

PARTICLE SWARM OPTIMIZATION FOR SOLVING  
DNA SEQUENCE DESIGN PROBLEM

NOOR KHAFIFAH BINTI KHALID

A thesis submitted in fulfilment of the  
requirements for the award of the degree of  
Master of Engineering (Electrical)

Faculty of Electrical Engineering  
Universiti Teknologi Malaysia

MARCH 2010

## ABSTRACT

Deoxyribonucleic Acid (DNA) has certain unique properties such as self-assembly and self-complementary in hybridization, which are important in many DNA-based technologies. DNA computing, for example, uses these properties to realize a computation, *in vitro*, which consists of several chemical reactions. Other DNA-based technologies such as DNA-based nanotechnology and polymerase chain reaction (PCR) also depend on hybridization to assemble nanostructure and to amplify DNA template, respectively. Hybridization of DNA can be controlled by designing DNA sequences properly. In this thesis, sequences are designed such that each sequence uniquely hybridizes to its complementary sequence, but not to any other sequences. This objective can be formulated using four objective functions, namely, *similarity*,  $H_{measure}$ , *continuity*, and *hairpin*. To achieve this, particle swarm optimization (PSO) for DNA sequence design is proposed to minimize the objective functions subjected to two constraints: melting temperature and  $GC_{content}$ . Two models are developed, namely the Continuous PSO and Binary PSO. Since DNA sequence design is a multi-objective optimization (MOO) problem, two methods to solve MOO are used in this thesis. These methods are the aggregation-based method and criterion-based method, particularly vector evaluated PSO (VEPSO). The implementation of PSO algorithm for DNA sequence design is first started with application of both proposed models to aggregation-based method. Then, the results between these models are compared. It is found that better results are produced by Binary PSO. Next, VEPSO is used to design DNA sequences based on Binary PSO. The results show that several set of good sequences are produced, which are better than other research works where only a set of DNA sequences is generated.

## ABSTRAK

Asid Deoksiribonukleik (DNA) mempunyai ciri-ciri unik tertentu seperti pasang sendiri dan pelengkap sendiri dalam penghibridan, yang penting dalam banyak teknologi berasaskan DNA. Pengkomputeran DNA, contohnya, menggunakan ciri-ciri ini untuk melaksanakan sesebuah pengkomputeran, *in vitro*, yang melibatkan beberapa tindakbalas kimia. Lain-lain teknologi berasaskan DNA seperti nanoteknologi berasaskan DNA dan *polymerase* tindak balas rantai (*PCR*) juga masing-masing bergantung kepada penghibridan untuk menggabungkan struktur nano dan menjana templat DNA. Penghibridan DNA boleh dikawal dengan merekabentuk urutan DNA dengan betul. Dalam tesis ini, urutan-urutan direkabentuk supaya setiap urutan dihibridkan secara unik kepada urutan pelengkapnyanya sahaja, tetapi bukan kepada urutan-urutan yang lain. Objektif ini boleh dirumuskan dengan menggunakan empat fungsi objektif, iaitu *similarity*,  $H_{measure}$ , *continuity*, dan *hairpin*. Untuk mencapai objektif tersebut, pengoptimuman kerumunan zarah (*PSO*) untuk merekabentuk urutan DNA dicadangkan bagi meminimumkan fungsi-fungsi objektif tersebut, tertakluk kepada dua kekangan: suhu cair dan  $GC_{content}$ . Dua model dibangunkan, iaitu *PSO* Berterusan dan *PSO* Perduaan. Oleh kerana masalah rekabentuk urutan DNA merupakan masalah pengoptimuman pelbagai objektif (MOO), dua kaedah penyelesaian MOO digunakan dalam tesis ini. Kaedah-kaedah ini adalah kaedah berasaskan penambahan dan kaedah berasaskan kriteria, iaitu penilaian vektor *PSO* (*VEPSO*). Pelaksanaan algoritma *PSO* untuk rekabentuk urutan DNA dimulakan dengan aplikasi kedua-dua model yang dicadangkan kepada kaedah berasaskan penambahan. Kemudian, keputusan antara model-model ini dibandingkan. Ternyata keputusan yang lebih baik dihasilkan oleh *PSO* Perduaan. Kemudian, *VEPSO* digunakan untuk merekabentuk urutan DNA berdasarkan *PSO* Perduaan. Keputusan menunjukkan bahawa beberapa set urutan-urutan yang bagus dihasilkan, yang mana lebih baik daripada kerja-kerja penyelidikan yang lain di mana hanya satu set urutan-urutan DNA dihasilkan.

## TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
	DECLARATION	ii
	DEDICATION	iii
	ACKNOWLEDGEMENTS	iv
	ABSTRACT	v
	ABSTRAK	vi
	TABLE OF CONTENTS	vii
	LIST OF TABLES	xi
	LIST OF FIGURES	xiii
	LIST OF ABBREVIATIONS	xvii
	LIST OF SYMBOLS	xviii
	LIST OF APPENDICES	xix
<b>1</b>	<b>INTRODUCTION</b>	<b>1</b>
	1.1 Particle Swarm Optimization in Brief	1
	1.2 Deoxyribonucleic Acid	2
	1.3 The Importance of DNA Sequence Design	5
	1.3.1 DNA Nanotechnology	6
	1.3.2 DNA Computing	8
	1.3.3 DNA Biotechnology	9
	1.4 Problem Statement and Objective	10
	1.7 Scope of Work	11
	1.8 Thesis Contributions	11
	1.9 Thesis Organization	12

<b>2</b>	<b>LITERATURE REVIEW</b>	<b>14</b>
2.1	Introduction	14
2.2	DNA Sequence Design Approaches	14
2.2.1	Simple Method	14
2.2.2	Evolutionary Method	16
2.2.3	Heuristic Method	16
2.2.4	Theoretical Method	18
2.2.5	<i>In vitro</i> Method	18
2.2.6	Population-based search Method	19
2.2.7	Hybrid Method	20
2.3	Particle Swarm Optimization	20
2.4	Chapter Summary	26
<b>3</b>	<b>OBJECTIVE FUNCTIONS AND CONSTRAINTS IN DNA SEQUENCE DESIGN</b>	<b>27</b>
3.1	Introduction	27
3.2	Design Criteria for DNA Sequence	28
3.3	Formulation and Fitness Functions for DNA Sequence Design	29
3.4	Objective Functions for DNA Sequence Design	33
3.4.1	$H_{measure}$	33
3.4.2	<i>Similarity</i>	34
3.4.3	<i>Hairpin</i>	36
3.4.4	<i>Continuity</i>	37
3.5	Constraints in DNA Sequence Design	38
3.5.1	$GC_{content}$	38
3.5.2	<i>Melting Temperature</i>	38
3.6	Example Calculations for Objective Functions	40
3.6.1	$H_{measure}$ Calculation	41
3.6.2	<i>Similarity</i> Calculation	44
3.6.3	<i>Hairpin</i> Calculation	45
3.6.4	<i>Continuity</i> Calculation	47
3.7	Chapter Summary	48

<b>4</b>	<b>PARTICLE SWARM OPTIMIZATION</b>	<b>49</b>
4.1	Introduction	49
4.2	Background	50
4.3	Particle Swarm Optimization	52
4.3.1	PSO Algorithm	54
4.3.2	PSO Control Parameters	57
4.3.4.1	Swarm Size, Number of Iterations	58
4.3.4.1	Velocity Clamping	58
4.3.4.1	Inertia Weight	59
4.3.4.1	Acceleration Coefficients and Random Values	61
4.3.3	Values Selection	64
4.4	Binary PSO	65
4.4.1	Formulation and Parameters in Binary PSO	66
4.4.2	Binary PSO Applications	66
4.5	Multi-objective Particle Swarm Optimization	68
4.5.1	Basic Concept	68
4.5.2	Classification of MOPSO	71
4.5.3	Aggregation-based Method	71
4.5.4	Criterion-based Method	72
4.5.5	Dominance-based Method	73
4.6	Discussions	75
4.7	Chapter Summary	77
 <b>5</b>	 <b>AGGREGATION-BASED DNA SEQUENCE DESIGN</b>	 <b>78</b>
5.1	Introduction	78
5.2	Aggregation-based Optimization	79
5.2.2	DNA Sequence Design Based on Binary PSO	80
5.2.3	DNA Sequence Design Based on Continuous PSO	82
5.3	Implementation	83
5.3.1	Experimental Parameters	83
5.3.2	PSOligo: A Program	85
5.4	Experimental Results and Discussions	87

5.5	Chapter Summary	99
<b>6</b>	<b>VECTOR EVALUATED PARTICLE SWARM OPTIMIZATION FOR DNA SEQUENCE DESIGN</b>	<b>100</b>
6.2	Introduction	100
6.2	Vector Evaluated Particle Swarm Optimization	101
6.3	DNA Sequence Design Based on Binary VEPSO Algorithm	103
6.3.1	Experimental Parameters	106
6.3.2	Binary VEPSOligo: A Program	106
6.4	Implementation and Experimental Results	108
6.5	Chapter Summary	115
<b>7</b>	<b>SUMMARY AND CONCLUSIONS</b>	<b>116</b>
7.1	Summary	116
7.2	Conclusion	118
7.3	Limitations	119
7.3	Suggestions for Future Research	120
	<b>REFERENCES</b>	<b>122</b>
	Appendices A-B	141 - 148

## LIST OF TABLES

TABLE NO.	TITLE	PAGE
2.1	Summary of the previous DNA sequence design approaches	14
3.1	Summary of the objective functions and constraints employed in previous works	29
3.2	Basic notations	30
3.3	$\Delta H$ and $\Delta S$ values of nearest-neighbour model	39
3.4	An example of a set of DNA sequences	39
3.5	The first and second sequences	40
3.6	The first sequence and the reverse of the second sequence	40
3.7	The values of $\Sigma bp$ , $\Sigma cbp$ , $h_{dis}$ , $h_{con}$ functions, and the $th_{dis\_h_{con}}$ value for all positions during $H_{measure}$ calculation. The bold bases are the hybridized bases	41
3.8	$H_{measure}$ for the first sequence. This sequence is compared with other sequences	43
3.9	$H_{measure}$ for a set of DNA sequences	43
3.10	<i>similarity</i> for a set of DNA sequences	44
3.11	The values of from $\Sigma bp$ , $pinlen(r, s)$ , and $T(\Sigma bp, pinlen(r, s))$ functions during hairpin calculation. The bold bases are the hybridized bases	45
3.12	<i>hairpin</i> for a set of DNA sequences	45



3.13	An example of a set of DNA sequences for <i>continuity</i> calculation	46
3.14	Sample set of DNA sequences for <i>continuity</i> calculation	46
4.1	Standard fitness functions	55
4.2	Values selection for PSO parameters	64
4.3	The summary of differences between PSO and EA	76
5.1	The values of PSO control parameters	83
5.2	DNA Sequence Parameters	84
5.3	Comparison of the sequences from Continuous PSO and the sequences generated by Binary PSO	87
5.4	Comparison of the sequences in (Deaton <i>et al.</i> , 1996) and the sequences generated by Binary PSO	91
5.5	Comparison results of the sequences in (Tanaka <i>et al.</i> , 2002) and the sequences by Binary PSO method	93
5.6	Comparison of the sequences in (Guangzhao <i>et al.</i> , 2007) and the sequences generated by Binary PSO	95
5.7	Comparison results of the sequences in (Zhao <i>et al.</i> , 2007) and the sequences by Binary PSO method	97
6.1	The values of control parameters for Binary VEPSO	105
6.3	The DNA sequences extracted from the first non-dominated solution	108
6.4	The DNA sequences extracted from the second non-dominated Solution	109
6.5	The DNA sequences extracted from the third non-dominated solution	109

## LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
1.1	Chemical structure of each base and phosphate in single-stranded DNA	3
1.2	A double-stranded DNA structure	3
1.3	Example of hybridization and denaturation of two single-stranded DNA	5
1.4	Example of hairpin formation	5
1.5	An example of four-armed junctions, assembled to create 2-D lattice (Seeman <i>et al.</i> , 1998)	7
1.6	Examples of DNA nanostructures (Seeman <i>et al.</i> , 1998)	8
1.7	Polymerase chain reaction (Vierstraete, 1999)	10
3.1	Schematic representation of the classic secondary structure of DNA	28
3.2	Example of $H_{measure}$ calculation (Kurniawan <i>et al.</i> , 2008)	32
3.3	Example of <i>similarity</i> measure (Kurniawan <i>et al.</i> , 2008)	34
3.4	A possible formation of a <i>hairpin</i> . In this example, $p = r = 6$	35
4.1	Flow chart representing a general PSO algorithm	53
4.2	The illustration to update $p_{best}$	56
4.3	The movement of one particle with $c_1 = c_2 = 0$	61

4.4	The movement of a particle with $c_1=0$	61
4.5	The movement of a particle with $c_2 = 0$	62
4.6	The movement of a particle with $c_1 = c_2 > 0$	63
4.7	The dominance concept for two objective functions and concept of Pareto Front	69
5.1	An example showing the proposed Continuous PSO model with two particles search for a set two DNA sequences	80
5.2	The example showing the proposed model, where the position of particle 1 represents one base DNA sequence in 2 dimensions binary search space	81
5.3	User interface program of Continuous PSOLigo for DNA sequence design problem	85
5.4	Average fitness comparison results between Continuous PSO and Binary PSO algorithms, for the case of 7 sequences with 20-mer length	87
5.5	Convergence pattern of $f_{DNA}$ for the Continuous PSO algorithm for the case of 7 sequences with 20-mer length	88
5.6	Convergence pattern of $f_{DNA}$ for the Binary PSO algorithm for the case of 7 sequences with 20-mer length	88
5.7	The position update for two iterations using Continuous PSO algorithm for a 20-mer DNA sequences	89
5.8	The position update for two iterations using Binary PSO algorithm for a 20-mer DNA sequences	90
5.9	Average fitness comparison results between (Deaton <i>et al.</i> , 1996) and the proposed approaches, for the case of 7 sequences with 20-mer length	92

5.10	Average fitness comparison results between (Tanaka <i>et al.</i> , 2002) and Binary PSO method, for the case of 14 sequences with 20-mer length	93
5.11	Average fitness comparison results between (Guangzhao <i>et al.</i> , 2007) and the proposed approach, for the case of 20 sequences with 20-mer length.	96
5.12	Average fitness comparison results between (Zhao <i>et al.</i> , 2007) and Binary PSO method, for the case of 14 sequences with 20-mer length	97
6.1	The basic concept of VEPSO	101
6.2	The ring migration scheme	102
6.3	The implementation of DNA sequence design based on VEPSO algorithm, with 4 objective functions	104
6.4	User interface program of Binary VEPSO for DNA sequence design problem	106
6.5	The control parameters setting for Binary VEPSOligo	107
6.6	A list of non-dominated solutions obtained based on Binary VEPSO. The 1 <sup>st</sup> , 2 <sup>nd</sup> , and 3 <sup>rd</sup> non-dominated solutions are in the marked circle	108
6.7	The objective functions space graph of $H_{measure} - similarity$ pair	110
6.8	The objective functions space graph of $H_{measure} - continuity$ pair	111
6.9	The objective functions space graph of $H_{measure} - hairpin$ pair	111
6.10	The objective functions space graph of $similarity - hairpin$ pair	112
6.11	The objective functions space graph of $hairpin - continuity$ pair	112
6.12	The objective functions space graph of $similarity - continuity$ pair	113

7.1	Difference between well distributed non-dominated solutions and solutions produced by VEPSO on objective function space	118
-----	--	-----

## LIST OF ABBREVIATIONS

A	-	Adenine
ACO	-	Ant Colony Optimization
AFSA	-	Artificial Fish Swarm Algorithms
ANN	-	Artificial Neural Network
BSGA	-	Bee Swarm Genetic Algorithms
C	-	Cytosine
DNA	-	Deoxyribonucleic Acid
G	-	Guanine
GA	-	Genetic Algorithms
MOO	-	Multi-objective Optimization
MOEA	-	Multi-Objective Evolutionary Algorithms
MOPSO	-	Multi-Objective Particle Swarm Optimization
PSO	-	Particle Swarm Optimization
SA	-	Simulated Annealing
SI	-	Swarm Intelligence
T	-	Thymine
$T_m$	-	Melting Temperature

## LIST OF SYMBOLS

$\Lambda$	-	bases of DNA {A, C, G, T}
$x, y$	-	a DNA sequence
$ x $	-	length of $x$ (DNA sequences)
$x_i$ ( $1 \leq i \leq  x $ )	-	$i$ -th nucleotide from 5'-end of sequence $x$
$\Sigma$	-	a set of $n$ sequences with the same length $l$
$\Sigma_i$	-	$i$ -th member of $\Sigma$
$\bar{a}$	-	complementary base of $a$
$l$	-	length of sequence
$n$	-	number of sequences
$C_T$	-	the total oligonucleotide strand concentration
$R$	-	the universal gas constant (Boltzmann's constant)
$\Delta H$	-	enthalpy changes of the annealing reaction
$\Delta S$	-	entropy changes of the annealing reaction

**LIST OF APPENDICES**

<b>APPENDIX</b>	<b>TITLE</b>	<b>PAGE</b>
A	List of Non-dominated Solutions	139
B	List of Publications	145



## **CHAPTER 1**

### **INTRODUCTION**

#### **1.1 Particle Swarm Optimization in Brief**

Swarm intelligence is a discipline that deals with natural and artificial systems composed of many individuals that coordinate using decentralized control and self-organization (Bonabeau, Dorigo and Theraulaz, 1999). In particular, this discipline focuses on the collective behaviours that result from the local interactions of the individuals with each other and also with their environment. Examples of systems studied by swarm intelligence are colonies of ants and termites, schools of fish, flocks of birds, and herds of land animals. Some human artifacts also fall into the domain of swarm intelligence, notably some multi-robot systems, and also certain computer programs that are written to tackle optimization and data analysis problems.

Particle swarm optimization (PSO) (Kennedy and Eberhart, 1995) is a population based stochastic optimization technique for the solution of continuous optimization problems. It is inspired by social behaviours in flocks of birds and schools of fish. In PSO, a set of agents called particles will search for good solutions to a given continuous optimization problem. PSO has been applied in many different problems and has successfully solved this problem better than other algorithms.

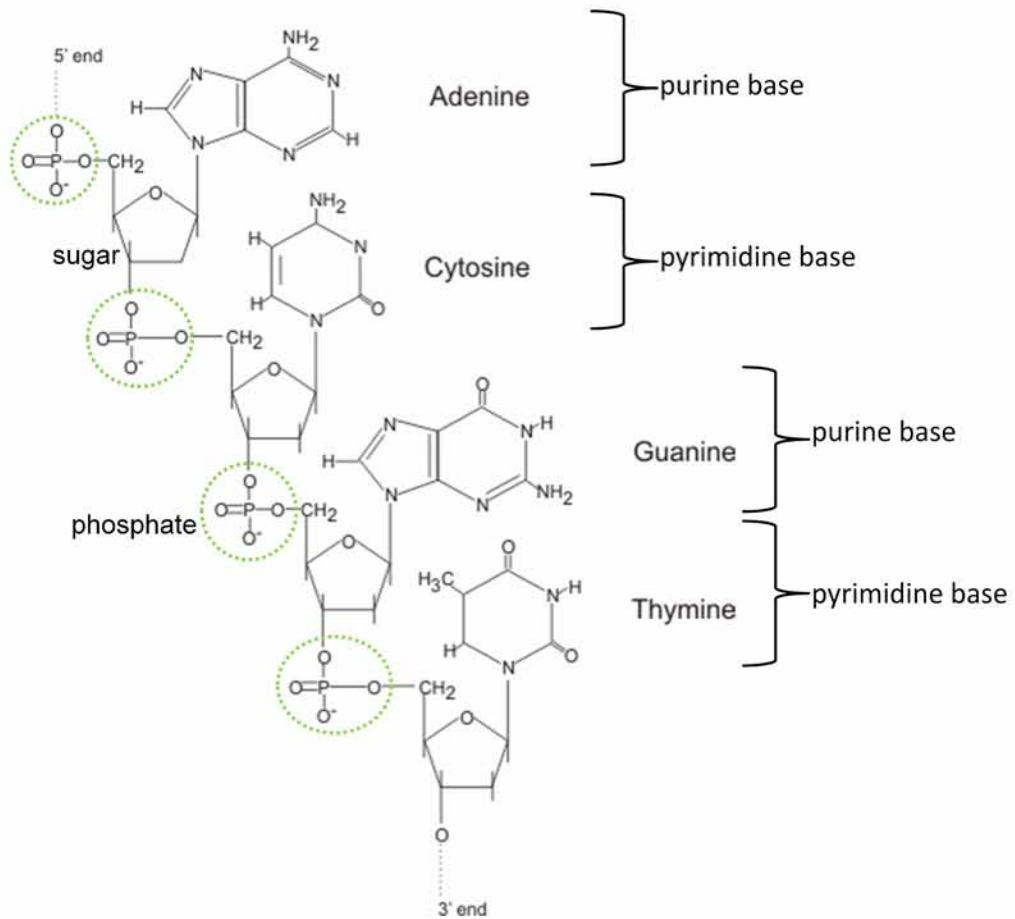
## 1.2 Deoxyribonucleic Acid

Deoxyribonucleic acid (DNA) is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms and some viruses. The main role of DNA molecules in living organism is the long-term storage of information.

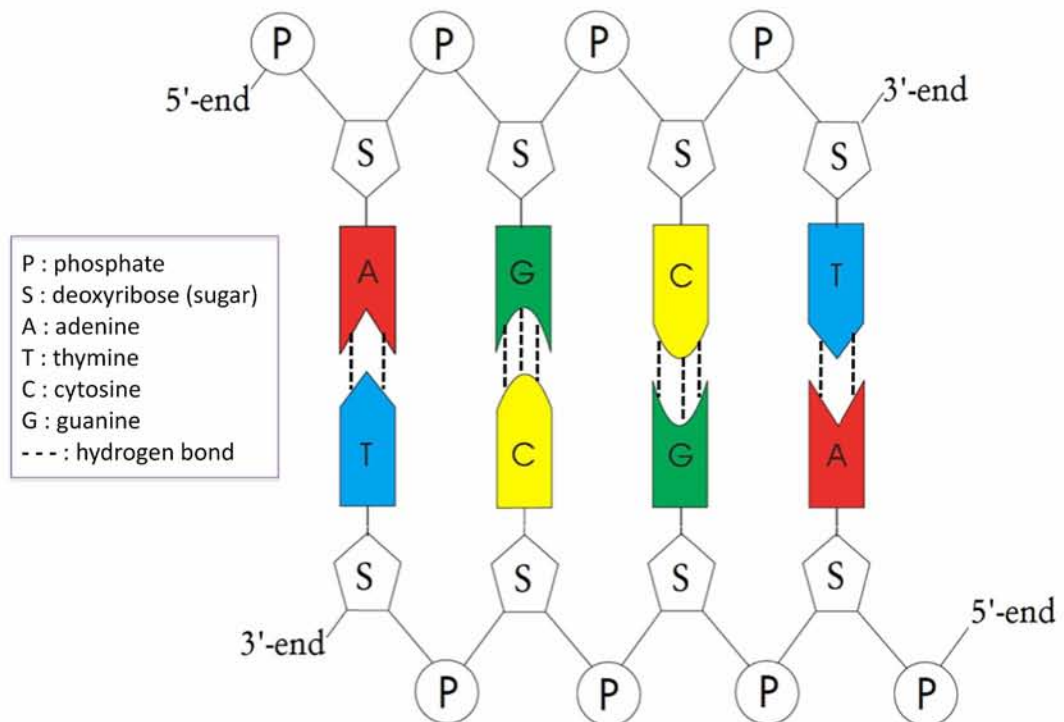
Chemically, DNA is a polymer, which is linked together from a series of monomers. Monomers, which form the structure of nucleic acids, are called nucleotides. Each nucleotide contains a sugar (deoxyribose), a phosphate group, and one of four bases: Adenine (A), Thymine (T), Guanine (G), or Cytosine (C). These bases are grouped into two. Adenine and Guanine are called purine bases because their structure consists of two rings of atoms. On the other hand, cytosine and Thymine are known as pyrimidine bases, since they have a single ring of atoms. Each base has a slightly different composition, or combination of oxygen, carbon, nitrogen, and hydrogen.

A single-stranded DNA consists of a series of nucleotides, which has a sense of direction, in which one end is chemically different than the other. The 5' end terminates in a 5' phosphate group (-PO<sub>4</sub>); the 3' end terminates in a 3' hydroxyl group (-OH), as shown in Figure 1.1. This is important because DNA strands are always synthesized in the 5' to 3' direction. Thus, single-stranded DNA is normally written according to the sequence of nucleobases, from 5' to 3'. As such, a single stranded DNA, 5'-ATCG-3' is normally written as ATCG.

The two single-stranded DNA are held together by hydrogen bonds between pairs of bases, which are called duplex or double-stranded DNA based on Watson-Crick complement. Each type of base on one strand forms a bond with just one type of base on the other strand. The nucleotides only form stable bonds in certain combinations: A pairs with T, and G pairs with C, as shown in Figure 1.2 (Seiffert and Huhle, 2008). Thus, A-T and G-C base pairs are said to be complementary. As shown in Figure 1.2, purines form hydrogen bonds to pyrimidines, with A bonding only to T, and C bonding only to G. This arrangement of two nucleotides binding together across the double helix is called a base pair.



**Figure 1.1:** Chemical structure of each base and phosphate in single-stranded DNA

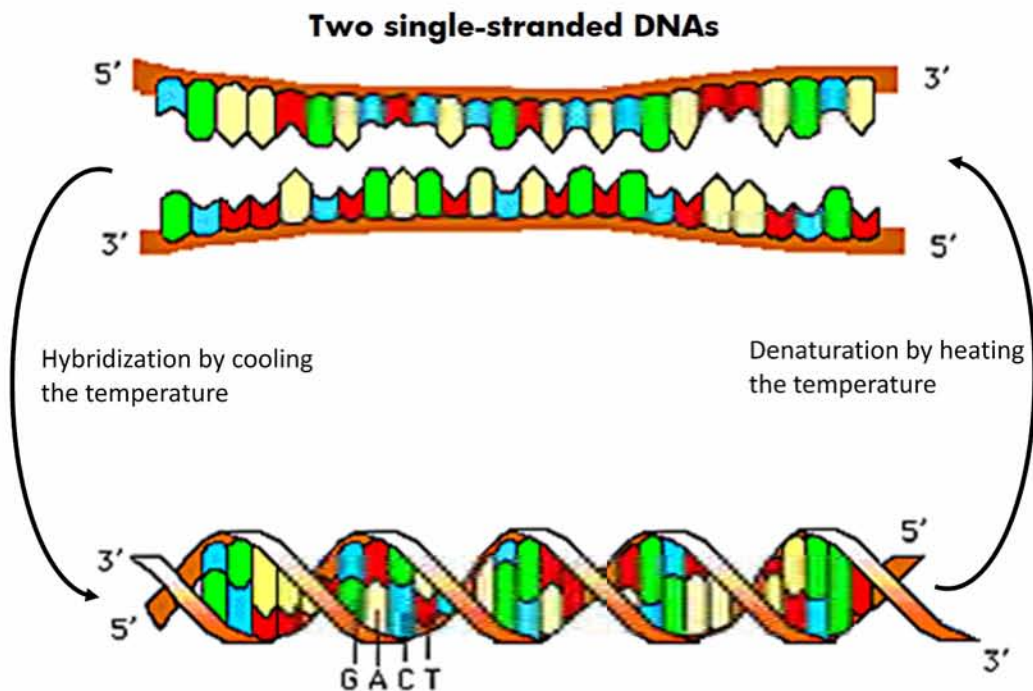


**Figure 1.2:** A double-stranded DNA structure.

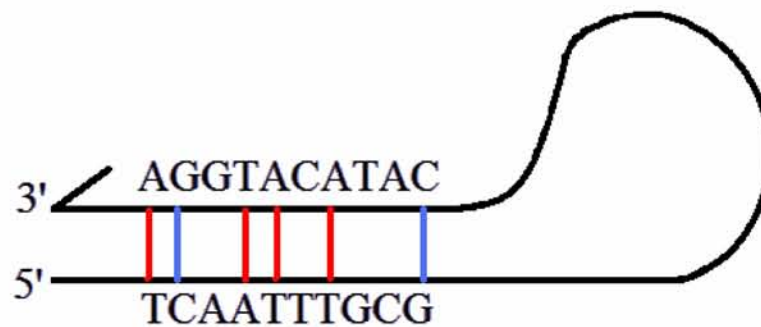
Hybridization is a technique or process that has been found by biochemist Roy Britten in the 1960s as a way to analyse the composition of genome. By understanding hybridization, the process which single DNA strands combine to form a double helix is fundamental to biology and central to DNA-based technologies such as DNA computing, DNA biotechnology, and DNA nanotechnology.

Denaturation is a process where double-stranded DNA strands uncoils and separate into single-stranded DNAs, while hybridization occurs when complementary DNA strands bind, or hybridize based on Watson-Crick base pairing. In denaturation, double-stranded DNAs can be separated by heating up the solution to about 85-95°C, while hybridization can be done by cooling down the test tube reaction solution (Ausubel and Struhl, 1995). There are three types of hybridization, namely bi-molecular hybridization, multi-molecular hybridization, and uni-molecular hybridization. Bi-molecular hybridization occurs when two kinds of single-stranded DNAs form a double helix structure of DNA as shown in Figure 1.3. Meanwhile, three or more strands are involved in the multi-molecular hybridization, which is the essence of Adleman DNA computing (Adleman, 1994). Uni-molecular hybridization or self-hybridization could develop a hairpin formation as shown in Figure 1.4. This would happen if a complementary subsequence exists in the same single-stranded DNA.

DNA hybridization is very sensitive to DNA sequence or composition. Knowledge of how the process occurs could enable researchers to more strategically design technologies. For example, if a researcher wants to design sequences that bind very rapidly or with high efficiency, he or she could place certain bases in specific locations, so that the hybridization reaction could occur faster or more reliably (<http://www.azonano.com/news.asp?newsID=13989>).



**Figure 1.3:** Example of hybridization and denaturation of two single-stranded DNA



**Figure 1.4:** Example of hairpin formation

### 1.3 The Importance of DNA Sequence Design

DNA molecule is presently used in many areas far beyond its traditional function. Unconventionally, the first DNA-based computation has been firstly demonstrated by Adleman in 1994 (Adleman, 1994). He manipulated DNA to solve a combinatorial problem. DNA molecules have also been used as information storage

media and three dimensional structural materials for nanotechnology (Seiffert and Huhle, 2008).

In DNA computing field, one of the major concerns is reliability. The reliability of the computation is determined by whether the oligonucleotides can be hybridized in a predetermined way. In DNA computing, information is encoded as DNA strands. Each DNA strand is composed of short sequences. DNA computing depends on the hybridization, which allows short single-stranded DNA sequences to self-assemble to form long DNA molecules. The key to success in DNA computing is the availability of a large collection of DNA sequences pairs that are unique from one another.

However, the necessity of DNA sequence design appears not only in DNA computing, but also in other fields, such as DNA nanotechnology and biotechnology (Reece, 2004). Due to the differences in experimental requirements, it seems impossible to establish an all-purpose library of sequences that effectively caters all requirements of laboratory experiments. Since the design of DNA sequences depends on the protocol of biological experiments, a method for the systematic design of DNA sequences is highly required (Kashiwamura *et al.*, 2003).

Designing good DNA sequences to control DNA reaction is a fundamental issue in the fields of DNA nanotechnology, DNA computing, and DNA biotechnology. It is important to ensure that DNA molecules react as desired because unexpected secondary structures of DNA sequences may cause unwanted results.

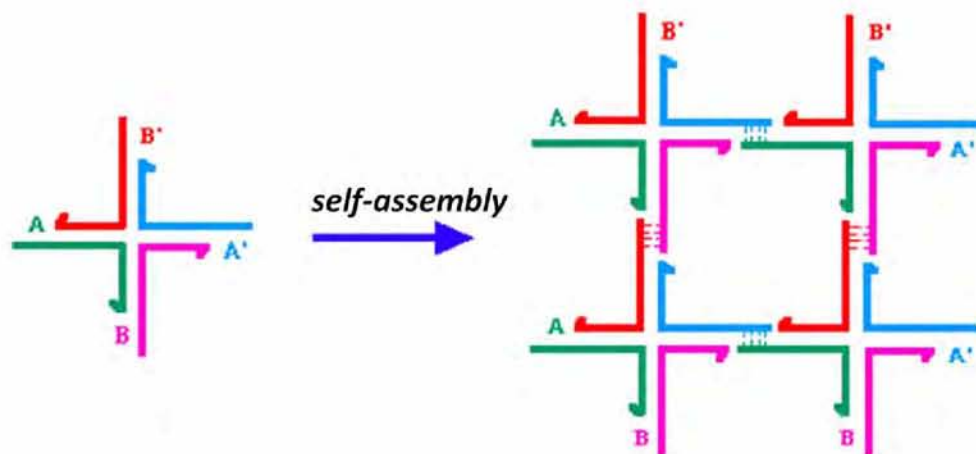
### **1.3.1 DNA Nanotechnology**

Currently, nanotechnology is a growing academic field that has high research potential for development, and has left nearly no part of natural and engineering sciences untouched. There are new materials that have remarkable properties because of their internal nanoscale structure. These remarkable properties have led to the

building of nanomechanical devices, and even the possibility to force single molecules and atoms into well defined arrangements (Seiffert and Huhle, 2008).

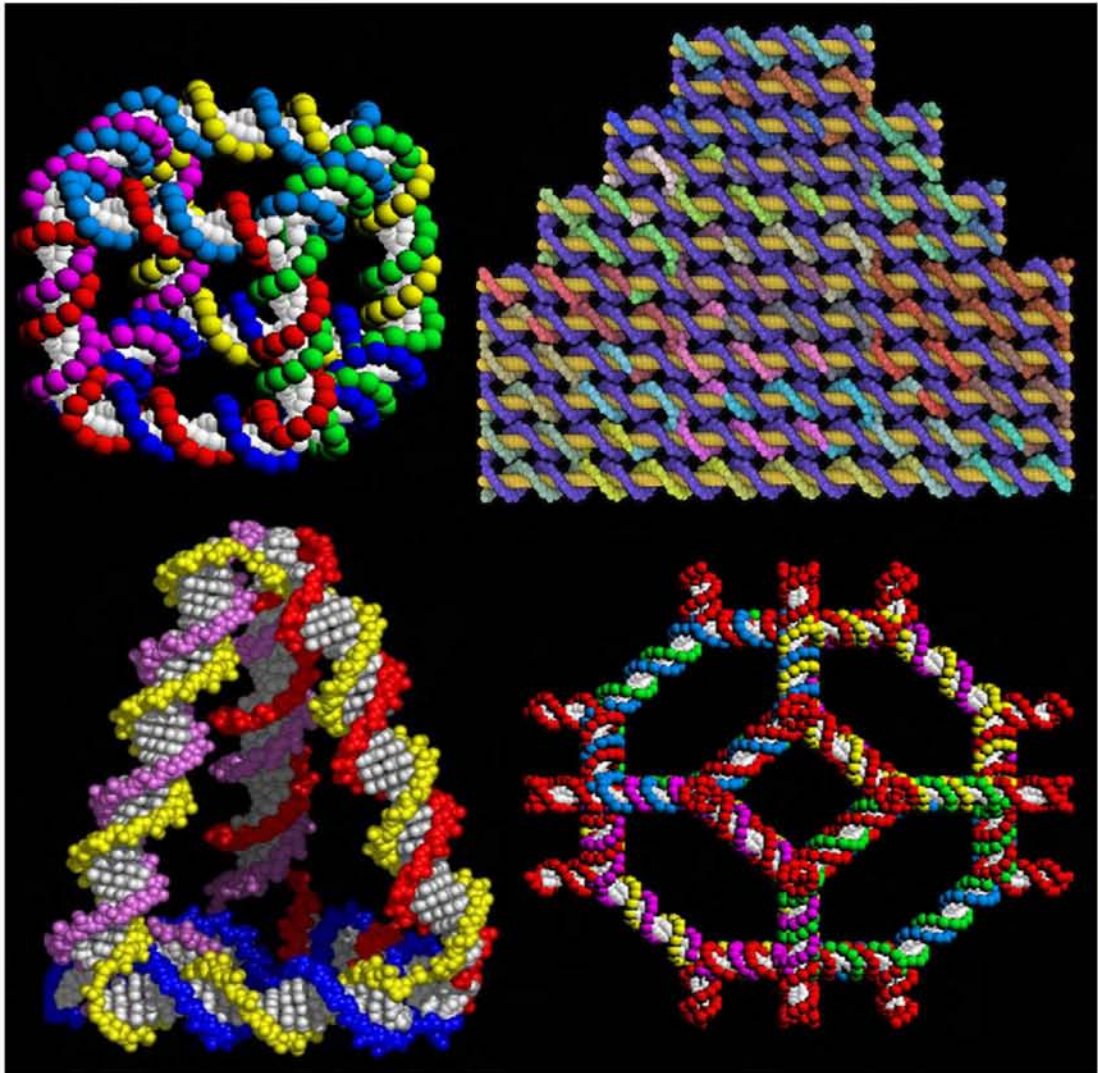
DNA nanotechnology makes use of branched DNA structures to create DNA complexes with useful properties. For example, an assembly of DNA branched four-armed junctions can be made into 2-D lattice, which are complementary to each other in the correct pattern. Based on Watson-Crick base pairing, only portions of the strands which are complementary to each other will hybridize to each other to form duplex DNA. Figure 1.5 shows an example of the lattice made by DNA sequences.

Many different DNA nanostructures have already been built in the laboratory. There are simple junctions, more complex tiles like double-crossover, two-dimensional lattices assembled from building blocks cubes, tubes, and even nanomechanical devices have also been created (Seiffert and Huhle, 2008). As an example, Figure 1.6 shows the DNA nanostructure of cubes, tetrahedra, tile, and truncated octahedron. All these DNA structures were produced by three major steps: (i) structure design (ii) sequence design, and (iii) building the structure by self-assembly in the laboratory. Among these three, sequence design is a very crucial step. The base sequences of the single strands determine the resulting DNA structure. Taking the wrong sequences would produce undesired structures. Therefore, many works have concentrated on producing good DNA sequences to avoid wrong results or structures (Seiffert and Huhle, 2008).



**Figure 1.5:** An example of four-armed junctions, assembled to create 2-D lattice  
(Seeman *et al.*, 1998)





**Figure 1.6:** Examples of DNA nanostructures (Seeman *et al.*, 1998)

### 1.3.2 DNA Computing

Since the Adleman's experiment of Hamiltonian path problem (HPP), DNA computing has become the attention for researchers to overcome the limitations of sequential silicon-based computing (Adleman, 1994). They have paid attention to its high storage density, massive parallelism, and bio-compatible capability (Maley, 1998). In showing of its computing power, DNA computing has been applied to various computational problems (Adleman, 1994), logical problem (Liu *et al.*, 2000),



Boolean circuit development (Owenson *et al.*, 2001), computational model (Mills, 2002), medical problem (Benenson *et al.*, 2004), and associative memory construction (Baum, 1995).

In DNA computing, single-stranded DNAs must hybridize correctly to produce a good solution. Otherwise, DNA computing fails to generate identical results for the same problem and algorithm. Also, DNAs could be wasted if the DNAs perform undesirable reaction. Usually, in DNA computing, the calculation process consists of several chemical reactions, where the successful lab experiment depends on the DNA sequences that have been used. Thus, DNA sequence design turns out to be one of the approaches to achieve high computation accuracy and becomes one of the most practical and important research topics in DNA computing.

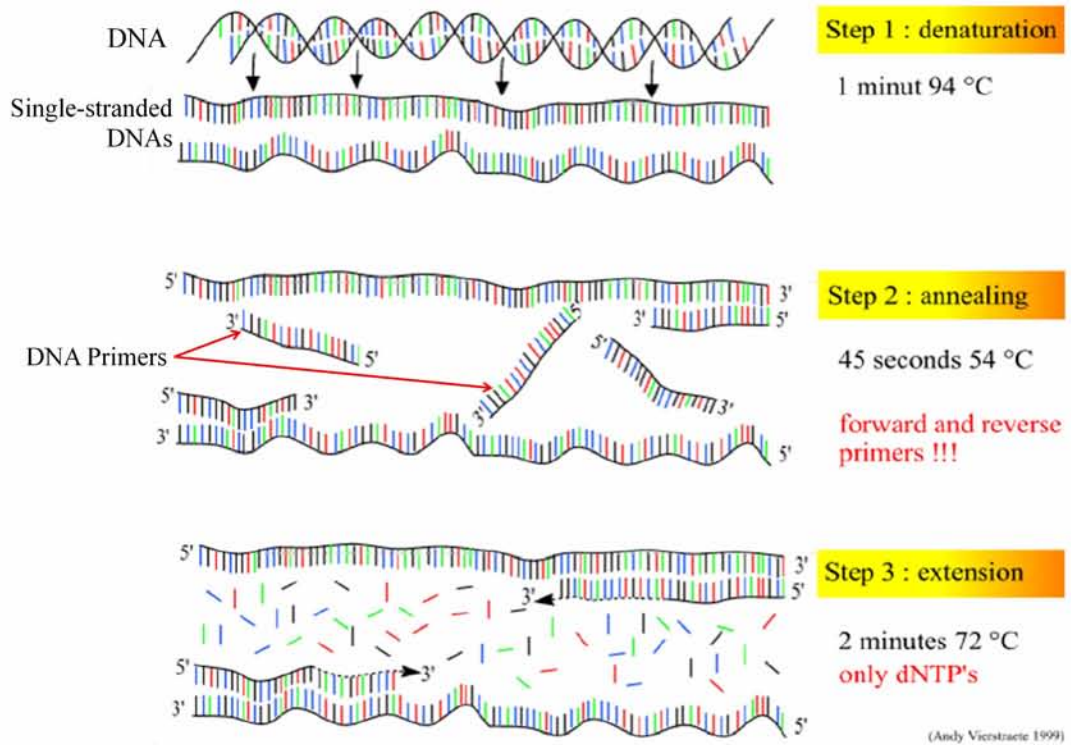
### 1.3.3 DNA Biotechnology

Biotechnology is a technology based on biology, agriculture, food science, and medicine. Biotechnology draws on pure biological sciences such as genetics, microbiology, molecular or DNA biology, biochemistry, and may also depends on knowledge and methods from outside the sphere of biology such as in chemical engineering, information technology, and biorobotics.

One of the popular and necessary techniques used in biotechnology is polymerase chain reaction (PCR). PCR techniques are applied in many areas of biotechnology including protein engineering, cloning, forensics (DNA fingerprinting) and for analysis of environmental samples. PCR is a method for amplifying segments of DNA by generating multiple copies of segments using DNA polymerase enzymes under controlled conditions. PCR can produce  $2^n$  copies of the same molecules in  $n$  cycle (Saaïd, 2008). ‘Primers’, which are usually about 20 bases long are attached to the specific start and end sites of the template for replication. PCR usually runs for 30-40 cycles of 3 phases: denaturation of DNA at about 95°C, annealing at 55°C, and extension at 74°C (Deaton *et al.*, 1996). Figure 1.7 shows the

process of PCR for a cycle, where the primers are needed to be designed to start the annealing process in step 2 (Vierstraete, 1999).

30 - 40 cycles of 3 steps :



**Figure 1.7:** Polymerase chain reaction (Vierstraete, 1999).

#### 1.4 Problem Statement and Objective

DNA sequence design problem is a multi-objective optimization (MOO) problem, where more than one objective needs to be optimized subjected to several constraints.

Given several short-stranded DNAs in a test tube, these DNAs tend to hybridize to other molecule in the tube subject to Watson-Crick complement when the temperature is lowered. The DNA sequence design problem is to avoid these hybridizations when the temperature is lowered. The probability of a sequence to hybridize with itself and other DNAs can be measured using  $H_{measure}$ , *similarity*,

*hairpin*, and *continuity*. These objectives are subjected to  $GC_{content}$  and melting temperature constraints.

Generally, given a number of objective functions and constraints in DNA sequence design, the objective of the problem is to design and produce sets of good DNA sequences with minimized values of the objective functions. If this condition is achieved, it can be said that the sequences in the set are unique and cannot hybridize to each other.

## 1.5 Scope of Work

For the design and development of PSO algorithm for DNA sequence design problem, the scope of this research has been defined as follows:

- (i) The proposed approach is developed by considering four objective functions namely:  $H_{measure}$ , *similarity*, *continuity*, and *hairpin* and two constraints, which are  $GC_{content}$  and *melting temperature* ( $T_m$ ).
- (ii) The research considers Continuous PSO and Binary PSO algorithms for optimizing of DNA sequences.
- (iii) The multi-objective optimization problem of DNA sequence design is solved using aggregation-based method and criterion-based method, particularly vector evaluated PSO (VEPSO).

## 1.6 Thesis Contributions

The first contribution of this work is the new model employed to design DNA sequences using Continuous PSO algorithm. This model represents a dimension of the search space as one sequence. Therefore, a particle in the algorithm carries more

than one sequence, which is a set of DNA sequences. Other researches have implemented PSO algorithm with different representation of model and search space.

The second contribution of this work is the development of Binary PSO algorithm for DNA sequence design. This work is believed to be novel since there is no research work has been carried out to implement Binary PSO to solve the DNA sequence design problem.

The third contribution in this work is the employment of VEPSO method to solve MOO problem of DNA sequence design. In this method, four objectives are represented using four swarms. Each swarm minimizes only one objective function, and by the end of each iteration, these swarms share their information with each other. It is also found that so far no research has employed VEPSO method for DNA sequence design.

## **1.7 Thesis Organization**

The thesis is organized as follows. Chapter 2 provides a review on the existing DNA sequence design approaches as well as the objective functions and constraints used. PSO algorithm is introduced briefly and the applications of PSO are also discussed in this chapter. In Chapter 3, four specific objective functions and constraints are decided and explained in detail.

Chapter 4 covers the concept of PSO algorithm for continuous and binary search spaces. Description on the fundamental of multi-objectives optimization (MOO) problem is also provided. Methods to solve MOO problems are discussed briefly. Chapter 5 is devoted for the implementation of aggregation-based method using Continuous PSO and Binary PSO algorithm for DNA sequence design problem. The results from both models is compared and discussed. This chapter also compares the results from Binary PSO with other existing method.

Chapter 6 provides an implementation of DNA sequence design using vector evaluated PSO (VEPSO) based on Binary PSO. Then, for further analysis, VEPSO is implemented for pairs of two objective functions only. Lastly, Chapter 7 summarizes and concludes this thesis. Future works are also presented.

## REFERENCES

- Abido, M. A. (2002). Optimal Power Flow Using Particle Swarm Optimization. *International Journal Electrical Power Energy System*. 24(7), 563–571.
- Adleman, L. (1994). Molecular computation of solutions to combinatorial problems, *Science*, 266 (11), 1021-1024.
- Adriansyah, A. (2007). *Particle Swarm Fuzzy Controller for Behavior-Based Mobile Robot*. Doctor Philosophy. University Teknologi Malaysia, Skudai.
- Anderson, C., and Franks, N.R. (2001). Teams in animal societies. *Behavioral Ecology*. 12(5), 534-540.
- Arita, M., and Kobayashi, S. (2002). DNA sequence design using templates. *New Generation Computing*, 20 (3), 263-278.
- Arita, M., Nishikawa, A., Hagiya, M., Komiya, K., Gouzu, H., and Sakamoto, K. (2000). Improving sequence design for DNA computing. *In Proceedings of Genetic and Evolutionary Computation Conference (GECCO 2000)*. 8-12 July. Las Vegas, Nevada, USA, 875–882.
- Ausubel, F. M., and Struhl, K. (1995). *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology*. (3<sup>rd</sup> ed.). New York: John Wiley & Sons.
- Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.
- Balci, H.H., Valenzuela, J.F. (2004). Scheduling electric power generators using particle swarm optimization combined with the Lagrangian relaxation method. *Int. J. Appl. Math. Comput. Sci.* 14(3), 411–421.

- Baum, E. B. (1995). Building an associative memory vastly larger than the brain. *Science*. 268 (5210), 583-585.
- Baum, E. B. (1996). *DNA Sequences Useful for Computation*, unpublished, available under <http://www.neci.nj.nec.com/homepages/eric/seq.ps>.
- Baumgartner, U., Magele, C., and Reinhart, W. (2004). Pareto Optimality and Particle swarm Optimization. *IEEE Transactions on Magnetics*. 40 (2), 1172-1175.
- Benenson, Y., Gil, B., Ben-Dor, U., Adar, R., and Shapiro, E. (2004). An autonomous molecular computer for logical control of gene expression. *Nature*. 249 (6990), 423-429.
- Brabazon, A., Silva, A., de Sousa, T. F., O'Neill, M., Matthews, R., Costa, E. (2005). Investigating strategic inertia using OrgSwarm. *Informatica*. 29 (2), 125–141.
- Braendler, D., and Hendtlass, T. (2002). The Suitability of Particle Swarm Optimisation for Training Neural Hardware. In *Proceedings of the 15th International Conference on Industrial and Engineering, Applications of Artificial Intelligence and Expert Systems (IEA/AIE)*. 17-20 June. Cairns, Australia, 190-199.
- Brenneman, A., and Condon, A., (2002). Strand design for biomolecular computation. *Theoretical Computer Science*. 287(1), 39-58.
- Breslauer, K. J., Frank, R., Blöcker, H., and Marky, L. A. (1986). Predicting DNA duplex stability from the base sequence. *Proc. Natl Acad. Sci. USA*. 83 (11), 3746–3750.
- Brits, R., Engelbrecht, A.P., and van den Bergh, F. (2002). A Niching Particle Swarm Optimizer. In *the Proceedings of Fourth Asia-Pacific Conference on Simulated Evolution and Learning*. November 18-22. Singapore, 692-696.
- Bonabeau, E., Dorigo, M., and Theraulaz, G. (1999). *Swarm intelligence: From natural to artificial systems*. New York, USA: Oxford University Press.

- Carlisle, A., and Dozier, G. (2001). An off-the-shelf PSO. *In Proceedings of the Workshop on Particle Swarm Optimization*. April 6-7. Purdue School of Engineering and Technology, Indianapolis, Indiana, USA, 1-6.
- Chang, B.C.H., Ratnaweera, A., Halgamuge, S.K., and Watson, H.C. (2004). Particle swarm optimization for protein motif discovery. *Genetic Programming and Evolvable Machines*. 5(2), 203–214.
- Chatterjee, A., and Siarry, P. (2006). Nonlinear inertia weight variation for dynamic adaptation in particle swarm optimization. *Computers and Operations Research*. 33 (3), 859-871.
- Chen, Y., Dong, J., Yang, B., and Zhang, Y. (2004). A local linear wavelet neural network. *In Proceedings of the fifth world congress on intelligent control and automation*. 15–19 June. Hangzhou, China, 1954–1957.
- Chunxia, X., Zhang, Q., Li, J., and Zhao, X. (2008). A Bee Swarm Genetic Algorithm for the Optimization of DNA Encoding. *3rd International Conference on Innovative Computing Information and Control (ICICIC'08)*. 18-20 June. Dalian, China, 35-38.
- Clerc, M. (1999). The Swarm and the Queen: Towards a Deterministic and Adaptive Particle Swarm Optimization. *Proceeding of the IEEE Congress on Evolutionary Computation*. July 6-9. Washington D.C, USA, 1951-1957.
- Clerc, M. (2000). Discrete Particle Swarm Optimization Illustrated by Traveling Salesman Problem, Technical Report, <http://clerc.maurice.free.fr/pso/>.
- Clerc, M., and Kennedy, J. (2002). The particle swarm - explosion, stability, and convergence in multi dimensional complex space. *IEEE Trans. Evolutionary Computation*. 6(1), 58-73.
- Coello, C.A. (1999). A Comprehensive Survey of Evolutionary-Based Multiobjective Optimization Techniques. *Knowledge and Information Systems*. 1(3), 269-308.



- Coello, C.A., Luna, H., Erika, H., and Aguirre, H. (2003). Use of Particle Swarm Optimization to Design Combinational Logic Circuits. *Lecture Notes in Computer Science, Springer*. 2606, 398–409.
- Coello, C.A., and Lechunga, M.S. (2002). MOPSO: A Proposal for Multiple Objectives PSO. *In Proceedings of the IEEE Congress on Evolutionary Computation*. May 12-17. Honolulu, Hawaii, USA, 1051-1056.
- Conradie, A., Miikkulainen, R., and Aldrich, C. (2002). Adaptive control utilising neural swarming. In: *Proceedings of the genetic and evolutionary computation conference*. 9-13 July. New York, USA, 60-67.
- Cui, G., Cao, X., Zhou, J., and Wang, Y. (2007). The Optimization of DNA Encoding Sequence Based on Improved AFS Algorithms. *In Proceedings of the IEEE International Conference on Automation and Logistics*. 18-21 August. Jinan, Shandong , China, 1141- 1144.
- Das, S., Konar, A., Chakraborty, U.K. (2005a). An efficient evolutionary algorithm applied to the design of twodimensional IIR filters. *In Proceedings of the Conference on Genetic and Evolutionary Computation (GECCO'05)*. June 25 – 29. Washington DC, USA, 2157–2163.
- Das, S., Konar, A., Chakraborty, U.K. (2005b). Improving particle swarm optimization with differentially perturbed velocity. *In Proceedings of the Conference on Genetic and Evolutionary Computation (GECCO'05)*. June 25 – 29. Washington DC, USA, 177–184.
- Deaton, R., Chen, J., Bi, H., and Rose, J. A. (2002a). A Software tool for generating non-crosshybridization libraries of DNA oligonucleotides. *In Proceedings of 8th International Workshop on DNA Based Computers*. June 10-13. Sapporo, Japan, 253-261.
- Deaton, R., Chen, J., Bi, H., Garzon, M., Rubin, H., and Wood, D. H. (2002b). A PCR-based protocol for in vitro selection of non-crosshybridizing oligonucleotides. *In Proceedings of 8th International Workshop on DNA Based Computers*. June 10-13. Sapporo, Japan, 196-204.

- Deaton, R., Garzon, M., Murphy, R. C., Rose, J. A., Franceschetti, D. R., and Steven, S. E. J. (1998). Reliability and efficiency of a DNA-based computation. *Physical Review Letters*. 80(2), 412-420.
- Deaton, R., Murphy, R. C., Garzon, M., Franceschetti, D. R., and Steven, S. E. J. (1996). Good Encodings for DNA-based Solutions to Combinatorial Problems. *Proceedings of the Second Annual Meeting on DNA Based Computers*. 10-12 June. Princeton University, 159–171.
- Deb, K. (2001). *Multi-objective Optimization Using Evolutionary Algorithms*. West Sussex, England: John Wiley & Sons, Ltd.
- Eberhart, R.C., and Shi, Y. (2000). Comparing inertia weights and constriction factors in particle swarm optimization. In *Proceedings of the Congress on Evolutionary Computing*. July 16-19. LA Jolla, California, USA, 84-89.
- Engelbrecht, A.P. (2005). *Fundamentals of Computational Swarm Intelligence*. West Sussex, England: John Wiley & Sons, Ltd.
- Faulhammer, D., Cukras, A. R., Lipton, R. J., and Landweber, L. F. (2000). Molecular computation: RNA solutions to chess problems. *Proc. Natl. Acad. Sci. USA*. 97(4),1385-1389.
- Feldkamp, U., Saghafi, S., Banzhaf, W., and Rauhe, H. (2001). DNA sequence generator—A program for the construction of DNA sequences. In *Proceedings of 7th International Workshop DNA Based Computer*. June 10-13. Tampa, Florida, USA, 179–188.
- Fieldsend, J.E., and Singh, S. (2003). A Multi-Objective Algorithm Based upon Particle Swarm Optimisation. In *Proceedings of the UK Workshop on Computational Intelligence*. September 1-3. University of Bristol, UK, 37–44.
- Fieldsend, J.E. (2004). Multi-Objective Particle Swarm Optimisation Methods, Technical Report # 419, Department of Computer Science, University of Exeter, Exeter, UK.

- Franken, N., and Engelbrecht, A.P. (2004). Evolving intelligent game-playing agents. *South African Computer Journal*. 32, 44-52.
- Frutos, A. G., Liu, Q., Thiel, A. J., Sanner, A. M. W., Condon, A. E., Smith, L. M., and Corn, R. M. (1997a). Demonstration of a word design strategy for DNA computing on surfaces. *Nucleic Acids Research*. 25(23), 4748–4757.
- Frutos, A. G., Thiel, A. J., Condon, A. E., Smith, L. M., and Corn, R. M. (1997b). DNA computing at surfaces: Four base mismatch word design. *In Proceedings of 3rd DIMACS Workshop DNA Based Computer*. June 23-25. University of Pennsylvania, Philadelphia, USA, 238.
- Garzon, M., Neathery, P., Deaton, R., Murphy, R. C., Franceschetti, D. R., and Stevens Jr., S. E. (1997). A new metric for DNA computing. *In Proceedings of Genetic Programming (GP97)*. 13-16 July. Stanford University, Stanford, California, 472–478.
- Garzon, M. H., and Deaton, R. J. (2004). Codeword design and information encoding in DNA ensembles. *Natural Computing*. 3 (3), 253-292.
- Gies, D., Rahmat-Samii, Y. (2004). Vector evaluated particle swarm optimization (VEPSO): optimization of a radiometer array antenna. *In Proceedings of antennas and propagation society - IEEE international symposium*. 3, 2297-3000.
- Guangzhao, C., Yunyun, N., Yangfeng, W., Xunca, Z., and Linqiang, P. (2007). A New approach Based on PSO algorithm to Find Good Computational Encoding Sequences. *Progress in Natural Science*. 17 (6), 712-716.
- Guangzhao, C., Cao, X., Zhou, J., Wang, Y. (2007). The Optimization of DNA Encoding Sequences Based on Improved AFS Algorithms. *IEEE International Conference on Automation and Logistics (ICAL 2007)*. August 18–21. Jinan, Shandong, China, 1141-1144.
- Hartemink, A. J., Gifford, D. K., and Khodor, J. (1998). Automated Constraint-based Nucleotide Sequence Selection for DNA Computation. *In Proceedings of 4th DIMACS Workshop DNA Based Computer*. June 16-19. University of Pennsylvania, USA, 227–235.

- Head, T. (2000). Relativised code concept and multi-tube DNA dictionaries. In Calude, C. S. and Paun, G. (Eds.) *Finite versus Infinite: Contributions to an Eternal Dilemma, Discrete Mathematics and Theoretical Computer Science*. 157-168.
- Heitsch, C.E., Condon, A. E., and Hoos, H. H. (2002). From RNA secondary structure to coding theory: A combinatorial approach. In *Proceedings of 8th International Workshop on DNA Based Computers*. June 10-13. Sapporo, Japan, 215-228.
- Heppner, F., and Grenander, U. (1990). *A Stochastic Non-linear Model for Coordinated Bird Flocks*. S. Krasner, editor, The Ubiquity of Chaos. AAAS Publications.
- Hendtlass, T. (2001). A Combined Swarm Differential Evolution Algorithm for Optimization Problems. *Proceedings of the 14th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE 2001)*. June 4-7. Budapest, Hungary, 11-18
- Hsiao, Y.T., Chuang, C.L., Jiang, J.A. (2005). Particle swarm optimization approach for multiple biosequence alignment. In *Proceedings of the IEEE international workshop on genomic signal processing and statistics*. 22–24 May. Rhode Island, USA.
- Hu, X., and Eberhart, R.C. (2002). Solving constrained nonlinear optimization problems with particle swarm optimization. *6th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2002)*. July 14 – 18. Orlando , Florida, USA.
- Hu, X., and Eberhart, R.C. (2002). Multiobjective optimization using dynamic neighborhood particle swarm optimization. In *Proceedings of the IEEE Congress on Evolutionary Computation*. May 12-17. Honolulu, Hawaii, USA, 1677-1681.
- Hu, X., Eberhart, R.C., and Shi, Y. (2003). Particle Swarm with Extended Memory for Multi-objective Optimization. In *Proceedings of the IEEE Swarm Intelligence Symposium*. April 24-26. Indianapolis, Indiana, USA, 193–197.

- Hussini, S., Kari, L., and Konstantinidis, S. (2003). Coding properties of DNA languages. *Theoretical Computer Science*. 290 (3), 1557-1579.
- Ismail, A., and Engelbrecht, A.P. (1999). Training product units in feedforward neural networks using particle swarm optimization. *In Proceedings of the international conference on artificial intelligence*. 24-26 September. Durban, South Africa, 36–40.
- Jime'nez, J.J., and Ceden˜o, J.R. (2003). *Application of particle swarm optimization for electric power system restoration*. The IASTED International Conference on PowerCon. December 10 – 12. New York, USA.
- Jin, Y., Olhofer, M., and Sendhoff, B. (2001). Evolutionary dynamic weighted aggregation for multiobjective optimization: Why does it work and how?. *In Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2001)*. July 7-11. San Francisco, California, 1042–1049.
- Kashiwamura, S., Kameda, A., Yamamoto, M., and Ohuchi, A. (2003). Two-step Search for DNA Sequence Design. *In Proceedings of the 2003 International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC 03)*. July 7-9, Phoenix Park, Kang-win Do, Korea, 1815–1818.
- Kannan, S., Slochanal, S.M.R., Subbaraj, P., Padhy, N.P. (2004). Application of particle swarm optimization technique and its variants to generation expansion planning problem. *Electric Power Systems Research*. 70(3), 203–210.
- Karpat, Y., O'zel, T. (2006). Swarm-intelligent neural network system (SINNS) based multi-objective optimization of hard turning. *Transactions of North American Manufacturing Research Institute*. 34, 9-16.
- Kassabalidis, I.N., El-Sharkawi, M.A., Marks, R.J.I., Moulin, L.S., Alves da Silva, A.P. (2002). Dynamic security border identification using enhanced particle swarm optimization. *IEEE Trans Power Syst*. 17(3), 723–729.
- Kauffman, S., Levin, S. (1987). Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*. 128, 11–45.

- Kendall, G., Su, Y. (2005). A particle swarm optimization approach in the construction of optimal risky portfolios. *In Proceedings of the 23rd IASTED international multi-conference artificial intelligence and application*. 14–16 Feb. Innsbruck, Austria, 140–145.
- Kennedy, J. (1998). The Behavior of Particles. *Evolutionary Programming*. 581-589.
- Kennedy, J., Eberhart, R.C., and Shi, Y. (2001). *Swarm Intelligence*. San Francisco: Morgan Kaufmann Publishers.
- Kennedy, J. and Eberhart, R. C. (1997). A discrete binary version of the particle swarm algorithm. *In Proceedings of the IEEE 1997 International Conference on Systems, Man and Cybernetics*. October 12-15. Orlando, Florida, USA, 4104-4109.
- Kennedy, J., and Mendes, R., (2003). Neighbourhood Topology in Fully-Informed and Best-of Neighbourhood Particle Swarm. *In Proceedings of the IEEE International Workshop on Soft Computing in Industrial Applications*. June 23-25. New York, USA, 45-50.
- Kennedy, J., and Eberhart, R.C., (1995). Particle Swarm Optimization. *Proceeding of IEEE International Conference on Neural Networks*. IV. Nov. 27 - Dec. 1. Perth, Australia, 1942-1948.
- Khemka, N., Jacob, C., Cole, G. (2005). Making soccer kicks better: a study in particle swarm optimization. *In Proceedings of the Conference on Genetic and Evolutionary Computation (GECCO'05)*. June 25 – 29. Washington DC, USA, 382–385.
- Knowles, J.D., and Corne, D.W. (2000). Approximating the Non-dominated Front using the Archive Evolution Strategy. *Evolutionary Computation*. 8(2), 149-172.
- Ko, P.C., Lin, P.C. (2004). A hybrid swarm intelligence based mechanism for earning forecast. *Asian Journal of Information Technology*. 3(3), 197–201.
- Kurniawan, T. B., Khalid, N. K., Ibrahim, Z., Khalid, M., and Middendorf, M. (2008). An Ant Colony System for DNA Sequence Design Based On

- Thermodynamics. *Proceedings of the Fourth IASTED International Conference Advances in Computer Science and Technology (ACST 2008)*. April 2-4. Langkawi, Malaysia, 144-149.
- Laskari, E.C., Parsopoulos, K.E., and Vrahatis, M.N. (2002). Particle Swarm Optimization for Integer Programming. *In Proceedings of the IEEE Congress on Evolutionary Computation*. May 12-17. Honolulu, Hawaii, USA, 1582-1587.
- Li, Y., Yao, D., Yao, J., Chen, W. (2005). A particle swarm optimization algorithm for beam angle selection in intensity-modulated radiotherapy planning. *Phys Med Biol*. 50, 3491–3514.
- Lin, C.J., Hong, S.J., and Lee, C.Y. (2006). The design of neuro-fuzzy networks using particle swarm optimization and recursive singular value decomposition. *Neurocomputing*. 71, 271-310.
- Liu, Q., Wang, L., Frutos, A. G., Condon, A. E., Corn, R. M., and Smith, L. M. (2000). DNA computing on surface. *Nature*. 403, 175-179.
- Maley, C.C. (1998). DNA computation: Theory, practice, and prospects. *Evolutionary Computer*. 6 (3), 201–229.
- Marathe, A., Condon, A. E., and Corn, R. M. (1999). On Combinatorial DNA Word Design. *In Proceedings of the 5th International Meeting on DNA Based Computers*. June 14 – 15. Cambridge, Massachusetts, 75-89.
- Mauri, G., and Ferretti, C. (2004). Word design for molecular computing: A survey. *In Proceeding of 9<sup>th</sup> DIMACS Workshop on DNA Based Computers*. June 1-3. Madison, WI, USA, 37-46.
- Mehran, R., Fatehi, A., Lucas, C., Araabi, B.N. (2006). Particle swarm extension to LOLIMOT. *In Proceedings of the Sixth International Conference On Intelligent Systems Design and Applications (ISDA'06)*. October 16-18. Jinan, Shandong, China, 969–974.
- Mendes, R., Cortez, P., Rocha, M., Neves, J. (2002). Particle Swarms for Feed Forward Neural Network Training. *In Proceedings of the International Joint*

- Conference On Neural Networks (IJCNN 2002)*. 27-30 August. Madrid, Spain, 1895–1899.
- Mills, A.P. (2002). Gene Expression Profiling Diagnosis through DNA Molecular Computation. *Trends in Biotechnology*. 20(4), 137-140.
- Moore, J., and Chapman, R. (1999). *Application of Particle Swarm To Multiobjective Optimisation*. Technical Report, Department of Computer Science and Software Engineering, Auburn University.
- Mostaghim, S., and Teich, J. (2003). The Role of  $\varepsilon$ -dominance in Multi-objective Particle Swarm Optimization Methods. *In Proceedings of the IEEE Congress on Evolutionary Computation*. December 9-12. Canberra, Australia, 1764–1771.
- Mostaghim, S., and Teich, J. (2003). Strategies for Finding Local Guides in Multi-Objective Particle Swarm Optimization (MOPSO). *In Proceedings of the IEEE Swarm Intelligence Symposium*. April 24-26. Indianapolis, Indiana, 26–33.
- Naka, S., Genji, T., Yura, T., Fukuyama, Y. (2003). A hybrid particle swarm optimization for distribution state estimation. *IEEE Trans Power Syst.* 18(1), 60–68.
- Nenortaite, J. (2005). Computation improvement of stockmarket decision making model through the application of grid. *Information Technology and Control*. 34(3), 269–275.
- Nenortaite, J., and Simutis, R. (2004). Stocks' trading system based on the particle swarm optimization algorithm. *In Workshop on computational methods in finance and insurance (ICCS 2004)*. 7–9 June. Kraków, Poland.
- Omkar, S.N., Mudigere, D., Naik, G.N., and Gopalakrishnan, S. (2007). Vector evaluated particle swarm optimization (VEPSO) for multi-objective design optimization of composite structures. *Computers and Structures*. 86 (1-2), 1-14.



- Omran, M.G., Engelbrecht, A.P., Salman, A. (2005). A color image quantization algorithm based on particle swarm optimization. *Informatica*. 29, 261–269.
- Owenson, G.G., Amos, M., Hodgson, D.A., and Gibbson, A. (2001). A DNA-based logic. *Soft Computing*. 5, 102-105.
- Pareto, V. (1927). *Economie Politique*. (2<sup>nd</sup> edition). Paris: Marcel Giard.
- Parsopoulos, K.E., Vrahatis, M.N. (2002). Particle Swarm Optimization for Imprecise Problems, D. Fotiadis and C. Massalas (eds.), *Scattering and Biomedical Engineering, Modeling and Applications*. 254-264. World Scientific.
- Parsopoulos, K.E., Vrahatis, M.N. (2002). Particle Swarm Optimization Method in Multiobjective Problems, *Proceedings of the Symposium on Applied Computing (SAC 2002)*. March 11-14. Madrid, Spain, 603-607.
- Parsopoulos, K.E., Vrahatis, M.N. (2002). Computing Periodic Orbits of Nonlinear Mappings Through Particle Swarm Optimization. *Proceedings of the 4th GRACM Congress on Computational Mechanics*. June 27-29. Patras, Greece, 2002.
- Parsopoulos, K.E., and Vrahatis, M.N. (2002). Recent Approaches to Global Optimization Problem through Particle Swarm Optimization. *Natural Computing*. 1 (2-3), 235-306.
- Parsopoulos, K.E., Vrahatis, M.N. (2004). UPSO: A Unified Particle Swarm Optimization Scheme. *Lecture Series on Computer and Computational Sciences, Proceedings of the International Conference of "Computational Methods in Sciences and Engineering" (ICCMSE 2004)*. 1, 868-873.
- Parsopoulos, K.E., Vrahatis, M.N. (2004). On the Computation of All Global Minimisers through Particle Swarm Optimization. *IEEE Transactions on Evolutionary Computation*. 8 (3), 211-224.
- Partridge, B.L. (1982). The Structure and Function of Fish School. *Scientific American*. 246, 114-123.

- Penchovsky, R., and Ackermann, J. (2003). DNA library design for molecular computation. *Journal Computer Biology*. 10(2), 215–229.
- Plagianakos, V.P., and Vrahatis, M.N. (2002). Parallel evolutionary training algorithms for “hardware–friendly” neural networks. *Natural Computing*. 1(2–3), 307– 322.
- Rahmat-Samii, Y., Gies, D., and Robinson, J. (2002). Particle Swarm Optimization (PSO): A Novel Paradigm For Antenna Designs. *The Radio Science Bulletin*. 305, 14-22.
- Reece, R. J. (2004). *Analysis of Genes and Genomes*. England: Wiley.
- Reyes-Sierra, M., and Coello, C.A. (2006). Multi-Objective Particle Swarm Optimizers: A Survey of the State-of-the-Art. *International Journal of Computational Intelligence Research*. 2 (3), 287-308.
- Reyes-Sierra, M., and Coello, C.A. (2007). A Study of Techniques to Improve the Efficiency of a Multi-Objective Particle Swarm Optimizer. *Evolutionary Computation in Dynamic and Uncertain Environments*. 269-296.
- Reynolds, C.W. (1987). Flocks, Herds, and Schools: A Distributed Behavioral Model. *Computer Graphics*. 21(4), 25-34.
- Saaïd, M.F.M. (2008). *Clustering Techniques for DNA Computing Readout Method Based On Real-Time Polymerase Chain Reaction*. Master Degree. Universiti Teknologi Malaysia, Johor.
- SantaLucia Jr., J. (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbour thermodynamics. *Proc. Natl Acad. Sci. USA*. 95, 1460–1465.
- SantaLucia Jr., J., Allawi, H. T., and Seneviratne, P. A. (1996). Improved Nearest-Neighbor Parameters for Predicting DNA Duplex Stability. *Biochemistry*. 35(11), 3555-3562.
- Schaffer, J.D. (1985). Multiple Objective Optimization with Vector Evaluated Genetic Algorithms. *Proceedings of the 1st International Conference on Genetic Algorithms*. July. Pittsburgh, PA, USA, 93-100.

- Seeman N.C., Wang H., Yang X., Liu F., Mao C., Sun W., Wenzler L., Shen Z., Sha R., Yan H., Wong M.H., Sa-Ardyen P., Liu B., Qiu H., Li X., Qi J., Du S.M., Zhang Y., Mueller J.E., Fu T.-J., Wang Y., and Chen, J. 1998). New Motifs in DNA Nanotechnology. *Nanotechnology*. 9, 257-273.
- Seeman, N. C. (1990). De Novo Design of Sequences for Nucleic Acid Structural Engineering. *Journal of Biomolecular Structure & Dynamics*. 8(3), 573–581.
- Seiffert, J., Huhle. A., (2008). A Full-automatic Sequence Design Algorithm for Branched DNA Structures. *Journal of Biomolecular Structure & Dynamics*. 25(5).
- Sharpe, F.A. (2000). *Social Foraging of Southeast Alaskan Humpback Whales*. Doctor Philosophy. Simon Fraser University, Burnaby, British Columbia.
- Shi, Y., and Eberhart, R.C. (1998a). A Modified Particle Swarm Optimizer. *IEEE International Conference on Evolutionary Computation*. May 4-9. Anchorage, Alaska, USA, 69-73.
- Shi, Y., and Eberhart, R.C. (1998b). Parameter Selection in Particle Swarm Optimization, *The 7th Annual Conference on Evolutionary Programming*. March 25-27. San Diego, C.A, 591-601.
- Shi, Y. and Eberhart, R.C. (1999). Empirical Study of Particle Swarm Optimization. *Proceeding of the IEEE Congress on Evolutionary Computation*. July 6-9. Washington D.C, USA, 1945-1950.
- Shi, Y. and Eberhart, R.C. (2001a). Fuzzy Adaptive Particle Swarm Optimization. *Proceeding of Congress on Evolutionary Computation*. May 27-30. Seoul, Korea, 101-106.
- Shi, Y. and Eberhart, R.C. (2001b). Particle Swarm Optimization with Fuzzy Adaptive Inertia Weight. *Proceeding of the Workshop on Particle Swarm Optimization*. April 6-7. Indianapolis, USA.
- Shi, Y. and Eberhart, R.C. (2001c). Particle swarm optimization: developments, applications and resources. *Proceeding of Congress on Evolutionary Computation*. May 27-30. Seoul, Korea.

- Shihua, Z., Zhang, Q., Zhao, J., and Li, J. (2007). DNA Encoding Based on Multi-objective Particle Swarm. *Journal of Computational and Theoretical Nanoscience*. 4, 1249-1252.
- Shin, S.Y., Lee, I.H., Kim, D., and Zhang, B.T. (2005). Multiobjective Evolutionary Optimization of DNA Sequences For Reliable DNA Computing. *IEEE Transaction on Evolutionary Computation*. 9 (2), 143-158.
- Spector, I.L., Goodman, E.D., Wu, A., Langdon, W.B., Voigt, H., Gen, M., Sen, S., Dorigo, M., Pezeshk, S., Garzon, M.H., and Burke, E. (2001). Dynamic Weighted Aggregation for Evolutionary Multi-Objective Optimization: Why Does It Work and How? *In Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2001)*. July 7-11. San Francisco, California, 1042–1049.
- Sugimoto, N., Nakano, S., Yoneyama, M., and Honda, K. (1996). Improved thermodynamic parameters and helix initiation factor to predict stability of DNA duplexes. *Nucleic Acids Res.* 24, 4501–4505.
- Sugisaka, M., Fan, X. (2005). An effective search method for neural network based face detection using particle swarm optimization. *IEICE Trans Inf Syst*. E88-D(2), 214.
- Tanaka, F., Nakatsugawa, M., Yamamoto, M., Shiba, T., and Ohuchi, A. (2002). Towards a general-purpose sequence design system in DNA computing. *In Proceedings of the IEEE Congress on Evolutionary Computation*. May 12-17. Honolulu, Hawaii, USA, 73-78.
- Tanaka, F., Nakatsugawa, M., Yamamoto, M., Shiba, T., and Ohuchi, A. (2001). Developing support system for sequence design in DNA computing. *In Proceedings of 7th International Workshop DNA Based Computer*. June 10-13. Tampa, Florida, USA, 340–349.
- Ting, T.O., Tao, M.V.C., Loo, C.K., and Ngu, S.S. (2003). Solving unit commitment problem using hybrid particle swarm optimization. *Journal Heuristic*. 9(6), 507–520.

- Trelea, I.C. (2003). The Particle Swarm Optimization Algorithm: Convergence Analysis and Parameter Selection. *Information Process. Letter.* 85(6), 317-325.
- Tulpan, D. C., Hoose, H., and Condon, A. (2002). Stochastic local search algorithms for DNA word design. *In Proceedings of 8th International Workshop on DNA Based Computers.* June 10-13. Sapporo, Japan, 229–241.
- Ujjin, S., Bentley, P.J. (2003). Particle swarm optimization recommender system. *In Proceedings of the IEEE Swarm Intelligence Symposium.* April 24-26. Indianapolis, Indiana, USA, 124–131.
- van den Bergh, F. (1999). Particle swarm weight initialization in multi-layer perceptron artificial neural networks. Development and practice of artificial intelligence techniques. IAAMSAD, Durban, South Africa, 41–45.
- van den Bergh, F., Engelbrecht, A.P. (2000). Cooperative learning in neural networks using particle swarm optimizers. *South Africa Computer Journal.* (26), 84–90.
- van den Bergh, F., and Engelbrecht, A.P. (2002). A New Locally Convergent Particle Swarm Optimizer. *In Proceedings of the IEEE International Conference on system, Man and Cybernetics.* October 6-9. Yasmine Hammamet, Tunisia, 96-101.
- van den Bergh, F., and Engelbrecht, A.P. (2001). Effects of Swarm Size on Cooperative Particle Swarm Optimizers. *In Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 2001).* July 7-11. San Francisco, California, 892-899.
- Veldhuizen, D., and Lamont, G. (2000). Multiobjective Evolutionary Algorithms: Analyzing the State-of-the-Art. *Evolutionary Computation.* 8(2), 121-147.
- Vierstraete, A., (1999). *Insertional RNA editing in metazoan mitochondria: The cytochrome b gene in the nematode Teratocephalus lirellus.* A Publication Of The RNA Society.

- Vlachogiannis, J.G., and Lee, K.Y. (2005a). Determining generator contributions to transmission system using parallel vector evaluated particle swarm optimization Power Systems. *IEEE Transactions*. 20 (4), 1765 – 1774.
- Vlachogiannis, J.G., and Lee, K.Y. (2005b). Reactive Power Control Based on Particle Swarm Multi-Objective Optimization. *In Proc. of 13<sup>th</sup> International Conference on Intelligent Systems Application to Power Systems*. November 6-10. Arlington, VA.
- Voss, M.S. (2003). Social programming using functional swarm optimization. *In Proceedings of the IEEE Swarm Intelligence Symposium*. April 24-26. Indianapolis, Indiana, USA.
- Voss, M.S., Feng, X. (2001). Emergent system identification using particle swarm optimization. *Complex adaptive structures conference*, Hutchinson Island, FL.
- Voss, M.S., Feng, X. (2002). A new methodology for emergent system identification using particle swarm optimization (PSO) and the group method of data handling (GMDH). *In Proceedings of genetic and evolutionary computation conference*. July 9-13. New York, NY.
- Wachowiak, M.P., Smolikova, R., Zheng. Y., Zurada, J.M., Elmaghraby, A.S. (2004). An approach to medical biomedical image registration utilizing particle swarm optimization. *IEEE Transaction of Evolutionary Computing*. 8(3), 289–301.
- Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T., and Itakura, K. (1979). Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. *Nucleic Acids Research*. 6(11), 3543–3557.
- Wetmur, J. G. (1991). DNA probes: applications of the principles of nucleic acid hybridization. *Critical Reviews in Biochemistry and Molecular Biology*. 26, 227–259.
- Xiao, X., Dow, E.R., Eberhart, R.C., Miled, Z., Oppelt, R.J. (2003). Gene clustering using self-organizing maps and particle swarm optimization. *In Proceedings*

*of second IEEE international workshop on high performance computational biology.* April 22. Nice, France.

- Xiao, J., Xu, J., Chen, Z., Zhang, K., and Pan, L., (2009). A hybrid quantum chaotic swarm evolutionary algorithm for DNA encoding. *Computer and Mathematic with Application.* 57, 1949-1958.
- Yen, G.G., and Lu, H. (2003). Dynamic Population Strategy Assisted Particle Swarm Optimization. *In Proceedings of the IEEE International Symposium on Intelligent Control.* October 5-8. Texas, USA, 697–702.
- Yoshida, H., Kawata, K., Fukuyama, Y., Takayama, S., Nakanishi, Y. (2001). A Particle Swarm Optimization For Reactive Power And Voltage Control Considering Voltage Security Assessment. *In Proceedings of Power Engineering Society Winter Meeting.* Jan 28- Feb 1<sup>st</sup>. Columbus, Ohio, USA, 498.
- Zhang, Y., and Huang, S. (2003). Multi-objective Optimization using Distance-based Particle Swarm Optimization. *In Proceedings of the International Conference on Computational Intelligence, Robotics and Autonomous System.* December 15-18. Singapore.
- Zhang, K., Xu, J., Geng, X., Xiao, J., and Pan, L. (2008). Improved taboo search algorithm for designing DNA sequences. *Science Direct Progress in Natural Science.* 18, 623-627.
- Zhang, L., Liu, C., Ma, Z., Ma, M., Liang, Y. (2003). Solving Multi-objective Optimization Problems Using Particle Swarm Optimization. *In Proceedings of the IEEE Congress on Evolutionary Computation.* 4, 2400-2405.
- Zhang, B.T., and Shin, S.Y. (1998). Molecular algorithms for efficient and reliable DNA computing. *In Proceeding of Genetic Programming (GP98).* July 22-25. Madison, Wisconsin, USA, 735-742.
- Zitzler, E., Deb, K., and Thiele, L. (2000). Comparison of Multi objective Evolutionary Algorithms: Empirical Results. *Evolutionary Computation.* 8(2), 173–195.

Zhou, S., Zhang, Q., Zhao, J., and Li, J. (2007). DNA Encodings Based on Multi-Objective Particle Swarm. *Journal of Computational and Theoretical Nanoscience*. 4, 1249-1252.