Unconstraint Assignment Problem: A Molecular Computing Approach

Ibrahim Zuwairie[1&2], Tsuboi Yusei[2], Ono Osamu[2], and Khalid Marzuki[3]
[1]Department of Mechatronics and Robotics, University Technology Malaysia, Malaysia
[2]Institute of Applied DNA Computing, Meiji University, Japan
[3]Center for Artificial Intelligence and Robotics (CAIRO), University Technology of Malaysia, Malaysia

Abstract: *Deoxyribonucleic Acid or DNA-based computing is an emerging field that bridging the gap between chemistry, molecular biology, computer science, and mathematics. This research area is a new paradigm whereby the computation can be done by the use of DNA molecules to encode the computational problem. During the massively parallel computation in a test tube, a series of bio-molecular reactions are employed and the output encoded also by DNA molecules can be printed and read out by electrophoretical fluorescent method. Since DNA computing is very suitable for combinatorial problems, in this paper, an idea on DNA-based computing algorithm for solving unconstraint assignment problem is proposed. The proposed approach basically consists of two phases; encoding phase and computational phase. During the encoding phase, a method to encode the computational problem is carried out by introducing four rules. On the other hand, for the computational phase, it is discovered that the complexity of the unconstraint assignment problem can be reduced to a path problem of a graph, and the possibility to solve the unconstraint assignment problem by DNA computing approach is shown in detail.*

Keywords: *DNA computing, assignment problem, optimization.*

Received December 25, 2004; accepted March 11, 2005

1. Introduction

In 1965, Moore [8] observed an exponential growth in the number of transistors per integrated circuit against time. Currently, this is the definition of Moore's Law, meaning that more and more transistors can be crammed into a chip until the silicon itself reaches its finite atomic scale limitation. Since the traditional silicon-based computer is restricted by its fundamental physical limitation, researchers have been searching for alternative medium for computation and Deoxyribonucleic Acid (DNA) which would turn out to be the answer.

A new computing paradigm based on DNA molecules had appeared in 1994 when Adleman [1, 2] launched a novel approach to solve the so-called Hamiltonian Path Problem (HPP) with seven vertices by DNA molecules. The goal of HPP is to determine whether a path exists that will commerce at the start city, finish at the end city and pass through each city of the remaining cities exactly once. While in conventional silicon-based computer, information is stored as binary numbers in silicon-based memory, he encoded the information of the vertices by a randomly DNA sequences. For the computation, gigantic memory capacity and massively parallelism inherent in DNA computing is exploited to make a brute force search on a big problem space in constant or polynomial-time. The output of the computation, also in the form of DNA molecules can be read and printed by electrophoretical fluorescent method.

DNA molecules are composed of single or double DNA fragments or often called oligonucleotides or strands. Nucleotides form the basis of DNA. A single-stranded fragment has a phospho-sugar backbone and four kinds of bases denoted by the symbols A, T, G, and C for the bases adenine, thymine, guanine, and cytosine respectively. These four nucleic acids, which can occur in any order combined in Watson-Crick complementary pairs to form a double strand helix of DNA. Due to the hybridization reaction, A is complementary with T and C is complementary with G. As an example, any sequence oligonucleotides, such as 5'-ACCTG-3' has a complementary sequence, 3'-TGGAC-5'. Digits 5' and 3' denotes orientation of a DNA oligonucleotides.

In unconstrained assignment problem, one is concerned with establishing a full one-to-one correspondence between two sets E and J, both of which have N elements as depicted in Figure 1. An assignment is a one-to-one mapping α: $E \rightarrow J$ [7]. The assignment problem is so fundamental in operations research as well as in engineering field. It is very useful because assigning n tasks to n people is a basic primitive in many applications [4].

Skiena [10] stated that an efficient algorithm for constructing matching is based on constructing *augmented path* in graph. Given a matching M in a

graph G, and augmenting path P is a path of edges where every odd-numbered edge (including the first and last edge) is not in M, while every even-numbered edge is. Further, the first and last vertices must not be already in M. By deleting the even-numbered edges of P from M and replacing them with the odd-numbered edges of P, we enlarge the size of the matching by one edge. Berge's theorem states that a matching is maximum if and only if it does not contain any augmenting path. Therefore, we can construct maximum-cardinality matching by searching for augmenting paths and stop when none exist.

However, the algorithm described in [10] performs efficiently only when the computation is done in a sequential silicon-based computer. At present, there is no molecular or DNA-based computing algorithm that has been proposed to solve the unconstraint assignment problem. Hence, this paper proposes a DNA-based computing algorithm for the unconstraint assignment problem and massively parallel computation can be done in a test tube. The computation involves several kinds of molecular biology operations such as ligation, polymerase chain reaction (PCR), affinity purification, and gel electrophoresis.

The proposed DNA-based computing algorithm for solving the unconstraint assignment problem basically can be divided into two phases; encoding and synthesize phase, and computation phase.

2. Encoding and Synthesize

The input graph in Figure 1 consists of two sets E and J, where each element in the set E and set J is presented as $E_x = \{E_1, E_2, \ldots\ldots\ldots, E_w\}$, and $J_y = \{J_1, J_2, \ldots\ldots\ldots, J_z\}$, respectively. Then, a unique 20 bases DNA sequences is assigned for each element in J and E, and are placed in Table 1. These DNA sequences are designed by using DNASequenceGenerator [11]. Every DNA sequences can be divided into half left sequences and half right sequences as shown in Figure 2.

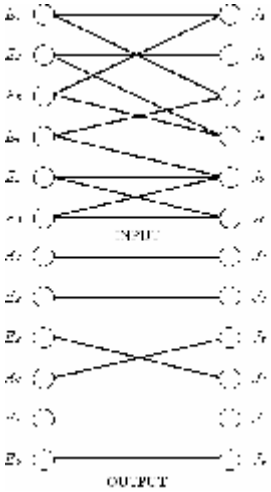

Figure 1. An example of input and output of unconstraint assignment problem.

Table 1. DNA sequences designed for each element in E and J.

Name	DNA Sequences	Complement Sequences
E_1	AAAGCTCGTCGTTTAAGGAA	TTCCTTAAACGACGAGCTTT
E_2	GAAGCCTACTGTACTCTGCG	CGCAGAGTACAGTAGGCTTC
E_3	TATCGTGATTTGGAGGTGGA	TCCACCTCCAAATCACGATA
E_4	CAGCCACGTAGTAGAGCTAG	CTAGCTCTACTACGTGGCTG
E_5	TACCCAATCGAACTGATAAG	CTTATCAGTTCGATTGGGTA
E_6	TCGGTCAACGGAGGGGGCTC	GAGCCCCCTCCGTTGACCGA
J_1	GCGTTTTTGCGAGGCATGTG	CACATGCCTCGCAAAAACGC
J_2	GCCTAAAGAATTGATCGCTT	AAGCGATCAATTCTTTAGGC
J_3	TAGGTGCGTGCATAACTGGG	CCCAGTTATGCACGCACCTA
J_4	CTAAGTGCGGCTGCATGACC	GGTCATGCAGCCGCACTTAG
J_5	TGGGGTTTTATCTTACGACC	GGTCGTAAGATAAAACCCCA
J_6	TGAGATTTTTAACGCCGTTA	TAACGGCGTTAAAAATCTCA

TATCGGATCG | GTATATCCGA

Half left sequences | Half right sequences

Figure 2. An example of 20 bases DNA sequences.

The oligonucleotides or oligos for short, for each edge in the input graph are synthesized based on four rules as follows:

1. If there is a connection between a vertex E_x to vertex J_1, synthesize the oligos as $J_1E_xJ_2\downarrow$.
2. If there is a connection between a vertex E_x to vertex J_y where $1 < y < z$, synthesize the oligos as $J_y\uparrow E_xJ_{y+1}\downarrow$.
3. If there is a connection between a vertex E_x to vertex J_y where $y = z$, synthesize the oligos as $J_z\uparrow E_xJ_1$.
4. Synthesize the oligos for complements of J_y.

Where the symbol ↑ and ↓ indicates the half right and half left sequences of oligos respectively. The oligos synthesized based on these four rules can be presented as depicted in Figures 3-6.

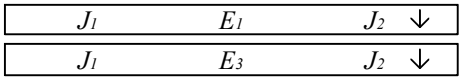

Figure 3. Graphical representation showing the oligos synthesized based on rule (1).

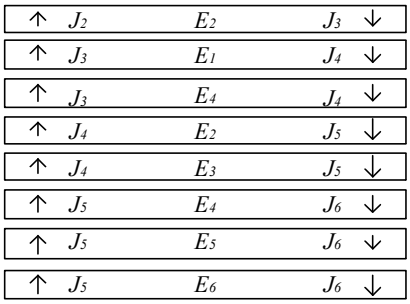

Figure 4. Graphical representation showing the oligos synthesized based on rule (2).

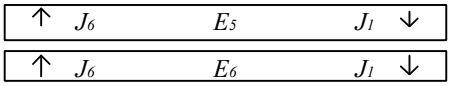

Figure 5. Graphical representation showing the oligos synthesized based on rule (3).

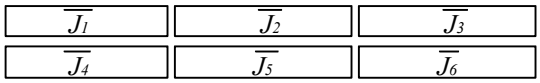

Figure 6. Graphical representation showing the oligos synthesized based on rule (4).

3. Computation Phase and Discussions

The beginning of this phase is the insertion of all the synthesized oligos into a test tube and DNA ligase reaction is performed in which every possible combination is generated by DNA molecules. Ligation often invoked after the single DNA strands is annealed and concatenated to each other. Many single-strand fragments will be connected in series and ligase is used as 'glue' to seal the covalent bonds between the adjacent fragments as shown in Figure 7 [13].

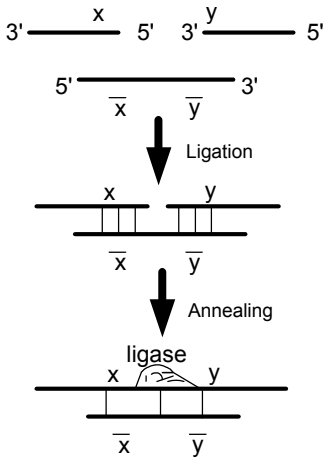

Figure 7. An example of ligation reaction.

For better understanding how all the combinations are formed, the input graph in Figure 1 can be presented in different way as depicted in Figure 8. According to Figure 8, as an example, it is clear that the vertex J_1 is assigned with either vertex E_1 or E_3 in the output graph. If the input graph is arranged in such a way that every edge E_x, which is connected with vertex J_y, where $y \neq z$ is connected to vertex J_{y+1}, and the edge E_x connected with vertex J_z is connected to vertex J_{end}, a new modified input graph is shown in Figure 9.

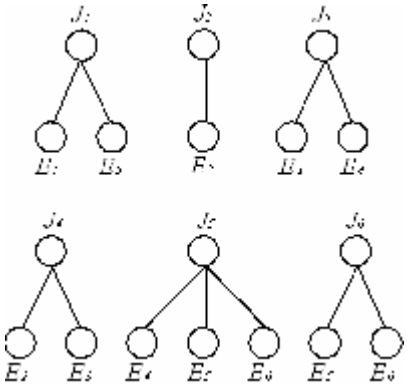

Figure 8. Input graph from J point of view.

Due to the modified input graph, the complexity of the unconstraint assignment problem now can be reduced to a path finding problem. As such, from the node J_1 to node J_2, one can choose either to pass through the node E_1 or node E_3. For a complete path, beginning from node J_1 and end at node J_{end}, there will be 2! x 1! x 2! x 2! x 3! x 2! = 96 number of combinations. Thus, this is clearly a combinatorial problem. In this case, the combination representing the answer of the unconstraint assignment problem is shown in Figure 10 while the example of a combination representing the wrong answer of the unconstraint assignment problem is shown in Figure 11. The combination shown in Figure 10 can be read as "*node E_1 is assigned to node J_1, node E_2 is assigned to node J_2, node E_4 is assigned to node J_3, node E_3 is assigned to node J_4, node E_5 is assigned to node J_5, and node E_6 is assigned to node J_6*". Note that all combinations can be read in the same manner. All of these combinations are formed in a test tube by annealing of previously synthesized oligos with respect to Watson-Crick complementary DNA molecules. One combination is encoded by one kind of DNA strands and only one kind of DNA strands out of 96 can be regarded as the answer of the unconstrained assignment problem.

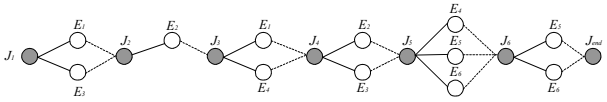

Figure 9. Modified input graph.

J_1 - E_1 - J_2 - E_2 - J_3 - E_4 - J_4 - E_3 - J_5 - E_5 - J_6 - E_6 - J_{end}

Figure 10. Combination representing the right answer.

J_1 - E_3 - J_2 - E_2 - J_3 - E_1 - J_4 - E_2 - J_5 - E_5 - J_6 - E_6 - J_{end}

Figure 11. An example of combination representing a wrong answer.

Figure 12 shows the combination of right answer encoded by a DNA molecule. The node sequence J_{end} shown in Figure 9, Figure 10, and Figure 11 is replaced by J_1 in the DNA molecules and this replacement is possible because of the oligos synthesized previously by using rule (3). Figure 12 also shows that the complement oligos synthesized by using rule (4) is important during the computation. It is true because during the concatenation in a test tube, these complement oligos will joint together two respective oligos beside each other before the ligase enzyme is pour into the test tube to seal the respective two oligos. Then the polymerase enzyme is occupied to produce the complete double stranded DNA molecules as well depicted in Figure 13.

Figure 12. Combination of right answer encoded by a DNA molecule.

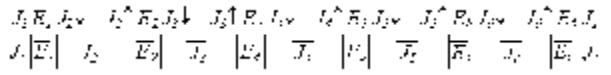

Figure 13. A complete double stranded DNA molecule representing the right answer.

At this moment, there will be a numerous number of DNA molecules concatenated representing various number of combination in the test tube. However, the quantity of DNA molecules representing the right answer is small compared to other DNA molecules representing the wrong answer. The right answer DNA molecules could be extracted but not by merely filtering the particular DNA molecules but rather amplifying the quantity of DNA molecules representing the right answer. The best method to amplify a particular DNA molecule is by employing Polymerase Chain Reaction (PCR) reaction.

PCR is an incredible sensitive copying machine for DNA. It also can be used for DNA detection. Given a site-specific single molecule DNA, a million or even billion of similar molecules can be created by PCR process. In n steps, it can produces 2^n copies of the same molecules. PCR needs a number of sub-sequence strands called 'primers', which is usually about 20 base long to signal a specific start and end site at a template for replication. PCR normally runs for 20-30 cycles of 3 steps as shown in Figure 14. Each cycle consists of three phases: Separating base pair strands of DNA at about 95°C, annealing at 55°C and extension at 75°C [6]. It takes about 3 hours normally to complete the cycles. Thus, in this paper, two kinds of primers will be used where the first one encodes the complement of the first node, CACATGCCTCGCAAAAACGC in and the latter encodes the first node, GCGTTTTTGCGAGGCATGTG.

Since 20 DNA bases are assigned for each E_x and J_y, then the actual length of the DNA molecules representing the right answer is known in advance. In the example introduced in this paper, the length of the complete double stranded DNA molecules is 260 base pair (bp). Thus, after the PCR operation is performed, the DNA solution in the test tube is brought to gel electrophoresis to extract only the DNA molecules with the length of 260 bp.

DNA strands in a solution can be separated in term of its length by means of gel electrophoresis. In fact, the molecules are separated according to their weight, which is almost proportional to their length [5]. This technique is based on the fact that DNA molecules are negatively charged [9]. Hence, by putting them in an electric field, they will move towards the positive electrode at different speed. The longer molecules will remain behind the shorter ones. The speed also depends on the gel porosity. Agarose gel is frequently used and by varying the porosity of the gel used, the sensitivity of this length separation operation can be altered. The precision is high, even molecules which differ by one nucleotide can still be distinguished between each other. An example of gel electrophoresis process [3] and its output [12] is well depicted in Figure 15 and Figure 16 respectively. This technique can be used to "print" the results of DNA computation as well. Normally, at the end of this process, the gel is photographed for convenience.

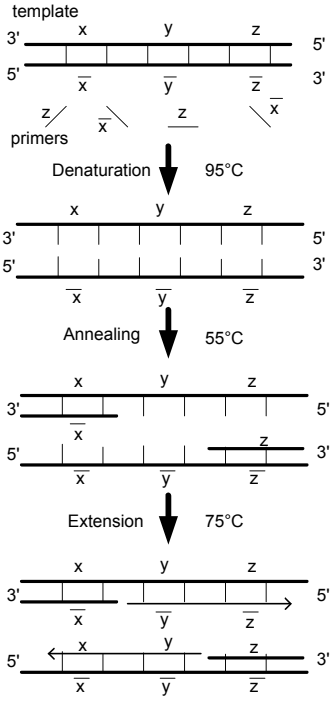

Figure 14. An example of one cycle of polymerase chain reaction.

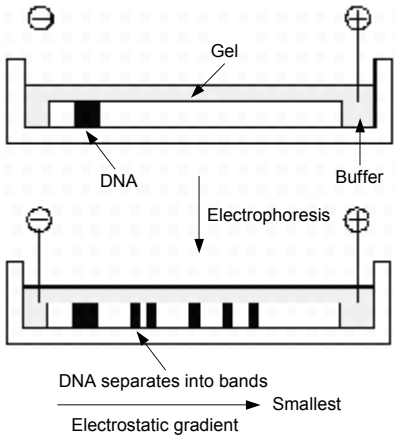

Figure 15. Gel electrophoresis process.

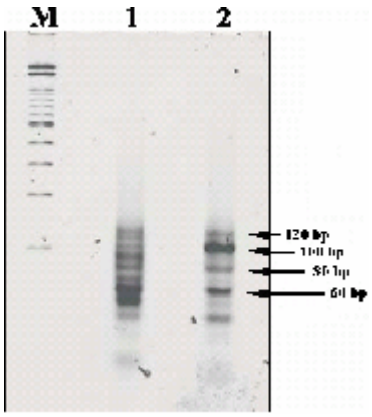

Figure 16. Gel electrophoresis output where lane M is DNA size marker. Lane 1 and 2 are used for the tested DNA molecules.

At this moment, one will have 260 bp DNA molecules at hand and these DNA molecules are formed by a various kind of combinations. The DNA molecules representing the right combination can be distinguished from others because the sequences encoding the node E_x, where the value x range from 1 to 6 occurs once only. The 260 bp DNA molecules are melted and the double stranded DNA molecules will come apart to become single stranded DNA molecules. Then the magnetic bead separation is executed.

The magnetic bead separation uses multiple copies of DNA probe molecules that encode the complement of a particular sequence node, E_x. The DNA probe molecules are attached to a microscopic iron ball. When the iron ball is placed in the test tube, by taking the advantages of the Watson-Crick complementary, only the single stranded DNA molecules containing the particular sequence will anneal to the probes and the remaining will be poured out. By applying this operation w times, the DNA molecules containing the sequences for all E_x could be extracted.

After the affinity separation is finished, one knows that each DNA molecule remaining in the test tube contains all the sequences encoding the node E_x. However, there is no information regarding the order of the sequences E_x. This order is crucial because it shows which node E_x should be connected to the node J_y, which is the answer of the unconstraint assignment problem.

The order of the node E_x encoded in the DNA molecules could be determined by additional PCR operation, followed by gel electrophoresis. This process, which is also called as graduated PCR is executed by first applying the PCR operation where the complement of node J_1 is employed as the right primer and the primers encoded as the node E_x are employed as the left primer. The graduated PCR operation is depicted graphically in Figure 17. Finally, the expected output of the graduated PCR operation can be read out as "*node E_1 is assigned to node J_1, node E_2 is assigned to node J_2, node E_4 is assigned to node J_3, node E_3 is assigned to node J_4, node E_5 is assigned to node J_5, and node E_6 is assigned to node J_6*", where the expected graduated PCR output indicates the answer of the unconstraint assignment problem.

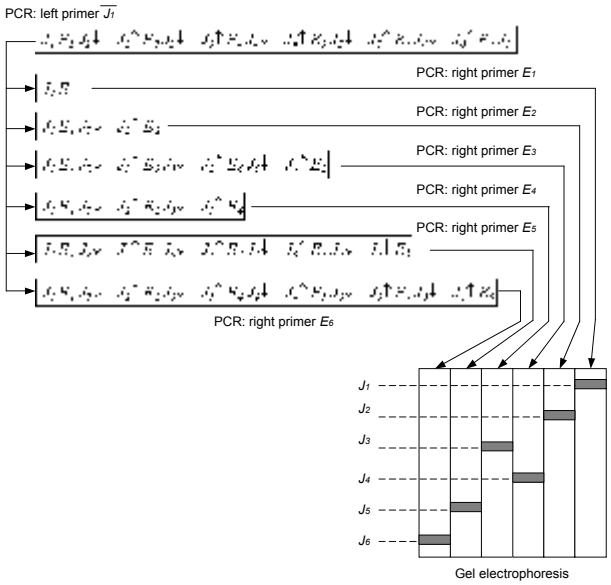

Figure 17. Graduated PCR.

4. Conclusion

This research is concerned with unconstraint assignment problem. A DNA computing approach is proposed in order to solve the unconstraint assignment problem at molecular level in a test tube in massively parallel fashion. Based on massive parallelism, many researchers in DNA computing so far tried to solve NP (nondeterministic polynomial time) problems. These are mathematical problems which have exponential complexity and no efficient solution has been found yet. Even though the unconstraint assignment problem is not a class of NP problems, it is important to solve them since this kind of problem occurs generally in many real world problems. As such, let E be a set of employees and J be a set of jobs, then, the solution of this problem is a matching between employees and jobs. In other words, which job should be assigned to a particular employee. Thus, it is expected that the proposed DNA computing approach is indispensable in many real world applications in future.

References

[1] Adleman L. M., "Computing with DNA," *Scientific American Journal*, vol. 279, no. 2, pp. 54-61, 1998.

[2] Adleman L. M., "Molecular Computation of Solutions to Combinatorial Problems," *Science Journal*, vol. 266, pp. 1021-1024, 1994.

[3] Amos M., "*DNA Computation*," *PhD Thesis*, The University of Warwick, 1997.

[4] Arora A., Frieze A., and Kaplan H., "A New Rounding Procedure for the Assignment Problem

with Application to Dense Graph Arrangement Problem," *in Proceedings of the 37th Annual Symposium on Foundation of Computer Science*, pp. 21-30, 1996.

[5] Calude C. S. and Paun G., *Computing with Cell and Atoms: An Introduction to Quantum, DNA, and Membrane Computing*, Taylor and Francis, 2001.

[6] Fitch J. P., *An Engineering Introduction to Biotechnology*, SPIE-The International Society of Optical Engineering, 2002.

[7] Liu G., and Haralick R. M., "Assignment Problem in Edge Detection Performance Evaluation," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 26-31, 2000.

[8] Moore G. E., "Cramming More Components onto Integrated Circuits," *Electronics Journal*, vol. 38, no. 8, pp. 114-117, 1965.

[9] Paun G., Rozenberg G., and Salomaa A., *DNA Computing: New Computing Paradigms*, Springer-Verlag, 1998.

[10] Skiena, S. S. and Skiena S., *The Algorithm Design Manual*, Telos Pr, 1997.

[11] Udo F., Sam S., Wolfgang B., and Hilmar R., "DNA Sequence Generator: A Program for the Construction of DNA Sequences," *in Proceedings of the 7th International Workshop on DNA Based Computers*, pp. 23-32, 2001.

[12] Yamamoto M., Kameda A., Matsuura N., Shiba T., Kawazoe Y., and Ohuchi A., "A Separation Method for DNA Computing Based on Concentration Control," *New Generation Computing Journal*, vol. 20, no. 3, pp. 251-262, 2002.

[13] Zucca M., "*DNA* Based Computational Model," *PhD Thesis*, Politecnico Di Torino, Italy, 2000.

Ibrahim Zuwairie received his BEng in mechatronics and his MEng in image processing from University Teknology Malaysia, Malaysia, in 2000 and 2002, respectively. Since 2002, he has been engaged with the Department of Mechatronics and Robotics, Faculty of Electrical Engineering, University Technology Malaysia as a lecturer. Currently, he is pursuing his PhD at the Institute of Applied DNA Computing, Meiji University, Kanagawa, Japan. He is a student member of Institute of Electrical and Electronics Engineers (IEEE), International Computational Intelligence Society (ICIS), and International Signal Processing Society (ISPC). His research interests include signal and image processing, automated visual inspection, evolutionary, and unconventional computing such as molecular or DNA computing.

Tsuboi Yusei received the BEng and Master's degree from Meiji University, Japan, in 2000 and 2002, respectively. Currently, he is pursuing the PhD degree also at Meiji University, Japan. Presently, his research interest includes robotics and DNA computing for artificial intelligence.

Ono Osamu received the Bachelor, Master and Doctor Degree in Engineering all from Waseda University, Tokyo, in 1974, 1976, and 1979, respectively. He is a professor at the Department of Electrical and Electronic Engineering, Meiji University, Japan. He is the director of Tokyo Branch of Japan Institute Electrical Engineering and a committee member of Japan Society for Simulation Technology (JSST). His research interest includes large scale industrial process, mechatronics, advanced mobile robotics and image processing. Currently, he is interested in the application of DNA computing in engineering field.

Khalid Marzuki is currently a professor in intelligent control and a director at the Centre of Artificial Inteligence and Robotics (CAIRO), University Technology Malaysia, Malaysia. His research interest is in the field of artificial intelligence with applications of control. He has been appointed to the editorial advisory board of the International Journal of Engineering Applications of Artificial Intelligence published by Elsevier Science and an associate editor of the Journal of Systems and Control Engineering published by the Institute of Mechanical Engineers, United Kingdom. Currently, he is the IEEE student activities chair for Region 10 (Asia Pacific) and also the founding member of the Asian Control Professors Association.