Constructing Boolean Matrices with Restriction Enzymes as Row/Column Indicators in DNA Computing

Nordiana Rajaee, Kismet Hong Ping, Asrani Lit, Dyg Nur Salmi Dharmiza Awg Salleh, and Liang Yew Ng

***Abstract*—The massive parallel computing capabilities of DNA computing is extremely attractive for solving matrices problems with large data sets such as in network fault monitoring, data mining and social studies. However, the design and strategy to encode these problems into DNA sequences for computation still proves a hurdle as DNA computing is a wet-lab process highly dependent on bio-molecular tools. Various designs, strategy and tools give different advantages, limitations and affects the performance of the computation including extractions of results. In this paper, we study the utilization of restriction enzymes as row/column indicators for matrices problem with DNA computing. We model and propose steps to compute Boolean matrices and discuss highlights, drawbacks and applicability of the restriction enzymes during the encoding of the problems in DNA computing. In general, the utilization of restriction enzymes enables information to be tagged directly to the DNA sequences. However, due to the cutting reactions, the formed path cannot be utilized as weight or distance.**

***Index Terms*—DNA computing, evolutionary computing, restriction enzymes, Boolean matrices.**

I. Introduction

Ever since the groundbreaking finding by Leonard M Adleman to execute computations using DNA, DNA computing has emerged as an alternative computation medium which attracts researches in various fields due to its properties. Its high density, massive parallel computing capabilities and low energy dispersion gives hope for its potential to rival silicon computers. However, much of DNA computing relies on developing algorithms that solve problems using the encoded information in the sequence of nucleotides that make up DNA double helix and then breaking and making new bonds between them to reach the answer [1]. The DNA computing itself is a wet-lab process dependent on bio-molecular tools such as hybridization, ligation, amplification by polymerase chain reaction method and extraction using gel electrophoresis process. Spurred by Adleman's works, subsequent proposals to implement DNA computation for basic computing operations were presented. Ogihara and Ray (1997) proposed to simulate Boolean circuits with DNA [2]. Leete *et al.* (1997), Gupta *et al.* (1997) and Orlian *et al.* (1998) proposed methods for basic operations such as arithmetic addition and subtraction by chaining the output into inputs to supplementary operations [3]-[5]. Rubin *et al.* (1998) presented a demonstration of a computation method for chained integer arithmetic and Guarneri and Bancroft (1999) developed DNA based addition algorithm employing successive primer reactions to implement carries and the Boolean logic required in binary addition [6], [7]. Ibrahim *et al.* (2004) proposed DNA computing for solving Shortest Path Problem while Bakar *et al.* (2008) proposed for clustering problems [8], [9]. DNA has also been used to design sequential circuits as proposed by Song *et al.* (2008) [10]. More diverse applications of DNA computing are proposed in the field of cryptography by Wang *et al.* (2009), in information security field by Cui *et al.* (2009) and biosensors by Wang *et al.* (2012) [11]-[13].

In this paper we discuss to compute a Boolean matrix multiplication problem with DNA computing. Although the Boolean matrix multiplication itself is a very simple problem, our focus lies in the design and strategies to model and compute the problem. Boolean matrices are widely applicable in computational architecture, digital signal processing, fault analysis, data mining, scheduling and clustering problems. The intuition behind using Boolean matrices is that sometimes the counts of the objects do not matter. Data in the form of 0's and 1's are highly interpretable in many systems, even in complex computation [14]. In our work, we focus on the utilization of restriction enzymes as row/column indicators for Boolean matrix multiplications. We compare their limitations and advantages and their possible applicability in solving more complex zero-one matrix problems.

II. DNA Computing

A DNA molecule is a long string composed of two strands wound around each other to form a double helix. There are four types of organic bases: adenine (A), cytosine (C), guanine (G) and thymine (T). A short single stranded DNA chain, usually less than 30 nucleotides long is called an oligonucleotide. The ends of a DNA strand are chemically polar, with the so called 5' end and the 3' end. Each base has a bonding surface, where the bonding surface of A is complementary to that of T, and that of G is complementary to that of C. This complementary rule is called Watson-Crick complementary. A single DNA strand can pair with another strand when their sequences of bases are mutually complementary and the chains have opposite polarity [15].

DNA strands are often quoted in 5' – 3' order and length of a DNA strand is denoted in mer, where one mer represents one DNA oligonucleotides. Two single stranded DNA under

Manuscript received March 12, 2014; revised June 26, 2014. This work was supported in part by Universiti Malaysia Sarawak.

N. Rajaee, K. Hong Ping, A. Lit, D. N. S. A. Salleh, N. Liang Yew are with Faculty of Engineering, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia (e-mail: {rnordiana, hpkismet, lasrani, asdnsdharmiza, ngliangy}@feng.unimas.my).

certain conditions form a double stranded DNA. The length of a double stranded DNA is denoted in base pairs (b.p.). Fig. 1 shows two 10 mer single stranded DNA molecules combining to form a 10 b.p. double stranded DNA [15].

```
5'                                  3'
  G – G – A – T – A – G – C – T – G – G
  |   |   |   |   |   |   |   |   |   |
3' C – C – T – A – T – C – G – A – C – C  5'
```

Fig. 1. Two single stranded DNA forming a double stranded DNA.

Bio-Molecular Tools

Hybridization is the annealing of complementary single stranded molecules to form a double stranded DNA. This is the basis for initial path formation during the reaction step and is subsequently employed during the extraction phase for the isolation of generated path molecules.

Ligation is a process often invoked after single stranded DNA are annealed and concatenated to each other. Many single stranded fragments are connected in series and ligase is used as "glue" to seal the covalent bonds between the adjacent fragments.

*Denaturatio*n is a melting process in vitro. Double stranded DNA molecules can be separated without breaking the single strands by applying heat to the solution. The double stranded molecules come apart because the hydrogen bonds between complementary nucleotides are much weaker than the covalent bond between the adjacent nucleotides in the same strands.

Cutting process is carried out by using Restriction Enzymes (RE). 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 enzyme at that point.

Parallel Overlap Assembly (POA) method was successfully applied by Kaplan *et al.* for initial pool generation consisting of binary numbers to solve a maximal clique problem with DNA computing. The initial pool is a combinatorial library containing numerical or indicative information represented by DNA sequences. Construction of computational DNA libraries is based on a DNA shuffling method consisting of two parts; one is the position string of fixed length and the other is value string (0 or 1) of various lengths. The 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 dNTPs incorporation 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 an amplification technique widely used in molecular biology. A pair of DNA sequences known as "primers" is used to signal the starting point and ending point for a specific target DNA sequence for amplification. The PCR process is capable of exponentially amplify a DNA strand into millions of its copies given a site-specific single molecule DNA and the process is usually carried out in three stages of different temperatures.

Gel electrophoresis is a technique used for separation of DNA strands according to their sizes using electric current applied to the gel containing the strands. The size of the DNA strands refers to the weight of the DNA strands which is proportional to the lengths of their sequences. This technique is based on the fact that DNA molecules are negatively charged. Since DNA molecules have the same charge per unit length, they all migrate at the same speed in aqueous solution. However, if electrophoresis is carried out in gel, the migration rate is affected by its size causing less weighted strands to migrate faster. Thus, sorting the strands by their sequence lengths is made possible using this technique. The results of gel electrophoresis process can be viewed by staining gel with fluorescent dye and photographed under UV light [15].

III. Modeling Boolean Matrix with DNA Computing

Kim (1982) noted that a Boolean matrix multiplication problem can be represented by a graph problem [16]. By representing the row/column identifiers for the matrices as vertices in a graph problem, the value of the elements in the Boolean matrices can be represented by directed edges between the vertices. In this case, the directed edges are only drawn for elements of value 1 ("no edge" is equivalent to element of value 0) [16].

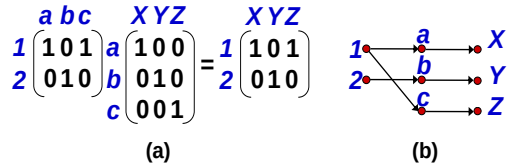

Fig. 2. Two single stranded DNA forming a double stranded DNA.

An example of two matrices and their multiplication product represented by a graph problem is shown in Fig. 2. By representing the row indicators for the first matrix as initial vertices and the column indicators for the first matrix as intermediate vertices, the elements of value 1 in the first matrix are represented by directed edges from the initial vertices to the intermediate vertices. Similarly, by representing the row indicators for the second matrix as intermediate vertices and the column indicators for the second matrix as terminal vertices, the elements of value 1 in the second matrix are represented by directed edges from intermediate vertices to terminal vertices (matrix multiplication rule whereby the column indicators for the first matrix must be equivalent to the row indicators for the second matrix).

To model the Boolean matrix multiplication problem with DNA, strands of oligonucleotides to represent all the vertices and directed edges for the graph problem are constructed. The construction of the vertices and directed edges are based on the idea that individual strands will hybridize to form a longer strand where the solutions to the problem exist in the form of elongated strands containing all oligonucleotides from an initial vertex to a terminal vertex. Solution strands are also known as "paths". Extraction of "paths" depends on the designs of the DNA sequences embedding the problem, and the bio-molecular tools to identify the extracted "path".

In this paper, we focus on the use of restriction enzymes as row / column indicators for the Boolean matrices and highlight their advantages and drawbacks in the DNA sequence designs.

Design and Strategy: Using Restriction Enzymes as row/column indicators for solving a Boolean matrix multiplication problem was proposed by John S. Oliver. All DNA strands representing Initial Vertices and Terminal Vertices are constructed from double stranded DNA with restriction enzyme "cutting" sites encoded in their sequences. For Initial Vertices strands, the "cutting" sites are encoded in the beginning of the sequences (5'-3') while for Terminal Vertices strands, the "cutting" sites are encoded at the end of the sequences (5'-3'). In this case, the row/column indicators for the product matrix are represented by restriction enzyme sites for Initial/Terminal Vertices. Intermediate Vertices are constructed from single stranded DNA which extend as overhangs at the other end (non-restriction-enzyme-site) of the double stranded DNA. For all elements of value 1 in the first matrix, the single strands are attached to their corresponding Initial Vertices double strands. For all elements of value 1 in the second matrix, the complements of the single strands are attached to their corresponding Terminal Vertices double strands. No extended overhang is generated for elements of value 0.

Coding: Consider the Boolean matrix multiplication problem in Fig. 1. We encode the Initial/Terminal Vertices double strands in Table II with restriction enzymes "cutting" sites as shown in Table I. Single strands and their complementary for Intermediate Vertices are shown as in Table III.

TABLE I: RESTRICTION ENZYMES FOR VERTICES

Vertex	Restriction Enzyme	Cutting Site
V1	EcoRV	GAT ¦ ATC CTA ¦ TAG
V2	EcoRI	G ¦ AATTC CTTAA ¦ G
VX	BamHI	G ¦ GATCC CCTAG ¦ G
VY	HindIII	A ¦ AGCTT TTCGA ¦ A
VZ	SmaI	CCC ¦ GGG GGG ¦ CCC

TABLE II: DNA SEQUENCES WITH EMBEDDED RESTRICTION ENZYMES SITES FOR INITIAL/TERMINAL VERTICES

Vertex	DNA Sequences (5'-3')	Length
V1	GAT ↓ ATCtagcacacgaaccc gggttcgtgtgctaGAT ↓ ATC	20 b.p
V2	G ↓ AATTCgtgagggaggagtg cactcctccctcacG ↓ AATTC	20 b.p
VX	agggttgctcttgtG ↓ GATCC G ↓ GATCCacaagagcaaccct	20 b.p
VY	ctgttactcattcggcggcA ↓ AGCTT A ↓ AGCTTgccgccgaatgagtaacag	25 b.p
VZ	gtaggcgtatttgaaagacgctgaCCC ↓ GGG CCC ↓ GGGtcagcgtctttcaaatacgcctac	30 b.p

Formation of Path: All constructed DNA strands representing the Initial, Intermediate and Terminal Vertices are poured into a single solution. An initial pool containing all possible solution is generated where the constructed DNA strands hybridize to form "paths". A "path" is formed when the single stranded Intermediate Vertices (attached to the Initial Vertices) hybridize with their complementary single strands (attached to the Terminal Vertices). Thus, a "path" consists of elongated strands containing Initial, Intermediate and Terminal Vertex sequences with two restriction enzyme sites at both ends. Fig. 3 shows the formation of "paths" for V1 – VX, V2 – VY and V1 – VZ.

TABLE III: DNA SEQUENCES FOR INTERMEDIATE VERTICES

Vertex	DNA Sequences (5'-3')	Length
Va	ttttcgtctgagtgtttcgc gcgaaacactcagacgaaaa	20 mer 20 mer
Vb	tgttccatttgattgcgtccagcta tagctggacgcaatcaaatggaaca	20 mer 20 mer
Vc	agggttgctcttgtctcggattaccgtacg cgtacggtaaaggctctgttctcgttggga	20 mer 20 mer

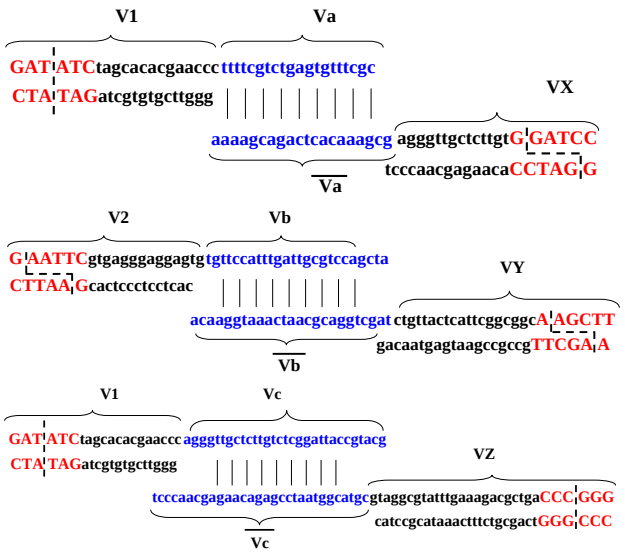

Fig. 3. Formation of "paths" for V1 – VX, V2 – VY and V1 – VZ.

Extraction: The results of the computation can be derived from the reactions of the restriction enzymes. The generated pool containing all possible solutions are divided into the exact number of test tubes representing the number of elements in the product matrix. As in Fig. 4, into each test tube, restriction enzymes corresponding to its row/column indicators are added.

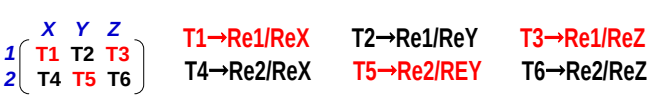

Fig. 4. Restriction enzymes added into test tube T1-T6.

A "path" which consisted of two restriction enzyme sites has two "cutting" reactions denoting a value 1 for the element in the product matrix. Strands which have none or only one "cutting" reaction denote an element of value 0 for the corresponding elements. The PCR process is conducted to

amplify the shortened strands for gel electrophoresis process. The expected result from the gel is shown in Fig. 5.

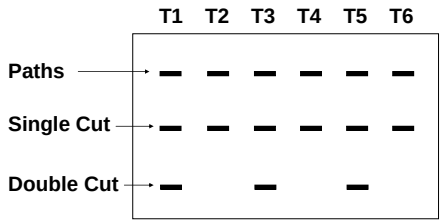

Fig. 5. Expected gel results.

IV. Results and Discussions

Using restriction enzymes as row/column indicators, the value 1 in the matrix product are denoted when a "path" is cut twice (both ends) while the value 0 are denoted by a no-cut or single-cut. Using restriction enzymes as row/column indicators for product matrix also requires coding of restriction enzyme sites in the Initial and Terminal Vertex sequences. Since the number of restriction enzymes is limited, this also limits the computational size for model. Types of restriction enzymes are estimated around 267 different kinds for TYPE II which is commonly used in DNA computing. Requirement to code with restriction enzyme sites also constrains the designs of the sequences. Proper measure has to be taken to ensure no overlap of restriction enzyme sites in other sequences to avoid unwanted "cutting".

Directed edge in the model was constructed from Intermediate Vertices which extended from an Initial Vertex as a single stranded overhang and extended from a Terminal Vertex as a complementary single stranded overhang. For a multiplication problem with two matrices, the Intermediate Vertex is easily represented by single stranded overhangs but if the same model is used to compute more than two matrices, the design for the Intermediate Vertex will have to be changed into individual single stranded DNA sequences. The individual single stranded DNA sequences have two components, one which has a complementary sequence representing the row of the intermediate matrix and another sequence which represents the column of the intermediate matrix.

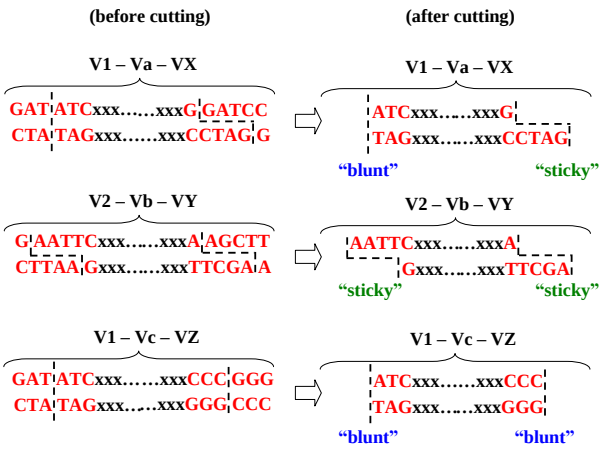

Fig. 6. Irregularity of remaining length of "path".

The actual length of an Initial/Terminal Vertex sequence consists of restriction enzyme sites and double strand unique sequence. After "cutting" reactions, a constructed "path" with Restriction Enzymes is cut short of two restriction enzyme sites. Different types of restriction enzymes produce either "blunt" ends or "sticky" ends after "cutting", causing irregularity in the remaining length of "path" (see Fig. 6).

The gel electrophoresis result for the model, the lengths of paths after "cutting" by restriction enzymes are shown in Fig. 7.

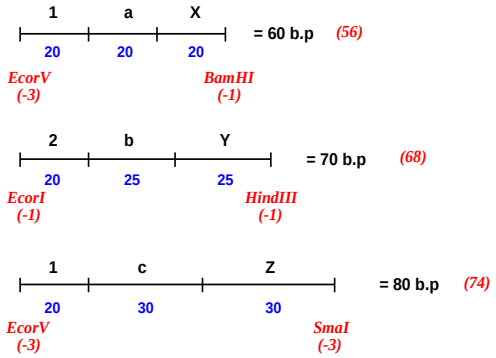

Fig. 7. Lengths of paths after "cutting" by restriction enzymes.

V. Conclusion

In matrices problem, the row/column indicators holds information which set the data contained the matrices in an order for computing/retrieval. These matrices deal with zero-ones large data set such as network fault monitoring, data mining and social studies. Using restriction enzymes as row/column indicators for Boolean Matrices in DNA Computing is advantageous during the encoding of the problems in DNA sequences as information can be tagged directly and represented by the restriction enzyme sites in the DNA sequences. The computational results are also easily interpretable based on the reactions to the specific restriction enzymes denoting different types of information tagged to the cutting sites. However, there are a few drawbacks which require extra precaution during utilizing restriction enzymes as row/column indicators in Boolean matrices such as, DNA sequences have to be synthesized to exclude repetitions of DNA sequences similar to specific restriction enzymes cutting sites. The information tagged in the DNA sequences also risk lost during extraction where a portion of the original data may be lost after "cutting". Due to its cutting reaction, the direct proportional lengths of the DNA sequences are unable to be used as weight in computing problems.

Acknowledgment

The authors would like to thank Universiti Malaysia Sarawak for supporting this research work.

References

[1] L. Adleman, "Molecular computation of solutions to combinatorial problems," *Science*, vol. 266, no. 5187, pp. 1021-1024, 1994.

[2] M. Ogihara and A. Ray, "Simulating boolean circuits on a DNA computer," in *Proc. First Annual Conference on Computational Molecular Biology*, ACM Press, Springer, 1997, pp. 326–331.

[3] T. Leete, J. P. Klein, and H. Rubin, "Bit operations using a DNA template," in *Proc. Third DIMACS Workshop on DNA-Based Computers*, 1997, pp. 159–171.

[4] V. Gupta, S. Parthasarathy, and M. J. Zaki, "Arithmetic and logic operation with DNA," in *Proc. Third DIMACS Workshop on DNA-Based Computers*, 1997, pp. 212-222.

[5] M. Orlian, F. Guarnieri, and C. Bancroft, "Parallel primer extension horizontal chain reactions as a paradigm of parallel DNA-Based computation," *DNA-based Computer III, DIMACS series in Discrete Math. Theoretical Comp. Sci.*, vol. 48, pp. 142-158, 1999.

[6] H. Rubin, J. P. Klein, and T. Leete, "A bimolecular implementation of logically reversible computation with minimal energy dissipation," in *Proc. 4th International Meeting on DNA-Based Computin*g, 1998.

[7] F. Guarnieri and C. Bancroft, "Use of horizontal chain reaction for DNA based addition," *DIMACS 44*, pp. 105-111, 1999.

[8] Z. Ibrahim, Y. Tsuboi, O. Ono, and M. Khalid, "Direct proportional length based DNA computing for shortest path problem," *International Journal of Computer Science and Applications*, pp. 46-60, 2004.

[9] R. Bakar, J. Watada, and W. Pedrycz, "DNA approach to solve clustering problem based on a mutual order," *BioSsystems*, pp. 1–12, 2008.

[10] T. Song, S. Wang, and X. Wang, "The design of reversible fate and reversible sequential circuit based on DNA computing," in *Proc. 3rd International Conference on Intelligent System and Knowledge Engineering*, 2008.

[11] X. Wang and Q. Zhang, "DNA computing based cryptography," in *Proc. 4th International Conference on Bio-Inspired Computing*, 2009, pp. 1-3.

[12] G. Cui, C. Li, H. Li, and X. Li, "DNA computing and its application to information security field," in *Proc. 5th International Conference on Natural Computation*, 2009.

[13] Q. Wang and X. Geng, "Application research of biosensor technology in DNA computing," in *Proc. IEEE International Conference on Computer Science and Automation Engineering*, 2012, pp. 639–643.

[14] R. Agrawal, F. Gustavson, and M. Zubair, "A high-performance matrix multiplication algorithm on a distributed memory parallel computer using overlapped communication," *IBM Journal of Research and Development*, vol. 38, no. 6, pp. 673-681, 1994.

[15] M. Amos, "Theoretical and experimental DNA computation," *Natural Computing Series*, Springer, 1998.

[16] J. S. Oliver, "Matrix multiplication with DNA," *J. Mol.*, vol. 45, pp. 161-167, 1997.

[17] Z. Ezziane, "DNA computing: Applications and challenges," *Nanotechnology*, vol. 1, pp. R27-R39, 2006.

[18] R. Deaton, M. Garzon, J. Rose, D. R. Franceschetti, and S. E. Stevens Jr, "DNA computing: A review," *Fundamenta Informaticae*, vol. 30, pp. 23-41, 1997.

[19] J. Y. Lee, H. W. Lim, S. I. Yoo, B. T. Zhang, and T. H. Park, "Efficient initial pool generation for weighted graph problems using parallel overlap assembly," in *Proc.10th International Meeting on DNA Based Computers*, 2004, pp. 357-364.

[20] N. Rajaee, H. Aoyagi, and O. Ono, "Single-stranded DNA and primers for boolean matrix multiplication with DNA computing," *ICIC-EL, International Journal of Research and Surveys*, vol. 3, no. 4, p. 1067, 2009.

[21] J. Zhu, X. Li, and Q. Wang, "Complete local search with limited memory algorithm for no-wait job shops to minimize makespan," *European Journal of Operational Research*, no. 198, pp. 378-386, 2009.

[22] N. Rajaee and O. Ono, "Self assembled DNA computing for directed graph modeling," in *Proc. 6th EUROSIM Congress on Modeling and Simulation*, Ljubjana, Slovenia, 2007.

[23] N. Rajaee, Y. Kon, K. Yabe, and O. Ono, "Matrix multiplication with DNA based computing: A comparison study between hybridization-ligation and parallel overlap assembly," in *Proc. the 4th IEEE International Conference on Natural Computation*, Jinan, China, 2008, pp. 527-530.

[24] J. Watada, S. Kojima, S. Ueda, and O. Ono, "DNA computing approach to optimal decision problems," *International Journal of Innovative Computing, Information and Control*, vol. 2, no. 1, pp. 273-282, 2006.

[25] N. K. Khalid, Z. Ibrahim, T. B. Kurniawan, M. Khalid, N. H. Sarmin and A. P. Engelbrecht, "Function minimization in DNA sequence design based on continuous particle swarm optimization," *ICIC Express Letters*, vol. 3, no. 1, pp. 27-32, 2009.

Nordiana Rajaee received her BEng (Hons) degree in electronic and information engineering from Kyushu Institute of Technology, Fukuoka, Japan in 1999, MSc degree in microelectronics in University of Newcastle Upon Tyne, United Kingdom in 2003 and PhD degree in DNA computing from Meiji University, Japan in 2011. Her current interests are in DNA computing, evolutionary computing and microelectronics.

Kismet Hong Ping received his BEng (Hons) degree in electronics and telecommunications from Universiti Malaysia Sarawak, MSc degree in digital communication systems from Lougborough University, United Kingdom and PhD degree in systems science from Nagasaki University Japan. His research interests are biomedical imaging and microwave and antenna technologies.

Asrani Lit received his B. Eng (Hon.) degree in computer engineering in 2007 and M.Eng degree in microelectronics & computer system in 2011 from Universiti Teknologi Malaysia. His research interests are in microelectronics, computer system and network-on-chip.

Dyg Nur Salmi Dharmiza Awg Salleh received her BEng (VLSI System Design) from Ritsumeikan University Japan and MEng (Electronics and Applied Physics) in Tokyo Institute of Technology Japan. Her research interests are VLSI design and microelectronics.

Liang Yew Ng received the B.S. degree in electrical engineering degree from University of Alabama, AL, USA in 1990 and M.S. degree in electrical engineering from Rochester Institute of Technology, Rochester, NY, USA in 1992. He is currently a lecturer in the Department of Electronic Engineering with Universiti Malaysia Sarawak. His research interests include signals processing, communication and integrated circuit design. He is a member of the Institute of Electrical and Electronics Engineers.