uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

u Ottawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Mengtao Wen

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.A.Sc. (Electrical Engineering)

GRADE / DEGREE

School of Information Technology and Engineering

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Optical design by Genetic Algorithm

TITRE DE LA THÈSE / TITLE OF THESIS

Jianping Yao

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Trevor Hall

Barry Syrett

Gary W. Slater

LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES /
DEAN OF THE FACULTY OF GRADUATE AND POSTDOCORAL STUDIES

Optical design by Genetic Algorithm

By

Mengtao Wen

A thesis submitted in partial fulfillment of
the requirements for
the degree of

Master of Applied Science

Ottawa-Carleton Institute of Electrical and Computer Engineering
School of Information Technology and Engineering
Faculty of Engineering
University of Ottawa

May 2005

© 2005, Mengtao Wen, Ottawa, Canada

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-494-11454-1
Our file Notre référence
ISBN: 0-494-11454-1

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

The design of holographic diffusers and birefringent filters by using the genetic algorithm is studied in this thesis.

A holographic diffuser is an optical device that distributes light from a light source with a desired spatial distribution pattern. For diffuse infrared wireless home networking, a holographic diffuser is used to diffuse the laser light to cover a large indoor area such as an office, and at the same time to solve the eye safety problem. In the thesis, a modified genetic algorithm is proposed to design holographic diffuser. The novel algorithm combines the conventional genetic algorithm and the simulated annealing algorithm, in which the simulated annealing algorithm is used to maintain a better diversity of chromosomes for the genetic algorithm. A better performance in locating the global minimum is demonstrated. A 4-level phase-only hologram that is designed by the modified genetic algorithm is fabricated on a quartz substrate. The fabricated hologram is experimentally verified. The results show that diffraction pattern generated by the fabricated hologram agrees well with the theoretically calculated pattern.

Fiber birefringent filters are very important optical devices that can find many applications in fiber optic systems. In the thesis, we propose to use the genetic algorithm to design birefringent filters. In the design, the section lengths and rotation angles between adjacent

sections are the two parameters to optimize. Conventional digital filter design algorithms can be used to design filters with one parameter to be optimized while the other is fixed, such as the design of a Solc filter and a Lyot filter. For fiber birefringent filters, however, the two parameters need to be optimized simultaneously. As a forward search algorithm, the genetic algorithm is well suited for this design problem. In this thesis, birefringent filters designed using the genetic algorithm is proposed. The performance of the birefringent filters designed using the genetic algorithm is better than those designed by the conventional digital filter design algorithms. A birefringent filter is constructed using polarization maintaining fibers. Experimental results are obtained and compared with the theoretical calculations.

Finally, a general discussion on the genetic algorithm for the optical design with large problem space is provided. In general, the performance of the genetic algorithm can be improved by increasing the population diversity and reducing the problem space. The increase of the population diversity is implemented in the holographic diffuser design which is achieved by incorporating the simulated annealing as a mutation operator in the genetic algorithm. The reduction of the problem space is implemented in the birefringent filter design, in which we use the processed fitness instead of original fitness, since the processed fitness contains the information of the already searched problem space.

TABLE OF CONTENTS

LIST OF FIGURES

Number ***Page***

LIST OF TABLES

Number ***Page***

ACKNOWLEDGEMENT

With a deep sense of gratitude, I wish to express my sincere thanks to my supervisor, Dr. Jianping Yao, for his immense help in planning the work. He has been providing constant encouragement, valuable suggestions, discussions and directions to my thesis work.

Special thanks are due to Mr. Damon Wong and Dr. George C. K. Chen from Nanyang Technological University of Singapore, who fabricated and tested the holographic diffuser of my design.

I would also like to thank Mr. Fei Zeng and Mr. Zhichao Deng from the Microwave Photonics Research Laboratory, University of Ottawa. During my experiment of the birefringent filter, they helped me with the experimental setup and provided me with valuable suggestions.

PUBLICATIONS:

M. Wen and J. P. Yao, D. Wong, and G. Chen "Holographic diffuser design using a modified genetic algorithm," SPIE Optical Engineering, vol. 44, no. 8, August 2005.

Chapter 1

Introduction

1.1 Background

In optical design, we always face problems of parameter optimization. One of the examples is the holographic diffuser design [1-8]. Holographic diffusers can be used to extend a collimated laser beam to cover a broad range, which solve a major issue of eye safety problem in the infrared wireless home networking [2]. Holographic diffusers are usually designed using the computer generated hologram (CGH) technique which was proposed by Brown and Lohmann in 1966 [1]. For the binary CGH we need to determine the phase of each pixel of a hologram in order to obtain maximum power output. In the past decades, different optimization methods have been applied to the design of CGHs [2] [4-8]. For example, Fienup [6] reported an iterative method that uses a succession of Fourier transforms to generate a hologram. For this method, a comparison is made between the reconstructed image and the original image to obtain the minimal difference while the required constraints in each domain are imposed. Yao et al. [2] investigated and implemented other methods including the error-reduction, input-output and simulated annealing to design a holographic diffuser. The research showed that simulated annealing achieves better performance than the error-reduction and input-output methods [2].

Another example of parameter optimization is the birefringent filter design [9-14]. Birefringent filters can transmit a narrow band optical signal centered at any selected wavelength and a have periodic spectral response. Therefore, they are widely used in fiber-optic systems and astronomical systems [9] [15-17]. The design of a birefringent filter is to find the section lengths of the birefringent crystals or fibers and the orientation angles between any adjacent sections to satisfy a desired spectral response [9-14]. Harris et al. [13] made a great contribution to this area by describing a generalized synthesis procedure for birefringent filters specified by their spectral response. Chu and Town [11] reported an inverse transform algorithm to design a birefringent filter in which they used Remez exchange algorithm to find a polynomial and used equations deduced from Jones Matrix to convert the polynomials to the orientation angles. The classical methods usually require a relatively complex mathematical model to describe the object function. However, in some cases the theoretical model is so sophisticated that one has to make some assumptions. For example, the algorithm proposed by Harris et al. [13] assumes the birefringent sections to have equal length and used a Fourier series approximation to calculate the filter coefficients. These assumptions may limit the applicability of the filters since we have less freedom to design the filter. Therefore, the classical methods usually require many birefringent sections to achieve a birefringent filter with low sidelobe. Furthermore, these methods are usually based on a specific mathematical model which is only applicable to a specific problem. For birefringent filter design, the method which is applicable to the Solc filters might not be applicable to the Lyot filters.

The common characteristic of these two examples is that although the object function is determined by different parameters, it is difficult to build a mathematical model to evaluate the parameters directly from the object function. The equations used in the classical method to evaluate the parameters have to use some approximations. These approximations might limit a further improvement of the performance. Even with these approximations, the derivation of those equations is an intensive job that requires lots of in-depth knowledge for those problems. Therefore, a more general algorithm which is more applicable and does not require intensive knowledge of the object problems is being pursued in the recent years. As discussed above, there should be two requirements for such an algorithm. First, it must be problem independent so it can be used under different environments even if we cannot describe our object functions very well. Second, the performance of the algorithm must match or surpass the classical methods. The simulated annealing (SA) is a general algorithm which is used widely in different areas [18] [19]. In some cases, it does not perform as well as the classical methods since it relies excessively on the cooling speed and it lacks sufficient method to search for the optimal solution in a large search space.

The genetic algorithm (GA) is another optimization algorithm that has been extensively studied in the past few years [20-23]. One of the advantages of the GA is that it provides more operators than the SA. It has more freedom to control the search for the parameters and usually performs better. Another advantage of the GA over the SA is that it can provide family solutions since it is implemented by a number of candidates instead of a

single candidate in the SA. In optical design such as birefringent filter design, a series of solutions can be used to realize one specified spectral response. Such a characteristic of the GA is especially beneficial when one deals with the multimodal problem which is common in optical design.

The principle of the GA is simple and direct. It imitates the process of nature evolution by maintaining a pool of population of candidate solutions [20]. Different individuals in nature compete for food generation after generation. It is with higher probability for the stronger individuals to survive and the weaker individuals to die out. The stronger individuals have better chance to mate and will have more offspring. Their offspring inherits a similar characteristic from their ancestors. Therefore, the offspring of the stronger individuals will be more competent to survive than the offspring of the weaker ones. The object of the nature evolution is to generate the individuals that are most adaptable to the environment. When the genetic principle of selecting the most adaptable individual is applied to our optimization problem, the fitness of the object function is similar to the adaptability of the individual of nature evolution. After iterations of genetic operators such as selection, cross-over and mutation, we will find the parameters that result in the object function with the highest fitness.

1.2 Objectives of this research

As we have introduced above, many algorithms have been applied to design holographic diffusers and birefringent filters. Those algorithms usually lack of multiple solutions and are specified for one kind of problems. The objectives of this research are to use the genetic algorithm to design holographic diffusers, and to design birefringent filters. Although the GA has been applied to solve different parameter optimization problems in other areas, the design in the optical area is quite different since the problem space in optical design is usually large. In this thesis, improved genetic algorithms for holographic diffuser and birefringent filter designs are investigated. In addition, by studying the two design problems we also investigate in detail the principle of the genetic algorithm for more general optical designs; some clues for the design of similar optical devices are presented. The holographic diffuser and the birefringent filter designed based on the GA are fabricated and their performance are characterized.

1.3 Major Contributions

1. Holographic diffuser design using a modified genetic algorithm is proposed. The algorithm combines the GA and the SA, in which the phase of the hologram is optimized by the genetic algorithm. The SA is used for the mutation operators to improve the population diversity of the GA. The performance of the modified GA and the SA is

compared. Two designed examples of 16×16 image and 64×64 spot pattern image are presented. A 64×64 holographic diffuser designed by the GA is fabricated and tested.

2. The synthesis of birefringent filters using an improved genetic algorithm is proposed. The shift angles and section lengths are optimized by the genetic algorithm. The binary encoding and the floating encoding scheme are presented and compared. Three birefringent filters of 4 and 8 sections with a side band suppression ratio up to 42 dB are presented. The genetic algorithm is further improved to achieve multiple solutions for a specified spectral response. A 4-section birefringent filter is fabricated and tested.

3. The design of a general form of a birefringent filter with both the section lengths and shift angles variable is proposed. The results demonstrate that birefringent filters with adjustable lengths and angles can produce better frequency response than those with only adjustable angles.

4. In this thesis, different methods for genetic operators are investigated. We demonstrate that by carefully controlling the parameter of genetic operators, we can obtain better performance. We also demonstrate that the performance of the genetic algorithm depends much on the search space. The smaller the search space, the better the performance. In this thesis, two methods are suggested to reduce the search space: (1) To divide the search space into sections and (2) Use the processed fitness to force the GA not to explore those already explored space.

1.4 Organization of the thesis

This thesis is organized into 4 chapters. Chapter 1 describes the background, a brief introduction of the genetic algorithm, research objectives and the major contributions. Chapter 2 explains the theory of the holographic diffuser and its encoding schemes including the detour and phase only encoding. A detailed description of using the genetic algorithm to realize the design of holographic diffuser is presented. A comparison between the normal GA, the SA and the modified GA is given. Finally, a holographic diffuser designed by the GA is fabricated and characterized. Chapter 3 explains the theory model of birefringent filters focusing on the Solc filters. A detailed description of using the GA for birefringent filter design is presented. Different encoding schemes including binary encoding and floating point encoding are discussed. Simulation results of 4- and 8-sections of birefringent filters are presented. Finally, a 4-section filter is fabricated and characterized. Chapter 4 makes a comparison between the GA and the digital filter design algorithms from two aspects: the performance and the ability to realize designs for a more general form of birefringent filters. Simulation results of 14-identical and 4-non-identical sections of birefringent filters are presented. Chapter 5 presents a general study of using GA for optical design, focusing on the analysis of the advantages and disadvantages of the GA. The work for the future research is also discussed.

Chapter 2

Holographic diffuser design

The theoretical model of holographic diffusers and the principle of the genetic algorithm are explained and discussed in this chapter. A modified genetic algorithm is proposed for the optimization of holographic diffusers for diffuse infrared wireless home networking. The novel algorithm combines the conventional genetic algorithm and the simulated annealing algorithm, in which the simulated annealing algorithm is used to maintain a better diversity of chromosome for the genetic algorithm. A better performance in locating the global minimum is demonstrated.

2.1 Introduction

Wireless home networking has been a hot area for both academia and industry for the last few years. Compared with RF-based wireless indoor communications, infrared wireless home networking provides many advantages such as broad and unregulated bandwidth, possible spatial reuse of the available bandwidth, absence of electromagnetic interference, and no multi-path fading. A major issue in the infrared wireless networking that has to be considered is the eye safety problem. Eye safety limits the amount of the power to be transmitted, and thus limits the coverage of an optical wireless home networking system. Holographic diffusers can be used to extend a collimated laser beam to cover a broad

range; the eye safety problem can thus be solved. Holographic diffusers are usually designed using the computer generated hologram (CGH) technique which was proposed by Brown and Lohmann in 1966 [1]. Different algorithms such as error reduction, input-output and simulated annealing have been studied for the CGH optimization [2]. It was reported recently that holographic diffusers generated based on the simulated annealing were found to have the best output compared to the error reduction and the input-output techniques [2].

2.1.1 Computer Generated Hologram (CGH)

In 1947, Gabor invented a method to store on photographic film a three-dimensional image by optical interference. The technique is termed holography [3]. In general, holography involves the recording and reconstruction of waves which contains both amplitude and phase information, while a hologram is the coded recording of waves. The principle of conventional holographic recording and reconstruction is shown in Figure 2.1. A coherent reference beam and a wave front diffracted by the object propagate to the hologram plane. The two beams interfere with each other and the resulting intensity pattern is recorded on a photographic film or a plate to form a hologram [2]. In the reconstruction procedure, the reference beam is used again to illuminate the hologram to decode the information from the hologram and reconstruct the object wave.

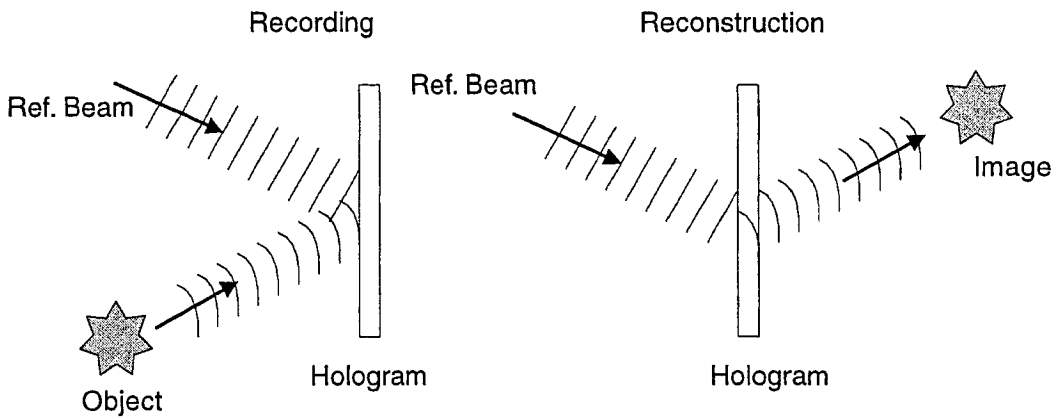


Figure 2.1 Holographic recording and reconstruction.

Unlike the conventional hologram technique that requires a physical object, computer generated hologram (CGH) technique proposed by Brown and Lohmann in 1966 [1] is more flexible in generating the holograms. A mathematical description is sufficient to describe the object and a physical object is not needed. In addition, different encoding techniques that enable physical recording of the complex wave front can be used [1]. For the CGH, Fresnel or Fraunhofer approximations can be employed to compute the wave propagation [2] [4]. The Fraunhofer approximation is applicable to a far-field computation while the Fresnel approximation is applicable to a near-field computation. In the far field situation, the complex amplitude is simply the Fourier transform of the object. For the computer generated hologram, since the distance between the observation plane and the hologram is much larger than the dimension of the pixel size, the computation of the wave propagation can be approximated to the Fourier transform of the object [1] [2].

2.1.2 Encoding Scheme

There are basically two techniques to encode the complex wave front on a physical medium. One encoding scheme proposed by Brown and Lohmann [1] is called detour phase encoding. The detour phase encoding encodes both the amplitude and the phase information of an image, as is shown in Figure 2.2. The hologram that uses the detour phase encoding is a binary (black or white) hologram constituted of rectangles written in cells with their height encoding the amplitude and their shift relative to the center of the cell encoding the phase.

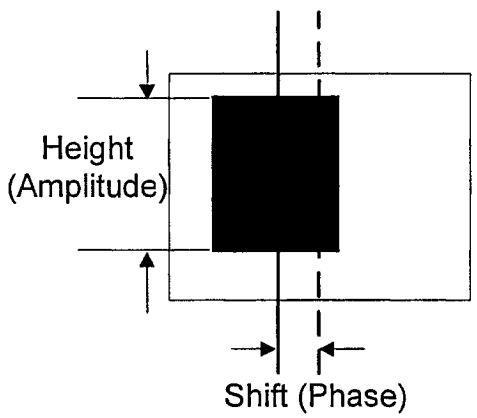


Figure 2.2 Detour phase encoding.

Phase-only encoding scheme is another encoding technique in which only the phase of the light is encoded while the magnitude of the light is set to unity. The phase-only hologram differs from the detour phase hologram in that all the light can pass the phase-only

hologram while the transmittance of some places of the detour hologram might be 0, thus the phase-only encoding can increase the light efficiency of the hologram. In theory the phase-only hologram can provide nearly 100% light efficiency [4].

In our holographic diffuser design, phase-only encoding scheme is used to increase the light efficiency. Phase-only CGHs can be implemented using kinoforms [4], a type of surface relief micro-optical element. Because of its high diffraction efficiency and its flexibility of design, the kinoform can be used in many applications, such as phase-only filters in optical information processing, optical pattern recognition and optical interconnection. Since it is a phase-only optical element generated by a computer, the amplitude of the transfer function is assumed to be unity. The phase information, consisting of ordinary quantized values, is usually represented by a relief profile on a recording material. Therefore, the kinoform can include the reconstruction noise caused by the unitized amplitude and phase quantization. Because reconstruction noise or error is a serious problem in many applications, it is necessary that the phase distribution of the kinoform be optimized to decrease the noise [2] [5] [6].

2.1.3 Cost Function

In the design of a holographic diffuser, the system to be optimized consists of variables and a cost function representing the system configuration. Using the probability process with appropriately controlled parameters, we can find the nearly global minimum of the cost function that corresponds to the optimum condition of the system. For the phase optimization of the kinoform it is assumed that the variable is the phases of the pixels and the cost function is the mean-square error between the reconstructed image and the desired image [2].

Figure 2.3 shows the relationship between the image and the hologram, where $F(m,n)$ is the hologram and $f(k,l)$ is the reconstructed image.

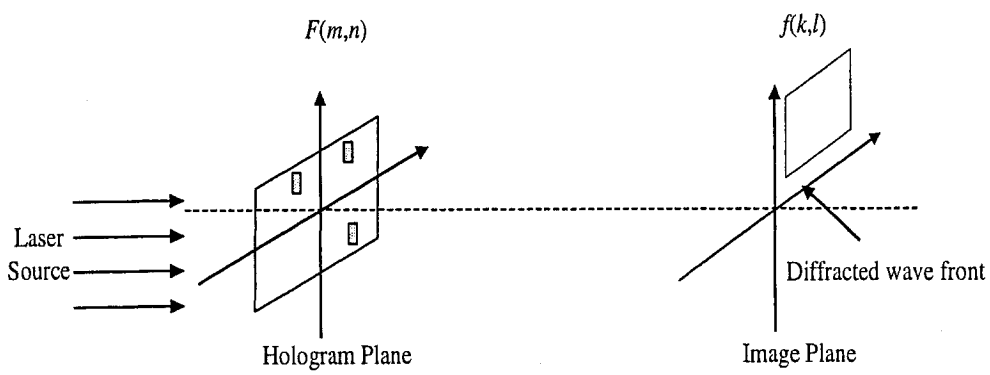


Figure 2.3 The hologram and image planes.

Assume that the hologram has a size of $N \times N$. The Fourier-transformed kinoform reconstructs the image in the far field. The reconstructed image in the image plane is given by [2]

$$f(k,l) = \frac{1}{N^2} \sum_{m=1}^{N} \sum_{n=1}^{N} F(m,n) \exp(2\pi i \frac{mk+nl}{N}). \quad (2.1)$$

In a phase-only hologram, with p equally spaced phase levels, the $F(m,n)$ takes the value of $\exp^{2(p-1)\pi i/P}$ $(p = 1,...,P)$. For example, in the binary and four-phase holograms, P are 2 and 4 respectively, then $F(m,n)$ takes the values:

Binary: ± 1

Four-phase: ± 1 and $\pm i$.

The cost function E can be expressed as [2]

$$E = \sum_{k} \sum_{l} (|\alpha f(k,l)|^2 - |ref(k,l)|^2)^2, \quad (2.2)$$

where α is the normalization factor given by [2]

$$\alpha = \sqrt{N_{ref} / \eta_t N_f}, \quad (2.3)$$

$$N_f = \sum_{k=1}^{N} \sum_{l=1}^{N} |f(k,l)|^2, \quad (2.4)$$

$$N_{ref} = \sum_{k=1}^{N}\sum_{l=1}^{N}\left|ref(k,l)\right|^2, \qquad (2.5)$$

where η_t is the desired diffraction efficiency of the hologram and $ref(k,l)$ refers to the desired image. N_{ref} and N_f measure the total power within the first period of ideal diffracted amplitude and the diffracted amplitude obtained from the hologram, respectively [2].

The optimization procedure is to find the values of $F(m,n)$ in a hologram to produce a reconstructed image $f(k,l)$ with a minimized cost function expressed by Equation 2.2. Therefore, the difference between the desired and reconstructed image is minimized.

2.2. Genetic Algorithm and Simulated Annealing

Several methods have been proposed for the optimization of a kinoform [2-8]. These algorithms can usually be classified into two categories according to their encoding directions. In the first category, the hologram is encoded directly from the object spectrum. The final hologram is obtained by iterations to minimize the errors between the original and the reconstructed images using different algorithms. (1) Sample the original image (2) encode a hologram using the original image (3) compare the reconstructed image with the original image (cost function) (4) use the information of the cost function to revise the

hologram. The error reduction algorithm proposed by Fienup [6] is such an algorithm that the original hologram is encoded by a Fourier transform from the object image and the iterative method is used to improve the hologram. In the second category, the holograms are designed with direct search, in which the original hologram is generated at random. The reconstructed image is compared directly with the original image to find the hologram with minimized cost function. Simulated Annealing [2] [7] is an example of such an algorithm. The advantage of the second approach is that it can avoid the problem of encoding the complex object spectrum. The disadvantage of the second approach is that it always takes longer computation time, especially when the dimension of the hologram is large. Recently, the second approach is more commonly used since it is reported to have better output than the first approach [2] [8].

In this chapter, the optimization of the phase distribution of a kinoform by a modified genetic algorithm (GA) is proposed. This approach can be classified into the second category. It is known the GA is a computational optimizing method that is analogous to the evolution of life [20-23]. It is a stochastic algorithm whose search methods model some natural phenomena: genetic inheritance and Darwinian strife for survival [20-23]: Individuals compete for the food and water; the stronger ones have the higher probability to survive and have more descendents, thus the good genes will propagate fast and the bad genes will disappear gradually. The genetic algorithm works on a population of encoded solutions called chromosomes and the chromosome is composed of a unique string of genes [20-23]. There are basically two types of representation methods for the GA: binary

encoding and floating point encoding. The binary encoding is normally used for discrete model parameter values while the floating point encoding is for continuous model parameter values. From the above discussion, we know that each pixel of the binary phase-only hologram is encoded as -1 or 1 while for the four level hologram is encoded as -1, 1, i or -i, which means that the diffuser design is a discrete problem. Therefore we adopt the binary encoding to represent each pixel of the hologram.

A chromosome is a vector w of n bits

$$w = [w_1, w_2, \cdots w_n] \in G, \tag{2.6}$$

where the genes w_i with $i = 1, 2, ..., n$ have binary values (0 or 1), n is the number of genes and G is the genotypic ensemble of the 2^n possible chromosomes. Each chromosome w corresponds to a unique non-coded solution of the optimization problem [20]. The chromosome is characterized by the performance with respect to objective function (fitness), which is the minimum cost function between the reference image and the hologram-generated image for our problem. The population of the next generation is formed by genetic operators such as selection, cross-over and mutation to their parent population. Chromosomes with higher fitness have higher probability to survive in the next generation. In this way, the GA can locate a relatively global optimum chromosome after certain iterations.

The simulated annealing (SA) was introduced by Kirkpatrick et al. in 1983, which was proposed based on an analogy with the annealing of solids: the object function to be minimized is analogous to the energy of solids and the control parameter is analogous to the temperature of solids [18]. The SA start from a random initial state as the GA does. Random change is generated for the state. The change is accepted if it results in better performance of the object function. If the change results in worse performance, it is accepted with certain probability that is a function of the temperature [18-19]. The temperature is initially high and decreases over time. This probability function is usually given by the Boltzmann distribution,

$$B = \exp(-\Delta E / T), \tag{2.7}$$

where ΔE is the change in the object function. In the diffuser design, ΔE is the cost function between the reference image and the hologram-generated image. T is the parameter of temperature that determines the cooling schedule of the function. From Equation 2.7 we know that the higher the temperature T, the higher the probability function B; which means the change is accepted with higher probability if the change results in a worse cost function. The temperature T is an important parameter that should be chosen carefully. In theory, a slow cooling schedule will have better output but at the cost of longer computation time.

In fact, the SA can be viewed as a special GA with the population size equal to one. Although the GA and the SA are similar in many ways, there are some important differences between them. The leading process in the SA is mutation while selection is the leading process in the GA. Mutation is the only operator used in the SA while 3 basic operators are used in the GA including the selection, mutation and cross-over. The advantage of the GA is that it maintains more solutions than the SA and can search in a broader range. Furthermore, the GA is parallel in nature. However, the GA has one critical shortcoming. The GA does not possess a formal proof of convergence to the global optimum while the SA possesses one as long as the cooling schedule is slow enough [24], although in practice there must be a trade-off between the cooling schedule and the speed of the program. For the GA, when a chromosome $W=[w_1, w_2, \cdots w_n]$ has a large number of genes n, the performance is usually poor. The reason is that the GA cannot maintain population diversity itself. The population diversity plays an important role in obtaining a better output of the hologram. As can be observed by the propagation of human beings, the cross-over between close relatives is more likely to produce worse descendents while the hybrid descendents are usually better. For the problems with large number of genes, some important genes may be lost during the search for global optimum which will result in the descendents to converge to conformity at last. Although large population size may result in better performance, we have to select proper population size to reduce the computation complexity. For an image with size of 64×64 and with the representation of quantization level 4, the size of the ensemble of solutions is 2^{8192}. It is a

huge value as compared with the number of candidates 100 (population size). Therefore, how to enable the candidates to maintain the population diversity so as to keep as many excellent genes as possible is crucial to solve such a problem. In this chapter, we propose a modified GA which combines the merit of the SA and the GA to design the holographic diffuser with large ensembles of solutions.

2.3 Description of the algorithm

The optimization procedure using the modified GA consists of the following six steps:

(1) Initialization

The phase-distribution of a kinoform with a size of $N \times N$ pixels is encoded as an individual chromosome, which is represented by a $N \times N$ matrix. Each gene of the chromosome is ranged within the maximum quantization level. At the beginning, all the chromosomes are generated at random. In our simulations, we use a quantization level of 4 to represent the kinoform. The quantization level must be greater than 2 to avoid the problem of conjugate-image generated by the CGH [2]. The higher the quantization level, the better the image quality. But holographic diffusers with higher quantization level are more difficult to fabricate. A good trade-off in the diffuser design is to use a quantization level of 4. The population size in the proposed modified GA is 100, a value that is commonly used in GAs [21].

(2) Selection

Selection bases on the fitness of object function and it is used to make those individuals with higher fitness to be selected with higher probability. The selection method can basically be classified into two categories: (1) proportionate schemes and (2) ordinal-based schemes [25]. The proportionate selection schemes select its candidate based on its relative fitness to the others in a population pool. One typical method of such a selection scheme is based on a roulette game [20]. The roulette game selection method states that the fitness Q of each chromosome c_i is ranked in an ascending order, $Q(c_1), Q(c_2), ..., Q(c_p)$. The chromosome c_i is selected for reproduction if

$$Q(c_{i-1}) < r = Q(c_i), \tag{2.8}$$

where r is a uniform random variable ranged from 0 to 1. $Q(c_i)$ is the cumulative selection probability, which depends on the chromosome fitness. In this process, the fitness depends on the cost function E and we can set the fitness to $1/E$ since the cost function is inverses to the fitness. The simulation results of the roulette game show that it results in the search to converge quickly and to be trapped to a local minimum even if the mutation probability is set large.

The ordinal-based scheme is a different selection scheme based on the rank of the fitness in a population pool. One example of the ordinal-based scheme is the tournament selection [25]. A different approach [5] is also investigated here: for each generation n ($n \geq 2$), select $(1-s)$ and s of all the individuals with the best fitness from the parents respectively to form the offspring. Therefore, s best individuals are selected twice and s worst individuals are replaced in the new generation. Here s is the selection probability which in our design is 18%. It is different from the roulette game; this approach is based on the rank of the fitness in a population pool instead of the relative fitness to other population. The advantage of this approach is that it is independent of the fitness distribution and the excellent individuals are treated equally, therefore the pressure for the best genes to propagate is reduced and the population diversity is better maintained. The disadvantage of this approach is that the convergence time is longer. For the diffuser design, the ensemble of solutions is much larger when compared with the population size we used, therefore the population diversity is important for the output performance and the ordinal-based selection scheme is used in this design.

The elitism principle of the selection process proposed to improve the performance of the GA [26] is adopted in our design. For the k-th generation, select n chromosomes with the best fitness from the (k-1)-th generation if the best fitness of the k-th generation is worse than the (k-1)-th generation. The elitism principle can reduce the genetic drift which are resulted from the mutation and cross-over by ensuring the best chromosome to copy its

traits to the next generation [22] [26]. The value of the n best chromosomes determines the selection pressure, the larger the n, the larger the selection pressure. In our design the n is 1 since we must make a trade-off between the selection pressure and population diversity.

(3) Cross-over

Two individuals are selected randomly with the probability

$$r < P_c \tag{2.9}$$

where P_c is the cross-over probability and r is a uniform random variable ranging from 0 to 1 [20]. Cross-over generates new individuals by combining genes from their parents. The methods of cross-over include single-point cross-over and double-point cross-over. The single-point cross-over is to divide each of the parents into two parts and exchange the two parts between the two individuals; double-point cross-over is to divide each of the parents into three parts and exchange the middle parts between the two individuals. Double-point cross-over is reported to have better performance [27] and is used in our diffuser design.

In our design, the cross-over probability P_c is 0.8, which means that P_c of all individuals are crossed over [21]. It is accepted that P_c can not be too low and the value below 0.6 is rarely used [24]. Note that a kinoform is a two-dimensional optical element; therefore for the double-point cross-over, four random variables are generated to divide the kinoform, as can be seen in Figure 2.4

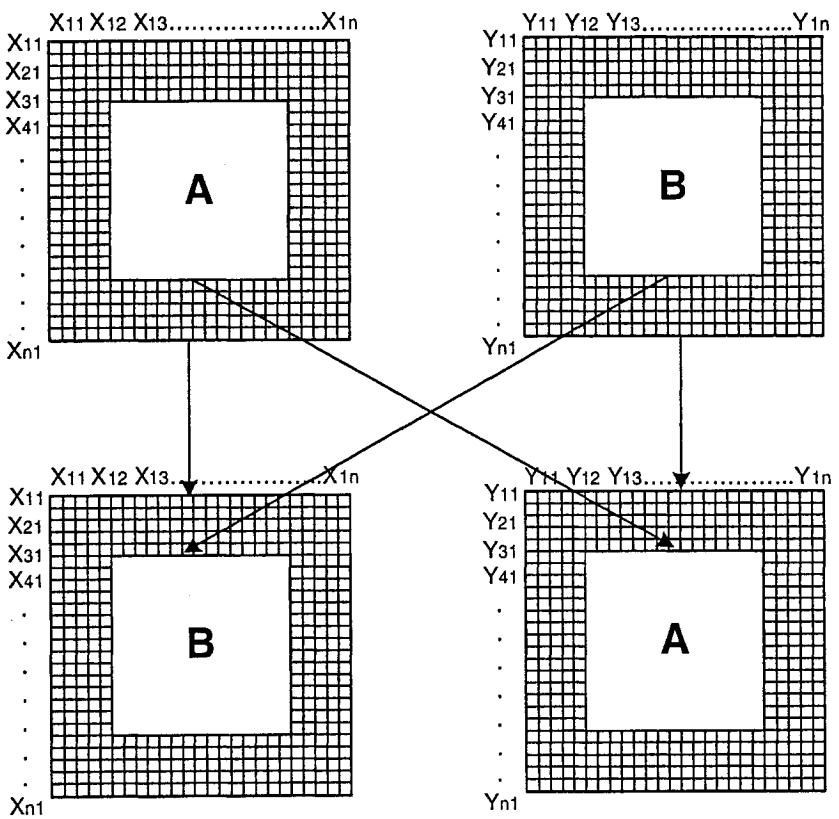


Figure 2.4 Double-point cross-over between two chromosomes.

(4) Mutation

To maintain a good genetic diversity of the genes in a population, the genes are mutated with a mutation probability P_m. The mutation probability is a very important parameter that should be chosen carefully. A premature convergence can result from a too low mutation probability, whereas a slow and poor convergence can result from a high mutation probability [20]. A random variable is generated and each gene is changed if the random variable is less than P_m. In our design, three mutation methods are investigated.

A widely used method to control P_m is to use the "optimally tuned" value, which means that P_m is set to an optimal value by experiments [20-21]. The commonly used value of P_m is between 0.001 and 0.01 [21]. In our simulations, we assign different values to P_m and compare the performance of the algorithm while keeping other parameters (population size, cross-over probability, etc.) the same. For the design problem here, the value of P_m of 0.002 is found to be the optimal value. The "tuned optimal" method is the simplest approach, but it lacks adaptability; since during the evolutionary process, P_m should be higher at first to get wider search range but lower afterwards to mitigate the deterioration of good genes. Furthermore, the effort to find the optimal value is time consuming and other parameters (cross-over probability, etc) have great impact on the optimal value of P_m.

Exponentially decreased P_m is a modification to the "tuned optimal" P_m, in which P_m is not a fixed value but decreases over generations. The value of P_m can be expressed as

$$P_m(i+1) = \alpha P_m(i) \tag{2.10}$$

where i represents the i-th generation. The initial P_m is set to 0.01 and $\alpha = 0.985$. The selection of the parameter of the exponentially decreased P_m is corresponding to the convergence tendency of the GA in this example.

The "1/5 success rule" can also be used to find the optimum value of P_m [20]. The "1/5 success rule" states that the ratio of successful mutations to all mutations should be 1/5. The process to adjust P_m is actually an evolutionary process. After each k generation, the cost functions of parents and descendents are compared. The P_m is increased by a factor of α if more than 1/5 descendents have better cost function than their parents and decrease otherwise. This can be expressed as:

$$P_m = \begin{cases} \alpha P_m, & \text{if } \varphi(k) < 1/5 \\ P_m / \alpha, & \text{if } \varphi(k) > 1/5, \\ P_m, & \text{otherwise} \end{cases} \tag{2.11}$$

where $\varphi(k)$ is the success ration of mutation.

For each of the above methods we use the 16×16 image to simulate for 20 times and their results are shown in Table 2.1.

Table 2.1 Normalized cost functions of different mutation methods

	Max cost function	Min cost function	Ave. cost function
tuned optimal P_m	1	0.7449	0.8153
Exponentially decrease P_m	0.9354	0.7305	0.7769
1/5 success rule of P_m	0.9453	0.7240	0.7772

From Table 2.1, we can see that the exponentially decreased P_m and "1/5 success rule" P_m perform better than the "tuned optimal" method. The performance difference between the exponentially decreased P_m and the "1/5 success rule" P_m is not apparent. However, the disadvantage of the "1/5 success rule" is that it is more computationally intensive since there is a monitoring process during the search. Therefore, as far as the speed and performance are concerned, the approach using exponentially decreased P_m provides the optimal trade-off.

(5) Mutation by SA

Although we choose P_m carefully, it is still with high probability that the search for best cost function converges to a premature optimum. The reason is what we have mentioned before: we have a large ensemble of genes but small population size. After certain iterations, the cost functions of almost all the chromosomes will be the same. To increase P_m can delay the speed of convergence and diversify the population, but the random mutation may destroy excellent genes when the chromosomes converge to nearly optimal. In order to maintain better diversity without destroying excellent genes; we incorporate simulated annealing as a supplemented mutation algorithm into the GA.

As we have discussed above, selection is the leading evolutionist process in the classical GA whereas mutation is the leading evolutionist process in the SA. In the GA, new genes can only be introduced in a population by occasionally random mutations which will lead to better or worse individuals at the same possibilities; whereas mutation in the SA will only lead to worse individuals occasionally. Therefore when the chromosome is close to an optimal solution, the mutation by the SA will lead to a worse individual in much lower probability than the random mutation. Furthermore, even if the mutation by the SA leads to a worse individual, the selection operation of the GA might discard such an individual in the next generation. In this way, the mutation operation is performed and the population diversity is maintained. In our algorithm, the SA acts not only as a way to maintain diversity of the chromosomes, but also acts as a local search for better genes [20] [28].

In order to measure the diversity of chromosomes, we introduce a parameter called variance of fitness (VOF). VOF is derived from the cost function in our design. If VOF equals to zero, that means all the chromosomes are the same. The VOF is large value if the population diversity is large. The process of using the SA is that after every n iterations from steps (2) to (4), we check the VOF. If the VOF is less than threshold value, we randomly select k chromosomes and use the SA to mutate these chromosomes. Therefore, better diversity of chromosomes can be maintained without destroying the excellent genes.

(6) Output the result

Steps (2) to (5) are performed repeatedly until the change of the total fitness of all the chromosome in certain iterations is less than a minimum value we define. The chromosome with the best cost function is then what we need. The process is shown in Figure 2.5.

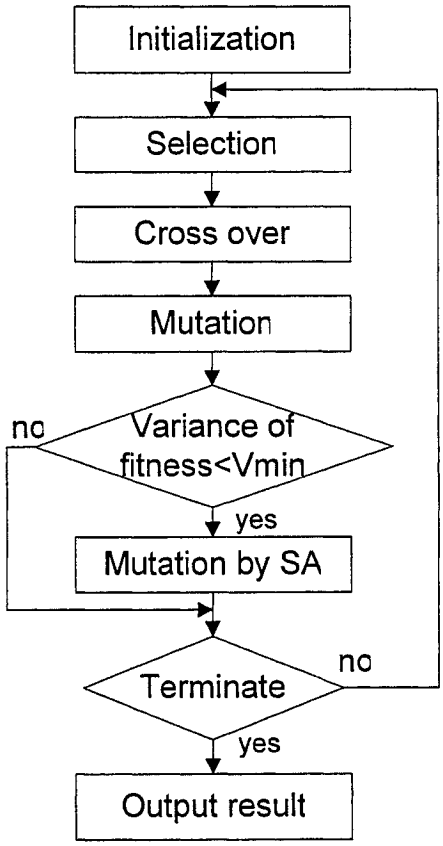
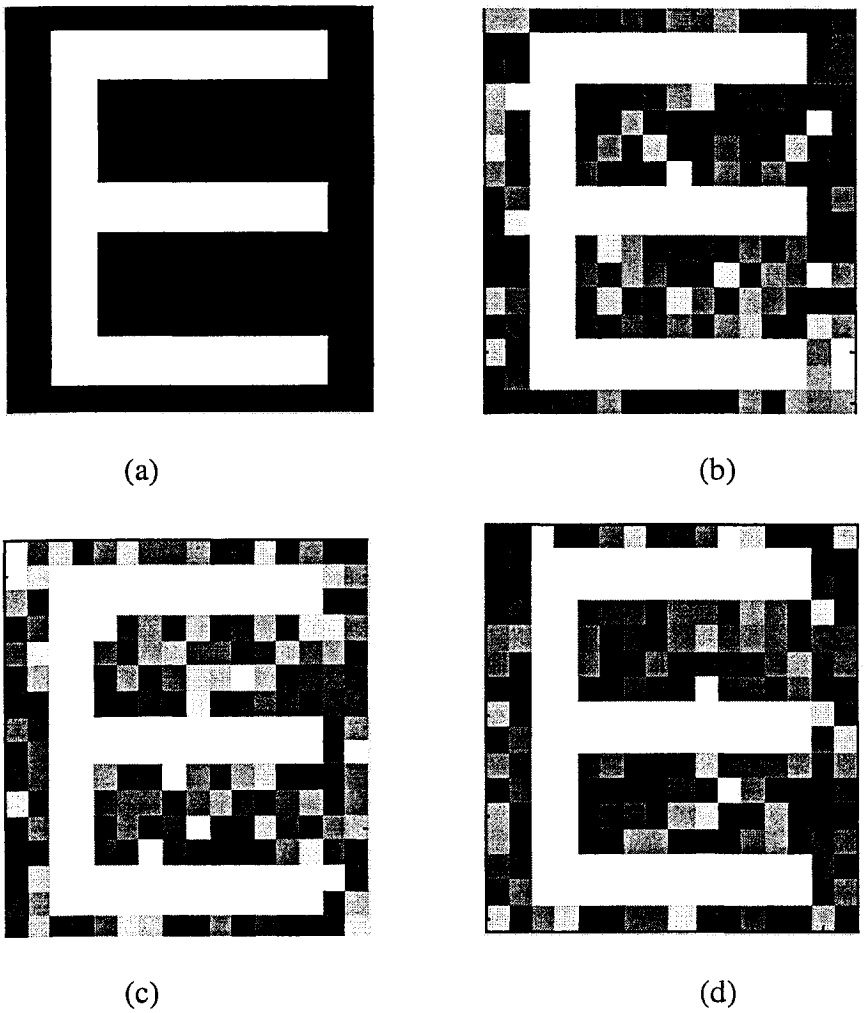


Figure 2.5 The flow chart of the modified GA for CGH design.

2.4. Simulations and Discussions

In the simulations, we compare the classical GA, the SA and the modified GA under different circumstances. 2 images with image size of 16×16 and 64×64 are used in the simulations. The 16×16 image is the alphabetic 'E' and 64×64 image is the spot pattern circle.

(1) Image size: 16×16



(a) (b)

(c) (d)

Figure 2.6 Reconstructed image of the kinoform for the size 16×16. (a) Original image, (b) image reconstructed by the classical GA, (c) image reconstructed by the SA (4) image reconstructed by the modified GA.

Table 2.2 Normalized cost functions of the classical GA, the SA and the modified GA.

Algorithms	Max cost function	Min cost function	Ave. cost function	Variance of cost function
GA	1	0.7449	0.8153	1
SA	0.8286	0.7325	0.7704	0.1776
modified GA	0.8827	0.7197	0.7665	0.2639

For the image with size of 16×16 pixels, we run the simulations to generate holograms with the GA, the SA and the modified GA 20 times, respectively. Table 2.2 summarizes the cost functions of the three algorithms. The table demonstrates that the performance among the algorithms is the modified GA > the SA > the GA. The smaller variance of cost function (VOF) of the SA than the other two algorithms demonstrates that the SA has the most consistent output, while the GA has the least consistent output. The better VOF of the modified GA than the GA shows that after the introduction of the SA as a mutation algorithm, the GA can maintain much better population diversity and therefore, we can reduce the convergence to a local minimum greatly. In addition, we also find that the performance of the best individual of the modified GA is better than the SA and normal GA. This is because the modified GA combines the merits of both the GA and SA. In compare with the normal GA, the modified GA has better mutation operator that can maintain more genes with the same population size. In compare with the SA, the modified

GA has a large population size (the population size is 1 for the SA) and can search in broader range.

From the above discussion, we also find that the SA produces more consistent output than the GA does and the VOF of the hologram by the SA is the smallest. One reason is that the SA will always accept mutation of an individual that results in better fitness while the cross-over and mutation of the GA might destroy the good genes and cause the gene drift. The search for optimal solution by GA is the cooperation by all the genetic operators of selection, mutation and cross-over. An excellent descendent that produced by the selection operator might be destroyed by the cross-over operator. The elitism principle used in our algorithm is to reduce the gene drift by ensuring the best chromosomes not been destroyed [26]. The introduction of mutation by the SA into the GA further mitigates such a problem by forcing the drift gene to return to a good gene. Therefore the VOF of the modified GA is better than that of the normal GA.

(2) Image size: 64×64

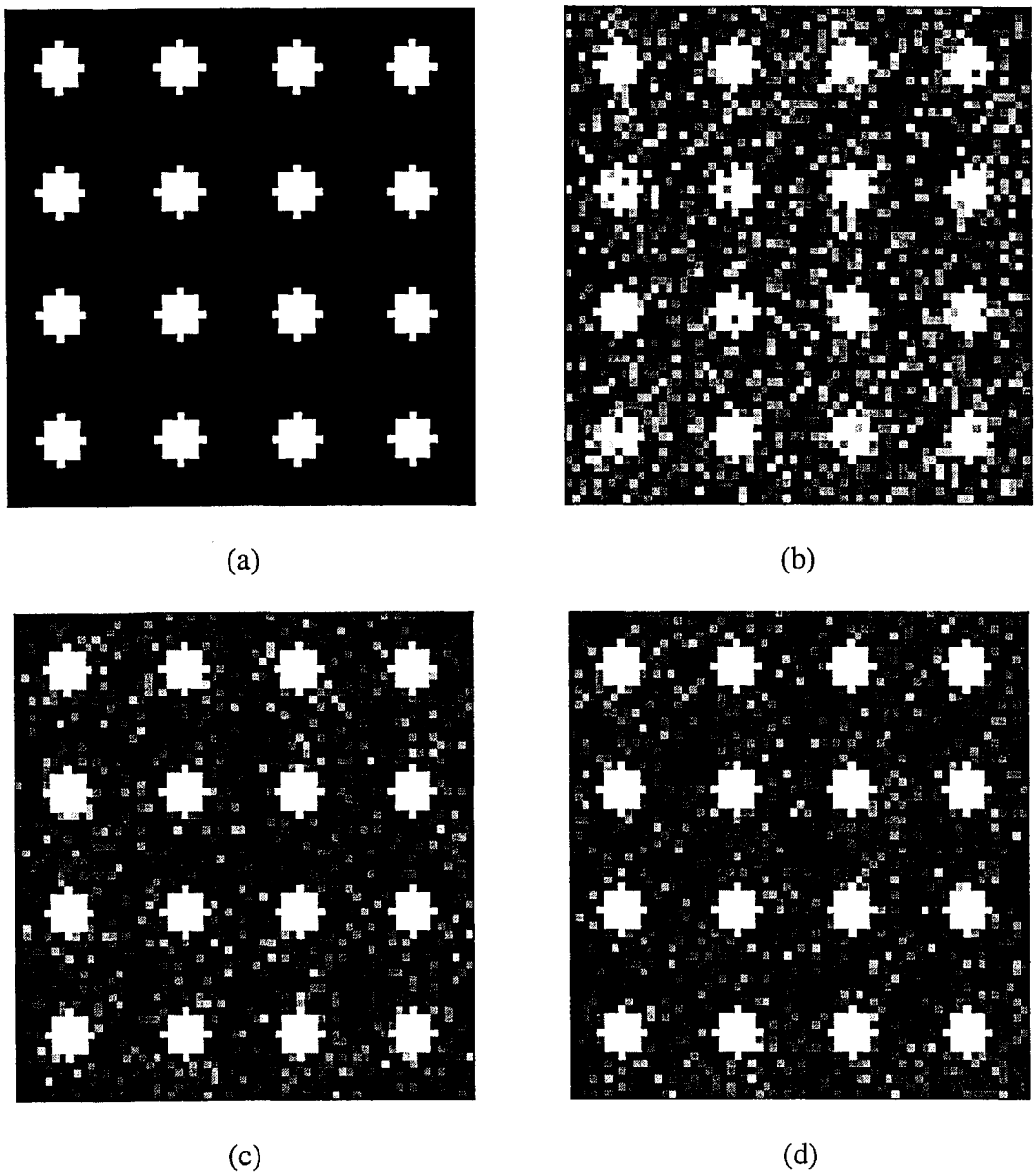


Figure 2.7 Reconstructed image of the kinoform for the size 64×64. (a) Original image, (b) image reconstructed by the classical GA, (c) image reconstructed by the SA (4) image reconstructed by the modified GA.

Figure 2.7 shows a 64×64 spot pattern to be used for the diffuser design. The diffuser designed based on this pattern will diffract the incoming single laser beam to 4 by 4 beams. In infrared wireless home networking systems, the diffracted beams are usually further diffused by the ceiling of the office or home which would finally provide a uniform diffuse infrared distribution to cover the entire space. For the 64×64 image, the same procedure is employed to design the diffuser. We find that in this case, the classical GA performs much poorer than the other two. If we take an average of the cost functions and normalized them, we find that the cost functions of the classical GA, the SA and the modified GA is 1, 0.8473 and 0.8461 respectively. This shows that the modified GA and the SA perform much better than the classical GA, and the modified GA provides the best performance.

The reason is that the solution ensemble of an image with 64×64 pixels is 2^{16} times larger than an image with 16×16 pixels. With the population size the same and only the random mutation to search for better genes, it is very difficult for the classical GA to find a global optimum in such a large solution ensemble. Therefore one normal method for the classical GA to guarantee the performance as the solution ensemble increases is to enlarge the population size. The introduction of the SA as a complementary mutation method can better diversify the population to reduce the population size needed. Furthermore, our simulation shows that when the GA approaches an optimal area, the SA can quickly locate the exact place by its internal search process but the cross-over and other operators of the

GA may drift the search out of that area, which means that the SA can act as a local search tool.

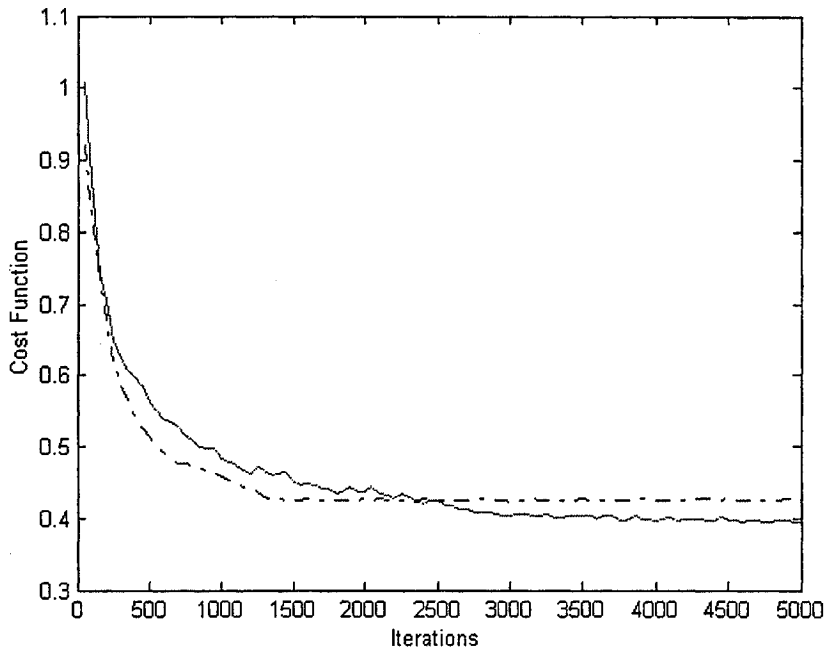


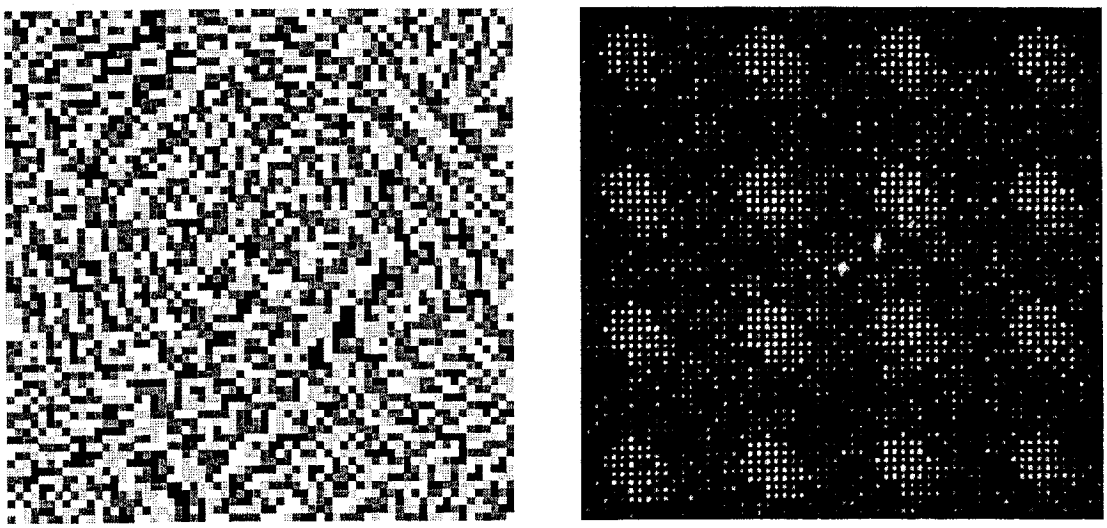
Figure 2.8 The cost functions of the classical GA (dash-dotted curve) and the modified GA (solid curve).

Figure 2.8 shows the convergence process of the cost functions of the classical GA and the modified GA. For the classical GA, it converges sharper than the modified GA at the beginning. But because its internal mutation algorithm cannot maintain good population diversity, it is trapped to a premature minimum after about 1500 iterations. As can be seen from the Figure 2.8, after 1500 iterations, the cost function does not improve much. We investigate the VOF of the classical GA and the modified GA, we find that the VOF of the

classical GA decreases quickly and after 1500 iterations it equals almost to zero. On the contrary, the VOF of the modified GA can always remain at a relatively higher value. Therefore, the better population diversity makes it possible for the modified GA to search through much wider area and get better results.

However, the modified GA will take longer computation time than the normal GA since the mutation process of the modified GA combines the SA. The computation time of different algorithms depends much on the inverse Fourier construction of a hologram because of its computation intensity. As our simulation shows, the inverse Fourier transform occupies around 70% of computation time in the normal GA and 95% computation time in the SA. The modified GA requires about 10% more inverse Fourier transform computation than the normal GA. The computation time of the GA also depends on the population size and the iterations needed. Normally the iteration of the GA is much less than that of the SA since the GA has large population size than the SA and the GA will converge much faster than the SA. For a GA process with population 100 and 2000 iterations and a SA process with 100,000 iterations, the ratio of the computation times of the modified GA, the normal GA and the SA is around 3:2.6:1.

2.5. Fabrication of the hologram



(a) (b)

Figure 2.9 (a) 64×64 pixel basic cell hologram mask obtained using modified GA (b) Reconstructed pattern using a fabricated hologram mask of a 16×16 array of the basic cell

The hologram designed by the modified GA has the better performance. A 4-level phase hologram, as shown in Figure 2.9(a), is calculated by the modified GA and fabricated at a pixel resolution of 2 μm on a quartz substrate, with a phase difference of 0, $\pi/2$, π or $3\pi/4$ radians at each pixel. To make full use of the 0.81 mm source laser beam width, the 64×64 pixel basic cell was stepped in a 16×16 array to create a total hologram size of 1024×1024 pixels, with a total physical size of 2.048 mm × 2.048mm. The fabricated hologram mask was illuminated using a 7-mW 633-nm laser source normally incident onto the mask, and the obtained reconstructed hologram is shown in Figure 2.9(b) above. The reconstructed

image was obtained by projecting the image onto a planar surface, such as a wall, in an enclosed dark room environment with no other sources of light besides the hologram. It can be observed that the obtained results are similar to the theoretically determined reconstruction pattern from Figure 2.9(a). However, a large central spot can be seen in the reconstructed pattern and this is attributed to fabrication limitations [29], containing almost 20% of the input laser power. By optimizing the fabrication techniques, the central spot power could be reduced to about 0.1%.

2.6. Summary

Holographic diffusers can be used to extend a collimated laser beam to cover a broad range thus solve the eye safety problem in the home wireless network. Holographic diffusers are usually designed using the computer generated hologram (CGH) technique and the theory model of CGH is explained and discussed in this chapter. The GA was proposed to design the CGH with the difference between the original image and constructed image minimized. How to control the values of various parameters and how to maintain the population diversity are two of the most important issues to be solved in the GA. Although the GA works well for a wide variety of applications in engineering and science, when facing the problems with a large ensemble of solutions such as the CGH design, the GA works poorly and may converge with a high probability to a local optimum. The modified GA proposed in this thesis was to solve this problem. In the modified GA, we used a selection process based on the rank of fitness, the exponentially decreased

mutation probability, and incorporating the SA as the mutation algorithm into the GA to maintain good population diversity. We compared the modified GA with the classical GA and the SA. The simulation results showed that for images with small and large size, even with a small population size, the modified GA could search the nearly global optimum with high probability. The spot pattern hologram designed with modified GA was used to fabricate on a quartz substrate with a pixel resolution of 2μm and the reconstructed image agrees well with the simulation result.

Chapter 3

Birefringent filter design

In this chapter, birefringent filter design by the genetic algorithm (GA) is investigated. The GA is used to optimize the shift angles and the element lengths of a birefringent filter. Unlike the conventional digital filter design algorithms for a particular birefringent filter such as Lyot filter or Solc filter, The GA can be used to design a more general birefringent filter with both the element lengths and shift angles variable. A better performance in locating the global optimum is demonstrated. The design of an 8-section birefringent filter with the side band suppression ratio up to 42 dB is presented.

3.1. Introduction

A fiber birefringent filter (BF) consists of a certain number of birefringent optical elements between a certain number of polarizers [12]. The elements usually have high birefringence (such as quartz, $LiNbO_3$ or $Nd:YVO_4$) [30] [31]. Phase shifts are introduced between the two orthogonal components of a linearly polarized light. The polarizer is used to interfere with the two orthogonal components such that the intensity of the linearly polarized light is strengthened or weakened depending on the shift angle between the polarizer and the polarized light [9-12] [31]. The transmission through a birefringent filter at a given wavelength is a function of the optical retardation exhibited by each of the filter

stages [9-11]. Adding the retardation to each stage in the proper fashion makes it possible to shift the pass bands and stop bands at will. Therefore, by rotating the relative angles between adjacent birefringent elements we can obtain different spectral responses. The design of a birefringent filter is to find a series of lengths and angles of the birefringent elements such that the designed spectral response best approximates the desired one [11] [13].

An important advantage of birefringent filters over other filters is their periodic spectral response and operation in transmission; therefore they are natural choices for applications to avoid the losses and expenses incurred with optical circulators [11]. Birefringent filters were first introduced to be an effective tool in astronomical research [9] [15-16] since they are capable of transmitting very monochromatic light from a broadband source. Now they have found applications in a wide variety of other disciplines such as spectroscopy, imaging and optical telecommunication [32-36]. There are two types of basic birefringent filters, Lyot and Solc birefringent filters, with retardation introduced by controlling the element lengths and adjacent element shift angles, respectively [9-11] [13].

3.1.1 Lyot Filter

The Lyot filter was first introduced by a French astronomer, Bernard Lyot in 1933 [9]. It consists of a number of alternating polarizers and birefringent plates. The length of each successive birefringent plate is twice that of the previous plate [9] [37]. The multiple pass

bands are spaced by the free spectral range (FSR) of the thinnest optical element [9] [38]. The tuning of a Lyot filter can be realized by different methods such as using the rotating element of achromatic retarders [39] or based on the optical rotatory dispersion effect [40]. The configuration of a Lyot filter is shown in Figure 3.1.

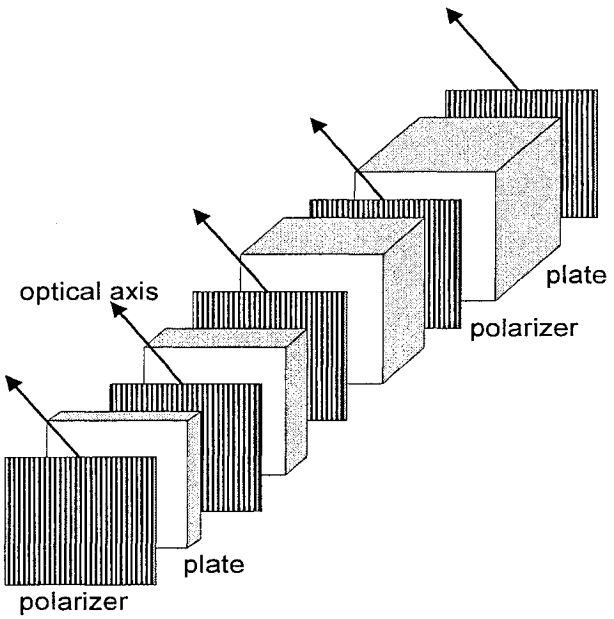


Figure 3.1 Lyot filter

The transmission T of each element between the polarizers is given by [14]

$$T = \cos^2 \Gamma / 2 , \qquad (3.1)$$

where $\Gamma = 2\pi\Delta nd/\lambda$, Δn is the birefringence, d is the length of the element and λ is the wavelength. Since the length of a successive plate is twice the previous one, the k-th Γ is given by [14]

$$\Gamma_k = 2\Gamma_{k-1} = 2^{k-1}\Gamma_1. \tag{3.2}$$

The overall transmission of an N element Lyot filter is given by

$$T = T_0 \prod_{k=1}^{N} \cos^2(\pi\Gamma_k), \tag{3.3}$$

where T_0 is the energy loss due to the reflection and absorption.

3.1.2 Solc Filter

The Solc filter is another type of birefringent filter with simpler structure than the Lyot birefringent filter [10]. Unlike the Lyot filter which has one birefringent section between a pair of polarizers, the Solc filter has a number of birefringent sections between a pair of polarizers. The Solc filter can be furthered classified into two categories, folded and fan Solc filters. A folded Solc filter consisting of alternating equal shift angles [14] [41] is shown by Figure 3.2.

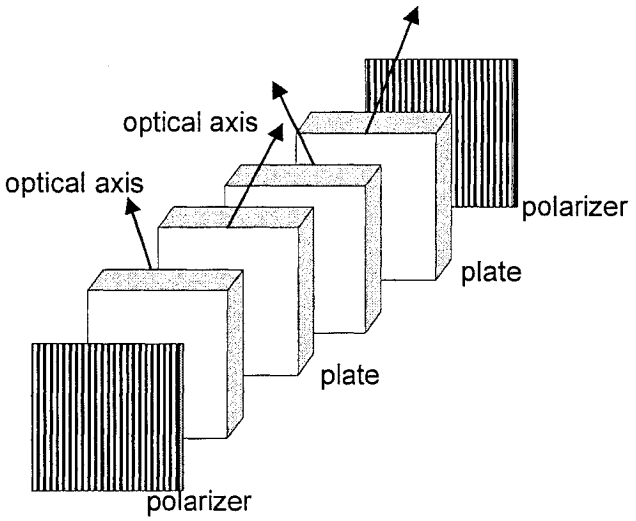


Figure 3.2 Folded Solc filter.

The angles in relation to the pre-determined x-axis are given by [14]:

$$\varphi_k = (-1)^{k-1}\rho\ , \tag{3.4}$$

where k is the k-th plate of the N section folded Solc filter and ρ is an arbitrary angle, typically $\rho = \pi / 4N$. The shift angle of the odd-number angle is $\pi / 4N$ while the shift angle of the even-number angle is $-\pi / 4N$. The transmittance of the folded Solc filter is given by [14]:

$$T = \left|\tan(2\rho)\cos(\chi)\frac{\sin(N\chi)}{\sin(\chi)}\right|^2, \tag{3.5}$$

with

$$\cos\chi = \cos(2\rho)\sin(\Gamma/2). \tag{3.6}$$

The fan Solc filter shown in Figure 3.3 differs from the folded Solc filter in that the shift angle increases in magnitude by a constant instead of alternating in sign. The kth angle φ_k in relation to the pre-determined x-axis is given by [14]:

$$\varphi_k = (2k+1)\rho. \tag{3.7}$$

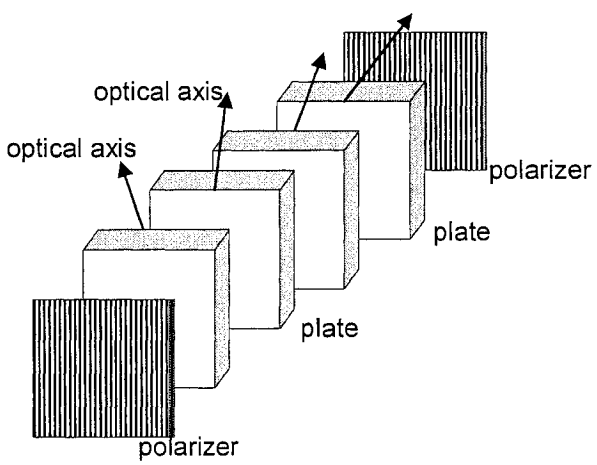


Figure 3.3 Fan Solc filter.

3.1.3 Jones Matrix description of birefringent filters

Both Lyot and Solc filters have symmetrical structures. However, as pointed out by Harris [13], the shift angles between the birefringent sections can be adjusted arbitrarily in order to further reduce the sidelobe levels of the filter spectral responses. The Jones Matrix is used to generalize the transmission characteristic of the Solc filter with arbitrarily angles [10] [42-43].

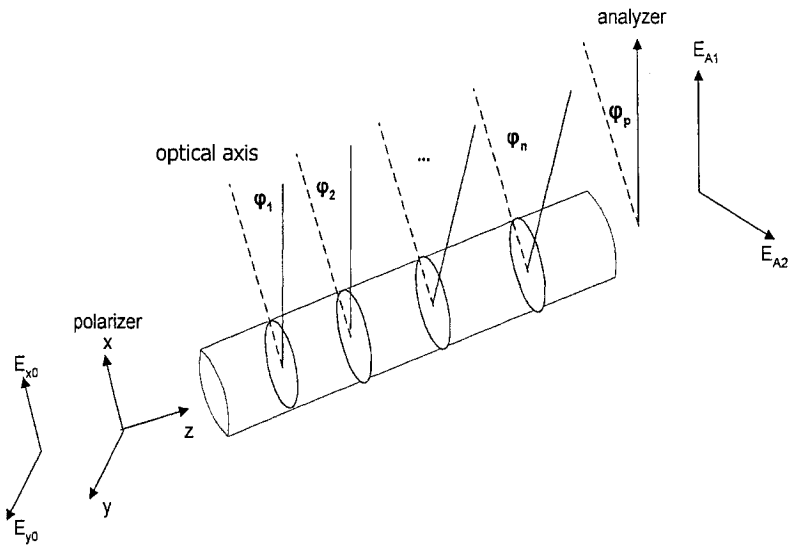


Figure3.4 Solc filter with arbitrarily angles.

Figure 3.4 shows the configuration of a Solc birefringent filter with arbitrarily angles. The light with the input E_{x0} and E_{y0} is polarized first and then travel along the z axis in a rectangular coordinate system. The face of each birefringent section is normal to the z axis. The electric field vectors of the light after the analyzer can be expressed as [11]

$$\begin{pmatrix} E_{A1} \\ E_{A2} \end{pmatrix} = R(\varphi_p) M(\varphi_i, \eta) P_x \begin{pmatrix} E_{x0} \\ E_{y0} \end{pmatrix}, \tag{3.7}$$

where E_{x0} and E_{y0} are the inputs, E_{A1} and E_{A2} are the transmitted and rejected components. $R(\varphi_p)$ is the matrix representing the rotation by φ_p. $R(\varphi_p)$ is defined as

$$R(\varphi_p) = \begin{bmatrix} \cos\varphi_p & \sin\varphi_p \\ -\sin\varphi_p & \cos\varphi_p \end{bmatrix}, \tag{3.8}$$

where P_x denotes the input polarizer transmitting the electric vector parallel to the x axis, defined as

$$P_x = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \tag{3.9}$$

$M(\varphi_i, \eta)$ is a matrix product of a series of rotations and retardations caused by a stack of birefringent sections, therefore

$$\begin{aligned} R(\varphi_p) M(\varphi_i, \eta) &= R(\varphi_p)\{R(-\varphi_n) R_\eta R(\varphi_n)\} \ldots \{R(-\varphi_1) R_\eta R(\varphi_1)\} \\ &= R(\varphi_p - \varphi_n) R_\eta R(\varphi_n - \varphi_{n-1}) R_\eta \ldots R(\varphi_2 - \varphi_1) R_\eta R(\varphi_1) \\ &= R(\theta_n) R_\eta R(\theta_{n-1}) R_\eta \ldots R(\theta_1) R_\eta R(\theta_0) . \end{aligned} \tag{3.10}$$

Here the relative rotation is

$$R(\theta_i) = \begin{bmatrix} \cos\theta_i & \sin\theta_i \\ -\sin\theta_i & \cos\theta_i \end{bmatrix}. \tag{3.11}$$

The retardation is

$$R_\eta = \begin{bmatrix} e^{j\eta/2} & 0 \\ 0 & e^{-j\eta/2} \end{bmatrix}, \tag{3.12}$$

and $$\eta = 2\pi f(\Delta nL/c), \tag{3.13}$$

where Δn is the birefringence, L is the length of the birefringent section and f is the frequency of the light

3.2 Design of birefringent filters by the GA

The design of birefringent filters is to find the shift angle and length of each birefringent section to meet the specific design requirement. Many algorithms have been proposed in the last decades [11] [13] [14]. Harris *et al.* [13] assumed that the birefringent sections could be equal in length and used a Fourier series approximation to calculate the filter coefficients. Chu and Town [11] used the Remez exchange algorithm and Pegis's method to find the polynomial, from which they used an inverse transform algorithm to find the angles. These algorithms usually require a large number of sections, by using inverse

transform to obtain a specified spectral response since they usually based on a particular case of birefringent filter (Solc or Lyot) and have less freedom to control the retardation of the optical elements. In this thesis, we propose an improved genetic algorithm that uses the forward transform of Jones Matrix directly to find the corresponding parameters for a given birefringent filter. Compared to the algorithms in [11] [13], the birefringent filter designed by the improved GA provides better performance in terms of the sideband suppression ratio.

Since the Solc filters are more extensively used by virtue of their low loss and simple configuration, we will first take the Solc filters as the example to describe our algorithm and assume the filter section length to be identical as Harris and Chu assumed [11] [13]. However, the GA does not require that the birefringent sections have identical length since the GA is a forward transform algorithm and can be applied to an arbitrary configuration of birefringent filters as long as the parameters (element lengths, angles) are encoded as their corresponding chromosomes.

3.2.1 Cost Function

To design the birefringent filter, we need to specify the desired spectral response. In our design, the desired spectral response of a filter can be expressed as *(0, fp, fs, 1)* in which

fp is the normalized pass frequency and *fs* is the normalized stop frequency. The desired intensity is:

$$I = \begin{Bmatrix} \geq I_{pass}, & 0 \leq f \leq f_p \\ Not\ Concern, & f_p < f < f_s \\ \leq I_{stop}, & f_s \leq f \end{Bmatrix} \tag{3.13}$$

The I_{stop} should be as low as possible to achieve the low sidelobe. For example, if the spectral response is $(0, 0.2, 0.5, 1)$, the desired spectral response is shown in Figure 3.5.

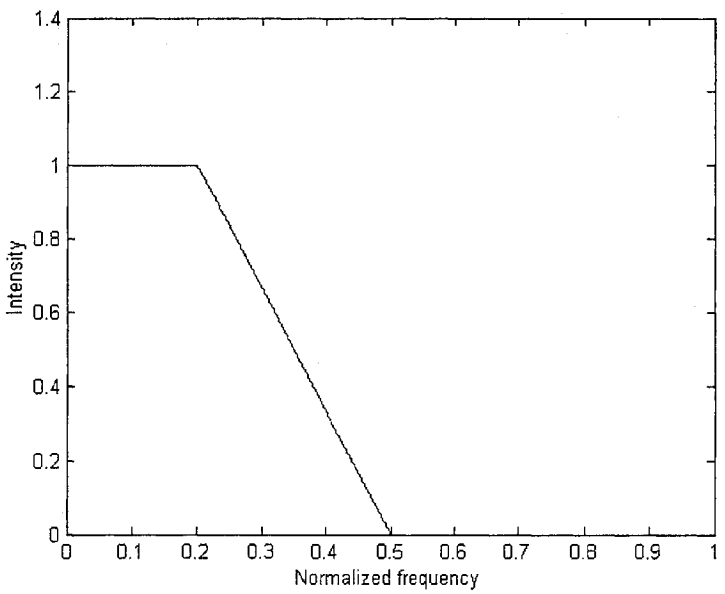


Figure 3.5 Desired spectral response of a low pass filter.

The design of a birefringent filter is to find a series of shift angles $\varphi_1, \varphi_2, ...\varphi_n$ so that the output spectral response best approximates the desired response. The difference between the output spectral and desired spectral responses is the cost function, which is the parameter that we will optimize as fitness in the GA. The cost function is represented as the least mean square error (LMS). Therefore, the design problem is reduced to find the shift angles between adjacent sections such that the LMS is minimized. The cost function can be expressed as:

$$\cos t(\varphi_1, \varphi_2, ...\varphi_n) = (\int W(\lambda)(\left|E_{A1}(\varphi_1, \varphi_2, ...\varphi_n; \lambda)\right|^2 - I_{desired})d\lambda)^{\frac{1}{2}}, \qquad (3.14)$$

where E_{A1} is the normalized transmitted component that comes from the Jones matrix $R(\varphi_p)M(\varphi_i, \eta)$ and $R(\varphi_p)M(\varphi_i, \eta)$ is a function of wavelength. $W(\lambda)$ is a weighted function which can be used to control the ripples of the designed spectral response. In particular we can control those important regions such as the pass band and the stop band and ignore the errors in transition band. For example, if we want to design a filter with maximum pass band deviation of $\pm 2dB$ and sidelobe level to be $-20dB$. Since the values of $-2dB$ and $-20dB$ are around 0.6 and 0.01 respectively, then if we want the pass band and stop band to pay the same significance in the fitness, the weight of the stopband should be $0.6/0.01 = 60$ if the weight of the passband is normalized to 1.

Equation 3.14 gives the integration form of the LMS. The spectral response is a numerical approach in our design and we sample the spectral response at discrete wavelengths. Therefore, the cost function is expressed as:

$$\cos t(\varphi_1, \varphi_2, ...\varphi_n) = (\sum_{\lambda} W(\lambda)(\left|E_{A1}(\lambda)\right|^2 - I_{desired}))^{\frac{1}{2}}. \qquad (3.15)$$

The GA is employed to find the largest fitness or the smallest cost function. Therefore the fitness can be expressed as:

$$fitness = \frac{1}{\cos t(\varphi_1, \varphi_2, ...\varphi_n)} \qquad (3.16)$$

The weight $W(\lambda)$ is an important parameter in the design. If we need to achieve lower sidelobe level then we can assign large value of $W(\lambda)$ to the stop band [14]. However, the large value of $W(\lambda)$ to the stop band might degenerate the performance of the pass band and the transition band [14]. This means that a large value of $W(\lambda)$ can lead to lower sideband level but might impair the flatness of the pass band. Therefore, there exists a trade off in choosing the weight $W(\lambda)$.

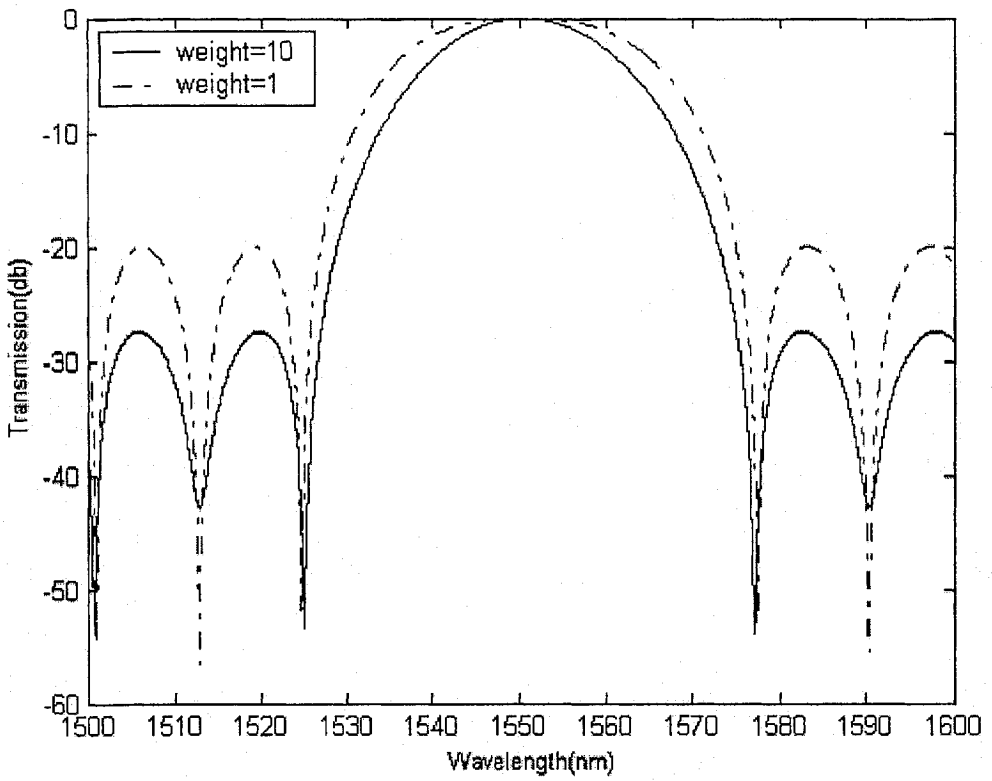


Figure 3.6 Spectral responses with different weights normalized to the passband.

Figure 3.6 shows the designed spectral responses of a 4-section Solc filter with the required spectral response of (0, 0.2, 0.5, 1) for two different weights. The dotted line shows the Solc filter with a weight of 1 in the stop band and the solid line shows the Solc filter with the weight of 10 in the stop band. As shown in the figure, the Solc filter with a weight of 10 in the stop band can achieve lower sidelobes; however, the pass band is not as flat as the one with the weight of 1 in the stop band.

3.2.2 Binary encoding and floating point encoding

Either the binary encoding or the floating point (FP) encoding of the GA can be used for the optimization problems [44-46]. The binary encoding that we use in the design of holographic diffusers differs from what we use for the design of birefringent filters since the shift angle is a floating point data. Therefore, to use the binary encoding the shift angle must be mapped to a binary string first. The mapping of value x with the precision of m and range $[lower, upper]$ to the binary string $(b_1 b_2 ... b_n)$ is completed in three steps:

Step 1: Find the length n of a binary string $(b_1 b_2 ... b_n)$ to represent the required precision. n is calculated such that

$$(upper - lower) * 10^m \leq 2^n - 1,$$

Therefore

$$n \geq \log_2\{(upper - lower) * 10^m + 1\}. \tag{3.17}$$

Step 2: Convert the binary string $(b_1 b_2 ... b_n)$ from base 2 to base 10, let

$$x' = (b_1 b_2 ... b_n)_2 = (\sum_{i=0}^{n-1} b_i * 2^i)_{10}. \tag{3.18}$$

Step 3: The real value of x is:

$$x = lower + x' * \frac{upper - lower}{2^n - 1}. \tag{3.19}$$

The binary encoding offers the maximum number of schemata per bit of information of any coding [44] [45] and consequently the bit string representation of solutions has dominated genetic research. This coding also facilitates theoretical analysis and allows elegant genetic operators. One drawback of binary encoding for real number is that the string length is very long for high precision. The longer the length, the wider the search space will be. For example, to design a 14-section birefringent filter with a precision of 6, the search space is 2^{23*14}. In order to achieve a better performance in such a design, we need a large population size. However, large population size means longer computation time. Floating point encoding mitigates such a problem by using a floating point data to directly represent a real number and the precision of the data depends only on the precision of the computer.

The floating point encoding gives the chromosomes more adaptability. The shift angle φ_i can be encoded as:

$$\varphi_i = \pi * (r - 0.5), \tag{3.20}$$

where r is a random variable ranging from 0 to 1 and φ_i is the shift angle ranging from $-\pi/2$ to $\pi/2$.

3.2.3 Genetic operators

The design of birefringent filters is a one-dimensional problem while the design of holographic diffusers is a two-dimensional problem. For the floating point encoding, the corresponding genetic operators differ much from those of the holographic diffuser design. The difference between them lies mainly on the cross-over and the mutation. In our design, we use the tournament selection [25] [47] and the arithmetic cross-over. The elitism principle is also used to reduce the genetic drift [22].

3.2.3.1 Cross-over

The arithmetical cross-over [20] is used for the floating point encoding to generate new offspring. If two chromosomes $\varphi^v = (\varphi_1^v, \varphi_2^v ... \varphi_n^v,)$ and $\varphi^w = (\varphi_1^w, \varphi_2^w ... \varphi_n^w,)$ are to be crossed, the resulting offspring are $(\varphi_1^v, \varphi_2^v ... \varphi_k^{v+}, ... \varphi_n^v)$ and $(\varphi_1^w, \varphi_2^w ... \varphi_k^{w+} ... \varphi_n^w)$. k is the position to be crossed and,

$$\varphi_k^{v+} = (1-a) \times \varphi_k^v + a \times \varphi_k^w,$$

$$\varphi_k^{w+} = (1-a) \times \varphi_k^w + a \times \varphi_k^v, \qquad (3.21)$$

where a is generated randomly ranging from 0 to 1 to determine how close the offspring to their parents.

3.2.3.2 Mutation

As what we have described in Chapter 2, we can use adaptive and fixed mutation probability. However, our simulation shows that P_m in the floating point encoding does not affect the performance as much as in the binary coding. What affects the performance more is the step size, which is shown in Equation 3.22. For the floating point coding, the mutation can be expressed as:

$$x_k^{t+1} = x_k^t \pm \Delta, \tag{3.22}$$

where x_k^t is the kth gene of the t-th generation, Δ is the step size that determines how a gene is mutated. There are various methods to control the step size [20] [21]. The principle of these methods imitates the natural mutation phenomenon in which wide step size is with low probability and narrow step size is with high probability. Therefore, a Gaussian distribution is widely used. Another method that was recommended by Michalewicz [20] is to use a non-uniformity decrease function, which is employed in our simulation. In this method, the next generation after the mutation is expressed as:

$$x_k^{t+1} = \begin{cases} x_k^t + \Delta(t, r(k) - x_k), & \textit{if random binary digit is } 0 \\ x_k^t - \Delta(t, x_k - l(k)), & \textit{if random binary digit is } 1 \end{cases}, \tag{3.23}$$

where $r(k)$ is the upper limit of x_k^t and $l(k)$ is the lower limit of x_k^t. In our design, $r(k)$ and $l(k)$ represent the range of a shift angle, which are $\frac{\pi}{2}$ and $-\frac{\pi}{2}$ respectively. $\Delta(t, y)$ is the step size which is expressed as [20]

$$\Delta(t, y) = y.r.(1 - \frac{t}{T})^b . \tag{3.24}$$

where t is the generation number and T is the maximum generation number. The step size $\Delta(t, y)$ will have a value in the range $[0, y]$. The probability that the step size $\Delta(t, y)$ close to 0 increases as the generation number t increases. b is a parameter to control the speed of $\Delta(t, y)$ which will converge to 0 finally. b is empirically set to 5 in our design.

3.3 Design examples by the GA

Before we start our design, we need to specify several parameters first. As shown in Equation 3.13, $\eta = 2\pi f(\Delta nL/c)$. That means the phase shift η is determined by the birefringence Δn and the section length L. At the center wavelength, the phase shift η should be $2k\pi$ (k is integer) to have the maximum output. Therefore,

$$\eta = 2\pi f(\Delta nL/c) = 2\pi(\Delta nL/\lambda) = 2k\pi$$

and

$$k\lambda = \Delta nL \tag{3.25}$$

Equation 3.25 means that the ΔnL must be integer times the center wavelength. In our design, the birefringence of the fiber is 7.95×10^{-4}. If we want the center wavelength to be 1550 nm, the element length can be 3.9 cm, 7.8 cm or etc. The free spectrum range FSR can be calculated from Equation 3.26 [11]

$$FSR = \frac{c}{\Delta n \times L}. \tag{3.26}$$

To determine the fiber length, we use equation 3.25 to obtain a number of fiber lengths that can produce the required center wavelength. Then for a given FSR, say, in the range from 1500 nm to 1600 nm with only one peak value, we will obtain an FSR range. In this case it is $50nm < FSR < 100nm$. If it is converged to the frequency domain, the range is $6.24\times10^{12}\text{ Hz} < FSR < 1.25\times10^{13}\text{ Hz}$. From equation 3.26, we will have the length range to be 3 cm to 6.1 cm. Since the center wavelength is 1550 nm, the length of 3.9 cm is our choice, which will produce an FSR of $9.67\times10^{12}\text{ Hz}$.

The GA starts to run by setting the population size to be 60, the cross-over rate to be 0.6 and the mutation probability to be 0.1. The mutation probability is higher than the holographic diffuser design since the floating pointing encoding uses step size to control the mutation and higher mutation probability will help to search a wider area.

3.3.1 4-section birefringent filter

For our design, we specify the required spectral response to be (0, 0.2, 0.5, 1), in which the pass frequency $f_{pass} = 0.2$ and the stop frequency $f_{stop} = 0.5$. The requirement for the pass band is $A_{pass} \pm \delta \leq 0 \pm 2db$

Table 3.1 Designed Shift angles of 4-section filter with binary and FP encoding.

Section Number	0	1	2	3	4
Shift angles of Fig 3.7 (binary, radians)	-0.1373	-0.0401	0.4562	0.7306	-1.1955
Shift angles of Fig 3.8 (FP, radians)	0.3684	0.9444	0.3706	-1.2731	0.4878

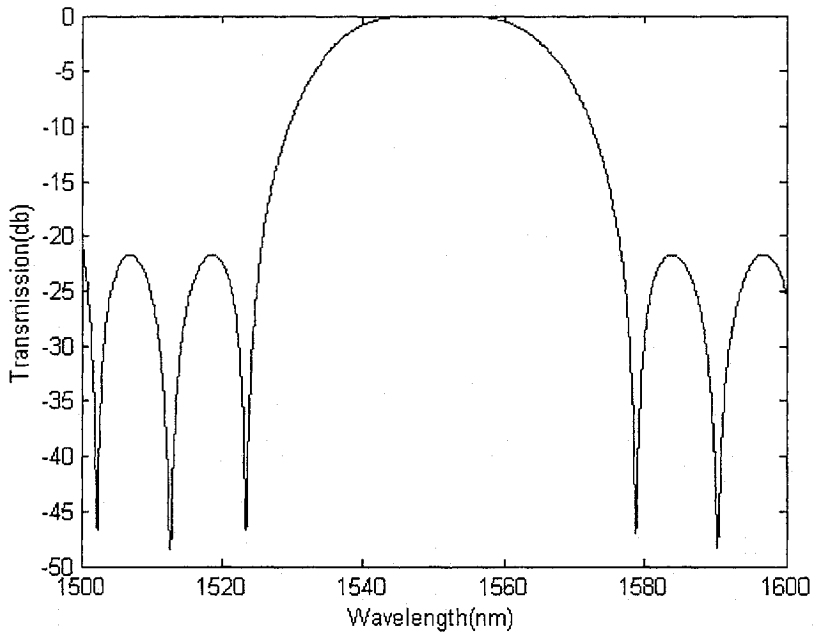


Figure 3.7 Spectral response by binary encoding.

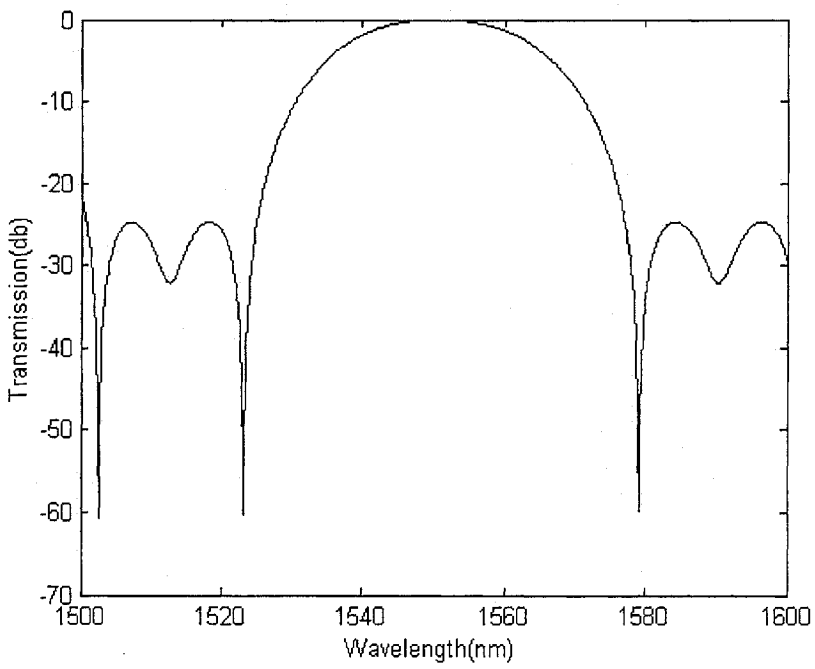


Figure 3.8 Spectral response by floating point encoding.

Table 3.1 gives the results of a 4-section birefringent filter design using binary encoding and floating point encoding. The spectral responses are shown in Figure 3.7 and Figure 3.8, respectively. As we discussed earlier, both the encoding schemes can be used to optimize parameters of a given problem, however, as our simulation results show, the floating point encoding performs better than the binary encoding in terms of the sideband suppression ratio. In order to further investigate this problem, we run the simulations 30 times for each encoding scheme and their cost functions are compared in Table 3.2

Table 3.2 Cost functions of different encoding methods.

	Max cost function	Min cost function	Ave. cost function
Binary encoding	1.9874	0.9163	1.4539
FP encoding	1.1849	0.7935	0.9267

As can be seen from Tabl 3.2, the average cost function of the floating point encoding is 0.9267, which is much less than that of the binary encoding (1.4539). This means that the spectral response obtained using floating point encoding provides a better approximation than that using binary encoding. In addition, the maximum and minimum cost functions of the floating point encoding is close to its average cost function, which implies that it provides more stable output.

The birefringent filter design is a multimodal problem in which many local and global minima exist. More than one solution can be applied to a required spectral response. One of the reasons that the floating point encoding performs better is that the genetic operator is more favorable for this kind of problem. Suppose that we have an angle value of 0.9834 to be represented by a 20 bit binary data, then its binary string is

0	1	0	1	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	1

If we select one bit to perform a mutation operation, the probability that the value of the offspring is far from its parents depends on the location of the bit to be mutated. If the bit locates in the front of the string, the probability is higher. This means that small difference in the chromosome may correspond to a large difference in the real parameter space. Therefore, when an individual is close to a local minimum, the probability that it can leave the local minimum is low because the probability that the mutation operation produces worse offspring is very high. Subsequently the population will converge quickly. The floating point encoding is conceptually close to the problem space since the value of its coding directly corresponds to the value of the parameter to be optimized. The mutation operation of the floating point encoding also performs more reasonably by using the step size to control the search direction and the convergence speed. Furthermore, in such a case the arithmetic cross-over of the floating point encoding is also better than the single or double point cross-over of binary coding since the latter usually alters greatly the gene of

an individual. The convergence history of both encoding methods in Figure 3.9 and Figure 3.10 also explains such a difference of performance.

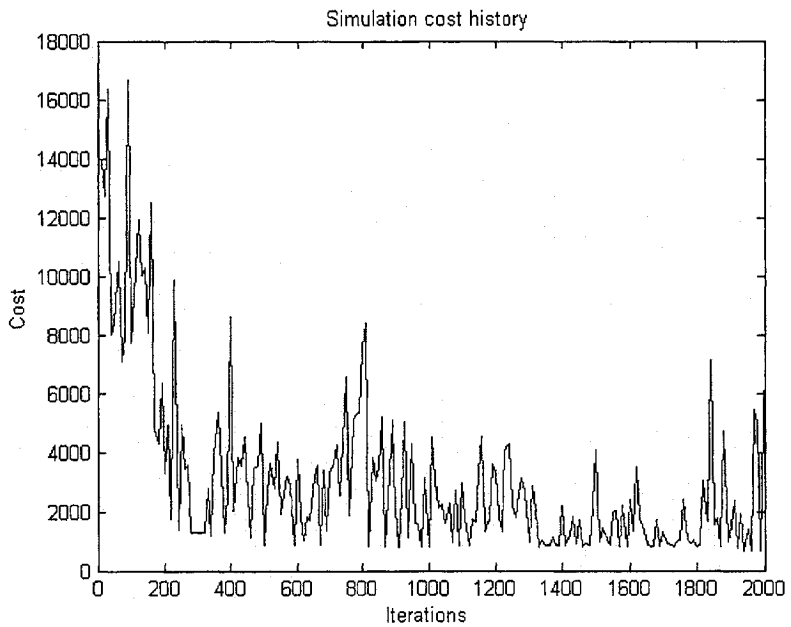


Figure3.9 Simulation history of the binary encoding.

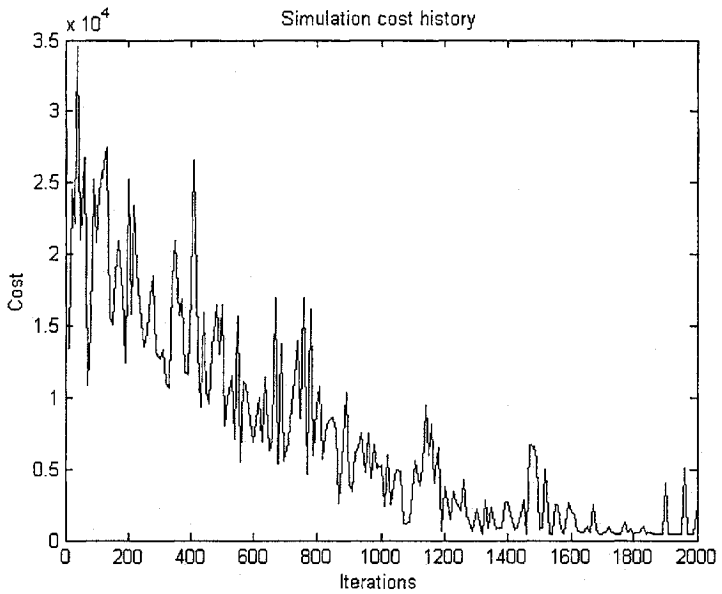


Figure 3.10 Simulation history of the floating point encoding.

Figure 3.9 and Figure 3.10 present the simulation history of the cost functions of the two encoding methods. Binary encoding presents a sharp decrease of the cost function in the first 200 iterations. After the 200 iterations, it fluctuates greatly but ceases to make much progress. The floating point encoding demonstrates a different convergence history by delaying the finding of solution to after about 1200 iterations. The extent of fluctuation of both encoding methods is high during the first 1000 iterations. The extent of fluctuation decreases gradually for the floating point method yet keeps almost the same for the binary one. This means that although the binary encoding tries to search a new area other than the already explored space, it is difficult for it to leave that explored space when it falls into that area.

In the design of the birefringent filter, we find that the cross-over using the two encoding methods play a less important role than in the diffuser design. If we remove the cross-over of the binary encoding, the performance will be even better. In birefringent filter design, the probability that the exchange of genes from the two parent binary strings to produce a better descendent is very low since the real value that it generates is similar to a mutation operation with large step size. The cross-over is supposed to combine the merit of its parents in a design problem; however, in this case it does not do well at that. Therefore, to use the binary encoding we can just remove the cross-over operator to achieve faster speed.

3.3.2 8-section birefringent filter

One of the methods to reduce the sidelobe level is to increase the number of sections of birefringent filters since we will have more freedom to control the filter [14]. In the 8-section birefringent filter design, we use the same required spectral response (0, 0.2, 0.5, 1).

Table 3.3 Designed shift angles of the 8-section filter.

Section number	Shift angles of Figure 3.11(radians)
0	0.9958
1	1.5709
2	0.3294
3	1.4726
4	-0.1300
5	0.9220
6	-1.5476
7	-0.4954
8	-0.1935

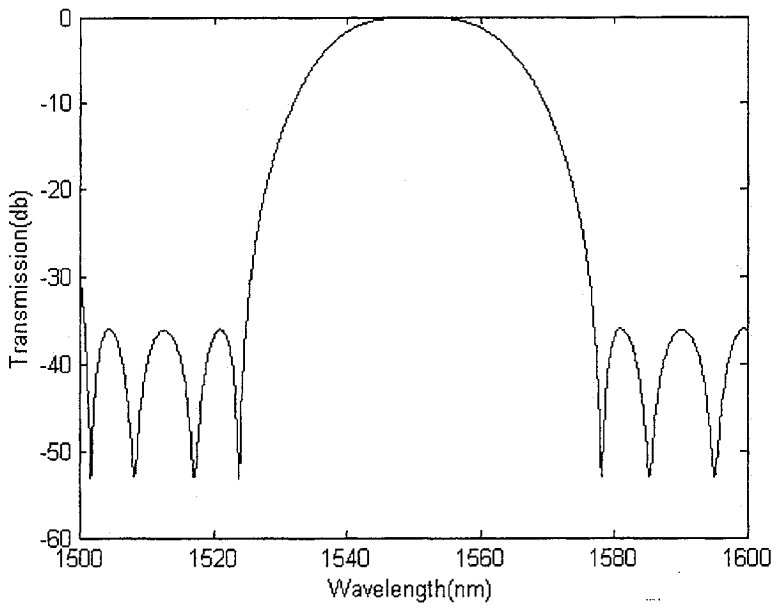


Figure 3.11 Spectral response of the 8-section birefringent filter.

Figure 3.11 shows that the spectral response of the 8-section birefringent filter. From the figure we can find that its sidelobe suppression ratio is 36 dB, much higher than that of the 4-section birefringent filter, which is 25 dB.

3.4 Improvements and Discussions

3.4.1 Reduce search space

The performance of the GA is affected greatly by the search space. The larger the search space, the poorer the performance might be. In our simulation, we found that the 4-section birefringent filter design performs more consistently than the 8-section one since it has less parameters thus less search space. To achieve comparative performance for the design with large search space, one of the methods is to enlarge the population size such that the GA has enough candidates to search in wider areas [20]. As we have discussed before, large population size will slow down the simulation time and in practice we have to make a compromise between the speed and performance.

Another method to improve the performance is to reduce the search space directly. One approach is to perform the simulation hierarchically. The first run we call it a rough run. The purpose of the first run is to locate the individual in a certain range of problem space. After the first run, we start the second run (delicate run). The second run is to accurately locate the individual. This method is similar to the delta coding [48]. In particular, for the

binary encoding, we start the first run by defining a low precision of the shift angles, i.e. 2 digits the floating point value. Then use the equation $n \geq \log_2\{(upper - lower) * 10^m + 1\}$ (3.17), the length n of the binary string will be 10. If the precision is changed to 4 floating point digits, the length of the binary string will be 16, which has a much larger space than the 2 digit floating point encoding. Then we can choose the 2 digits floating point data to represent the shift angles in the first run. After the first run, suppose that we have one of the shift angles to be φ_i, then the final angle should be located in the range of $[-0.1 + \varphi_i, 0.1 + \varphi_i]$. Therefore in the second run, one can just encode the distance from the angle φ_i, which is from -0.1 to 0.1. Then based on Equation 3.16 and the 4 digit floating point representation, we can use the binary strings with a length 11 to encode the distances of the angles. This approach increases the runs of simulation, but reduces the search space from 16 binary digits to 11 binary digits. Although it has multiple runs, since the GA converges much faster in a small search space, the total running time is greatly reduced.

The method to reduce the search space can also be applied to the floating point encoding. In the first run, we have a large step size. In the second run we have a small step size and the chromosomes to be encoded are the distance from the results that we obtain in the first run. The number of runs can be extended to more than two. After each run, the angles will be closer to its precise position.

3.4.2 Multiple outputs

The search for the optimum by multiple runs depends on one assumption: the best area of the first run includes the global best individual. However, this is not always the case. In many circumstances the optima are not distributed evenly. The optimal area found in the first run might not include the best individual in the successive runs. Therefore another approach to reduce the search space is to divide the search space directly into a certain number of sections. These sections are also called niches [47] [49] [50]. The search for optimum is confined to each section. The more sections there are, the less the search space for each run.

As our simulation shows, the solution to a given birefringent filter is not unique. Different combinations of shift angles can be applied to realize the same specified spectral response with similar performance. Although the GA has a large population size and can be used to produce multiple solutions, its individuals usually converge to an identical individual finally. Therefore in practice we usually obtain only one solution for each run. One way to obtain multiple solutions is to run the simulation many times. Since the GA starts the search from random individuals, it might reach different outcomes with different runs.

The search for multiple solutions by multiple runs of the GA may also produce the same results by different runs since the GA does not know what it has explored in the previous run and may plunge into the same area. Consequently simple repeat of runs is ineffective

and time consuming. One way to solve this problem is to make the GA memorize what it has done and not explore those already explored areas. From the above discussions, we know that we can divide the search space into different sections and each section will produce an optimum, then we may obtain our desired multiple outputs out of those sections. In order that the algorithm knows what it has explored, we must record those optimal angles that the algorithm has found. Then when the successive runs approach that area, the algorithm can force it to leave that area.

Suppose that we have two series of angles $(\alpha_1, \alpha_2, ... \alpha_n)$ and $(\beta_1, \beta_2, ... \beta_n)$, the Hamming distance can be used to measure the distance between $(\alpha_1, \alpha_2, ... \alpha_n)$ and $(\beta_1, \beta_2, ... \beta_n)$ for the binary encoding and the Euclidean distance can be used for the floating point encoding [47]. If the distance between them is smaller than a certain value (section radius), we may consider that $(\beta_1, \beta_2, ... \beta_n)$ is now exploring the already explored area of $(\alpha_1, \alpha_2, ... \alpha_n)$ and we can force $(\beta_1, \beta_2, ... \beta_n)$ to leave this area. The Euclidean distance between $(\alpha_1, \alpha_2, ... \alpha_n)$ and $(\beta_1, \beta_2, ... \beta_n)$ is defined as [47]:

$$\sqrt{\sum_{i=1}^{n} (\alpha_i - \beta_i)^2} \,. \tag{3.27}$$

To force an individual to leave an area can be implemented by reducing the opportunities for the individual to survive when it is close to that area. Since the probability of an

individual to survive depends on its fitness; the higher the fitness the higher the chance to survive. Hence, when an individual approaches the already explored area, we can lower its fitness in accordance with the distance to the best individual in that area. Therefore the new fitness (processed fitness) of an individual x can be expressed as:

$$F'(x) = F(x) * G(x, s_1, s_2, ...s_n), \tag{3.28}$$

where $F'(x)$ is the processed fitness, $F(x)$ is the original fitness and $G(x, s_1, s_2, ...s_n)$ is an evaluation function that corresponds to the distance between the individual x and those already found best individuals $(s_1, s_2, ...s_n)$ [47] [51]. The closer the distance, the smaller the $G(x, s_1, s_2, ...s_n)$ will be. The evaluation function used in our design can be expressed as [47]:

$$G(x, s_1, s_2, ...s_n) = \begin{cases} (d(x, s_1, s_2, ...s_n)/r)^{\alpha}, d(x, s_1, s_2, ...s_n) < r \\ 1, otherwise \end{cases}, \tag{3.29}$$

where $d(x, s_1, s_2, ...s_n)$ is the minimum Euclidean distance of the individual x and those already found best individuals $(s_1, s_2, ...s_n)$, α is a system parameter and it is 5 in our example. From Equation 3.29, we know that the distance is inverse to the number of sections. r is the section radius which is given by [47]:

$$r = \frac{\sqrt{n}}{2 * \sqrt[n]{p}}, \tag{3.30}$$

where n is the number of genes. In our design, n is the number of birefringent sections and p is the number of the divided problem space.

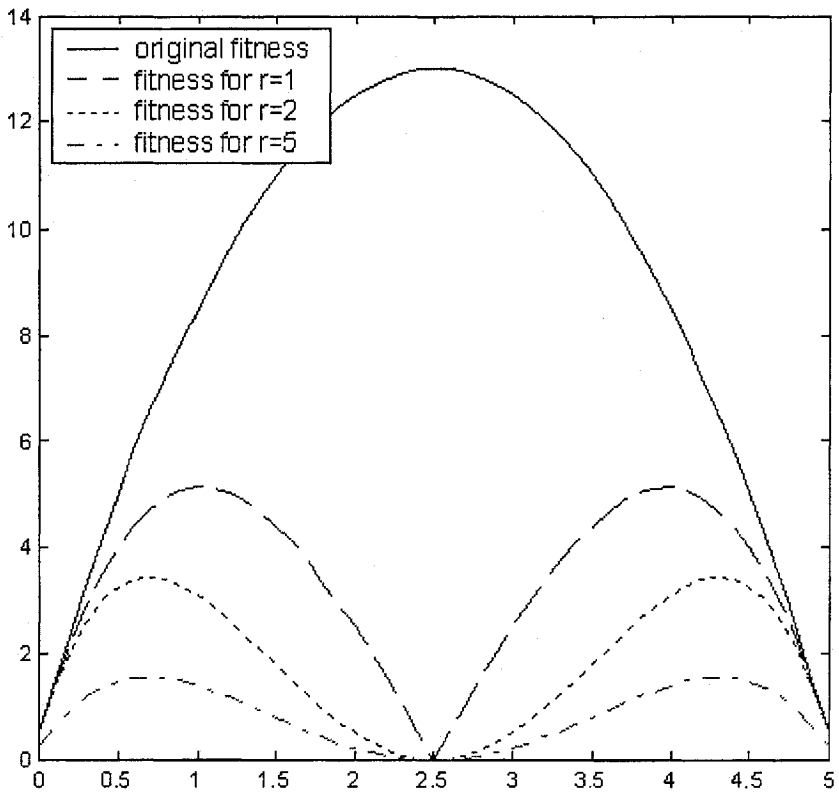


Figure 3.12 Original fitness and processed fitness.

Figure 3.12 shows the processed fitness with different parameter α . We can see that α actually defines the decrease speed of the processed fitness when an individual is close to the best individual.

3.4.3 Reduce search space for birefringent filter design

The search space can be further reduced in birefringent filter design. For an n-section (n is an even integer) birefringent filter, the shift angle $(\theta_0,\theta_1,\theta_2,...\theta_n)$ represents a certain spectral response. We can prove that for $(\theta_0,\theta_1,\theta_2,...\theta_n)$, every two angles that are symmetrical to the center angle $\theta_{n/2}$ can be exchanged without altering the output spectral response. We briefly describe the derivation of this theory in two steps:

Steps1: For a 2-section filter, use Equation 3.10, the output is

$$R(\theta_2)R_\eta R(\theta_1)R_\eta R(\theta_0)=\begin{bmatrix}\cos\theta_2 & \sin\theta_2\\ -\sin\theta_2 & \cos\theta_2\end{bmatrix}\begin{bmatrix}e^{j\eta/2} & 0\\ 0 & e^{-j\eta/2}\end{bmatrix}\begin{bmatrix}\cos\theta_1 & \sin\theta_1\\ -\sin\theta_1 & \cos\theta_1\end{bmatrix}\begin{bmatrix}e^{j\eta/2} & 0\\ 0 & e^{-j\eta/2}\end{bmatrix}\begin{bmatrix}\cos\theta_0 & \sin\theta_0\\ -\sin\theta_0 & \cos\theta_0\end{bmatrix}.$$

$$\text{Set}\begin{bmatrix}e^{j\eta/2} & 0\\ 0 & e^{-j\eta/2}\end{bmatrix}\begin{bmatrix}\cos\theta_1 & \sin\theta_1\\ -\sin\theta_1 & \cos\theta_1\end{bmatrix}\begin{bmatrix}e^{j\eta/2} & 0\\ 0 & e^{-j\eta/2}\end{bmatrix}=\begin{bmatrix}A & -B^*\\ B & A^*\end{bmatrix},$$

where $A=e^{j\eta}\cos\theta_1$ and $B=-\sin\theta_1$, then the intensity can be expressed as

$\left|A\cos\theta_2\cos\theta_0 - A^*\sin\theta_2\sin\theta_0 + B\sin\theta_2\cos\theta_0 + B^*\cos\theta_2\sin\theta_0\right|^2$, therefore the intensity must be the same if the $\theta_2 and \theta_0$ are exchanged.

Steps2: Use the induction method. If the n-section birefringent filter agrees with this theory. Then for an (n+2)-section filter, since the output of the n-section filter can also be expressed as $\begin{bmatrix} A & -B^* \\ B & A^* \end{bmatrix}$, the intensity can be expressed as $\left|A\cos\theta_{n+2}\cos\theta_0 - A^*\sin\theta_{n+2}\sin\theta_0 + B\sin\theta_{n+2}\cos\theta_0 + B^*\cos\theta_{n+2}\sin\theta_0\right|^2$, which also satisfies the theory.

This theory tells us that for each solution of angles $(\theta_0,\theta_1,\theta_2,...\theta_n)$, we will have $2^{n/2}$ solutions which can realize the same spectral response. Therefore, for every solution we find, we can disable the search for the other $2^{n/2}$ areas.

3.4.4 Multiple outputs of 4-section and 8-section birefringent filters

The purpose that we divide the search space into sections is not only to discover multiple designs but also to increase the performance of the GA. The separation of the search space reduces the search space the GA is supposed to search for each run. Furthermore, since the GA memorizes those already explored spaces it can make the successive iterations not

repeat the search for the already explored areas, thus improving the convergence time. The investigation of the number of sections for a specified spectral response is beyond the scope of this thesis. In our design example, we just simply assign a fixed value to the number of sections p. Normally, the larger the p is, the better the performance will be. In the same way, we have to make a trade off between the performance and the speed.

Table3.4: Multiple designs for a 4-section birefringent filter.

Section Number	0	1	2	3	4
angles of fig3.13.a	-0.1568	-0.4014	-0.3698	-0.1222	-1.5359
angles of fig3.13.b	-0.5379	1.2600	-0.2003	-0.7740	-0.3316
angles of fig3.13.c	0.3642	1.1707	0.7733	-1.3469	0.2594
angles of fig3.13.d	-0.2984	-0.9388	-0.7899	1.3151	-0.1106

We use the method that we have mentioned above to redesign the 4-section birefringent filter and Table 3.4 presents the 4 design outputs. Figure 3.13 shows their respective spectral responses (the desired spectral response is (0, 0.2, 0.5, 1)).

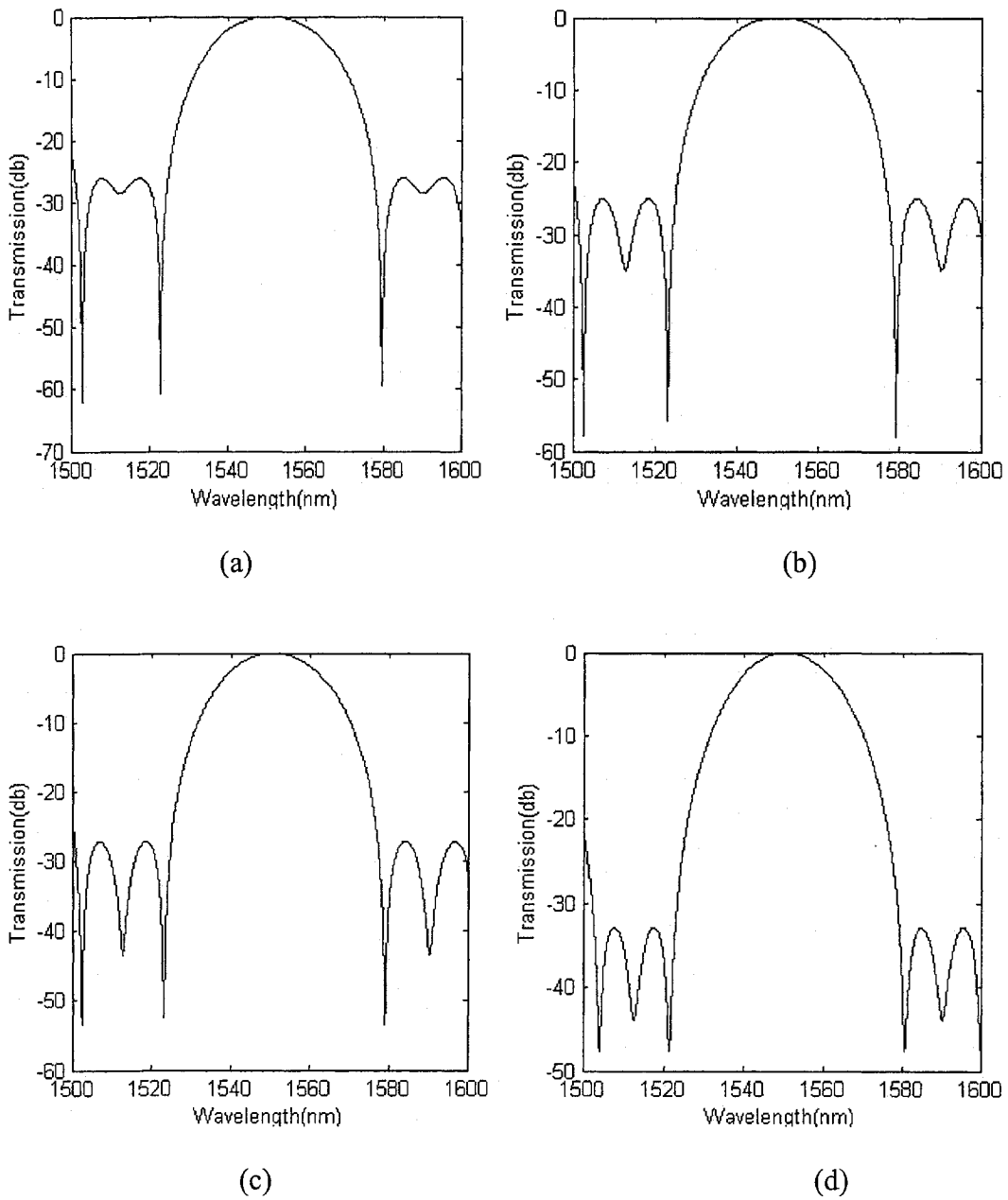


Figure 3.13 Multiple designs of a 4-section birefringent filter.

Among the 4 designs given in Table 3.4, the design in Figure 3.13(d) presents the best sideband suppression ratio, up to nearly 33dB. The sideband suppression ratios in Figure 3.13(a) and Figure 3.13(c) are around 27 dB, which are also better than the previous design by a normal GA. Figure 3.13(b) shows the spectral response with the lowest sideband suppression ratio of 25 dB. However, Figure 3.13(b) presents the best passband flatness. As these data shows, the technique of reducing the search space and the processed fitness in the GA can obviously improve the performance of the design.

Table 3.5 Multiple designs for an 8-section birefringent filter.

Section number	Angles of Figure 6 a	Angles of Figure 6 b	Angles of Figure 6 c	Angles of Figure 6 d
0	1.1127	-1.1941	0.9402	-1.0310
1	-0.1014	0.3451	0.5356	0.1949
2	1.2369	-0.2822	-1.5526	-0.8540
3	-0.0663	1.1229	0.5889	-1.1232
4	1.4689	-1.2107	1.2154	1.2004
5	1.4942	-0.9924	0.1726	-0.0335
6	-0.7399	0.0241	-0.7203	-0.0698
7	0.8166	-1.0186	1.0324	-1.2765
8	0.6282	0.5483	0.8727	0.0819

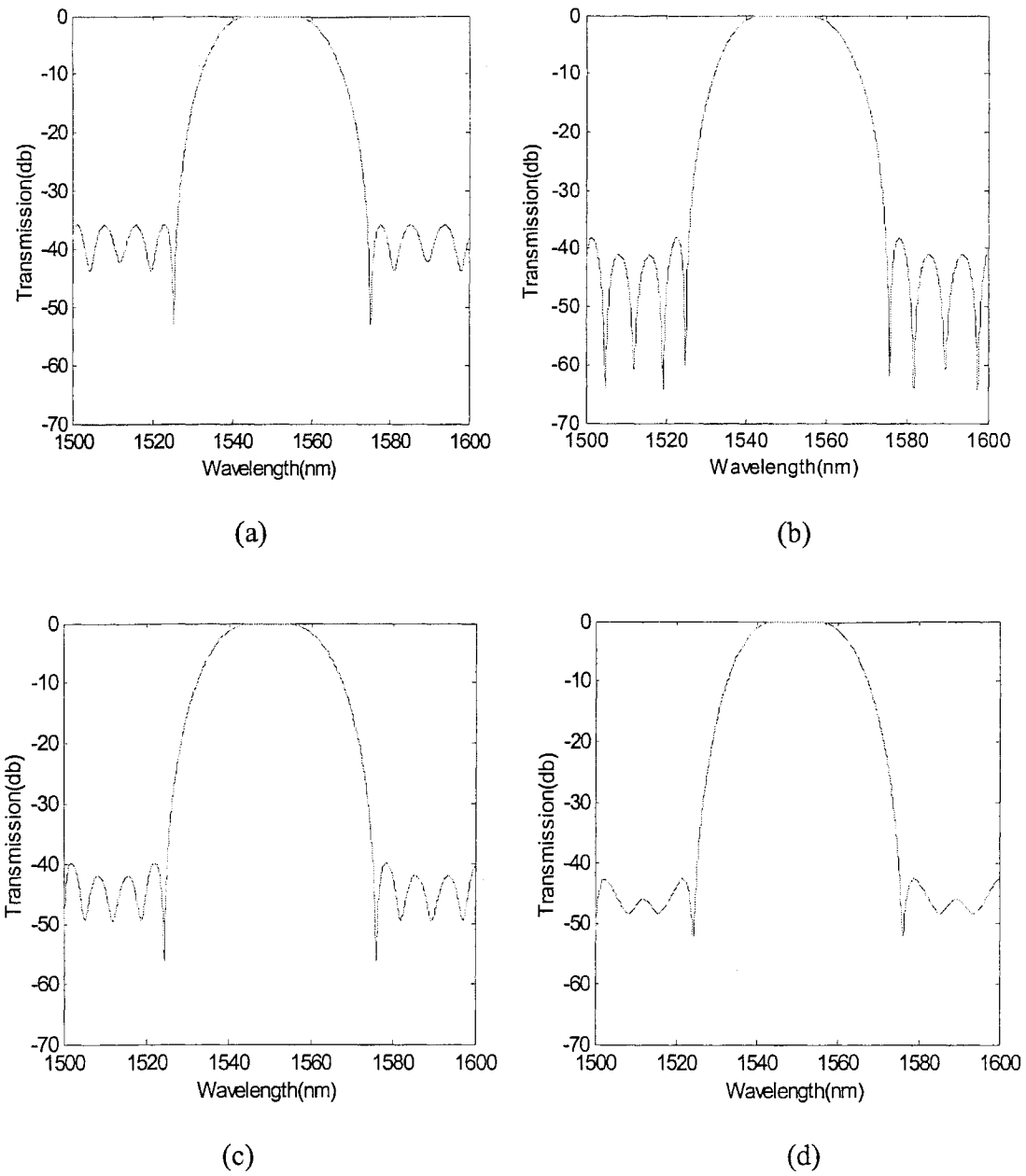


Figure 3.14 Multiple designs of an 8-section birefringent filter.

Table 3.5 and Figure 3.14 present the designs and their corresponding spectral responses of an 8-section birefringent filter by multiple runs of the GA. The response shown in

Figure 3.14(a) has a similar sideband suppression ratio as that in Figure 3.11 achieved by a normal GA. But the spectral response of the passband shown in Figure 3.16(a) is within 1.1 dB, which is much better than that in Figure 3.11. Figures 3.14(b) and (c) has higher sideband suppression ratio, up to 40 dB but the designs are totally different from that shown in Figure 3.14 (a). Figure 3.14 (d) achieves a side suppression ratio of 42 dB and the passband deviation within 1dB. Therefore, as far as both the passband and stopband are considered, the design shown in Figure 3.14 (d) has the best performance. The results indicate again that the increase of birefringent sections can improve greatly the performance of birefringent filter. Furthermore, the results also indicate that the multiple run of the GA is necessary since not only it improves the performance but also it provides different output that we can choose as our desired solution.

3.4.5 Local Search by the SA

In our holographic diffuser design, we incorporate the SA [18] [19] into the GA to maintain the population diversity of individuals. In the birefringent filter design, since we employ the floating point encoding and the processed fitness, the genes will disappear not as fast as in the holographic diffuser design. Therefore, the population diversity can be maintained much better in this case. We also use the SA to design the birefringent filter independently; we find that the SA does not demonstrate a good capacity in locating an optimum solution in such a case. The fact that we obtain several designs for a given spectral response reveals that there exist many optima in the problem space. When we try

to alter a shift angle up to ± 0.01 radians we find that it might influence the sideband suppression ratio to up to 10 dB. However, ± 0.01 radians occupy only 0.6% of the problem space $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, which indicates that an optimum exists in a very narrow area. A little step out of that area will cause the fitness to decrease sharply. Although we can employ a very low cooling schedule in the SA in order to better locate a nearly global optimum, it will take longer time and it is impractical in our simulations. The GA compensates such a problem by employing a large population size to start the search toward different directions and it places its candidates into different narrow areas. Therefore, it performs better than the SA in such a case. Our simulations also demonstrate that the SA does not perform as well as the GA in the case of birefringent filter design. However, one advantage of the SA over the GA is that it can finely locate the optimum in a narrow area since the operator of the SA is simply mutation while the operator of the GA is complicated and those operations might cause the GA to deviate the precise position. Hence, in our algorithm we finally employ the SA as a delicate search and local search [20] [28] algorithm to precisely locate the final designs.

Because in this case the GA has already found the nearly optimum solution, the cooling temperature for the SA must be very low to avoid the drift from that area. The simulation results demonstrate that the local search by the SA can improve the sideband suppression ratio to up to around 0.2 dB.

3.4.6 Flowchart of the algorithm

By summarizing all the steps we discussed above, we have the flowchart for our algorithm:

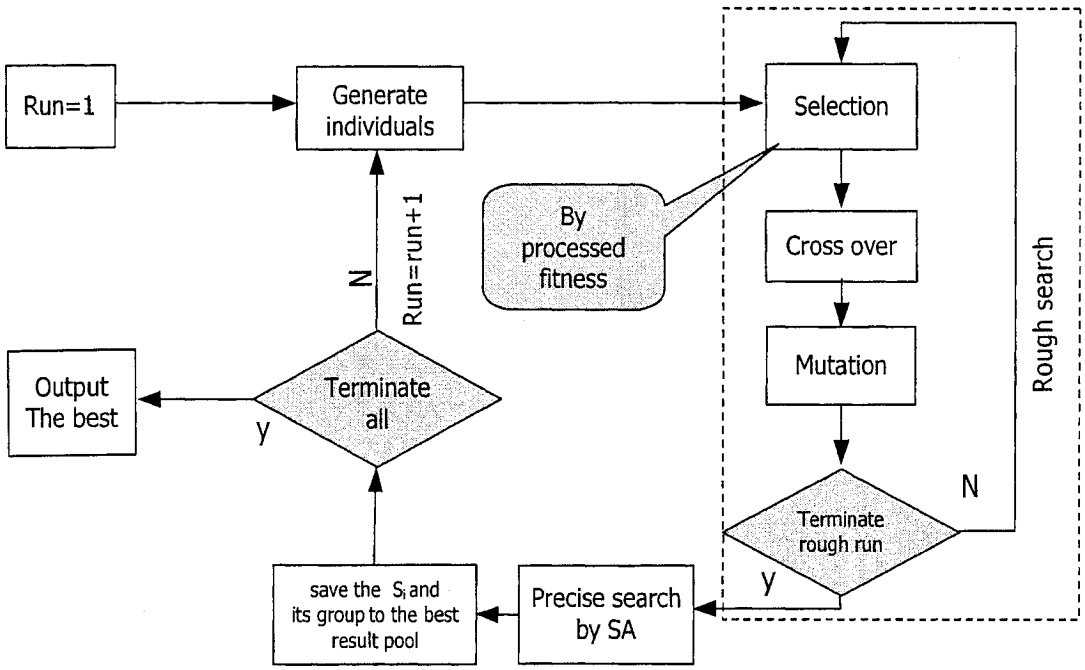


Figure 3.15 Flow chart of birefringent filter design by genetic algorithm.

3.5 Experiments

3.5.1. Experimental procedure and theory model

Base on the theoretical model and the simulation, a 4-section identical length birefringent filter is fabricated and tested. The experimental setup is shown in Figure 3.16.

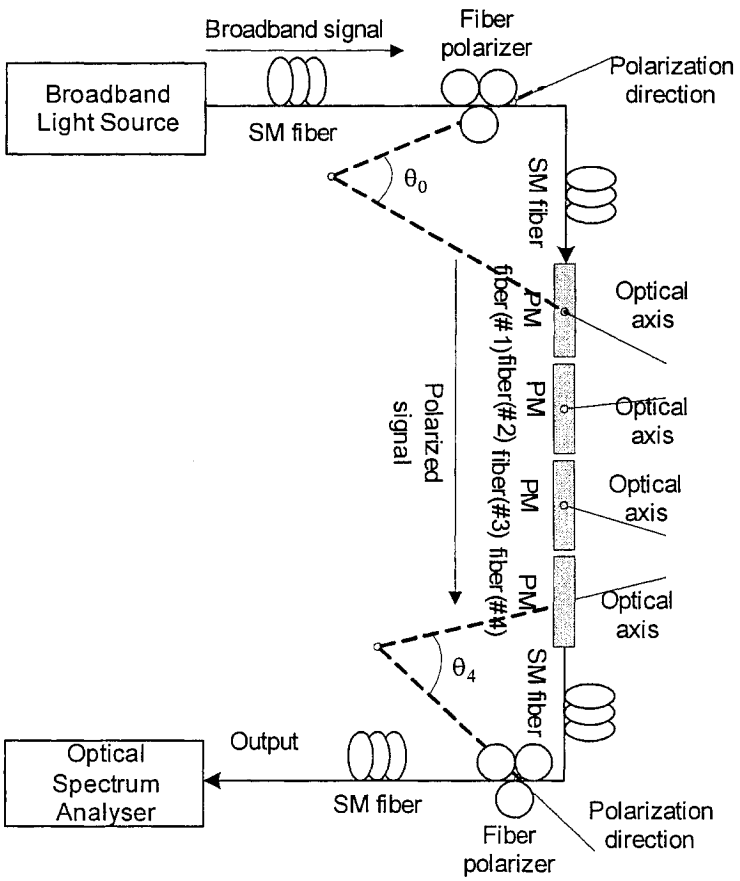


Figure 3.16 Experimental Setup.

The filter is fabricated by 4-section PM fibers and we use a splicer to splice them together. Each end of the PM fiber is spliced to a single mode fiber which is connected to a fiber polarizer. The fiber polarizers are used to control the input and output shift angles of the birefringent filter. In addition, the input polarizer converts the un-polarized light from the broadband source to linearly polarized light. The 3 shift angles and the input and output shift angles are determined by the GA. The spectral response is observed on an Optical Spectrum Analyzer.

The angular resolutions of the polarizers and the splicer that we use for our experiments are one degree; therefore we have to round the design results to integer numbers. Furthermore, the spectrum range of the broadband source is from 1535 nm to 1565 nm; in order to observe more periods in this range we must use longer fiber sections to get smaller FSR. The parameters of the PM fiber that we used in our experiments are:

Birefringence: 4×10^{-4}

Section length: 33.3 cm

The FSR calculated based on Equation 3.26 is 2.292×10^{12} Hz.

Table 3.6 Birefringent filter shift angles in the experiment

Section Number	0	1	2	3	4
Shift angles of filter (degree)	-54	-42	70	-11	19

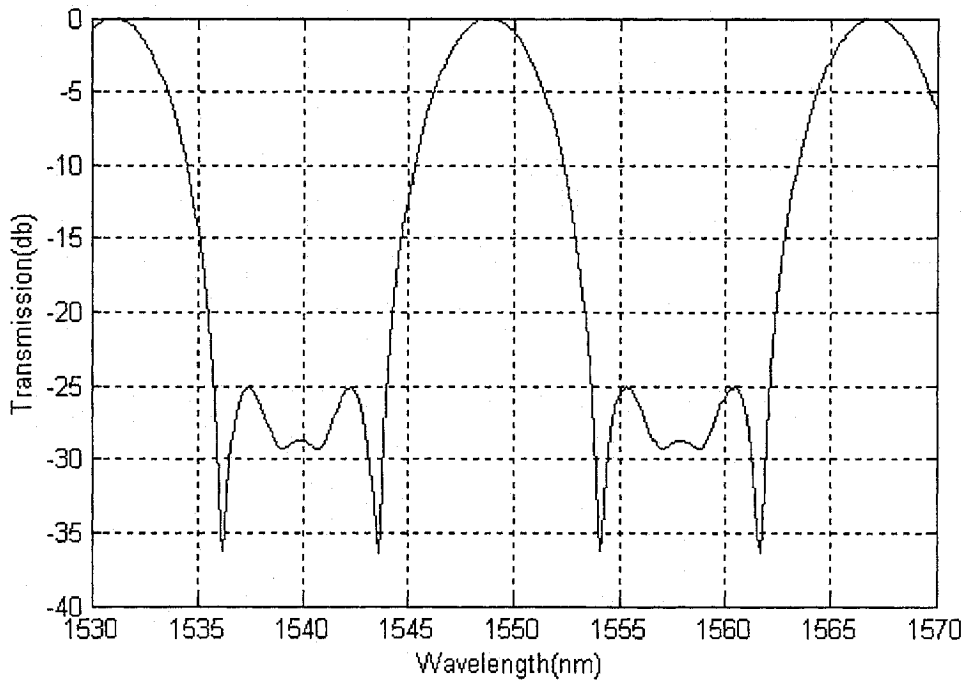


Figure 3.17 Theoretical spectral response.

The shift angles of the filter are shown by Table 3.6 and its theoretical spectral response are shown in Figure 3.17. The rounding of angles will affect the output spectrum. Based on our simulation, the extent of the influence on the stop band suppression ratio usually ranges from 1-3 dB. This issue will be further explained in Figure 3.19

3.5.2 Experiment results and discussions

To do the experiments, we must first splice the PM fibers according to the shift angles shown in Table 3.6. The angles between each section are angle 1, 2 and 3 degrees. At the two ends of the filter, we splice the PM fibers to the SM fibers. The splicing angle can be controlled from 0^0 to 90^0 with a resolution of 1^0. For those shift angles that are from -90^0 to 0^0, the only thing we need to do is to change their sequence of splicing. For example, if we want to splice a shift angle of -30^0 between the elements 1 and 2, we can inverse the sequence of splicing of the elements 1 and 2. Since one cannot control the shift angle between a PM fiber and an SM fiber, we use two polarizers to adjust the angles of angle θ_0 and θ_4 shown in Figure 3.16. However, since we cannot know the shift angle when adjusting the two polarizers, so what we can do is to tune the two polarizers to find the output response that best approximates our theoretical response. The tuning range of the polarizers used in the experiments is from 0^0 to 360^0. Because the period of the shift angle is 180^0, we can fix one polarizer and tune the other from 1^0 to 180^0. Then we increase the fixed polarizer by a small step size (i.e. 1^0) and tune the other polarizer from 1^0 to 180^0 again. By this way, we can observe all the spectral outputs of all the combinations of the shift angles θ_0 and θ_4

In order to make sure that the shift angles θ_0 and θ_4 that we tune agrees with our theoretical design, we perform an additional simulation: Fix angle θ_0 at angles from 1^0 to 180^0 while increasing angle θ_4 from 1^0 to 180^0. We find that our theoretical results shown in Table 3.6 provide the lowest sidelobe level. We might conjecture that the angles with the lowest output sidelobe level will approximate out theoretical angles. Therefore during the experiments we will trace and record the data that corresponding to the lowest sidelobe output.

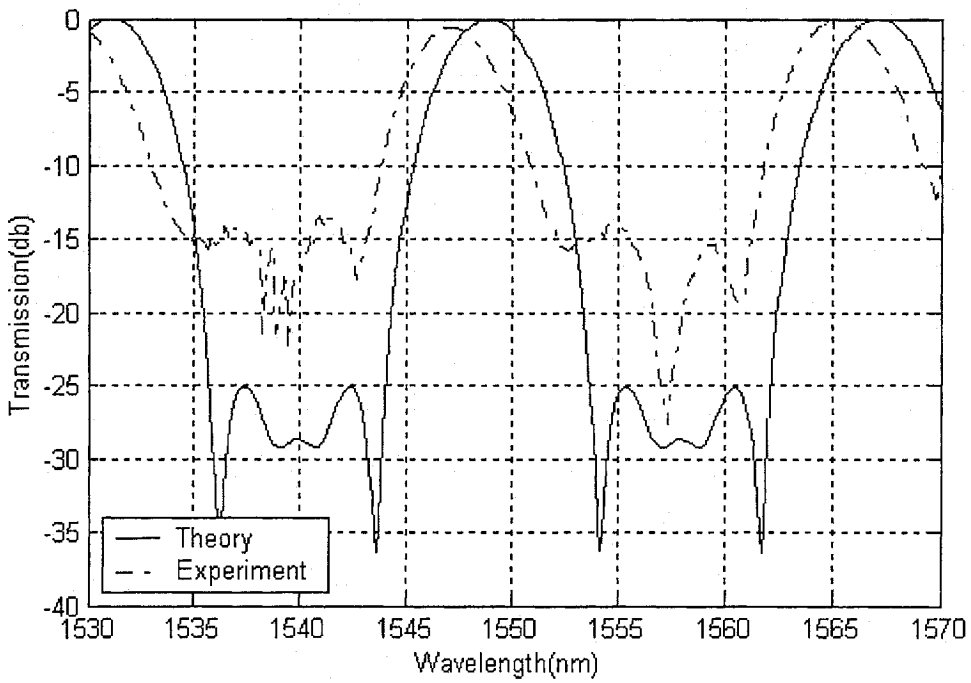


Figure 3.18 Spectral responses of theory (solid line) vs. experiment (dotted-dashed line).

The experiment results are shown in Figure 3.18. Form the figure, we can find that the period of the experimental filter agrees well with the theory one. The shape of the

experimental response is also close to the theory one. However, the theoretical sideband suppression ratio is 25 dB while the sideband suppression ratio obtained experimentally is 14 dB, about 11 dB less then the theoretical result. Furthermore, the center wavelength has an offset of around 3 nm from the theory center wavelength.

We then fabricate a few more filters based on the same design parameters in Table 3.6. The experimental results are similar: The FSR is close to the theoretical value while the center wavelength is drifting a little away from the theoretical center wavelength and the sideband suppression ratio is more than 10 dB less than the theoretical value. To find the reasons that lead to these discrepancies, we first intentionally introduce some errors into the shift angles. Since the resolution of the polarizer is 1^0, we randomly select the shift angles from 0 to 4 and add or subtract 1^0 to the original angles. The responses with and without shift angle errors are shown in Figure 3.19.

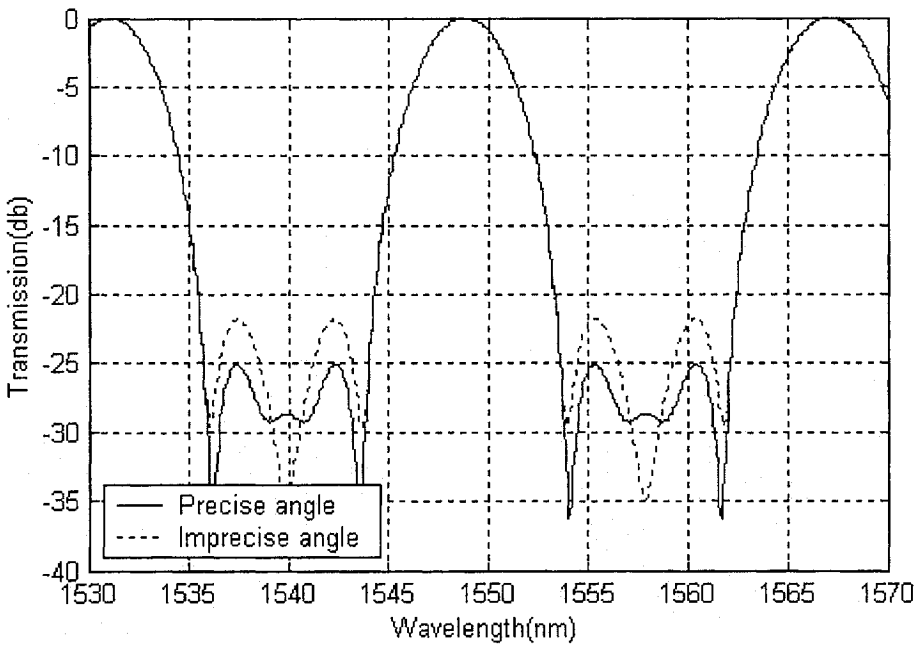


Figure 3.19 Precise angles (solid line) vs. angles with up to 1^0 error (dashed line).

From Figure 3.19 we find that the angle errors will influence the sideband suppression ratio, but the centre wavelength is not affected.

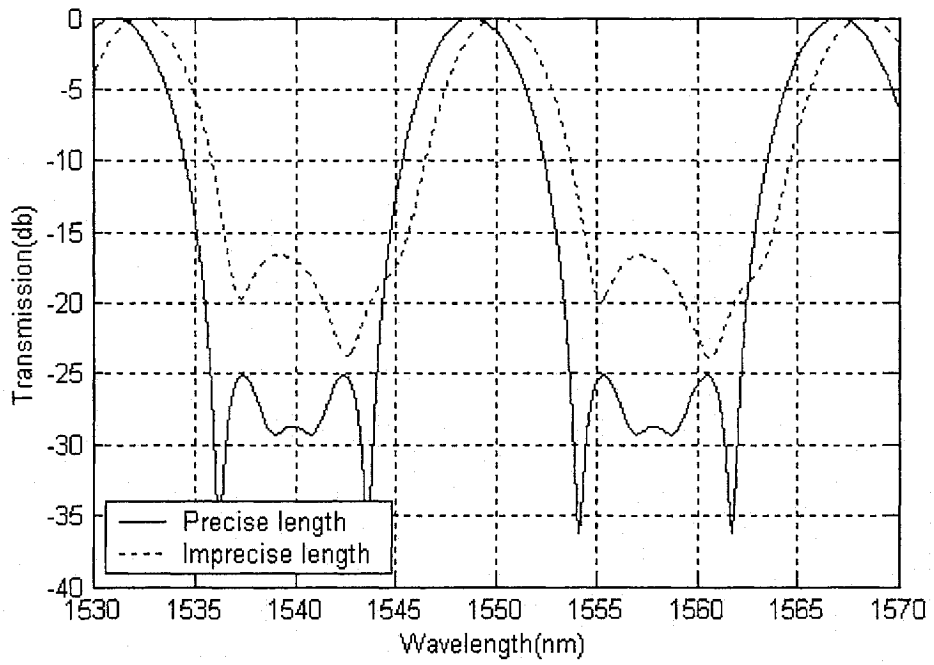


Figure 3.20 Precise lengths (solid line) vs. imprecise lengths up to 1mm error (dashed line)

In the same way, we intentionally introduce errors to the fiber length. We randomly add or subtract 1mm to the original design length of sections 1 to 4. Figure 3.20 shows the spectral responses of the filter with precise section lengths and lengths with errors. From the Figure we find that the length errors have a great impact on the output spectral response. It will affect not only the sideband suppression ratio but also the center wavelength. From Equation 3.24, we know that the section length is inversely propositional to the fiber birefringence. To reduce the influence of length errors on the center wavelength, PM fiber with smaller birefringence may be used. Another solution to this problem is to use integrated optical waveguides. Since the waveguide length can be

controlled very precisely, the spectral response of the birefringent filter based optical waveguides can provide precise spectral response.

3.6 Summary

The GA has been proposed to design birefringent filters. The floating point encoding and the binary encoding schemes were compared. The results showed that the floating point encoding can provide a more steady output. The step size was used to control the mutation probability and the arithmetical cross-over was used to improve the cross-over performance. An improved GA was proposed to use processed fitness to memorize those already explored areas, to get multiple outputs and improve the performance. The simulation results showed that the improved GA can search in a wider range and get lower sidelobe levels compared with the normal GA. Finally, the errors introduced to the shift angles and the section lengths were investigated. It was found that the errors in sections lengths would not only affect the center wavelength but also the sidelobe suppression ratio.

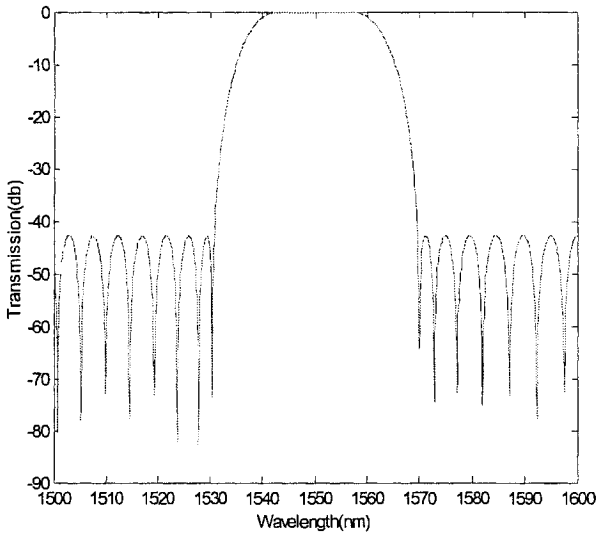
Chapter 4

Comparison of the GA with other algorithms for filter design

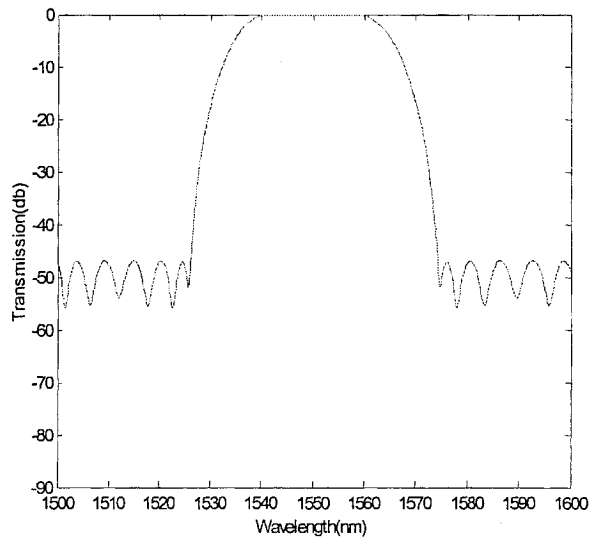
A comparison between the GA and other established algorithms for the design of birefringent filters is investigated in this chapter. As we discussed above, the design of a birefringent filter is to find the section lengths and shift angles between adjacent elements such that the output best approximates the desired spectral response. Many algorithms have been proposed [11] [13] [14]. Based on our investigation, there are two major advantages of the GA over the other algorithms: the ability to obtain better output and a more flexible structure.

4.1 Comparison of the GA with digital filter design algorithms

Chu and Town [11] proposed to use digital filter algorithm (Remez algorithm) to determine an optimal polynomial approximation to obtain a specified finite impulse response. In their example they used a 14-sections birefringent filter to design a bandpass filter with a desired spectral response $(0, 0.2, 0.5, 1)$ and they achieved a sideband suppression ratio of up to 42 dB. To compare the performance, we use the GA to design a 14-section birefringent filter with all the other parameters identical to those in [11] .



(a)



(b)

Figure 4.1 Comparison of the GA and the digital filter design algorithm. (a) Spectral response obtained by digital filter design algorithm. (b) Spectral response obtained by GA.

As Figure 4.1 shows, he birefringent filter that designed by the GA has a sideband suppression level of about 47 dB. The passband flatness of the filter by the modified GA is also better than that in the example of [11] and the deviation of the passband is within 0.3 dB while the deviation of the passband of the example [11] is 1 dB. The sideband suppression ratio of the birefringent filter designed by the digital filter design algorithm is around 42 dB, which means that the GA improve the side band suppression level to 5 dB and the passband flatness to 0.7 dB. Furthermore, the 8-section filter shown in Figure 3.14 (d) has a sideband suppression ratio of 42 dB, which is identical to that of a 14-section filter designed using the digital filter algorithm [11]. This indicates that filters designed by the GA can reach the same performance with fewer sections.

4.2 Birefringent filters with non-identical section length

The design of the Lyot filter is to find the length of each section while the shift angle between any two adjacent sections is identical. On the other hand, the design of the Solc filter is to find the shift angles between two adjacent sections while keeping the section lengths identical. The Lyot filter and Solc filter are two special examples of birefringent filters. Our designs presented above are Solc birefringent filters with identical section lengths and arbitrarily angles. A more general birefringent filter is the one that both the section lengths and the shift angles are adjustable. However, it is very difficult for the digital filter design algorithm to design such a birefringent filter since the digital filter design algorithms are backward algorithm and it is difficult to induce both the lengths and

angles directly from the desired output spectral response. One major advantage of the GA over other algorithms is that it is a forward search algorithm and the design process of it is "try and compare". Therefore, with some modifications, the GA can be applied to a more general birefringent filter design. The modifications are summarized as below:

1: For an N-section birefringent filter, the chromosome is represented as $(\varphi_1, \varphi_2 ... \varphi_{n+1}, l_1, l_2, ... l_n)$, here φ_i is the shift angle between the (i-1)-th element and the i-th element, while l_i is the length of the i-th element. l_i is integer times of the shortest element length which is determined by Equation 3.25.

2. The genetic operator of selection is the same as the Solc filter design while the cross-over and mutation operators are categorized into 2 groups: the group of shift angles and the group of lengths. They are characterized by different step sizes and different ranges.

3. The calculation of section radius is also modified to adapt to the group range. An individual is considered to be in the same section only when its fiber element length is the same as the individual in that area.

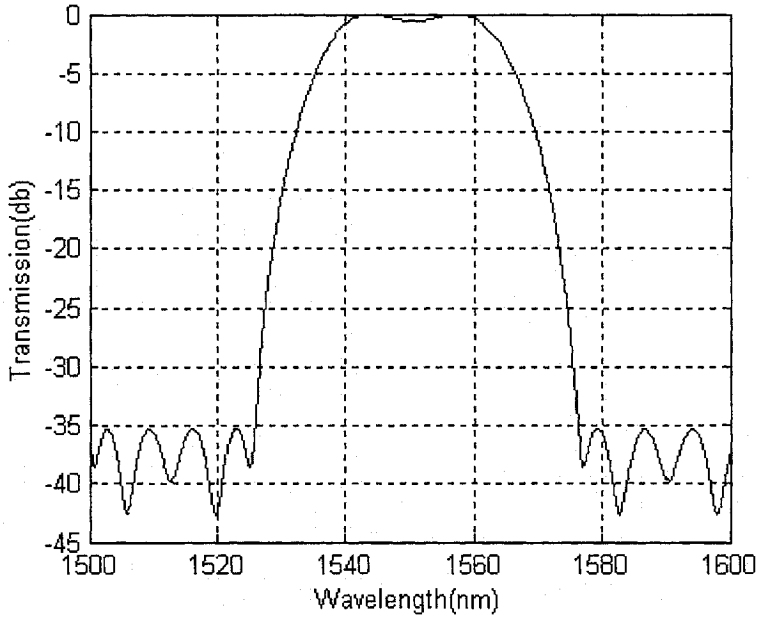


Figure 4.2 A 4-section birefringent filter with non-identical section lengths.

Figure 4.2 shows the output spectral response of a 4-section birefringent filter with non-identical section length. The passband is flat and the sideband suppression ratio is more than 35 dB. When it is compared with the 4-section birefringent filter with identical length in Figure 3.13, the sideband suppression ratio is improved by about 7 dB. The simulation results demonstrate that birefringent filters with adjustable lengths and angles can produce better output than those with only adjustable angles. The filters with both adjustable lengths and angles can be viewed as a special type of birefringent filters. The fiber elements with the length integer times as long as the shortest element can be viewed as formed by several shortest elements and the shift angles between these shortest elements

are 0. Therefore the improvement of performance is actually realized by the increase of section numbers.

4.3 Summary

The GA has been compared with digital filter algorithms from two perspectives: the performance and the ability to realize a more flexible structure design. The results showed that the GA has better performance. The GA was also used to design filter with both section lengths and shift angle variable. The results showed that better performance could be achieved with less filter sections.

Chapter 5

Conclusions and future work

5.1 Conclusions

The objectives of this work are to use the genetic algorithm to design two important optical components: the holographic diffuser and the birefringent filter. From the two design examples, we investigated the process to use the GA to optimize parameters under two different conditions: discrete and continuous. We also employed different methods to improve the algorithm for the two examples.

In chapter 1, the background of this study was presented. The objectives and major contributions of the work were also presented in this chapter.

In chapter 2, the theory model of the holographic diffusers was discussed and two designed examples of 16×16 and 64×64 image were given. The algorithm was compared with other established algorithms such as the normal GA and the simulated annealing. In this chapter, we found that the convergence performance of the GA depends much on its population diversity. In the normal GA the population diversity depends mainly on its population size. To improve the performance of the GA, we proposed to combine the simulated annealing as a complementary mutation operator to diversify the population. A

holographic diffuser designed by the modified GA with 64×64 spot pattern fabricated on a quartz substrate was experimented.

In chapter 3, the theory models of birefringent filters including the Lyot filter and the Solc filter were discussed. Two designed examples including the 4-, 8-section filters were provided. In order to obtain multiple outputs and improve the spectral response of the birefringent filter, the search space of the GA was reduced by dividing the search space to smaller sub-space. As a result, the population size needed was reduced. A 4-section filter with identical section length designed by the improved GA was used to fabricate a birefringent filter. The filter was tested and the results were analyzed.

In chapter 4, the GA was compared with digital filter design algorithms and their performance was investigated. The GA was also proposed to design birefringent filters with both angles and lengths variable. The simulation results of a 14-identical section and 4-non-identical section filters were presented.

In conclusion, the GA is well suited for optical design problems, especially in the condition that the inverse transform of the object problem can not be well specified. Based on the two design examples, in the following sections we will summarize the advantages, existing problems and possible solutions when using the GA for optical design.

5.1.1 Advantages of the GA

One and perhaps the most important advantage of the GA is its flexibility. Unlike other algorithms, the GA does not require intensive knowledge of the problem we are to optimize. The thing it must know is how to map the source parameters to the object functions, which means that a forward theory model is enough for the GA. For example, in the birefringent filter design; as long as the GA knows how to transform the section lengths and shift angles to the output spectral then it can start to search the optimal solution for a required problem. Other algorithms such as the digital filter design algorithms must establish a relatively complete mathematical model to start the optimization work. Only the forward model is usually not enough to use a digital filter design algorithm. Therefore it is difficult for the digital filter design algorithm to design a birefringent filter with both section lengths and shift angles variable while for the GA it does not make much difference whether the section lengths the same or not.

Another advantage of the GA is that it can generate family designs by ways of its large population size and random search process. We have used some mathematical functions with known multiple peak values to test the GA. We find that with proper usage of the GA operators, the GA can find all the peak values. In the birefringent filter design, we use this feature of the GA to find family solutions for a given spectral response by the multiple runs of the GA. We can also find the family solutions by a single run of the GA. In this case, we will divide the population pool into a number of sections and confine those

genetic operations between the individuals to a certain section. Such method can reduce the runs of the GA but increase the population size. As far as speed is concerned, the former method that used in this thesis is faster.

The other advantage of the GA is that it is theoretically capable of finding the best solution for a problem as long as the population size and number of iterations are large enough since the leading process of the GA is to select a more competitive candidate for the next generation. Other algorithms sometimes have to use approximation to represent an object function and acquire only approximate optimal solution. From the theoretical point of view, if with meticulous design, the GA should at last find the best solution for a given problem. In the birefringent filter design, we compare the performance of the GA with other algorithm and demonstrate that the GA outperforms the other algorithm in a given design example.

5.1.2 Existing problems and some possible solutions

As an algorithm that imitates the phenomenon of nature evolution, the GA also has some shortcomings for the optical design problems that need further investigation. One of the shortcomings is that the GA always takes longer calculation time than the other algorithms. In the holographic diffuser design we compare the GA with the SA and find that the time of the GA is longer. In the birefringent filter design the GA also takes longer calculation time than the digital filter design algorithm does. The problem can be mitigated by

running the GA in parallel since it is an inherently parallel algorithm. Another approach is to increase the population diversity so that we can use less population. Another problem regarding the simulation time is that the speed of the GA also depends on the order of our problem. For example, the simulation time of the 64×64 hologram diffuser and the 14-section birefringent filter is much longer than that of the 16×16 holographic diffuser and the 4-section birefringent filter respectively. Therefore in practice, we have to use fewer iterations and sacrifice the performance to improve the speed. This is one of the reasons that the GA might not always outperform other algorithms.

Another shortcoming of the GA is that it does not always work consistently and cannot completely prevent itself from trapping into a local optimum, especially when the order of the problem space is large since we find that the GA works more consistently in the 4-section birefringent filter synthesis than in the 14-section birefringent filter synthesis. In the holographic diffuser design we compare the variance of cost function of the GA with that of the SA and find that the variance of the cost function of the normal GA is worse than that of the SA. This problem can be improved by introducing the SA as a mutation operator. Another technique of the processed fitness that adopted in the birefringent filter design can be used to prevent the GA from searching the already explored area then reduce the probability of trapping into a local optimum, hence make the GA work more consistently.

The other issue that leaves us for further investigation is that there is not a best theoretical way to find the various parameters for the genetic operators. In the holographic diffuser design we used the adaptive parameter control techniques to find the parameters such as mutation probability and cross-over probability. However, the methods will change when the given problems change. For example, in the holographic diffuser design, the cost function of an individual does not change much if a pixel is mutated and the mutation probability should be low. The birefringent filter design is on the contrary, since the optimum locates in a very narrow range, the cost function will change greatly with only a little change in the shift angle, so the mutation probability should be much larger. Therefore the finding of these parameters is sometimes empirically.

5.2 Future work

One future work will be focused on the existing problems discussed above. For the speed problem of the holographic diffuser design, one approach is to divide the desired diffuser into certain sections; then the higher order diffuser problem can be transformed to a number of lower order diffuser problems. Since the object image is an inverse Fourier transform of the diffuser and each pixel in the object image is determined by all the pixels in the diffuser, how to divide the holographic diffuser and compose the image is a problem that leaves to us for further investigation.

The design of holographic diffusers and birefringent filters has some similar characteristics: large ensemble of solutions, which means large search space in the GA. Other optical components also has similar feature. The performance of the GA is dependent on the search space, normally the large the search space the poorer the performance. Therefore the higher order design might performs poorer than the lower order design. Those methods that we use in the holographic diffuser and the birefringent filter design to improve the performance of the GA are essentially to enlarge the population diversity and to reduce the search space. Although the GA is a random search process and does not require intensity knowledge between the parameters of a problem, if we know more knowledge between those parameters, we can add some limitations to the problem space to reduce the search space. In the birefringent filter design with non-identical section length, since we know that the section length should be the integer times of the shortest section length to acquire a periodic output, we reduce greatly the range of the section length. A further study and employment of the limitations between the parameters will be very helpful to the increase of speed and the performance of the algorithm.

In the GA, the selection of an individual depends simply on the fitness of the individual. An individual will disappear even if it possesses some important gene. This characteristic of the GA might result in the quick disappearance of some important gene and the convergence of population. In the real world, the adaptability of an individual is fixed at birth. If a man is not strong when he was born, he can exercise more and become stronger.

Other excellent genes of this man such as intelligence can be inherited by his descendents. Therefore how to make the individual improve itself from the study is a useful topic. For example, an individual learns not to explore the already explored space of other individuals is part of the job of this aspect. How to make the algorithm possess more capability for self improvement will be very helpful for the increasing the speed and the diversifying the population for the GA, which will lead to a further improvement for the problems regarding the optical design.

Bibliography

[1] J. R. Brown and A.W. Lohmann, "Complex Spatial Filtering with Binary Masks," Applied Optics, 5(6), 967-969 (1966)

[2] J. P. Yao, G. C. K. Chen, and T. K. Lim, "Holographic diffuser for diffuse infrared wireless home networking," Optical Engineering, 42(2), 317-324 (2003)

[3] G. Dennis, "Holography, 1948-1971," Science, 177(46), 299-313,1972

[4] T. Yatagai, "Phase only computer generated hologram produced by an ion-exchange technique," Optics Letters, 13(11), 952-954 (1988)

[5] N. Yoshikawa, M. Itoh, and T. Yatagai, "Quantized phase optimization of two-dimensional Fourier kinoforms by a genetic algorithm," Optics Letters, 20(7), 752-754 (1995)

[6] J. R. Fienup, "Iterative method applied to image reconstruction and to computer-generated holograms," Optical Engineering, 19(3), 297-305 (1980)

[7] M. A. Seldowitz, J. P. Allebach, and D. W. Sweeney, "Synthesis of digital holograms by direct binary search," Applied Optics, 26(14), 2788-2798 (1987)

[8] F. Wyrowski and O. Bryngdahl, "Iterative Fourier-transform algorithm applied to computer holography," Journal of the optical society of America, 5 (7), 1058-1065 (1988)

[9] J. W. Evans, "The birefringent filter," Journal of the Optical Society of America, 39(3), 229-242 (1949)

[10] J. W. Evans, "Solc birefringent filter," Journal of the Optical Society of America, 48(3), 142 -146 (1958)

[11] R. H. Chu and G. Town, "Birefringent Filter Synthesis by Use of a Digital Filter Design Algorithm," Applied Optics, 41(17), 3412-3418 (2002)

[12] F. K. von Willisen, "A tunable birefringent filter," Applied Optics, 5(1), 97 -104 (1966)

[13] S. E. Harris, E. O. Ammann, and I. C. Chang, "Optical Network Synthesis Using Birefringent Crystals. I. Synthesis of Lossless Networks of Equal-Length Crystals," Journal of the Optical Society of America, 54(10), 1267-1279 (1964)

[14] Z. Zalevsky, D. Mendlovic, and E. Marom, "Tunable birefringent filters - optimal iterative design," Optic Express, 10(24), 1534-1541 (2002)

[15] J. Katzenstein and S. Ward, "Realization of a new type of birefringent filter," Applied Optics, 10(5), 1119-1124 (1971)

[16] J. M. Becker, "High-Resolution Measurements of Photosphere and Sun-Spot Velocity and Magnetic Fields using a Narrow-Band Birefringent Filter," Solar Physics, Vol. 3, 258-263 (1968)

[17] J. M. Beckers, Larry Dickson, and Randy S. Joyce , "Observing the sun with a fully tunable Lyot-Ohman filter," Applied Optics, 14(9), 2061-2066 (1975)

[18] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," Science, 220(4598), 671-680 (1983)

[19] S. Kirkpatrick, "Optimization by simulated annealing: Quantitative studies," Journal of Statistical Physics, Vol. 34, 975--986 (1984)

[20] Z. Michalewicz, "Genetic Algorithms+Data Structures = Evolution Programs," second edition, Springer-Verlag, Berlin (1994)

[21] A. E. Eiben, R. Hinterding, and Z. Michalewicz, "Parameter control in evolutionary algorithms Evolutionary Computation," IEEE Trans on Evolutionary Computation, 3(2), 124-141 (1999).

[22] J. A. Vasconcelos, J. A. Ramirez, R. H. C. Takahashi, and R. R. Saldanha, "Improvements in genetic algorithms," IEEE Journal on Magnetics, 37(5), 3414-3417 (2001)

[23] R. L. Haupt and S. E. Haupt, "Practical genetic algorithms," Wiley, New York, 1998

[24] G. Rudolf, "Convergence properties of canonical genetic algorithms," Journal of IEEE on Neural Networks, 5(1), 96-101 (1994).

[25] K. Sastry and D. E. Goldberg, "Modeling tournament selection with replacement using apparent added noise," Intelligent Engineering Systems Through Artificial Neural Networks, Vol. 11, 129—134 (2001)

[26] J. N. Gillet and Y. L. Sheng, "Multiplexed computer-generated holograms with polygonal-aperture layouts optimized by genetic algorithm," Applied Optics, 42(20), 4156-4165 (2003)

[27] K. P. Wong and Y.W. Wong "Genetic and genetic/simulated-annealing approaches to economic dispatch," IEEE Journal on Generation, Transmission and Distribution, 141(5), 507-513 (1994).

[28] G. Zhou, Y. Chen, Z. Wang, and H. Song, "Genetic local search algorithm for optimization design of diffractive optical elements," Applied Optics, 38(20), 4281-4290 (1999)

[29] P. L. Eardley, D. R. Wisely, D. Wood, and P. McKee, "Holograms for optical wireless LANs," IEEE Journal on Proc.-Optoelectron., 143(6), 365-369 (1996)

[30] A. J. Kemp, G. J. Friel, T. K. Lake, R. S. Conroy, and B. D. Sinclair, "Polarization effects, birefringent filtering, and single-frequency operation in lasers containing a birefringent gain crystal," IEEE Journal on Quantum Electronics, 35(2), 228-235 (2000)

[31] B. M. Schiffman and L. Young, "Birefringent Filter for Millimeter Waves," Journal of IEEE on Microwave Theory and Techniques, 16(6), 351-360 (1968)

[32] D. Bonaccini and F. Stauffer, "High resolution solar bidimensional spectroscopy with a universal birefringent filter in tandem with a Fabry-Perot interferometer - Tests and experimental results," Astronomy and Astrophysics, 229(1), 272-278 (1990)

[33] T. Kimura, M. Saruwatari, and K. Otsuka, "Birefringent branching filters for wideband optical FDM communications," Applied Optics, 12(2) 373-379 (1973)

[34] Y. Fujii and J. Minowa, "Four-channel wavelength multiplexing composed of phase plates and polarizing beam splitters," Applied Optics, 28(7), 1305-1308 (1989)

[35] A. Carballar, M. A. Muriel, and J. Azafi, "Fiber grating filter for WDM systems: an improved design," IEEE Journal on Photonics Technology Letters, 11(6), 694-696 (1999)

[36] P. Yeh, W. Gunning, R. Hall, and J. Tracy; "Dispersive birefringent filters for laser communications," IEEE Journal on Quantum Electronics, 17(12), 2424-2424 (1981)

[37] G. Holtom and O. Teschke, "Design of birefringent filter for high-power dye lasers," IEEE Journal on Quantum Electronics, 10(8), 577-579 (1974)

[38] L. J. November and F. Stauffer, "Derivation of the Universal Wavelength Tuning Formula for a Lyot Birefringent Filter," Applied Optics, Vol. 23, 2333-2341, (1984)

[39] J. M. Beckers, "Achromatic linear retarders," Applied Optics, Vol. 10, 973-975, (1971)

[40] C. Ye, "Wavelength-Tunable Spectral Filters Based on the Optical Rotatory Dispersion Effect," Applied Optics, 42(22), 4505-4513 (2003)

[41] Y. Zhou, L. R. Liu, J. Zhang, D. Liu, and Z. Luan, "Nearly-off-axis transmissivity of Solc birefringent filters," Journal of the Optical Society of America A, 20(4), 733-740 (2003)

[42] A. Lien, "A detailed derivation of extended Jones matrix representation for twisted nematic liquid crystal displays," Liquid Crystal, 22(2), 171-175 (1997)

[43] S. Y. Lu and R. A. Chipman, "Homogeneous and inhomogeneous Jones matrices," Journal of the Optical Society of America A, 11(2), 766-774 (1994)

[44] J. R. Perez and J. Basterrechea, "Near to far-field transformation using GA based optimization: real versus binary encoding schemes," Antennas and Propagation Society Symposium 2004 of IEEE, Vol. 1, 1122-1125 (2004)

[45] S. Rochet, M. Slimane, and G. Venturini, "Epistasis for real encoding in genetic algorithms," Australian and New Zealand Conference on Intelligent Information Systems, 268 – 271 (1996)

[46] V. S. Gordon and T. J. Slocum, "The knight's tour - evolutionary vs. depth-first search," Congress on Evolutionary Computation, Vol. 2, 1435-1440 (2004)

[47] D. Beasley, D. R. Bull, and R. R. Martin, "A Sequential Niche Technique for Multimodal Function Optimization," Evolutionary Computation, 1(2), 101-125, (1993)

[48] K. E. Mathias and L. D. Whitley, "Initial performance comparisons for the delta coding algorithm," IEEE World Congress on Computational Intelligence, Proceedings of the First IEEE Conference, Vol.1, 433 – 438 (1994)

[49] D. E. Goldberg, "Genetic Algorithms with Dynamic Niche Sharing for Multimodal Function Optimization, International Conference on Evolutionary Computation," 1-14 (1996)

[50] K. Deb, M. Kaufmann, and D. E. Goldberg, "An Investigation of Niche and Species Formation in Genetic Function Optimization," Proceedings of the Third International Conference on Genetic Algorithms, 42-50 (1989)

[51] H. Sakanashi and Y. Kakazu, "Co-evolving genetic algorithm with filtered evaluation function," IEEE Symposium on Emerging Technologies and Factory Automation, 454–457 (1994)