3]	CCC↓GGGtcagcgtctttcaa ttgaaagacgctgaCCC↓GGG	20
VA [VA]	agggttgctcttgtG↓GATCC G↓GATCCacaagagcaaccct	20
VB [VB]	caggctaagggtgaTTT↓AAA TTT↓AAAtcacccttagcctg	20
VC [VC]	actcattcggcggcA↓AGCTT A↓AGCTTgccgccgaatgagt	20

For each element of value 1 in X, a single stranded intermediate vertex sequence extends as an overhang to the respective initial vertex sequence. For each element of value 1 in Y, a complementary single stranded intermediate vertex sequence extends as an overhang to the respective terminal vertex sequence. DNA sequences for intermediate vertices and their complements are shown in Table 3.

Table 3
DNA Sequences for Intermediate Vertices and their reverse complements

Vertex	DNA Sequence (5' – 3')	Length (mer)
Va [Va]	ttttcgtctgagtgtttcgc gcgaaacactcagacgaaaa	20
Vb [Vb]	tgttccatttgattgcgtcc ggacgcaatcaaatggaaca	20
Vc [Vc]	ctacatctcgtgcccgttca tgaacgggcacgagatgtag	20

A "path" is constructed when a single stranded intermediate vertex (attached to an initial vertex sequence) anneals to its complementary strand (attached to a terminal vertex sequence). Figure 2 shows formation of "path" from V1 to VB and from V1 to VC.

V. Single Stranded DNA and Primers (Method II)

For the second method, random single stranded DNA sequences are generated to represent all vertices in the graph. Table 4 shows unique single stranded DNA sequences generated to represent all initial, intermediate and terminal Vertices.

Table 4
DNA Sequences for all Vertices (Single Stranded)

Vertex	DNA Sequence (5'-3')	Length (mer)
V1	ctcttttagcacacgaaccc	20
V2	cagtttgtgagggaggagtg	20
V3	acatcctcagcgtctttcaa	20
Va	ttttcgtctgagtgtttcgc	20
Vb	tgttccatttgattgcgtcc	20
VA	agggttgctcttgtctcgga	20
VB	gtaggatttagcctcaaagtcttagaagcg	30
VC	gaaggagccgtaagcaaatctccaagaactctgtaaggta	40

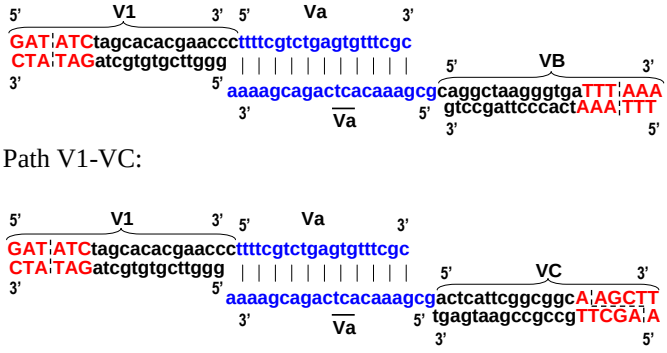
Directed edges are constructed for all elements of value 1 in the matrices X and Y for the respective intersecting vertices. The directed edges are constructed as connectors between the vertex sequences and anneal partially to both vertex sequences. The directed edges sequences are as shown in Table 5.

Table 5
DNA Sequences for all Directed Edges

Edges	DNA Sequence (5'-3')	Length (mer)
E1a	acacgaacccttttcgtctg	20
E2b	gggaggagtgtgttccattt	20
E3a	cgtctttcaattttcgtctg	20
EaB	agtgtttcgcgtaggattta	20
EbA	agggttgctcagggttgctc	20
EaC	agtgtttcgcgaaggagccg	20

A "path" is constructed when an initial vertex sequence is linked to a terminal vertex sequence by intermediate vertex and directed edges sequences. Figure 3 shows formation of "path" from V1 to VB and from V1 to VC.

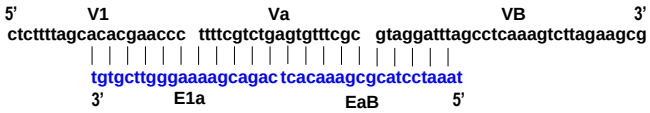
Path V1-VB:



Path V1-VC:

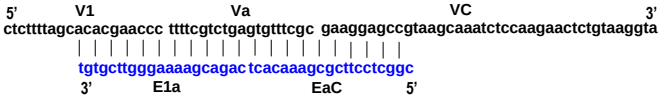
Figure 2: Formation of "path" for double stranded DNA with Restriction Enzymes

Path V1-VB:



(a)

Path V1-VC:



(b)

Figure 3: Formation of "path" for single stranded DNA with primers

VI. Results and Discussions

The outputs of the computation are determined by read-out analysis in the gel electrophoresis process. The major difference between both approaches is the mechanism for detecting a "path" which is equivalent to a value 1 in the product matrix. Using Restriction Enzymes, a value 1 is denoted by a double "cutting" reaction corresponding to "cutting" sites at both ends of a "path". Meanwhile, primers are used to directly detect a "path" containing sequences from an initial vertex to a terminal vertex.

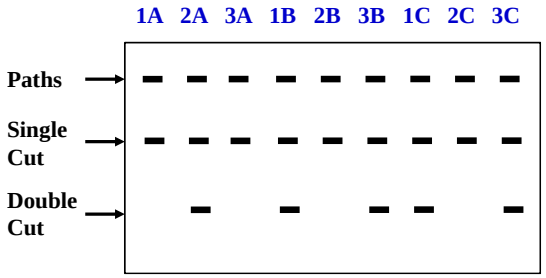


Figure 4: Gel electrophoresis results for double stranded DNA and Restriction Enzymes

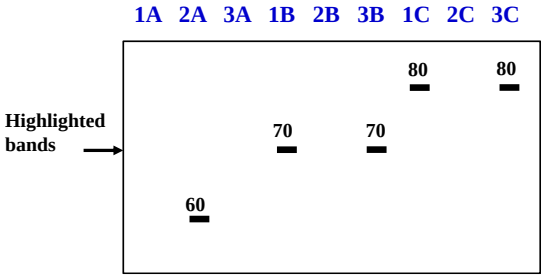


Figure 5: Gel electrophoresis results for single stranded DNA and primers

A constructed "path" with Restriction Enzymes sites at both ends is truncated. Different types of Restriction Enzymes produce either "blunt" ends or "sticky" ends after the "cutting" reactions, causing irregularity in the remaining length of "path", as shown in Figure 6(a) and Figure 6(b).

Figure 6(a): "blunt" ends

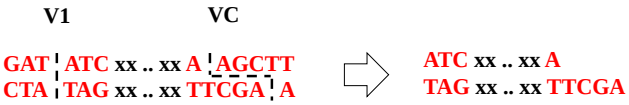


Figure 6(b): "sticky" ends

Constructed "paths" with single stranded DNA and primers are direct proportional to the total sum of vertex sequences length in the "path". Results of the computation not only verify the binary values of product matrix but can also be used to classify the results accordingly to their vertices.

Path V1 – VB :
V1 (20) + Va (20) + VB (30) = 70 base pair
Path V1 – VC :
V1 (20) + Va (20) + VC (40) = 80 base pair

VII. CONCLUSION

We presented two approaches to compute Boolean matrix multiplication problem with DNA computing and compare the efficiency and performance for both approaches. From the results, the first method which utilizes double stranded DNA sequences and restriction enzymes, requires restriction sites to be embedded into the DNA strands. This increases the complexity of the designing the DNA sequences for encoding the problem. Meanwhile, the second approach which utilizes single stranded DNA sequences and primers eliminates the requirement to embed restriction sites as direct encoding of the problem is carried out with primers. This also allows the length of the paths to be directly proportional to sum of components (vertices and edges) which gives the advantage to the second approach by having easier to be analysed gel electrophoresis results.

ACKNOWLEDGMENT

We are grateful for the Universiti Malaysia Sarawak for Small Grant Scheme F02(S153)/1154/2014(18).

REFERENCES

[1] R. Agrawal, F. Gustavson, M. Zubair, A high performance matrix multiplication algorithm on distributed memory parallel computer using overlapped communication, IBM Journal of Research and Development, 38, 6, 673-681 (1994)
[2] J. Watada, RA. Bakar, DNA computing and its applications, Proc 8th International Conference on Intelligent Systems Design and Applications, 288 – 294 (2008)
[3] L. Adleman, Molecular computation of solutions to combinatorial problem, Science, 266, 1021 – 1024 (1994)
[4] JS. Oliver, Matrix multiplication with DNA. J Mol. Evo 45, 161 – 167 (1997)
[5] M. Amos, Theoretical and Experimental DNA Computation, Natural Computing Series. Springer (2004)
[6] N. Rajaee, Y. Kon, K. Yabe, O. Ono, Matrix multiplication with DNA based computing: A comparison study between hybridization-ligation and parallel overlap assembly, Proc of the 4th IEEE International Conference on Natural Computation, 527 – 530 (2008)
[7] Z. Ibrahim, Y. Tsuboi, O. Ono, Hybridization-Ligation vs Parallel Overlap Assembly: An experimental comparison of initial pool generation for direct proportional length based DNA computing, IEEE Transactions on Nanobioscience, 5, 103 – 109 (2006)
[8] NK Khalid, Z. Ibrahim, TB Kurniawan, M. Khalid, NH Sarmin, Function minimization in DNA sequence design based on continuous particle swarm optimization, ICIC Express Letters, 3, 1, 27 – 32 (2009)
[9] CC. Chang, T. Lu, YF. Chang, CT. Lee, Reversible data hiding schemes for deoxyribonucleic acid (DNA) medium, International Journal of Innovative Computing, Information and Control, 3, 5, 1145 – 1160 (2007)