Optical Transmission Properties of Dielectric Aperture Arrays

YANG, Tao

A Thesis Submitted in Partial Fulfillment

Of the Requirements for the Degree of

Doctor of Philosophy

in

Electronic Engineering

The Chinese University of Hong Kong

April 2010

ABSTRACT OF THESIS ENTITLED:

Optical Transmission Properties of Dielectric Aperture Arrays

Submitted by YANG Tao

For the degree of Doctor of Philosophy in Electronic Engineering

at The Chinese University of Hong Kong in October 2009

Optical detection devices such as optical biosensors and optical spectrometers are widely used in many applications for the functions of measurements, inspections and analysis. Due to the large dimension of prisms and gratings, the traditional optical devices normally occupy a large space with complicated components. Since cheaper and smaller optical devices are always in demand, miniaturization has been kept going for years. Thanks to recent fabrication advances, nanophotonic devices such as semiconductor laser chips have been growing in number and diversity. However, the optical biosensor chips and the optical spectrometer chips are seldom reported in the literature. For the reason of improving system integration, the study of ultra-compact, low-cost, high-performance and easy-alignment optical biosensors and optical spectrometers are imperative. This thesis is an endeavor in these two subjects and will present our research work on studying the optical transmission properties of dielectric aperture arrays and developing new optical biosensors and optical spectrometers.

The first half of the thesis demonstrates that the optical phase shift associated with the surface plasmon (SP) assisted extraordinary optical transmission (EOT) in nano-hole arrays fabricated in a metal film has a strong dependence on the material refractive index value in close proximity to the holes. A novel refractive index

sensor based on detecting the EOT phase shift is proposed by building a model. This device readily provides a 2-D biosensor array platform for non-labeled real-time detection of a variety of organic and biological molecules in a sensor chip format, which leads to a high packing density, minimal analyte volumes, and a large number of parallel channels while facilitating high resolution imaging and supporting a large space–bandwidth product (SBP). Simulation (FDTD Solutions, Lumerical Solutions Inc) results indicate an achievable sensitivity limit of 4.37×10^{-9} refractive index units (RIU) and a dynamic range as large as 0.17 RIU.

Subsequently, optical transmission properties through a self-mixing interferometer array are studied and a novel high-resolution cost-effective optical spectrometer is proposed. The miniature interferometer-based spectrometer is made of polymethyl methacrylate (PMMA) with a CCD as the detector. The detected intensity of each CCD pixels contains the spectral information. Since each frequency component in the incoming beam corresponds to a unique phase difference of the two beam portions of each optical interferometer, the total intensity received by each CCD pixel, which is resulted from the addition of the interference signals from all the frequency components in the beam, should also be unique. Therefore, the spectrum calculation is a problem to solve an ill-posed linear system by using Tikhonov regularization method. Simulation results show that the resolution can reach picometer level. Apart from the choice of path difference between the interfering beams, the spectral resolution also depends on the signal-to-noise ratio and analogue-digital conversion resolution (dynamic range) of

the CCD chip. In addition, the theory of uniform waveguide scattering is explored to expand the possibility of using such mini-interferometers for performing free-space spectral analysis of waveguide devices. At the same time, the method of least squares is used to correct the pixel non-uniformity of the CCD so as to improve the performance of the spectrometer.

The sensor chip and spectrometer chip introduced here are based on the interference of light transmitted through dielectric aperture arrays. Their compact feature renders these devices ideal for miniaturization and integration as the systems in microfluidics architectures and lab-on-chip designs.

摘要

光學檢測設備，如光學生物傳感器和光譜儀被廣泛用於諸多檢測領域。而傳統的設備由於採用了大尺寸的棱鏡和光柵，通常佔用較大空間。鑒於實際應用中通常需要更便宜體積更小的設備，小型化已成爲近年來設備發展的趨勢。得益于現代加工工藝的成熟，納米光子器件如半導體雷射器已經在數量和種類上得到迅猛發展。然而集成化的光學生物感測器和光譜儀卻鮮有報到。爲了提高光學系統的集成度，研究超小型、價格低廉、高性能、光路簡單的光學生物感測器和光譜儀就顯得很有必要。這篇學位論文就是在研究光學介質陣列透射特性的基礎上致力於新型感測器和光譜儀的研究。

在這篇論文的上半部分，我們論證了基於表面等離子體的金屬納米孔超強透射所產生的相位變化與孔周圍的材料折射率變化是緊密聯繫的。我們通過建立模型提出了基於測量增強透射光相位變化的新型折射率感測器。這個器件將用於組成一個二維的生物感測器陣列晶片，實現不加入其他物質並即時地檢測多種有機生物分子。它具有集成度高，體積小，並行檢測等優點，便於高解析度成像以及寬頻檢測。我們採用了 Lumerical Solutions 公司的軟體 FDTD Solutions 進行模擬，模擬結果顯示該器件解析度能夠達到 4.37×10^{-9} RIU。其動態檢測範圍爲 0.17 RIU。

在本論文的第二部分，我們研究了亞微米干涉儀陣列的光學透射特性並提出了一種新型高解析度、低成本的光譜儀。這個基於微型干涉儀的光譜儀是用 PMMA 作爲材料，以 CCD 作爲探測裝置。每一個 CCD 單元的探測強度包含了入射光的光譜資訊。因爲入射光中的每一個頻率對應於每一個干涉儀中兩束光之間特定的相位差。那麼每一個 CCD 單元所接收到的總的強度就是所有頻

率單元他們的干涉信號迭加的結果，這也是唯一確定的。所以復原光譜的問題就轉換成用正則化的方法解一個病態線性方程組的問題。而我們的模擬結果顯示，波長解析度可以達到皮米量級。所以我們實驗中所用 CCD 的靈敏度和信噪比以及模擬數字轉換效率是波長解析度的重要參數。爲了拓展自由空間光譜儀的應用範圍，我們提出了用它來測量光纖或波導光的檢測方法，相應的均勻發光的理論也作了研究。與此同時，我們也提出了用最小二乘法校正 CCD 圖元不均勻性的方法，這樣可以進一步提高該光譜儀的性能。

本文所研究的感測器晶片以及光譜儀晶片都是基於介質孔陣列透射光的光學干涉特性。體積小的特性將使得這兩個器件在微流控體系的高集成度系統中得以廣泛應用。

Acknowledgements

It is my pleasure to thank the many people who made this thesis possible.

It is difficult to overstate my gratitude to my supervisor, Prof. H. P. Ho, who shared with me a lot of his expertise and research insight. Without his help at my difficult time, I would have been lost. I really appreciate his helpful suggestions, important advice, constant encouragement and financial support. His passion to the research work and kindness to other people have deeply moved me, which inspire me how to be a good teacher in the future. He becomes for me the role model of a successful researcher in the field.

I am indebted to my former supervisor, Prof. K. T. Chan, for his valuable instructions on Terahertz project research.

I wish to express my appreciation to Prof. J. B. Xu, Prof. C. T. Shu, Prof. E. Pickwell-MacPherson, Prof. C. L. Pan and Prof. Chinlon Lin for their helpful suggestions on my research work.

I would like to express my thanks to Dr. J. Pond and Mr. C. Kopetski in Lumerical Solutions Inc, Mr. S. Chan and Mr. J. Law in ITSC, Mr. W. K. Chan and Dr. N. Ke in Solid-State Electronics Lab for their technical support.

Special thanks are due to Mr. C. C. Li for taking intense academic interest in the spectrometer study as well as providing numerous ideas and valuable suggestions on the design, which indeed greatly improve the thesis. Without his help, I cannot finish my work so soon.

Special gratitude goes to Mr. Z. W. Wang in East China Institute of Technology, on behalf of devoting his precious time and making valuable suggestions on solving linear systems.

My keen appreciation goes to Prof. H. C. Chui and Mr. C. W. Huang in National Cheng Kung University for nanohole structure fabrication.

Sincere thanks are extended to all my colleagues in Solid State Electronics & Photonics Group, who made many valuable suggestions and gave constructive

advice.

I am grateful to my friends in Department of Electronic Engineering especially people in room 310B for their kindness and affection making my days more enjoyable and worth living, and for never letting me feel that I am away from my hometown.

I wish to express my cordial appreciation to my colleagues in Nanjing University of Posts and Telecommunications for the trust and support that they gave me in order to study in the HK. I am tempted to individually thank Prof. Z. C. Liang. Due to his instruction during my master study, I had a chance to develop some of my knowledge before this study commenced. He has been a cornerstone in my professional development.

I wish to thank my entire extended family for providing a loving environment for me. My special appreciation goes to my aunt L. Z. Yang, who is particularly supportive.

Finally, I would like to express heartfelt gratitude to my mother R. Y. Yang, who always kept me away from family responsibilities and encouraged me to concentrate on my study. Without her support and encouragement, this study would not have been completed. I also dedicate this thesis to my father L. F. Yang. He raised me, supported me, taught me and loved me. I regret that he left me so early and I cannot do anything for him now except to use this thesis in honor of him.

The study was funded under the research studentship provided by The Chinese University of Hong Kong. The funding support from the Hong Kong Research Grants Council under the Competitive Earmarked Research Grant (CERG) projects # 411907 and # 412208 is also much appreciated.

Table of Contents

Acknowledgements vi
Table of Contents viii
List of Tables and Figures xi
List of Abbreviations xiv
Chapter 1: Introduction 1
1.1 Background 1
1.1.1 Light transmission through arrays of nanoholes in a metallic film 1
1.1.2 Light transmission through a phase object array 3
1.2 Aim of this work 4
1.3 Organization of the thesis 5
Chapter 2: Principles of SPR Sensors 8
2.1 Introduction 8
2.2 Basic theory of surface plasmon 9
2.2.1 Definitions of some terms 9
2.2.2 SPPs dispersion relation 11
2.2.3 SP propagation length and decay length 12
2.3 Optical excitation of surface plasmon modes 13
2.3.1 Prism coupling 14
2.3.2 Grating coupling 15
2.4 Interrogation methods for SPR sensors 16
2.4.1 Angular modulation 17
2.4.2 Spectrum modulation 17
2.4.3 Intensity modulation 18
2.4.4 Phase modulation 18
2.4.5 Comparisons of different interrogation method 20
2.4.6 Phase and intensity sensitivities affected by noises [63] 22
2.5 EOT based SPR phase sensors 28
2.5.1 The traditional EOT based sensors 29
2.5.2 The structure of the EOT based SPR sensor using phase shift 30
2.5.3 Operation principle of the nanohole sensor device 32
2.5.4 Optical setup for SPR phase shift measurement 35
2.6 Summary 37
Chapter 3: Simulation Experiments for EOT Based Sensors 39
3.1 Introduction 39

3.2 Simulation theory and results 39
3.2.1 Introduction of FDTD methods 39
3.2.2 Merits of FDTD Solutions 46
3.2.3 Boundary conditions and symmetries 50
3.2.4 Parameter design for the sensors 52
3.2.5 Square hole period simulation 54
3.2.6 Metal layer thickness simulation 57
3.2.7 Square hole width simulation 59
3.2.8 Predicted sensitivity and dynamic range 61
3.3 Other design considerations 63
3.3.1 Curved corners at hole edges 63
3.3.2 Surface roughness of glass substrate 65
3.3.3 Chemical sensing 65
3.3.4 Mesh size consideration 68
3.4 Summary 69
Chapter 4: Spectrometer Designs 71
4.1 Introduction 71
4.2 Dispersive spectrometers 72
4.3 Fourier transform spectrometers 74
4.3.1 Basic theory of Fourier transform spectrometer 75
4.3.2 Spectrum resolution of Fourier transform spectrometer 77
4.3.3 Sampling theory of Fourier transform spectrometer 79
4.3.4 Advantages of Fourier transform instruments 81
4.4 Self-mixing interferometers 81
4.4.1 Principles of the self-mixing interferometer 82
4.4.2 Simulation results of the self-mixing interferometer 84
4.4.3 Correlation induced spectral changes 87
4.5 Self-mixing interferometer spectrometers 89
4.5.1 Principles of self-mixing interferometer spectrometers 89
4.5.2 Reconstruction methods 93
4.5.3 Uniformity of the incoming light 95
4.5.4 Correction method for the nonuniformity of CCD pixel response 98
4.5.5 Fabrication and measurement error 99
4.5.6 Possible applications 101
4.6 Summary 102
Chapter 5: Simulation Experiments for Step Interferometer Based Spectrometers 103
5.1 Introduction 103
5.2 High resolution of self-mixing interferometer 104

5.3 Crosstalk consideration 113
5.4 Reconstructions for nonlinear phase delay 115
5.5 Simulation results of pit structures 117
5.6 Broadband reconstruction 121
5.7 Few interferometer reconstruction 127
5.8 Reconstruction procedure 130
5.9 Impact factors for satisfactory reconstructions 134
5.10 Fabrication error analysis 135
5.11 Summary 137
Chapter 6: Conclusions and Future Work 138
6.1 Conclusions 138
6.2 Future work 139
6.2.1 Experimental test for the sensors and spectrometers 139
6.2.2 Other possible applications of the self-mixing interferometers 142
Publication List 147
References 150
Appendix: Properties of Some Polymers 164

List of Tables and Figures

Table 2-1 Comparisons of four interrogation techniques 20

Fig. 1-1 Zero-order transmission spectrum of an array in silver film 2
Fig. 2-1 Cold plasma and surface plasmon 10
Fig. 2-2 Excitation of surface plasma 11
Fig. 2-3 SP propagation length and decay length 12
Fig. 2-4 Dispersion relation of surface plasmon 13
Fig. 2-5 Prism coupling to SPPs using attenuated total internal reflection in the Kretschmann (left) and Otto (right) configuration 14
Fig. 2-6 The dispersion relation of bulk plasmon, surface plasmon and light wave 15
Fig. 2-7 Phase-matching of light to SPPs using a grating 16
Fig. 2-8 Angular modulation 17
Fig. 2-9 Spectrum modulation 18
Fig. 2-10 Experimental setup for a phase-shift based sensor 19
Fig. 2-11 Phase response as a function of incident angle 20
Fig. 2-12 Intensity noise of the (a) He-Ne laser normalized to shot noise for different discharge currents; (b) Nd laser for the pumping frequency of 1 mW 25
Fig. 2-13 Approximate diagram of intensity and phase noises of various light sources 27
Fig. 2-14 Normalized transmission spectra of normally incident white light 30
Fig. 2-15 Schematic of nanohole array 31
Fig. 2-16 Cross section of proposed phase-sensitive SPR sensor 33
Fig. 2-17 Optical setup for SPR phase shift measurement 36
Fig. 3-1 Comparison of simulation results and experimental results 50
Fig. 3-2 Square hole period versus phase change 54
Fig. 3-3 logarithmic intensity distribution in near field 56
Fig. 3-4 Metal layer thickness versus phase change 58
Fig. 3-5 Square hole width versus phase change (red line for gold and black line for silver) 60
Fig. 3-6 Influence of hole shape on phase change (sensor layer material: gold) 60
Fig. 3-7 Refractive index versus phase change with all parameters optimized 62
Fig. 3-8 Effect of curved corners at hole edges 64
Fig. 3-9 Surface roughness on glass substrate 65
Fig. 3-10 Chemical sensing 67
Fig. 3-11 the relation between mesh size and phase change 68
Fig. 4-1 Dispersive spectrometer 73
Fig. 4-2 Fourier transform spectrometer 75

Fig. 4-3 Diagram of the spectrum resolution 78
Fig. 4-4 Sampling of the interference pattern 79
Fig. 4-5 Setup of classic Young's experiment 83
Fig. 4-6 Modified setup of Young's experiment 83
Fig. 4-7 Cross-sectional view of a self-mixing interferometer 83
Fig. 4-8 Cross-sectional view of a self-mixing interferometer 85
Fig. 4-9 Field intensity distribution of E_y in log scale 86
Fig. 4-10 Phase difference dependence of the interference intensity for the self-mixing interferometer (black line) and a classic ideal interferometer (red line) 86
Fig. 4-11 The normalized spectra in the interference field 87
Fig. 4-12 Schematic of the optical spectrometer 89
Fig. 4-13 Cross section of the optical interferometer 90
Fig. 4-14 Source spectrum used in simulation 91
Fig. 4-15 Nonuniformity of maximum power for a single wavelength illumination with difference phase delay 95
Fig. 4-16 Two lens system for large, uniform, collimated beam 96
Fig. 4-17 The first design for uniform illumination 96
Fig. 4-18 The second design for uniform illumination 97
Fig. 4-19 The third design for uniform illumination 98
Fig. 5-1 Reconstruction in the ultra violet spectra range with different n 106
Fig. 5-2 Reconstruction in the infrared spectra range with different n 109
Fig. 5-3 Distortion factors comparison 112
Fig. 5-4 Optical interference effect with side illumination 113
Fig. 5-5 Reconstruction with a large space interval (10 microns) 115
Fig. 5-6 Step height of each interferometer 116
Fig. 5-7 Reconstruction for nonlinear phase delay 117
Fig. 5-8 Schematic of the pit structure spectrometer 118
Fig. 5-9 Cross section of the pit structure interferometer 118
Fig. 5-10 Reconstruction for the pit structure 119
Fig. 5-11 The torus pit structure interferometer 120
Fig. 5-12 Reconstruction for the torus pit structure 121
Fig. 5-13 Broadband reconstruction 122
Fig. 5-14 Broadband reconstruction with high resolution 124
Fig. 5-15 Broadband reconstruction for pit structures 125
Fig. 5-16 Nonlinear phase delay reconstruction for pit structures 126
Fig. 5-17 Broadband reconstruction with 100 interferometers 128
Fig. 5-18 Broadband reconstruction with large distortion for few interferometers 130
Fig. 6-1 Modeling experiment for spectrometer 141

Fig. 6-2 Other step design for the interferometers 141
Fig. 6-3 Other pit design for the interferometers 142
Fig. 6-4 Bio-chemical sensor designs with the self-mixing interferometer 144
Fig. 6-5 Super-resolution imaging with the self-mixing interferometer 144

List of Abbreviations

2D	two-dimensional
3D	three-dimensional
ASAP	advanced systems analysis program
ASCII	the American standard code for information interchange
BC	boundary condition
BPM	beam-propagation method
BRO	breault research organization
CAMFR	cavity modelling framework
CW	continuous-wave
EBL	electron beam lithography
EOT	extraordinary optical transmission
FDTD	finite-difference time-domain
FEM	finite-element methods
FFT	fast Fourier transform
FIB	focused ion beam
FT	Fourier transform
FTIR	Fourier transform infrared
FTS	Fourier transform spectrometer
GDSII	graphic data system II
ILS	instrumental line shape
IR	infra-red
LOD	limit of detection
MIM	metal/insulator/metal
MPB	MIT photonic-bands
PML	perfectly matched layer
PMMA	polymethyl methacrylate
RIU	refractive index units
S. D.	standard deviation
SBP	space-bandwidth product
SNR	signal to noise ratio
SP	surface plasmon
SPPs	surface plasmon polaritons
SPR	surface plasmon resonance
TDS	time-domain spectroscopy
TFSF	total- field scattered-field
THz	terahertz
UV	ultra-violet

Chapter 1: Introduction

1.1 Background

Existing methods for the detection and measurement of biological and chemical material are often hampered by issues associated with ease of use, small size, high performance and low costs. As a result, there remains a need for inexpensive, compact, real-time sensors and spectrometers in emerging biotech fields of biophotonics and integrated optical detection. However, the commercial devices with high resolution are normally bulky, which are limited by traditional technologies and difficult to break through. In order to satisfy the requirements of industry, we need to explore new technologies. Optical transmission properties of dielectric aperture arrays are studied in this thesis. The theory can be used to design compact and high performance bio-sensors and spectrometers respectively. According to the applications, we consider two situations, i.e. dielectric arrays with metal and without metal.

1.1.1 Light transmission through arrays of nanoholes in a metallic film

When we consider the situation of dielectric arrays in metal film, we should mention the original paper published in 1998, which reported that surface plasmons could enhance the transmission of light through subwavelength holes in a metallic film. That seminal paper reported that, when subwavelength holes are disposed on a

metallic film forming a two-dimensional (2D) array, the transmission of light through this structure is greatly enhanced at some particular wavelengths as shown in Fig. 1[1].

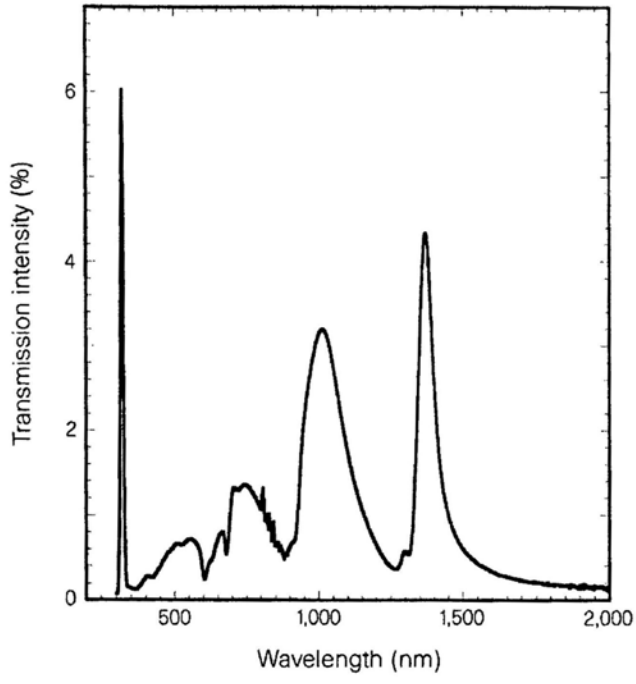


Fig. 1-1 Zero-order transmission spectrum of an array in silver film

The locations of the transmission peaks appearing in the experimental spectra could be approximately found from the dispersion relation of SPs modes running on the metal surface. Then, from the beginning of the history of this subject, it was clear that there was a close connection between extraordinary optical transmission (EOT) and the excitation of SPs. Since 1998, several experimental and theoretical groups around the world have reproduced the main features found in the first set of experiments. The dependence of this phenomenon with the type of metal (noble metals show larger enhancements), type of lattice (square or triangular), shape of the holes (circular, elliptical, square or rectangular) and frequency regime (optical,

THz or microwave) have been thoroughly analyzed [2-13]. Four years after the discovery of EOT in 2D hole arrays, it was also reported [14] that EOT phenomenon also appears in a single aperture (hole or slit) flanked by periodic corrugations in the side of the metal film the light is impinging on. Moreover, it was also found that very strong directional emission (beaming) is possible through single apertures if the corrugation is placed on the exit side [15].

In this thesis, we concentrate on the explanation of the fundamental physics behind the phenomenon of EOT with a model, which can be used to design high performance and compact bio-sensors.

1.1.2 Light transmission through a phase object array

Let's move our concentration to another subject, i.e. light interference. If there is no metal in the route of the light, we do not need to consider the surface plasmon effect. Then, the wave theory is important to analysis of the situation. The interference experiment of Thomas Young [16, 17], performed in 1801, was one of the pillars on which the wave theory of light was erected. Fresnel made use of it in 1818 in his classic memoir in which he placed the wave theory of light on a sound foundation [18]; and Zernicke in 1938 in a basic paper in which he introduced the concept of the degree of coherence of light [19]. Zernkie showed theoretically that the Young interference experiment also reveals more subtle features of light, namely some of its spatial coherence properties, a prediction which was later confirmed by experiment [20]. Zernike's work also provided a new and physically more basic

interpretation of Michelson's classic technique for measuring stellar diameters [21]. It was later found that Young's and Zernike's investigation do not exhaust the amount of information which can be deduced from the analysis of the light produced by superposing two beams. In particular the spectral analysis of the light in the region of superposition was found to provide information about the spectral coherence properties of light incident on the two pinholes [22, 23]. Papers published by Mandel [24] on cross-spectral purity and by Wolf [25] on the effect of source correlations show that the spectrum of light changes when two beams are superposed. On the basis of above theoretical and experimental results, we proposed a novel spectrometer, which consists of a self-mixing optical interferometer array. Light transmission properties through an optical interferometer array are studied in the thesis and the corresponding theory can be used to design the novel spectrometer. The array design can dramatically improve the spectral resolution of the spectrometer.

1.2 Aim of this work

Optical biosensors and optical spectrometers have become important optical detection devices, which are widely used in many applications such as bio-sensing, chemical inspections and material analysis. However, the traditional optical devices normally occupy a large space with complicated components due to the usage of large dimension of prisms and gratings. Since cheaper and smaller optical devices are always in demand, the study of ultra-compact, low-cost, high-performance and

easy-alignment optical biosensors and optical spectrometers are imperative. This thesis is an endeavor in these two subjects and introduces our research work on studying the optical transmission properties of dielectric aperture arrays and developing new optical biosensors and optical spectrometers, which have small size and high performance at the same time.

1.3 Organization of the thesis

The structure of this thesis is arranged as follows:

In Chapter 1, the history information of SPs based EOT and self-mixing interference theory is briefly introduced. The aim of the work is proposed. An overview of the thesis is also included.

In Chapter 2, the basic theory of surface plasmon is described. The principles of all kinds of SPR sensors are introduced. Traditional SPR sensors generally operate in total internal reflection mode using the Kretschmann (reflection) configuration. But this configuration has difficulty in realizing large dynamic range, small probing area and high throughput sensing. Here we report a high sensitivity biosensor based on detecting the phase change of EOT light by using a heterodyne technique. The nanohole SPR configuration is capable of detecting surface binding of organic and biological molecules at a high sensitivity level with large dynamic range.

In Chapter 3, various design parameters of the EOT based sensors have been studied by simulation. Our results indicate that the scheme has a shot-noise limited sensitivity threshold of 4.37×10^{-9} RIU and a dynamic range of 0.17 RIU, which compare favorably with typical SPR sensors, particularly in terms of achieving high resolution and wide dynamic range sensor attributes. The phase change is also quite linear over the entire refractive index detection range. Curved hole edges and substrate roughness due to fabrication limitations have been quantitatively discussed. Application of the proposed device for chemical sensing is also demonstrated.

Chapter 4 presents the structure and principles of an ultra compact and high resolution optical spectrometer and compares the advantages and disadvantages with dispersive spectrometers and Fourier transform spectrometers. The miniature interferometer based spectrometer is a series of self-mixing phase objects on a polymethyl methacrylate (PMMA) film with a CCD as the detector. The spectrum is obtained by solving a system of simultaneous linear equations. The Tikhonov regularization method is used to achieve high resolution. The methods for the uniformity of incoming beam intensity and CCD pixel response are discussed respectively. Compared with conventional spectrometers, the proposed device is low-cost and easy to fabricate for its simple structure.

Chapter 5 describes the performance of the ultra compact and high resolution free space optical spectrometer and demonstrates it by using FDTD simulations. Simulation results show that the spectrometer can realize a very wide measurement

range from UV to IR and a very high resolution on picometer level. Reconstructions for the designs of a large space interval and a parabola phase delay indicate that the problems of crosstalk and fabrication error do not affect the results very much. Designs for extrude and the pit structures show that the proposed device is low-cost and easy to fabricate as well.

Chapter 6 gives a summary on the whole work and launches a brain storm on the future work. Some ongoing research works are also introduced in this part. Chapter 6 is then followed by a list of References, then the Appendices.

Chapter 2: Principles of SPR Sensors

2.1 Introduction

Surface plasmon resonance biosensors have become a central tool for characterizing and quantifying biomolecular interactions in both life sciences and pharmaceutical research in the past two decades [26-28]. The main reason for SPR sensors to become a powerful tool for characterization of biomolecules interaction is their capability of real-time monitoring and label-free sensing with high detection sensitivity and wide dynamic range. Recently, research attention in SPR sensing has shifted to measuring the SPR phase shift [29, 30], as the resonant phase behavior offers the potential of achieving extremely high detection sensitivity. Traditional SPR phase sensors generally operate in total internal reflection mode using the Kretschmann (reflection) configuration [26]. However, this configuration has difficulty in realizing large dynamic range, small probing area and high throughput sensing.

Recently, enhanced transmission through subwavelength hole arrays accompanied by strong field localization has applications in several fields such as quantum information processing [31], microscopic optical components [32], wavelength conversion [33], and nanolithography [34]. Such phenomena may also be exploited for biosensor applications due to their potential for decreasing the interrogation volumes while operating at normal illumination [35-38]. This leads to

desirable attributes including high packing density, minimal analyte volumes, and large number of parallel channels, while facilitating dense integration in a sensor chip. These advantages may enhance the preference of using such devices in a number of applications despite their low spectral resolution as reported in the literature [39]. Here we report a high sensitivity biosensor based on detecting the phase change of EOT light by using a heterodyne technique [40-43]. The sensitivity of the nanohole SPR configuration is capable of detecting surface binding of organic and biological molecules at a high sensitivity level with large dynamic range. This chapter aims to introduce to the principles of the traditional SPR sensors and the novel EOT based SPR phase sensor.

2.2 Basic theory of surface plasmon

2.2.1 Definitions of some terms

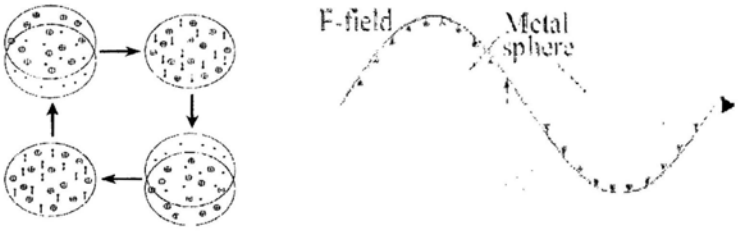
One may confuse about the difference between plasma and plasmon. Here the definitions of several corresponding terms are introduced in order to clarify their difference [44-51].

- Plasma (cold plasma) is periodic oscillation of charge density in conducting media such as metals or semiconductors.
- Plasmon is the particle resulting from the quantization of these oscillations. When plasmons couple with a photon, the resulting particle is called a polariton.
- Surface Plasmon (SP) or Surface Plasmon Polaritons (SPPs) are charge density waves on the surface of materials with free electrons propagating along the interface

of this conductor and a dielectric medium.

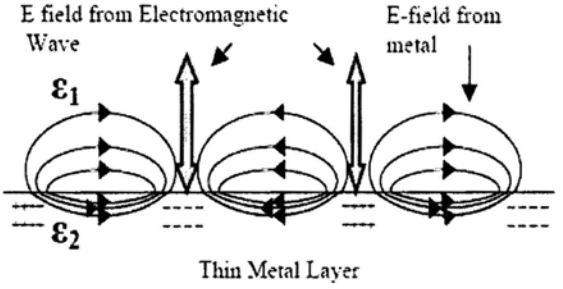
- Plasmonics is a new branch of photonics using SPPs. One way to produce SPPs is to use the electric field of TM-polarized waves to excite the free electron oscillation. So it is possible to transform photons into plasmons, use their special properties to manipulate the original light wave properties, and reconvert the plasmons into photons.

Above definitions indicate that plasma is just oscillation while SPPs are resonant oscillations. Fig. 2-1 (a) and (b) [46] show their oscillation processes respectively. The resonant excitation of surface plasmon is shown is Fig. 2-1 (c). In the figure, the coupled optical wave (TM mode) has an oscillating electrical field that induces oscillating electron charge on the metal surface. The charge densities produce an electrical field as to oppose the electrical field in the optical wave.



(a)

(b)



(c)

Fig. 2-1 Cold plasma and surface plasmon

2.2.2 SPPs dispersion relation

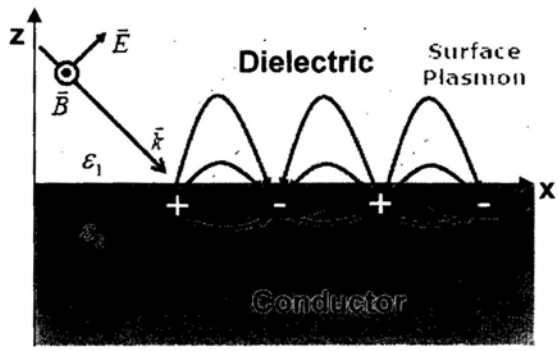


Fig. 2-2 Excitation of surface plasma

Assuming the produced electrical field of surface plasmon is expressed as [52]:

$$E = E_0^{\pm} \exp[+i(k_x x \pm k_z z - \omega t)] \qquad (3.1)$$

where + for z > 0, – for z < 0. Solving Maxwell's equations under the appropriate boundary conditions yields the SPPs dispersion relation:

$$k_x = \frac{\omega}{c}\sqrt{\frac{\tilde{\varepsilon}_1\tilde{\varepsilon}_2}{\tilde{\varepsilon}_1+\tilde{\varepsilon}_2}} = A + Bi \qquad (3.2)$$

$$k_{z1} = \frac{\omega}{c}\sqrt{\frac{\tilde{\varepsilon}_1^{\,2}}{\tilde{\varepsilon}_1+\tilde{\varepsilon}_2}} = C + Di \qquad (3.3)$$

$$k_{z2} = \frac{\omega}{c}\sqrt{\frac{\tilde{\varepsilon}_2^{\,2}}{\tilde{\varepsilon}_1+\tilde{\varepsilon}_2}} = E + Fi \qquad (3.4)$$

where k_x, k_{z1}, and k_{z2} are wavenumber in x direction, $+z$ direction and $-z$ direction respectively. c is vacuum velocity of light. ω is the angular frequency of SPPs. $\tilde{\varepsilon}_1 = \varepsilon_1 + i\varepsilon_1'$ and $\tilde{\varepsilon}_2 = \varepsilon_2 + i\varepsilon_2'$ ($\varepsilon_1 > 0$, $\varepsilon_2 < 0$) are the permittivity of the dielectric and conductor medium respectively. A, C and E are phase constants; B, D and F are decay constants. The phase of the surface plasmon can be calculated from the phase constants by $A \times l$, where l is the distance that the surface plasmon passing through. The propagation length as well as the decay lengths can be obtained from

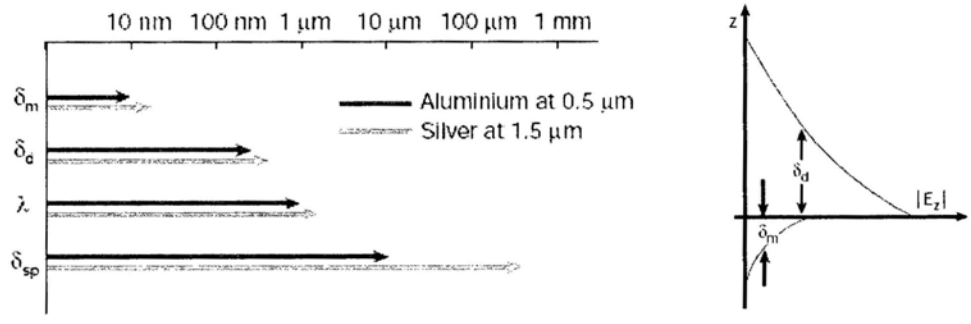
the decay constants.

2.2.3 SP propagation length and decay length

The propagation length of surface plasmon is defined as δ_{sp} after which the intensity of the surface plasmon decreases to 1/e. The propagation length is given by:

$$\delta_{sp} = \frac{1}{|2B|} \tag{3.5}$$

where B is the decay coefficient, which can be calculated with the aid of Eq. (2.2). For a relatively absorbing metal such as aluminium the propagation length 2 μm at a wavelength of 500 nm. For a low loss metal, for example, silver, it is increased to 1 mm at a wavelength of 1500 nm as shown in Fig. 2-3 [44]. The propagation length sets the upper size limit for any photonic circuit based on SPs.



(a) (b)

Fig. 2-3 SP propagation length and decay length

As shown in Fig. 2-3, the decay lengths of surface plasmon are defined as δ_d and δ_m after which the electrical fields of the surface plasmon decrease to 1/e in the dielectric material and the metal medium respectively. The decay lengths are given by:

$$\delta_d = \frac{1}{|D|} \tag{3.6}$$

$$\delta_m = \frac{1}{|F|} \tag{3.7}$$

where D, F are the decay coefficients, which can be calculated with the aid of Eq. (2.3) and Eq. (2.4). The decay length in the dielectric material, δ_d, is typically of the order of half the wavelength of light involved and dictates the maximum height of any individual features, and thus components, that might be used to control SPs. The decay length in the metal, δ_m, determines the minimum feature size that can be used, this is between one and two orders of magnitude smaller than the wavelength involved, thus highlighting the need for good control of fabrication at the nanometre scale. The combinations chosen give an indication of range from poor (Al at 0.5 mm) to good (Ag at 1.5 mm) SP performance.

2.3 Optical excitation of surface plasmon modes

Before moving on to discuss the SPR sensors we need to study how best to couple optical radiation to the SPR. The use of photons to excite SPs meets the difficulty that the dispersion relation curve lies right of the light line as shown in Fig. 2-4.

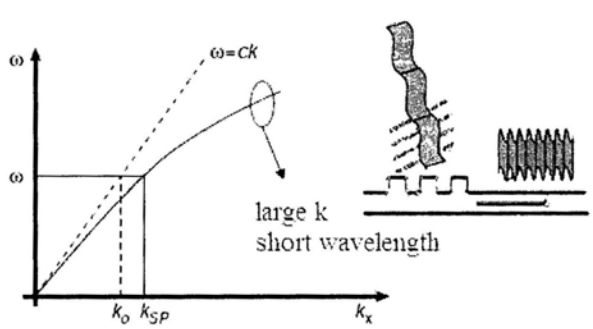


Fig. 2-4 Dispersion relation of surface plasmon

At a given photon energy, the wave vector has to be increased by a Δk value in order to "transform" the photons in to SPs. There are mainly two methods:

2.3.1 Prism coupling

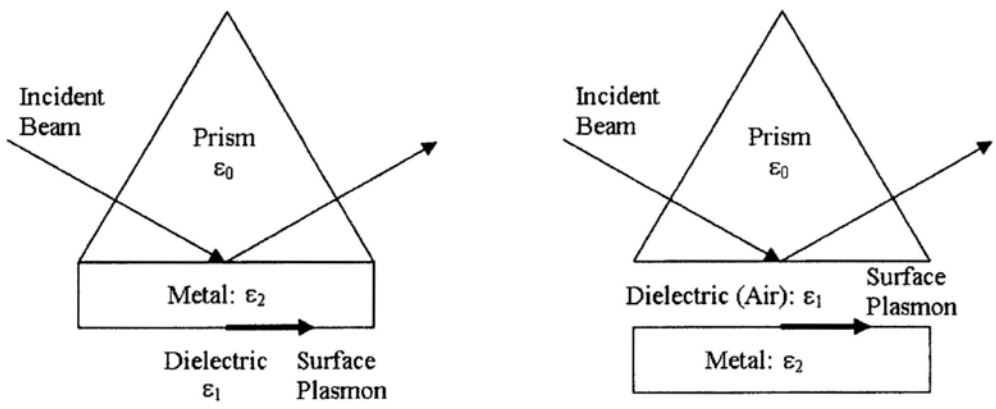


Fig. 2-5 Prism coupling to SPPs using attenuated total internal reflection in the Kretschmann (left) and Otto (right) configuration

There are two different methods of using a prism to couple to a SPP. Fig. 2-5 (a) and (b) [45] illustrate the geometries involved. The method must be used in conjunction with certain conditions to generate the SPPs. The first of these conditions is that the incident light must be p-polarized, that means the light must be polarized with the electrical vector parallel to the plane of incidence (the plane perpendicular to the surface which contains both the incident and reflected wave vectors). This is necessary as *p*-polarization (TM mode) will create time-dependent polarization charge at the interface. *S*-polarization (TE mode) would not create charge at the interface. The Otto method uses the air gap for an evanescent wave to tunnel across, coupling to the surface of the metal exciting the plasmons on the air (dielectric)/metal interface. Kretschmann and Raether realized that a much simpler

method was to use the metal itself as the evanescent tunnel barrier. Therefore the metal is evaporated onto the surface of the prism at an optimum thickness (approximately 50nm) for resonance to occur. By varying the angle of incidence of the p-polarized light at the dielectric/metal interface the momentum in the *X*-direction can be varied. This allows us to experimentally tune through the resonance. The intersection of dispersion relation of modified light line and surface plasmon in Figure 2-6 [50] indicates the excitation of surface plasmon, the resonance condition is described by equation $k_x^{'} = k_{SP}$, or

$$\frac{\omega}{c}\sqrt{\varepsilon_0}\sin\theta = \frac{\omega}{c}\sqrt{\frac{\varepsilon_1\varepsilon_2}{\varepsilon_1+\varepsilon_2}} \qquad (3.8)$$

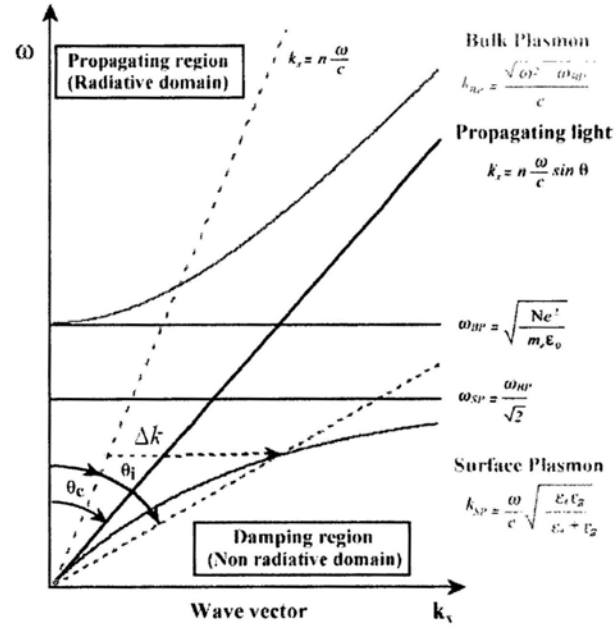


Fig. 2-6 The dispersion relation of bulk plasmon, surface plasmon and light wave

2.3.2 Grating coupling

The mismatch in wave vector between the component $k_x = k \sin \theta$ of impinging

photons and k_{SP} can also be overcome by patterning the metal surface with a shallow grating of grooves or holes with lattice constant a. For the simple one-dimensional grating of grooves depicted in Fig. 2-7 [51], phase-matching takes place whenever the condition

$$k_x \pm m \cdot \frac{2\pi}{a} = k_{SP} \qquad (3.9)$$

is fulfilled, where m is an integer.

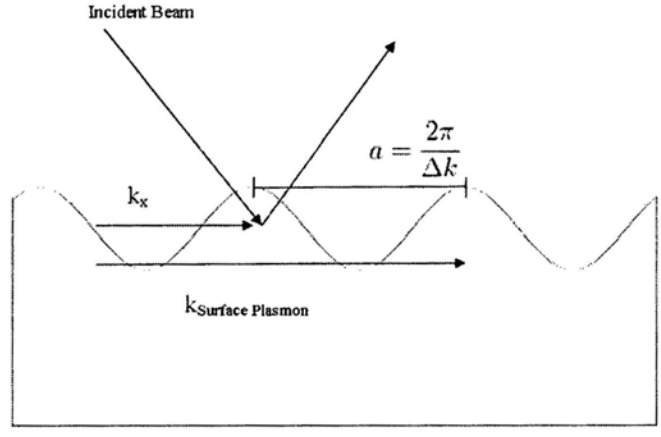


Fig. 2-7 Phase-matching of light to SPPs using a grating

If incident beams normally illuminate the grating, although there is no k_x component in the original beams, the scattering lights scattered on the edge of the grating can also be served to generate the surface plasmon.

2.4 Interrogation methods for SPR sensors

Surface plasmon resonance (SPR) was used for the first time for biosensing in 1983 and for imaging applications in 1987 [53]. With imaging SPR, also referred to as SPR microscopy, new applications are possible, e.g. label-free real-time multispot biochemical analysis, which can increase sample throughput tremendously. The angular [54, 55], spectrum [56] and intensity [57, 58] modulation interrogations are

widely used in the early stage of SPR sensors. Many companies such as BIAcore and Texas Instrument have made these sensors commercialized. However, the sensitivities of these interrogation methods are not very high. In order to realize high sensitivity detection, some designs based on phase modulations [53, 59, 62] have also been proposed.

2.4.1 Angular modulation

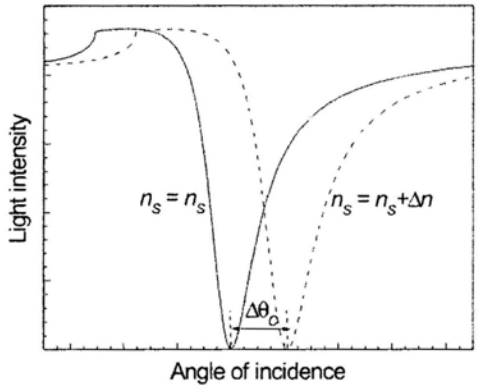


Fig. 2-8 Angular modulation

In SPR sensors with angular modulation a monochromatic light wave excites a surface plasmon. The strength of coupling between the incident wave and the surface plasmon is observed at multiple angles of incidence of the light wave and the angle of incidence yielding the strongest coupling is measured and used as a sensor output (Fig. 2-8). The sensor output can be calibrated to refractive index.

2.4.2 Spectrum modulation

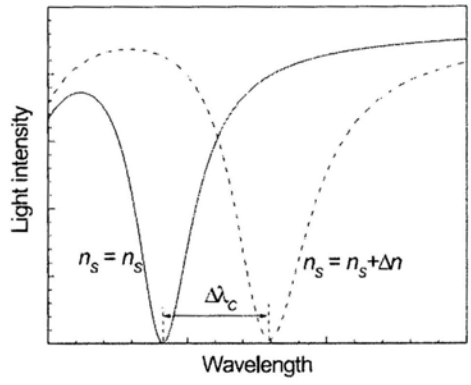


Fig. 2-9 Spectrum modulation

In SPR sensors with spectrum modulation, a surface plasmon is excited by a collimated light wave containing multiple wavelengths. The angle at which the light wave is incident onto the metal film is kept constant. The strength of coupling between the incident wave and the surface plasmon is observed at multiple wavelengths and the wavelength yielding the strongest coupling is measured and used as a sensor output (Fig. 2-9).

2.4.3 Intensity modulation

SPR sensors with intensity modulation are based on measuring the strength of the coupling between the incident light wave and a surface plasmon at a single angle of incidence and wavelength and the intensity of light wave serves as a sensor output. The setup for intensity modulation is simple, but it suffers the disadvantage of low SNR.

2.4.4 Phase modulation

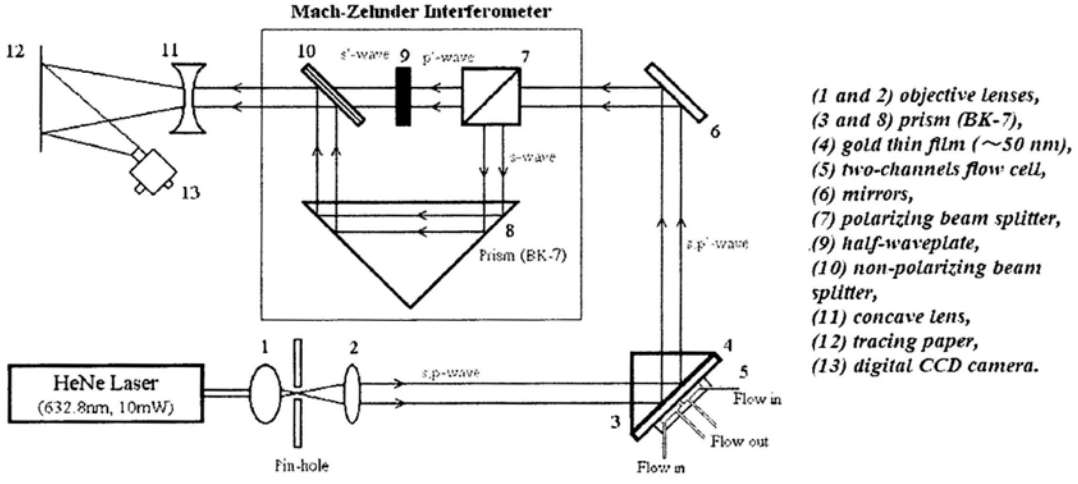


Fig. 2-10 Experimental setup for a phase-shift based sensor

In SPR sensors with phase modulation, the shift in phase of the light wave interacting with the SP is measured at a single angle of incidence and wavelength of the light wave and used as a sensor output. Because only *p*-polarized wave can excite SPs, the amplitude and the phase of the s-polarized wave do not change during resonance, so s-polarized light can be used as a reference beam in the design as shown in Fig. 2-10 [53].

The conditions for surface plasmon excitation at the interface between the metal and dielectric layers in a glass-metal-dielectric stack are given by Eq. (2.8). From Fresnel's equation, the reflection coefficients of the *p*- and *s*-polarized light (r_p and r_s) can be expressed as [60, 61]:

$$r_p = |r_p| e^{i\varphi_p} \text{ and } r_s = |r_s| e^{i\varphi_s} \qquad (2.10)$$

Since the phase angle exhibits a large change across the resonant peak, a small variation of ε_{sample} will lead to a sizable change of r_p of the *p*-polarization, which is accompanied by a phase shift $\Delta\varphi$ ($\varphi_p - \varphi_s$) in the reflected light as shown in Fig. 2-11 [50].

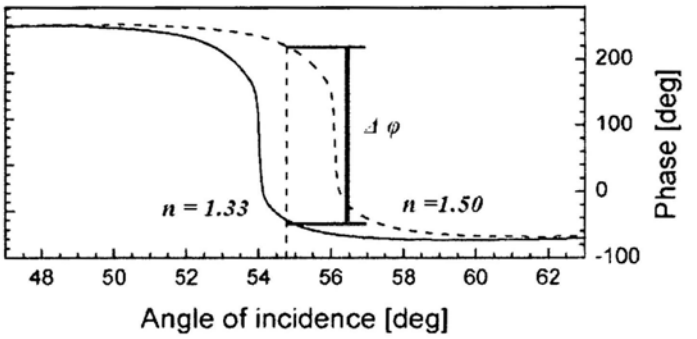


Fig. 2-11 Phase response as a function of incident angle

2.4.5 Comparisons of different interrogation method

The methods to interrogate these variations of reflected light characters with Kretschmann configuration lead to four kinds of different interrogation techniques, which are widely employed for SPR biosensors and have their own characteristics:

Intensity interrogation: monitor reflected intensity variation at constant incident angle and interrogation wavelength;

Angular interrogation: monitor the incident angle variation at a constant interrogation wavelength;

Spectrum interrogation: monitor reflection spectrum and its variation of resonance wavelength at a constant incident angle;

Phase interrogation: monitor the phase variation of reflected light at constant incident angle and interrogation wavelength.

Table 2-1 Comparisons of four interrogation techniques

Configuration	*Kretschmann*			
Interrogation method	Intensity	Angular	Spectrum	Phase

Classical S.D.	0.2% of Reflection	0.0001 deg.	0.02nm	0.01deg.
Sensitivity (RIU)	$1\sim5\times10^{-5}$	$10^{-6}\sim10^{-7}$	$10^{-5}\sim10^{-6}$	$10^{-8}\sim10^{-9}$
Dynamic range (RIU)	2×10^{-1}	5×10^{-2}	1×10^{-1}	$10^{-3}\sim10^{-4}$

Table 2.1 shows that four methods have their own advantages and disadvantages respectively as sensing techniques, which are basically reflected on their performance of sensitivity and dynamic range [50]. Sensitivity in the table is a concept similar with limit of detection (LOD) and defined as $S(\Delta n)=\frac{\Delta n}{\Delta x}.(S.D.)$, where $\Delta\chi$ is the corresponding response, i.e. R, θ, λ and δ induced by refractive index variation Δn, and standard deviation (S. D.) is experimental signal fluctuation at static environment and stands for the system resolution to intensity, phase, wavelength or angle. Dynamic range is defined as a refractive index range which is applicable and sensible by sensing system with a reasonable sensitivity.

In terms of sensitivity, which depends on both the methods itself and also the resolution adopted by system, the phase interrogation normally can offer one or two order higher sensitivity than other three techniques due to the following reasons [62, 63]: (i) Maximal phase changes occur in the very dip of the SPR curve where the vector of probing electric field is maximal, whereas maximal amplitude changes are observed on the resonance slopes: this provides a one order of magnitude larger sensitivity of phase to refractive index variations; (ii) Under a proper design of a detection scheme, phase noises can be orders of magnitude lower compared to amplitude ones, which results in a much better signal-to-noise ratio; (iii) Phase

offers much better possibilities for signal averaging and filtering, as well as for image treatment. We will elucidate the physical sense of intensity and phase sensitivities of SPR techniques and compare them based on the analysis of noises in the next section.

2.4.6 Phase and intensity sensitivities affected by noises [63]

We start our analysis by noting that any SPR technique requires three important ingredients: a source of light, SPR detection cell and a photodetector. Each of these elements contributes some noise into the system which ultimately defines the sensitivity and LOD of a SPR scheme. For the purpose of comparing phase and intensity modes of SPR measurements one needs to compare phase and intensity noises of these three elements. Noises of the detection cell are related to environmental factors, which can cause variations of the refractive index in the controlled liquid volume inside the cell. Here, refractive index drifts related to temperature variations are considered as the most important. The change of temperature of water by 1 Deg. C° is accompanied by a change of refractive index by 10^{-4} RIU. Therefore, measurements of refractive index changes at the level of 10^{-8} RIU require the stability of temperature inside the cell of the order of 10^{-4} Deg. C°. Commercial active thermostats provide the stabilization of the level of 10^{-3} Deg. Celsius (see, e.g.,[64]), while a combination of passive thermal stabilization with time filtering of thermal drifts can lead to the same level. Furthermore, the application of differential reference schemes provides even better level of thermal

stability [65]. In simplest approach, the required level of thermo stabilization (10^{-4} Deg. C°) can be achieved through a combination of thermo stating and reference channel strategies. Since the detection cell noises are related to the measured parameter (they are "signal" noises), they are identical for intensity and phase measurements and require tools to stabilize environmental parameters affecting refractive index inside the cell.

Drifts of characteristics of light source and photodetector present instrumental noises. There most common sources of light for SPR measurements are lasers. Lasers usually have excellent phase noise characteristics. The phase noise of a common laser can be evaluated as $\Delta\varphi/\varphi \approx \lambda/l_c \approx \Delta f_l/f_l$, where λ is the laser wavelength and l_c is the coherence length, Δf_l is the laser line width and f_l is the laser frequency [66]. The coherence length of typical laser is well above 5 m which yields the magnitude of relative phase noise for a laser working at the wavelength of 500nm at the level of about $\Delta\varphi/\varphi \approx 10^{-6}$. The single frequency solid state lasers can easily have even larger coherence length at the level of 10 km which makes phase noise of these lasers negligible [66]. At the same time, relative intensity noise of common lasers is significantly higher and is usually at the level of $\Delta I/I \approx 10^{-2}$. There are several important factors which contribute to higher intensity noise of laser sources: instability in pumping, spot burning due to spatial distribution of laser modes, photon statistics during photon emission and absorption, beam pointing fluctuations, etc. All these factors result in considerable drift of laser intensity at low frequencies ($1/f$ noise), a peak of the relaxation oscillation noise, quantum noise,

etc. Fig. 2-12 [67, 68] shows the frequency dependence of the intensity noise of the most common gas and solid state lasers normalised on shot noise $S(f)= 2Phf$, where P is the laser power. Fig. 2-12 clearly demonstrates that in a large frequency range from 0 to 10^6 Hz the intensity noise of lasers is much larger than the shot noise and is governed by the factors others than the photon statistics. It is easy to check that a similar situation takes place for other types of lasers. In fact, it means that the fundamentally conditioned shot noise can dominate only in an idealized situation of the absence of "technological" noises arising during the lasing process, which never take place in practice. We can write the laser intensity noise as $\sigma^2_{noise} = n_m^2 I^2 + n_s I + n_b^2$, where I is average intensity, n_m, n_s, n_b is the constant describing modulation, shot and background noise, respectively. The modulation noise (drift of laser intensity) is proportional to laser intensity I with coefficient $n_m \approx \frac{\chi}{\sqrt{B}}$ where χ is the numerical coefficient $\chi \approx 10^{-2}$ and B is the total bandwidth and the shot noise constant is $n_s \approx 2hf/A$, where A is the size of the laser spot. From here we conclude that the modulation noise overcomes shot noise for a laser without an intensity stabilization scheme at relatively low powers of above $P > \frac{2hfB}{\chi^2} \approx 1\mu W$ (at a generous bandwidth of 100MHz). Therefore the intensity modulation noise of the common lasers is the most important type of noise which limits operation of the SPR measurement techniques. (One needs to use different smart types of laser intensity stabilization in order to be in the limit of shot noise!).

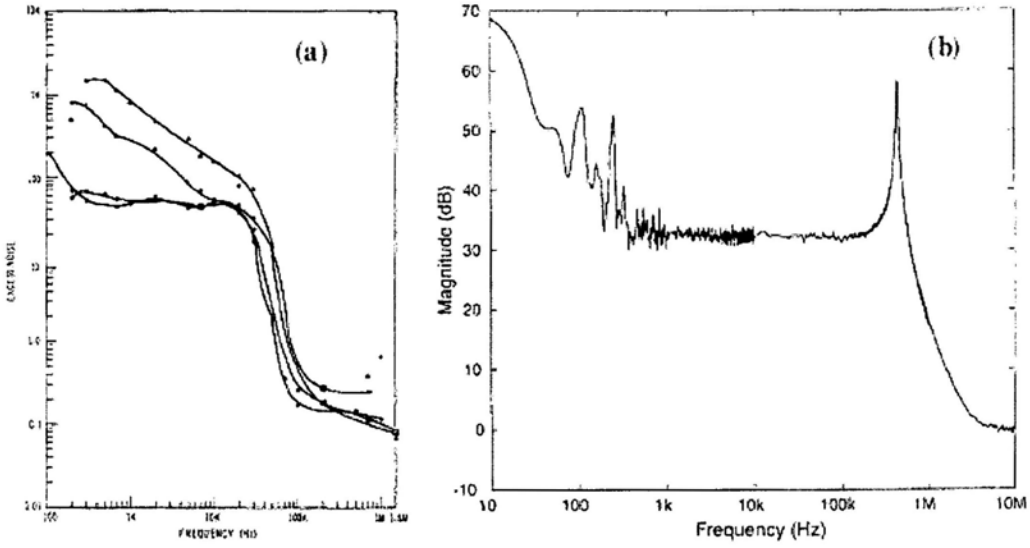


Fig. 2-12 Intensity noise of the (a) He-Ne laser normalized to shot noise for different discharge currents; (b) Nd laser for the pumping frequency of 1 mW

Noises related to the detector can be divided into three categories: shot noise, connected with the photon statistics, dark current noise, produced in a photodetector even in the absence of photons, and amplifier noise, e.g. Johnson-Nyquist thermal noise [69]. The correlation between these noises depends on the light intensity, bandwidth of measurements, temperature of the photodiode, electrical parameters of photodiode and amplifier, etc. It is worth stressing, however, that noises of a photodetector working at optimal parameters are usually quite small compared to noise induced by laser intensity fluctuations and noise of the detection cell. For example, using the same arguments as above, we find that shot noise of a photodetector become essential (comparable with laser power fluctuations) only when light power at the photodiode is very small, $P_{ph} < 1\text{nW}$ for a typical amplifier bandwidth of 100 kHz [70], which is well below the light power at which most of

SPR measurements take place. The dark current noise and thermal noise depend on a particular scheme of the detection and are also quite small. They can be decreased even further using different cooling methods and low-noise amplifiers [70]. Notice that the excitation of plasmons leads to the decrease of laser intensity in the plasmon-affected *p*-polarized beam. In theory, this decrease can be down to zero, but in practice the minimum of intensity is limited by parameters of the SPR-supporting film (roughness, uniformity, thickness optimization) and cannot be lower than 1-4%. Such a drop of intensity may require an increase of laser power to still be in the range of maximal photodetector sensitivity (P_m). Here, the increase of laser power should not exceed the critical value P_{cr}, which provides refractive index drifts because of heating effects (detection cell noises). Normally, there is a wide gap between P_m and P_{cr}, which makes possible efficient phase-sensitive control of events in the detection cell. Thus, if detection cell noises are minimized to a sufficient level through thermo-stabilization of the system, drifts related to laser source normally present main instrumental noise limiting the LOD of measurements. In this case, since amplitude noises of laser sources are orders of magnitude more pronounced compared to the phase ones (see Fig. 2-13 [63]), phase characteristics can potentially provide much lower detection limit. Therefore, in the design of phase-sensitive schemes it is important to avoid or subtract the above-stated amplitude noises related to the source. In interferometric schemes, it can be done e.g., through the application of fringe mode when phase information is extracted from the position of interference fringes, while any variation of light intensity can

only affect the interference pattern contrast and not this position [71,72]. In contrast, in SPR polarimetry amplitude noises can be subtracted or eliminated through the application of differential or filtering schemes.

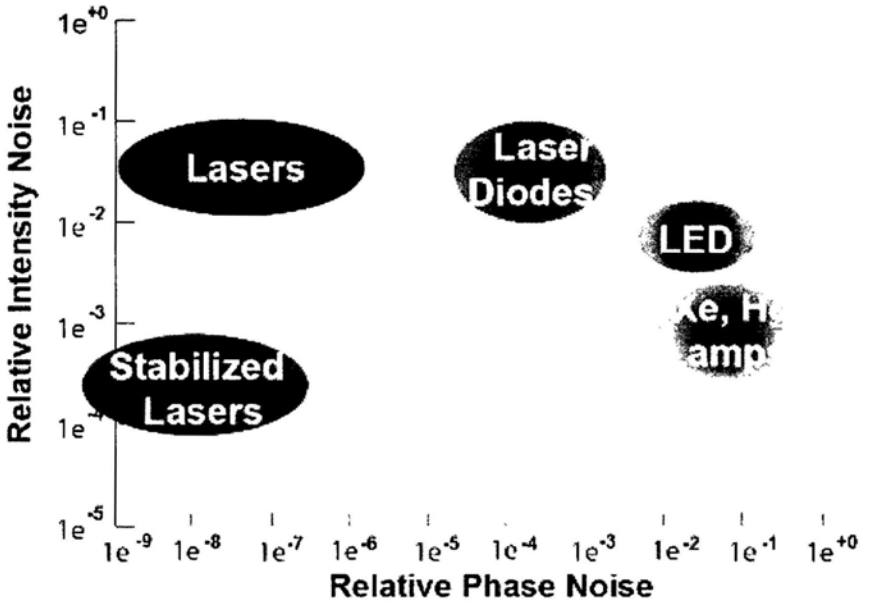


Fig. 2-13 Approximate diagram of intensity and phase noises of various light sources

We have to admit that phase measurements must be performed by measuring light intensities in some way, and therefore both modes of operation are limited in similar ways by photon statistics, which has been demonstrated in Ran and lipson's paper [62]. It is also important to note that the discussion on detection limits of phase and intensity measurements under SPR continued in many recent works, which appeared after Ran and lipson's paper, illustrating a persistent (and increasing) interest to this subject from photonics, analytical chemistry and biological communities. We believe that we need to clearly distinguish three different types of works: (i) works, in which the LOD is determined (and limited) by the level of instrumental and environmental (temperature and inertial drifts etc.) noises in a

single intensity or phase measurement; (ii) works reporting the LOD after the application of additional tools such as differential schemes, averaging, mathematical treatment of noises etc., which can significantly lower the LOD value in both phase- and intensity-sensitive modalities; (iii) works, declaring unrealistic LOD values, which can only be explained by artificial facts taking into account the anticipated level of environmental noises. To avoid ambiguities in interpretation, in the above analysis we compare LODs in phase and intensity measurements under the same level of instrumental noises without any additional signal treatment. Therefore, we can conclude that phase can provide at least two orders of magnitude lower LOD due to: (i) a much stronger response to refractive index variations as a result of probing of medium properties in the very dip of SPR dip where electric field is maximum; (ii) possibilities for a much lower level of phase noises compared to intensity ones under a proper design of phase-sensitive schemes; (iii) possibilities for additional filtering/averaging tools taking advantage of essentially relative nature of phase measurement. Providing a better sensitivity, phase-sensitive methodologies are expected to improve current SPR-based bio- and chemical sensing technology.

2.5 EOT based SPR phase sensors

Although the sensitivity with phase technique is quite high in Kretschmann configuration, the dynamic range is rather narrow. So we urgently need to design a new SPR sensor, which can provide a reasonably dynamic range and sensitivity for

macro molecular related applications [73]. Another major drawback with Kretschmann methods is the difficulty in incorporating the sensor elements in high NA aperture imaging systems to increase spatial resolution and the corresponding number of resolvable spots. Operating on a prism or with a first-or higher-order diffraction mode places severe constraints on the depth of focus in the imaging system needed for large arrays of assays. Massive parallelism, and hence high throughput, is of primary importance in many potential SPR sensor applications, but they are severely limited by most of the Kretschmann configurations. In order to overcome these disadvantages, EOT based sensors are proposed.

2.5.1 The traditional EOT based sensors

Several authors [73-75] have suggested and demonstrated the use of subwavelength hole arrays for sensing applications by intensity and spectrum methods. This technique is sensitive to surface binding events because it is based upon the resonant surface plasmon enhanced transmission through the array of nanoholes. For example, Arrays of nanoholes in a gold film were used to monitor the binding of organic and biological molecules to the metallic surface as shown in Fig. 2-14 [73]. The sensitivity was found to be 400 nm per refractive index unit, which is comparable to other SPR devices. The array of nanoholes is well suited for dense integration in a sensor chip. Furthermore, the optical geometry is collinear, which simplifies the alignment with respect to the traditional Kretschmann (reflection) arrangement for SPR sensing. These advantages may make such devices preferable

in a number of applications despite the fact that the ultimate spectral resolution is lower than the prism-based equivalent because of SPR broadening due to both radiative and material damping.

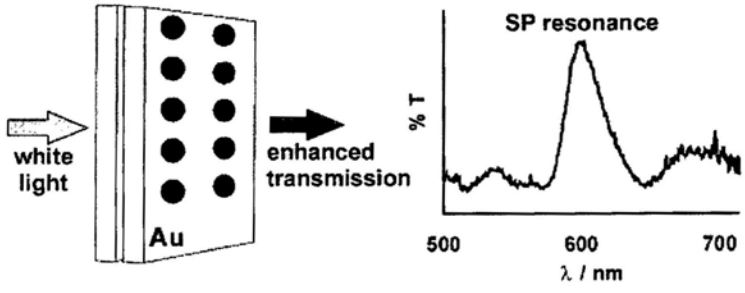


Fig. 2-14 Normalized transmission spectra of normally incident white light through an array of sub-wavelength (200-nm diameter) holes on a 100-nm-thick gold substrate deposited on a glass slide

2.5.2 The structure of the EOT based SPR sensor using phase shift

In order to further improve the sensitivity, we propose a compact and wide dynamic range biosensor based on detecting the phase change of EOT light from a hole array. A periodic square array of sub-wavelength holes on a metal film is utilized as the SPR phase sensor. Nanostructured systems of noble metals, such as silver and gold, are of great interest since their localized surface plasmons resonate at the optical wavelength of electromagnetic radiation. Compared with the typical Kretschmann configuration, the hole array configuration operates in transmission geometry allowing for a simpler collinear optical arrangement, thus providing smaller probing area and high throughput sensing. Moreover, heterodyne technique is used for phase detection so as to realize even high sensitivity. The unique advantages of the

sensors may enhance the preference of using such devices in a number of applications especially integration with lab-on-chip platforms

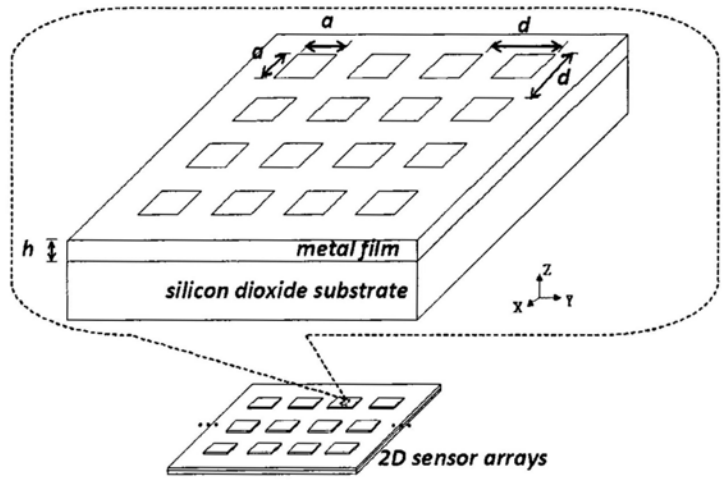


Fig. 2-15 Schematic of nanohole array

(a = hole width, d = hole period, h = hole depth)

The proposed sensor array device is shown in Fig. 2-15. A section of the sensor is also highlighted in the enlarged figure. The device contains a top metal layer, typically made of gold or silver, in which the nanohole arrays are located, and a glass (silicon dioxide) substrate. The array periodicity is *d* and the width of the square holes is *a*. The thickness of the metal is *h*. These sub-wavelength holes can be fabricated by focused ion beam (FIB), electron beam lithography (EBL), or UV interference techniques. The metal film is hydrophobic, thus requiring the use of a linker layer to enable direct immobilization of target molecules onto the surface. More importantly, since immobilization of target molecules occurs only at the sensor surface region where the optical field has been enhanced due to the presence of surface plasmons, one can readily appreciate that only a small number of

immobilized target molecules will trigger a sizeable phase change response. During real-time sensing operation, the nanohole device is immersed in an analyte solution. As the population of immobilized molecules on the surface increases, the effective refractive index of the thin surface layer increases accordingly. The change of refractive index will then result in a phase shift in the transmitted beam, thus leading to the possibility of quantitative detection of the amount of immobilized molecules in the far field without the use of any florescence tag.

2.5.3 Operation principle of the nanohole sensor device

As schematically shown in Fig. 2-16, the extraordinary optical transmission effect may be separated into two contributions. The first one corresponds to the direct transmission of incoming field through the holes, i.e. the Bethe-type diffraction regime and the transmission coefficient presented by T_{Bethe}. This component is wavelength dependent and proportional to an identity matrix. The second contribution, with the transmission coefficient described by T_{Plasmon}, corresponds to the resonant part of the transmission matrix and is related to the plamonic effect. In the present case, T_{Plasmon} is the main item that provides the phase change in our device. This resonant transmission process may be described by a four-step process [76-81]. (i) The incident plane wave is converted into a surface wave at a given point scatterer. Normally the scatterer is the edge of the nanohole array. (ii) The surface wave propagates on the surface of the hole array and builds up several constructive interference modes according to the propagation distance. (iii) The

surface wave is coupled into one of the holes and goes back and forth several times inside, thus resulting in some constructive interference effect. (iv) The surface waves within the hole array also produce constructive interference between them, and some energy will re-emit from the system as a plane wave. The four-step process indicates that when the effective refractive index of the dielectric on the surface of the metal experiences some change, the resonance parameters will also shift accordingly, thus leading to a sharp change in the phase of the radiation that drives the resonance. Therefore maximum change of the phase should occur near the resonant region, while the resonance parameters are dependent on the choice of hole periodicity, shape and size of the holes and thickness of the metal.

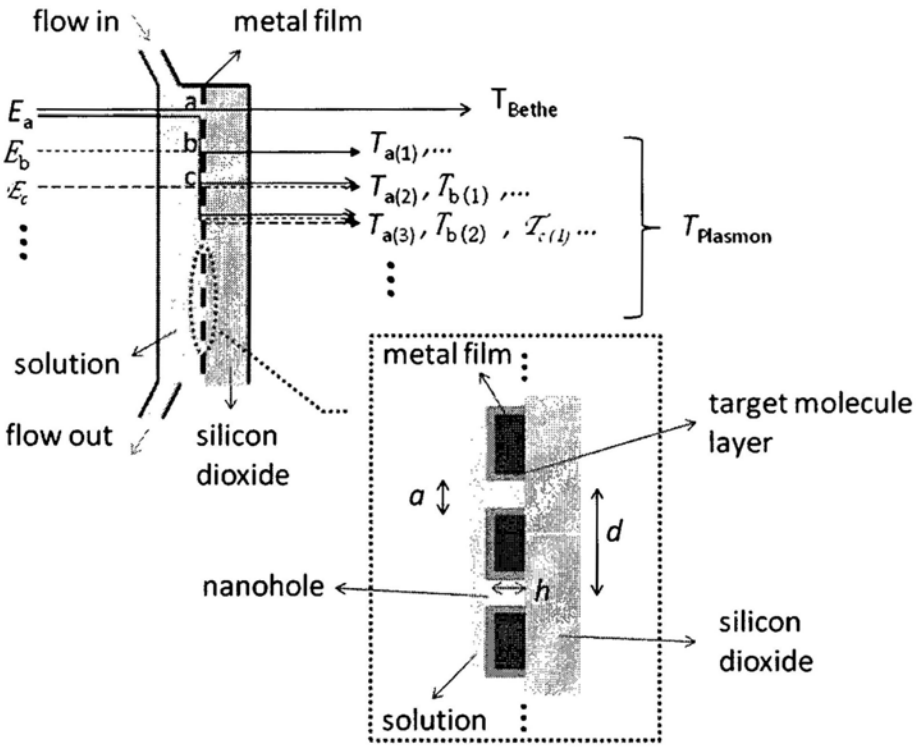


Fig. 2-16 Cross section of proposed phase-sensitive SPR sensor

We also notice that this configuration is a multi-resonance system. Even the resonance in the steps (ii) and (iii) is somewhat broken due to the change of effective refractive index of the dielectric, the constructive interference condition in step (iv) is always satisfied. The detailed explanation is as follows. Assuming there is an incident light E_a scattered at hole *a*, after passing through a distance of one hole period (*d*) and one hole depth (*h*), the generated surface plasmon at last re-emit as $T_{a(1)}$ from the adjacent hole b. The same process occurs for E_b , which is scattered at hole *b* and re-emit as $T_{b(1)}$ from the adjacent hole *c* or *a*. The same process is for $T_{c(1)}$, $T_{d(1)}$ and so on. Because $T_{a(1)}$, $T_{b(1)}$, $T_{c(1)}$,$\cdots$ have the same phase, they form constructive interference. The surface plasmon generated at hole a can also re-emit from hole *c* as $T_{a(2)}$ after passing through two hole period (2*d*) and one hole depth (*h*). The same process is for $T_{b(2)}$, which comes from hole *b* and re-emit at hole *d*. Therefore, $T_{a(2)}$, $T_{b(2)}$,$\cdots$ also have the same phase, they form constructive interference too. From the above analysis, we conclude that the surface plasmon passing through m_0 hole periods with a distance m_0 * *d* and going back and forth n_0 times inside *a* hole with a distance ($2n_0$-1) * *h* will always result in constructive interference and form a plane wave, so the electric field of the generated plane wave in the far field can be expressed as:

$$\begin{aligned} Ee^{i\phi} = E_0\{&T_{11}e^{i[\omega t-(k_1 d+k_2 h)+\phi_0]} + T_{21}e^{i[\omega t-(2k_1 d+k_2 h)+\phi_0]} + T_{31}e^{i[\omega t-(3k_1 d+k_2 h)+\phi_0]} + \cdots \\ &+ T_{12}e^{i[\omega t-(k_1 d+3k_2 h)+\phi_0]} + T_{22}e^{i[\omega t-(2k_1 d+3k_2 h)+\phi_0]} + T_{32}e^{i[\omega t-(3k_1 d+3k_2 h)+\phi_0]} + \cdots \\ &+ T_{13}e^{i[\omega t-(k_1 d+5k_2 h)+\phi_0]} + T_{23}e^{i[\omega t-(2k_1 d+5k_2 h)+\phi_0]} + T_{33}e^{i[\omega t-(3k_1 d+5k_2 h)+\phi_0]} + \cdots \\ &+\cdots\} \end{aligned} \tag{3.10}$$

where E_0 and E are the amplitude of the incident plane wave and the transmitted plane wave respectively. ϕ is the phase of the transmission light in the far field. ω is the angle frequency of the incident light. T_{mn_0} is the amplitude transmission coefficient of corresponding mode. ϕ_0 is the phase item that includes the initial phase of the source, the phase change from source to the front surface of the device and the phase change from the device to the point in the far field. k_1 and k_2 is the propagation constants of the surface plasmon on the front surface of the metal film and in the hole respectively. We can just approximately calculate the k_1 and k_2 by assuming metal and aqueous medium are infinite, i.e.

$$k_1 \approx k_2 \approx \mathrm{Re}[\frac{\omega}{c}(\frac{\varepsilon_{metal}\varepsilon_{eff}}{\varepsilon_{metal}+\varepsilon_{eff}})^{1/2}] \tag{3.11}$$

where $\varepsilon_{\mathrm{metal}}$ is the permittivity of the metal, and $\varepsilon_{\mathrm{eff}}$ is the effective permittivity of the dielectric containing receptors, target molecules and water. Therefore, if we neglect high mode due to their week amplitudes, Eq. (2.13) can be simplified as:

$$Ee^{i\phi} \approx E_0 T_{11} e^{i\{\omega t - \mathrm{Re}[\frac{\omega(d+h)}{c}(\frac{\varepsilon_{metal}\varepsilon_{eff}}{\varepsilon_{metal}+\varepsilon_{eff}})^{1/2}]+\phi_0\}} \tag{3.12}$$

Therefore if the effective refractive index of the dielectric $\varepsilon_{\mathrm{eff}}$ is changed, the phase of the transmitted light will also shift accordingly.

2.5.4 Optical setup for SPR phase shift measurement

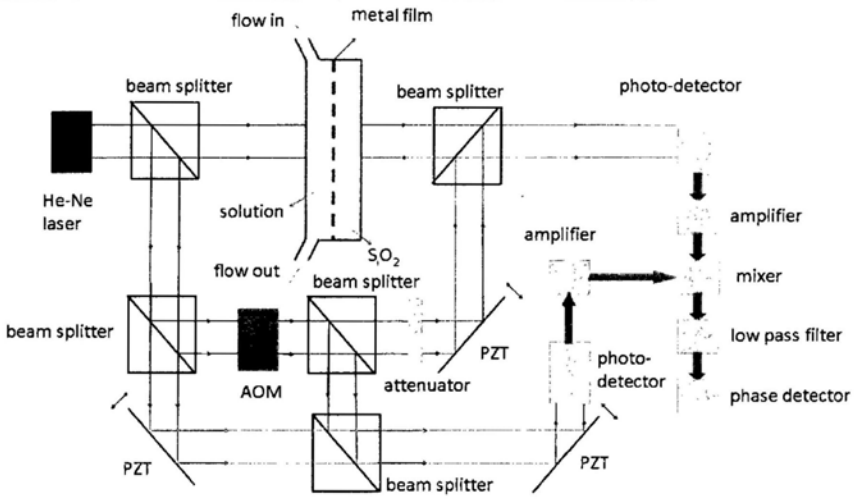


Fig. 2-17 Optical setup for SPR phase shift measurement

Fig. 2-17 demonstrates an experimental phase interrogation scheme based on a heterodyne interferometric system, which offers the benefits of detecting time-varying signals only and high noise rejection capability upon using long integration time [82-85]. A typical heterodyne interferometer uses an acousto-optic modulator to impose a frequency shift in the input laser beam. When the frequency-shifted reference beam interferes with the signal beam, the phase of the beat frequency, which may be readily measured by the lockin technique, will provide the phase reading introduced by the sensor device. In other words, when the target receptor molecules in the sample solution are immobilized on the nanohole sensor surface that has been functionalized by the conjugate ligands, the resultant refractive index change will shift the optical phase of the EOT accordingly. The heterodyne interferometer will then process the optical beams and produce the required phase reading as desired.

2.6 Summary

There is a need for detection and analysis of biochemical substances in many important areas including medicine, environmental monitoring, biotechnology, drug and food monitoring, military and civilian airborne biological and chemical agent testing, and real-time chemical and biological production process monitoring. SPR sensors hold potential for applications in these areas, hence attracting a great deal of attention. A myriad of research reports have appeared describing advances in SPR technology and its sensing applications. Over the last two decades, SPR sensors have made great strides in terms of optical systems, data processing, and microfluidics. At present, SPR biosensors have become a central tool for characterizing and quantifying biomolecular interactions due to their special advantages of label-free and real-time quantification capability. But conventional SPR biosensing techniques are normally based on Kretschmann configuration. One major drawback with the conventional methods is the difficulty in incorporating the sensor elements in high NA aperture imaging systems to increase spatial resolution and the corresponding number of resolvable spots. Operating on a prism or with a first- or higher-order diffraction mode places severe constraints on the depth of focus in the imaging system needed for large arrays of assays. Massive parallelism, and hence high throughput, is of primary importance in many potential SPR sensor applications, but they are severely limited by most of the traditional sensing methods.

The array of sub-wavelength holes investigated here are only a few micrometers in length, and the detection is performed in the transmission mode at normal incident angle. These features render that the device substrate is ideal for miniaturization as well as integration with lab-on-chip platforms. With the holes in sub-wavelength scales, there exists an inherent advantage of high packing density, thus leading to the implementation of 2-D biosensor arrays. While the optical geometry of this system is collinear, its operation in a transmission mode as opposed to the traditional Kretschmann (reflection) configuration will greatly reduce the beam alignment complexity for setting the phase interrogation Mach-Zehnder interferometer. The proposed nanohole sensor chip therefore offers a practical solution for a range of potential SPR applications in biological and chemical sensing. Heterodyne technique is used for phase detection so as to realize even high sensitivity.

Chapter 3: Simulation Experiments for EOT Based Sensors

3.1 Introduction

We have performed a series of simulation experiments by varying the hole period, hole shape, metal layer thickness and metal material respectively in order to find the optimized parameters for EOT based sensors. The finite-difference time-domain (FDTD) method [88-93] is implemented for computational electromagnetism. This is a widely used technique in which space is divided into a discrete grid and then the fields are evolved in time using discrete time steps—as the grid and the time steps are made finer and finer, this becomes a closer and closer approximation for the true continuous equations, and we can simulate many practical problems essentially exactly [86].

In this chapter, we introduce the theories and the simulation results of a FDTD method named FDTD Solutions [87]. Simulation results show that nanoholes fabricated in a metal film may lead to substantially phase change because of plasmonic effects, hence indicating the possibility of biosensing devices.

3.2 Simulation theory and results

3.2.1 Introduction of FDTD methods

Computational electromagnetism is implemented using the FDTD method in the thesis. The aim is to solve the system's Maxwell's equations in complex geometries

[88]. In particular, the equations for describing the fields in non-magnetic materials are:

$$\frac{\partial \bar{D}}{\partial t} = \nabla \times \bar{H} \tag{3.1}$$

$$\bar{D}(\omega) = \varepsilon_0 \varepsilon_r^*(\omega) \bar{E}(\omega) \tag{3.2}$$

$$\frac{\partial \bar{H}}{\partial t} = -\frac{1}{\mu_0} \nabla \times \bar{E} \tag{3.3}$$

where ***H***, ***E***, and ***D*** are the magnetic, electric, and displacement fields respectively, while $\varepsilon_r^*(\omega)$ is the complex relative dielectric constant ($\varepsilon_r^*(\omega)=n^2$, where n is the refractive index).

FDTD methods divide space and time into a finite rectangular grid. As described below, FDTD methods try to hide this discreteness from the user as much as possible, but there are a few consequences of discretization that it is good to be familiar with [86]. Perhaps the most important thing need to know is this: if the grid has some spatial resolution Δx, then our discrete time-step Δt is given by $\Delta t = S\Delta x$, where S is the Courant factor, in order for the method to be stable (not diverge). This means that when we double the grid resolution, the number of time steps doubles as well (for the same simulation period). Thus, in three dimensions, if we double the resolution, then the amount of memory increases by 8 and the amount of computational time increases by (at least) 16. The sampling in space is at sub-wavelength resolution set by the user to properly sample the highest near-field

spatial frequencies thought to be important in the physics of the problem. Typically, 10-20 samples per wavelength are needed. The sampling in time is selected to ensure numerical stability of the algorithm.

The second most important thing we should know is that, in order to discretize the equations with second-order accuracy, FDTD methods store different field components at different grid locations. This discretization is known as a Yee lattice, and is described in more detail at: Yee lattices. As a consequence, FDTD methods must interpolate the field components to a common point whenever we want to combine, compare, or output the field components (e.g. in computing energy density or flux). Most of the time, we don't need to worry too much about this interpolation since it is automatic. However, because it is a simple linear interpolation, while $\boldsymbol{E}$ and $\boldsymbol{D}$ may be discontinuous across dielectric boundaries, it means that the interpolated $\boldsymbol{E}$ and $\boldsymbol{D}$ fields may be less accurate than we might expect right around dielectric interfaces.

Although FDTD methods inherently use discretized space and time, as much as possible the methods attempt to maintain the illusion that we are using a continuous system. At the beginning of the simulation, we specify the spatial resolution, but from that point onwards we generally work in continuous coordinates in our chosen units.

For example, we specify the dielectric function as a function $\varepsilon(x)$ of continuous x, or as a set of solid objects like spheres, cylinders, etcetera, and a FDTD software is responsible for figuring out how they are to be represented on a discrete grid. Or if we want to specify a point source, we simply specify the point x where we want the source to reside — the software will figure out the closest grid points to x and add currents to those points, weighted according to their distance from x. If we change x continously, the current in FDTD software will also change continuously (by changing the weights). If we ask for the flux through a certain rectangle, then the software will linearly interpolate the field values from the grid onto that rectangle.

However, FDTD method is, of course, not the only computational method in electromagnetism, nor is it always the best one. In general, we advocate having several tools in our toolbox, and selecting the most convenient one for each task.

For example, although FDTD can be used to compute electromagnetic eigenmodes, in lossless structures it is often quicker, easier, and more reliable to use a specialized eigenmode solver such as MIT Photonic-Bands (MPB) package

For computing the field pattern or response of a structure at a single frequency, it may be more efficient to directly solve the corresponding linear equation rather than iterating in time. Especially in cases where there are large differences in scale, it may be better to use a method that allows a variable resolution in different spatial regions, such as a finite-element or boundary-element method. Boundary-element

methods are especially powerful when we have a large volume-to-surface ratio, such as for scattering calculations over small objects in a large (∞) volume.

A merit of time-domain methods is their ability to obtain the entire frequency spectrum of responses (or eigenfrequencies) in a single simulation, by Fourier-transforming the response to a short pulse or using more sophisticated signal-processing methods. Finite-element methods can also be used for time-evolving fields, but they suffer a serious disadvantage compared to finite-difference methods: finite-element methods, for stability, must typically use some form of implicit time-stepping, where they must invert a matrix (solve a linear system) at every time step.

In systems that are composed of a small number of easily-analyzed pieces, such as a sequence of constant-cross-section waveguides, a collection of cylinders, or a multi-layer film, transfer-matrix/scattering-matrix methods may be especially attractive. These methods treat the individual simple elements in some analytic or semi-analytic fashion, enabling the entire structure to be simulated with great speed and accuracy. There are too many such techniques to easily summarize here, but one useful free tool that can handle a wide variety of structures is CAMFR (CAvity Modelling FRamework) [89].

Another common technique is the beam-propagation method (BPM), but BPM is only well-suited for problems where the structure is slowly varying in one direction.

Finally, I want to mention another very different simulation method, i.e. rigorous coupled-wave analysis (RCWA). RCWA utilizes the Bloch theorem (known as the Floquet theorem in the one-dimensional case) that the eigenfunctions of the wave equation for a periodic potential have the form of a plane-wave multiplied by a function that has the periodicity of the potential. Thus the electric and magnetic field can be represented by a Fourier series where the fundamental frequency corresponds to the periodicity of the grating. RCWA solves for the field by finding a modal (Fourier) decomposition for the field in the grating region and in the incident and transmitted regions, then finding the amplitudes of these modes that satisfy Maxwell's equations and electromagnetic boundary conditions. The mathematics involved in this process include finding matrix eigenvalues and eigenvectors (which correspond to modes), and solving a linear system of equations (to find the mode amplitudes).

In the past, EM distributions in periodic structure were usually calculated by RCWA, which analyzes general three-dimensional grating diffraction problems by solving Maxwell's equations rigorously. It requires that the permittivity distribution in the grating region is represented by a Fourier expansion. In our thesis, we need to simulate for round hole, curve edge and other complicated structures such as interferometers which have nonperiodic structures. However, these structures cannot be calculated by simply solving a linear system of equations with RCWA. Some structures may be built up, but it is not always as convenient as FDTD and it may not be implemented in some codes.

All in all, there are five primary reasons for the choice of FDTD method in our simulations:

(1) FDTD is direct solution method for Maxwell's curl equations. This method employ no potentials. Rather, it is based upon volumetric sampling of the unknown electric and magnetic fields within and surrounding the structure of interest, and over a period of time. hence FDTD can obtain the frequency solution by exploiting Fourier transforms, thus a full range of useful quantities can be calculated, such as the transmission / reflection of light.

(2) FDTD is accurate and robust. The sources of error in FDTD calculations are well understood, and can be bounded to permit accurate models for a very large variety of electromagnetic wave interaction problems.

(3) FDTD treats impulsive behavior naturally. Being a time-domain technique, FDTD directly calculates the impulse response of an electromagnetic system. Therefore, a single FDTD simulation can provide either ultrawideband temporal waveforms or the sinusoidal steady-state response at any frequency within the excitation spectrum.(4) Computer memory capacities are increasing rapidly. While this trend positively influences all numerical techniques, it is of particular advantage to FDTD methods which are founded on discretizing space over a volume, and therefore inherently require a large random access memory.

(5) Computer visualization capabilities are increasing rapidly. While this trend positively influences all numerical techniques, it is of particular advantage to FDTD methods which generate time-marched arrays of field quantities suitable for use in

color videos to illustrate the field dynamics.

3.2.2 Merits of FDTD Solutions

FDTD Solutions (Lumerical Solutions, Inc) is a high performance microscale optics simulation software using the FDTD method [87]. Employing the FDTD method, FDTD Solutions empowers device and components designers to confront challenging optical design problems. Rapid prototyping and highly-accurate simulations reduce reliance upon costly experimental prototypes, leading to quicker assessment of design concepts and reduced product development costs. FDTD Solutions with minimum 1nm mesh size is used for studying the structure of the SPR sensors in view of its advantages in nanoscale simulation. The following is the summary of merits of the software.

Overview:

- 2D and 3D simulation software capabilities
- Design parameterization and hierarchical layout via structure groups and analysis groups
- Nonuniform mesh and automesh algorithms
- Simulation convergence autoshutoff
- Parallel computation on multi-core, multiprocessor, and high-performance computing clusters
- Movie (.mpg) generation of simulation dynamics
- Scripting software language to customize simulation and analysis and perform

powerful post-processing of simulation results

- Lorentz, Drude, Debye and anisotropic materials
- Data import/export with Breault Research Organization's (BRO's) The Advanced Systems Analysis Program (ASAP) ray-tracing software package
- Data export to Matlab (meaning "matrix laboratory") software, or The American Standard Code for Information Interchange (ASCII) file formats
- Structure import from Graphic Data System II (GDSII) files
- Extensive online help

Features:

✧ Boundary conditions (BC):

absorbing (PML), periodic, Bloch, symmetric, asymmetric, and metal boundaries

Simulation objects:

primitives that can be rotated and placed in three dimensions, and structure definition from imported SEM/image files; primitives include triangles, rectangular blocks, cylinders, conic surfaces, polyons, rings, user-defined (parametric) surfaces, spheres and pyramids

✧ Radiation sources:

waveguide sources, dipoles, plane waves, focused beams and diffraction-limited spots, total-field scattered-field (TFSF) sources, and source import/export from/to BRO's ASAP ray-tracing program

✧ Measurement monitors:

refractive index monitors, time- and frequency-domain monitors to measure pulsed

or continuous-wave (CW) field profiles and power flow, and movie monitors to generate .mpg movies of field dynamics

Parallel/clustered performance:

The parallel option of FDTD Solutions allows for large-scale and rapid simulation of optical components, by distributing computational load and memory requirements across multiple nodes. Measured data shows that parallel FDTD Solutions offers significant performance enhancements as measured by the speed increase of simulations performed on multi-core and multiprocessor computing systems. Access to an 8-computer cluster will allow us to simulate a structure 8 times larger, or allow us to simulate the original structure approximately 6 times quicker.

Surface plasmon simulation:

In three dimensions, Maxwell equations have six electromagnetic field components: E_x, E_y, E_z and H_x, H_y, and H_z. If we assume that the structure is infinite in the z dimension and that the fields are independent of z, specifically that

$$\varepsilon_r^*(\omega, x, y, z) = \varepsilon_r^*(\omega, x, y) \qquad (3.4)$$

$$\frac{\partial \vec{E}}{\partial z} = \frac{\partial \vec{H}}{\partial z} = 0 \qquad (3.5)$$

then Maxwell's equations split into two independent sets of equations composed of three vector quantities each which can be solved in the x-y plane only. These are termed the TE (transverse electric), and TM (transverse magnetic) equations. We can solve both sets of equations with the following components:

TE: E_x, E_y, H_z

TM: H_x, H_y, E_z

In the TM case, Maxwell's equations reduce to:

$$\begin{aligned} &\frac{\partial D_z}{\partial t} = \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} \\ &D_z(\omega) = \varepsilon_0 \varepsilon_r^*(\omega) E_z(\omega) \\ &\frac{\partial H_x}{\partial t} = -\frac{1}{\mu_0}\frac{\partial E_z}{\partial y} \\ &\frac{\partial H_y}{\partial t} = \frac{1}{\mu_0}\frac{\partial E_z}{\partial x} \end{aligned} \tag{3.6}$$

The TM mode can be used to generate SP which can be simulated by FDTD Solutions. The simulation results are very close to the experimental results [14, 80, 90] as shown in Fig.3-1. The left figures are the experiment results copied from publication papers. The right figures are our simulation results. There are three examples in the figure. In the first two examples, we only simulate for the largest peaks. All the themes are related to EOT. We also notice that although the transmission results of the simulations are quite similar to the experimental results, they are not identical. Most of the differences are due to the material properties. It is challenging to get an exact same simulation results for some special experiments, since we must know the material properties very exactly to get good simulation results.

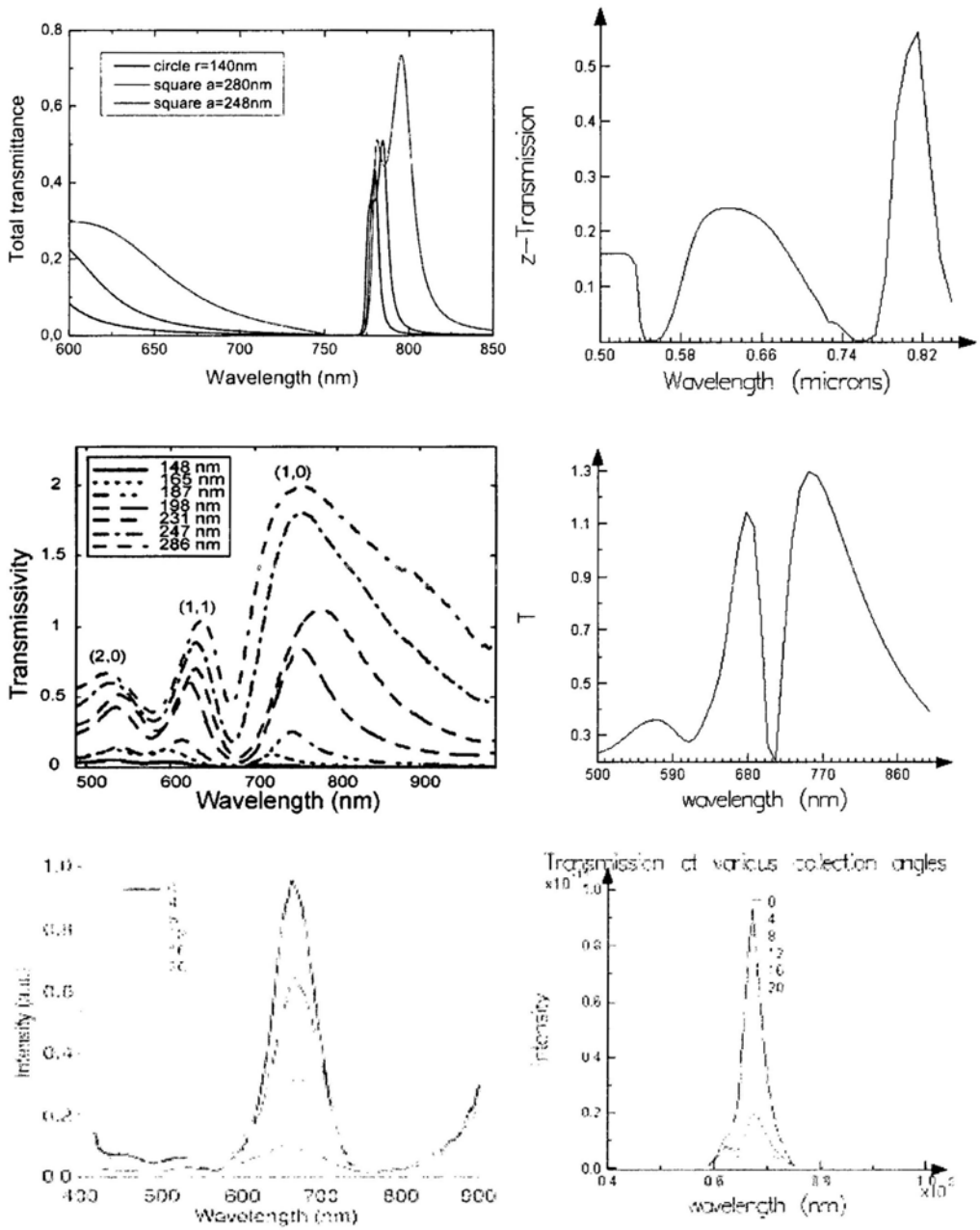


Fig. 3-1 Comparison of simulation results and experimental results

(Left figures: experimental results; Right figures: corresponding simulation results)

3.2.3 Boundary conditions and symmetries

On a computer, we can only simulate a finite region of space, which means that we must terminate our simulation with some boundary conditions [86, 88]. As reported

above, six basic types of terminations are supported by FDTD Solutions: absorbing (PML), periodic, Bloch, symmetric, asymmetric, and metal boundaries, which are explained in detail below.

PML

Many simulations employ absorbing boundary conditions that allow radiation to propagate out of the computational area without interfering with the fields inside. Perfectly matched layer (PML) boundaries absorb electromagnetic energy incident upon them. PML is most effective when absorbing radiation at normal incidence; it can have significant reflection at grazing incidence. Increasing the PML thickness (number of layers) will reduce reflections, but increase the simulation time and memory requirements.

Metal

Metal boundary conditions are used to specify boundaries which are perfectly reflecting, allowing no energy to escape the simulation volume along that boundary.

Periodic

Periodic BC should be used when both the structures and EM fields are periodic. Periodic boundary conditions can be used in one or more directions (i.e. only in the X direction) to simulate a structure which is periodic in one direction but not necessarily other directions.

Bloch

Bloch BC should be used when the structures are periodic, and the EM fields periodic, except for a phase shift between each period. Bloch boundary conditions

are used predominantly for the following two simulations:

- Launching a plane wave at an angle to a periodic structure – in this situation, accurate reflection and transmission data can be measured at a single frequency point for a given simulation.
- Calculating the bandstructure of a periodic object – in this situation, a broadband pulse is injected via a dipole source into a periodic structure.

Symmetric

Symmetric boundary conditions are used when the user is interested in a problem that exhibits one or more planes of symmetry. Both the structure and source must be symmetric. Symmetric boundaries are mirrors for the electric field, and anti-mirrors for the magnetic field. A visual explanation of a symmetric boundary condition is shown in the figure below. Careful consideration must be given to whether symmetric or asymmetric boundary conditions are required, given the vector symmetry of the desired solution. For meaningful results, the sources used must have the same symmetry as the boundary conditions.

Asymmetric

Asymmetric boundary conditions are used when the user is interested in a problem that exhibits one or more planes of symmetry. Asymmetric boundaries are anti-mirrors for the electric field, and mirrors for the magnetic field.

3.2.4 Parameter design for the sensors

Employing the FDTD algorithm, FDTD Solutions enable us to calculate the phase

changes in the far field one by one through the analysis of different array structures and their electromagnetic behavior. In order to shorten simulation time, an area with just one square hole is meshed and periodic boundary conditions are used around the hole in each simulation. We use a plane wave with a wavelength of 632.8nm and the propagation direction is normal to the structure. Perfectly matched layer boundary conditions with 200 layers are used in the source directions. We assume the total thickness of receptor layer and immobilized molecules is 5nm, and the refractive index of the surrounding aqueous medium is taken as 1.33. If the effective refractive index of this composite layer is changed from 1.33 to 1.50, the phase variation of the transmission beam in the far field after passing through a hole array (typically 100×100) can be calculated accordingly. Because the hole period and the metal thickness are two main parameters affecting the resonances on the film surface and in the holes respectively, we can fix one parameter and change another to obtain the optimal values of these two parameters for the largest phase change when the effect refractive index is varied from 1.33 to 1.50. Therefore, in the present case, we first fix the metal thickness at 108nm and find the relationship between the hole period and the phase change as shown in Fig. 3-2. Here we can see from the sharp resonance peaks that the optimal values of hole period for gold and silver are 427.6nm and 446.0 nm respectively.

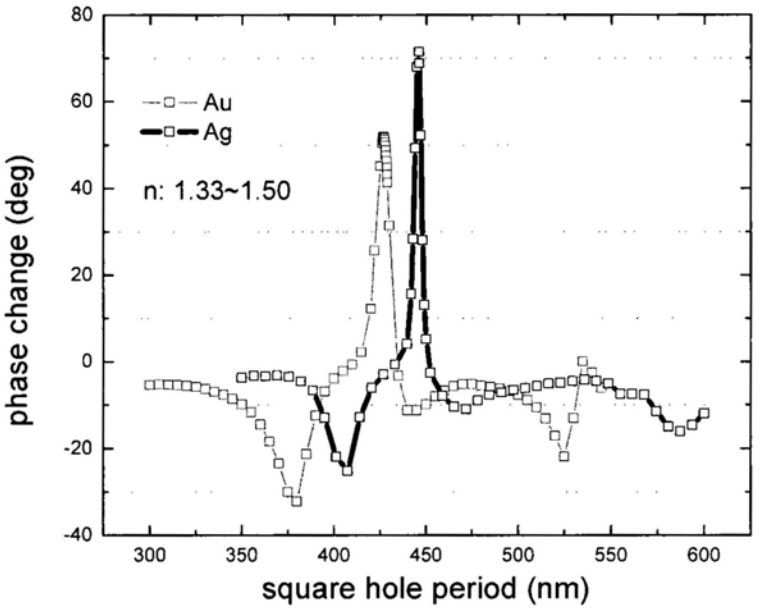


Fig. 3-2 Square hole period versus phase change

(red line for gold and black line for silver)

3.2.5 Square hole period simulation

As reported in Chapter 2, maximum phase change in the SPR case occurs near the resonant region, and for normal incident of light the equation for resonance condition is given by [76]

$$(i^2 + j^2)^{1/2} \lambda_{spp} = d \,\mathrm{Re}[(\frac{\varepsilon_m \varepsilon_d}{\varepsilon_m + \varepsilon_d})^{1/2}] \qquad (3.7)$$

where i and j are mode numbers in X and Y directions of Fig. 2-13, λ_{spp} is the wavelength of the incident light, d is the periodicity of the array, ε_m is the effective permittivity of the metal layer, and ε_d is the effective permittivity of the dielectric

layer containing receptors, target molecules and water. The effective permittivity of the dielectric can be estimated by performing a weighted averaging within the extension l of the evanescent surface plasmon mode into the dielectric (Z direction), according to Ref [77]

$$\varepsilon_d = \frac{2}{l}\int_0^\infty \varepsilon(z)\exp(-\frac{2z}{l})dz \qquad (3.8)$$

As an example, for a device using a gold sensor layer, when λ_{spp} equals to 632.8nm, according to Eq. (3.7), resonant peak occurs when the hole period is 436.0nm when we assume ε_d equals to $(1.33)^2$ and ε_m equals to -10.85+1.25i. One should note that the effective permittivity of the dielectric ε_d must be slightly larger than $(1.33)^2$ due to the existence of receptors and target molecules. On the other hand, the absolute permittivity value of the gold sensor layer should be slightly smaller than that of pure gold due to the existence of the holes [78]. Therefore, the resonant period should be close to 436.0nm, and the result of 427.6nm as obtained from our simulation is reasonable. The same calculation procedures also reveal that the resonant period should be close to 451.3nm for silver (ε_m equals to -17.64+0.50i), and our simulation result of 446.0nm is in line with this expectation.

We should mention that the actual device fabrication using FIB or EBL may be difficult due to the stringent requirement on periodicity accuracy, as the two resonance peaks shown in Fig. 3-2 occur within a hole-period range of about 10nm. Nonetheless, we believe that advancement of instrumentation technology should soon overcome this problem. Moreover, the fact that we are dealing with periodic

structures with periodicity in the region of 400~450nm may have alleviated the difficulty somewhat, as one might be able to fabricate the device using a UV interference process.

The intensity distribution of electromagnetic field in the near field by using the optimal period for golden device is shown in Fig. 3-3. Because periodic boundary condition is used, the figure only shows the intensity distribution of one hole. The intensity is shown on a logarithmic scale and the source beam propagates in the direction of –Z in this figure. Because of the large intensity of the incoming beam and reflected beam, the nodes and antinodes of SP on the front surface of the metal are blurry in the picture. However, there are two obvious nodes and antinodes on the back surface of the metal in one periodic structure. Therefore, (1,0) or (0,1) mode should be the prevalent mode.

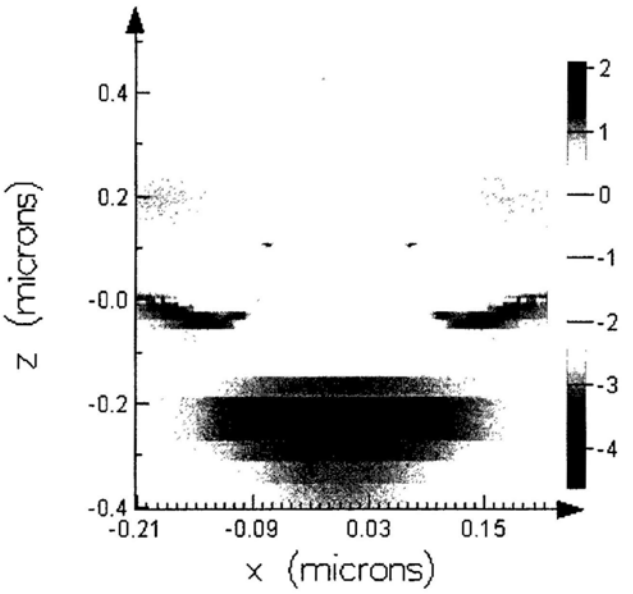


Fig. 3-3 logarithmic intensity distribution in near field (one hole)

This is a localized EM mode, confined to the interface. It is not the only EM mode as there also exists the zero-parallel-momentum propagating mode in the hole. when we calculate the SP of the periodically perforated metal slab, we find that the single interface SPs on the two metal surfaces combine to form a "SP molecule" in much the same way as electronic state of isolated atoms combine to form molecular levels. Because the dielectric constants on the two metal surfaces are different, the SP molecule here is in fact a diatomic heteropolar molecule but the picture still applies in the nonsymmetric case. So, the molecule levels are fully formed for the small metal layer thickness, which is smaller than the propagation length of the SP. The photon then goes back and forth several times inside the hole, building up coherent constructive interference in the forward direction much as would occur in electron resonant tunneling [79]. That is the reason why metal layer thickness is also an important parameter, which will be discussed in next section.

3.2.6 Metal layer thickness simulation

The metal thickness also plays an important role in the calculation of phase change. We first fix the period at the optimized value (i.e. 427.6nm for gold and 446.0 nm for silver), then we vary the metal layer thickness and monitor the phase change. Our results are shown in Fig. 3-4. The figure indicates that the largest phase change occurs at 108nm and 110nm for gold and silver respectively. The physical mechanism for such a resonant peak to occur at a certain thickness, for example 108nm for gold, is due to the fact that the photon trapped inside the hole goes back

and forth several times, which leads to the building up of constructive interference [79].

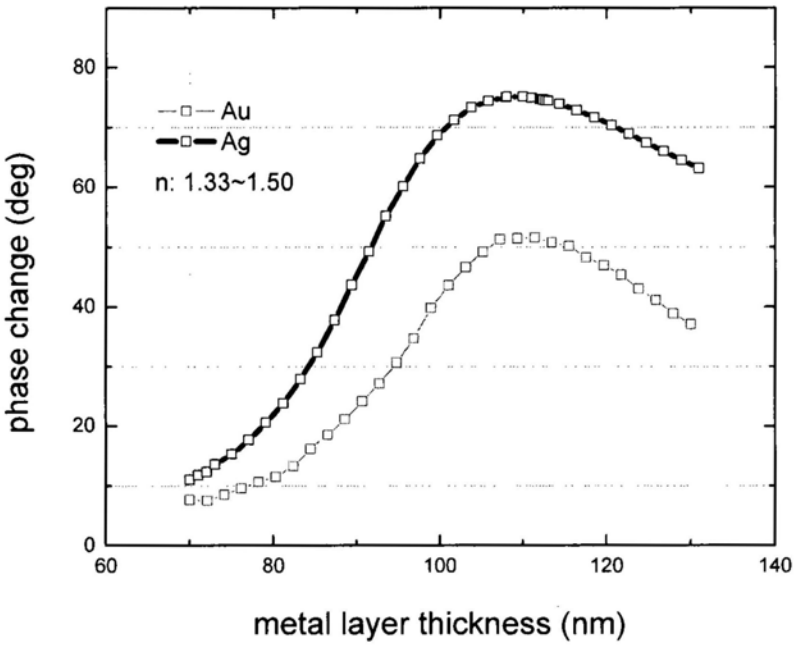


Fig. 3-4 Metal layer thickness versus phase change

(red line for gold and black line for silver)

If we assume that the surface plasmon polaritons in the hole propagates along an interface between an infinite insulator and a metal despite that it is in fact a metal/insulator/metal (MIM) heterostructure, then the propagation length of a surface plasmon with 2π phase change in a hole of golden film is calculated as 436nm. This assumption is valid when the thicknesses of the gold and water of the hole are larger than the intensity decay lengths of the surface plasmon. It is easy to be satisfied because theoretical results show that the intensity decay lengths of SP into water and gold are 86.39nm and 13.97nm respectively when the wavelength of

the incoming beam is 632.8nm, ε_d equals to $(1.33)^2$ and ε_m equals to -10.85+1.25i.. Since fourfold of 108nm is 432nm which is very close to 436nm, we can conclude that the plasmon in the hole should experience constructive interference if the metal thickness is 108nm. This also means that when constructive interference is disrupted due to the change of refractive index in the surrounding medium, the phase changes rapidly.

3.2.7 Square hole width simulation

We also explore the effect of varying the width of the holes, which has the obvious effect of changing the effective permittivity of the MIM heterostructure. The phase change in the range of refractive index between 1.33 and 1.50 is shown in Fig. 3-5. In addition, one may readily expect that square holes may have more localized electric field than circular holes because of their sharp corners [80, 81]. As shown in Fig. 3-6, square holes indeed provide a more pronounced phase change peak than circular holes when we vary the hole width or diameter. In order for us to investigate this aspect of the holes, we have chosen to adopt a 3D simulation model rather than the 2D version.

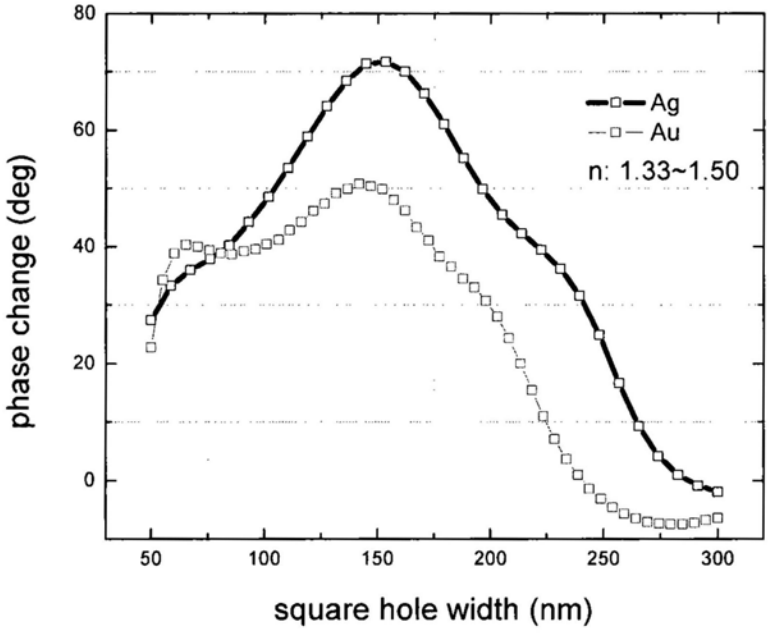


Fig. 3-5 Square hole width versus phase change (red line for gold and black line for silver)

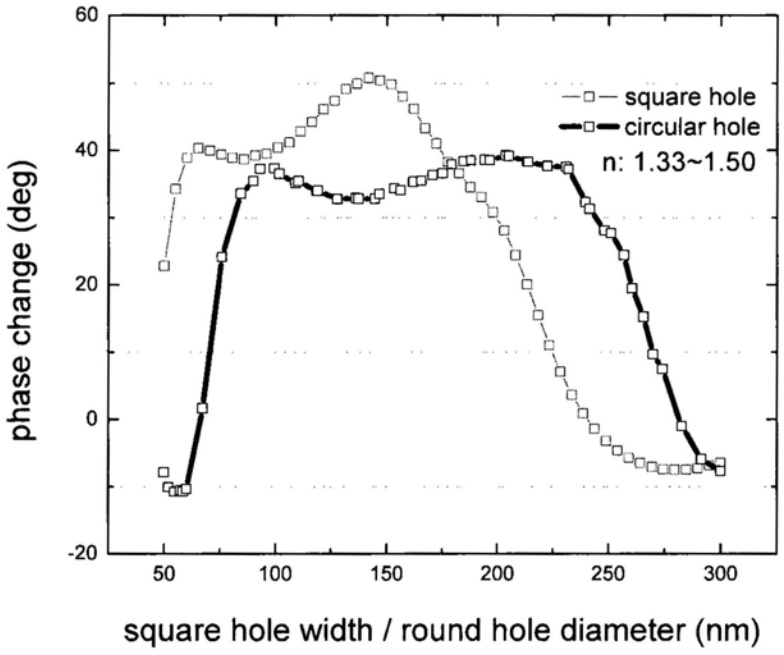


Fig. 3-6 Influence of hole shape on phase change (sensor layer material: gold)

3.2.8 Predicted sensitivity and dynamic range

In Fig. 3-7, we show the calculated phase change versus refractive index in the range of 1.33 and 1.50 when all the device parameters are optimized. For a gold sensor device with square holes, we use 142nm for the hole width, 428nm for the hole period and 108nm for the film thickness. For a maximum refractive index shift of 0.17, the phase change of 51.8°.

For the case of using silver for the sensor layer, we use a hole width of 153nm, hole period of 446nm, film thickness of 110nm, and the maximum phase change is 75.1°. The phase change from silver device is higher than gold device in light of its higher free electron density. An attribute as revealed from Fig. 3-7 is that the phase change is quite linear over the entire refractive index range of 0.17, whereas the lack of phase linearly and measurement dynamic range are known problems for the common Kretschmann configuration. Such drawbacks are resulted from the fact that the surface plasmon waves excited in an inverted prism containing a single metal layer has only one resonance peak, where all the phase change happens very abruptly within a small refractive index range. On the other hand, the nanohole device that we are proposing here is a complex multi-resonance system, as explained in Chapter 2. A continuous change of the phase over a wide range of refractive index values may become possible.

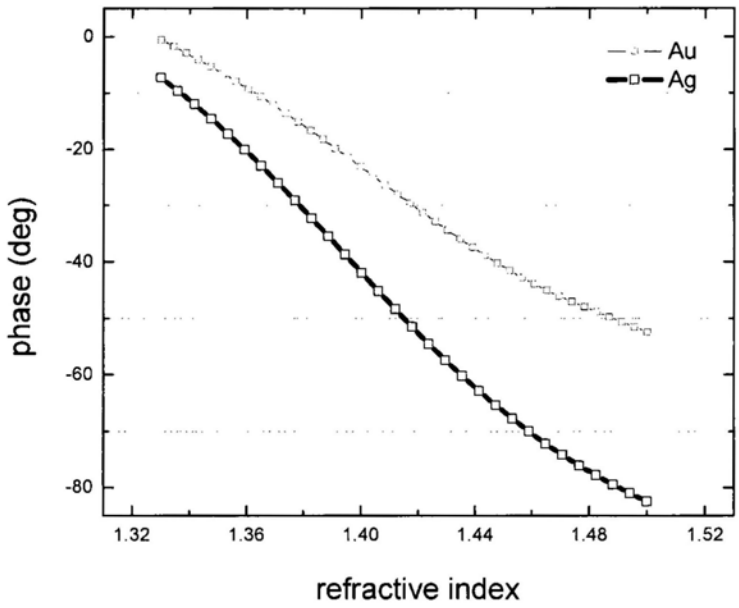


Fig. 3-7 Refractive index versus phase change with all parameters optimized (red line for gold and black line for silver)

As for the actual phase detection implementation using a heterodyne technique, assuming that the interferometer is operating at its most sensitive point (which corresponds to zero output for the balanced detector arrangements) the photon noise equivalent displacement δx_N can be calculated by the following formula [94]

$$\delta x_N = \frac{\lambda}{4\pi}\left(\frac{h\upsilon\Delta f}{\eta W_s}\right)^2 \qquad (3.9)$$

where we assume the wavelength λ is 632.8nm, the quantum efficiency η is 70%. For a signal power W_s of 2.5μW (Assuming that the power of the incoming beam is 0.25mW and the transmission coefficients are 1%, although simulation results indicate that the transmission coefficients are 1.7509% and 1.3233% for silver and gold respectively.) and a bandwidth Δf of 1Hz, we find δx_N=0.0213pm. Therefore,

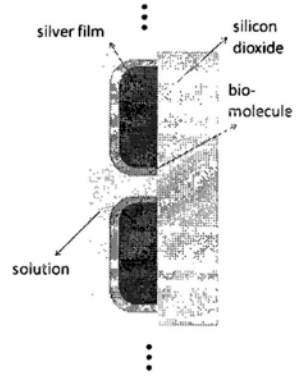
the detection resolution is 1.93×10^{-6} degree. The predicted sensitivity limit of an optimal gold device and an optimal silver device are 6.33×10^{-9} RIU and 4.37×10^{-9} RIU respectively. As shown in Table 1, this value is comparable to those calculated for other SPR sensor schemes using the Kretschmann configuration [95-97]. Although the phase interrogation SPR sensor using the Kretschmann configuration can also offer similar sensitivity limits, its refractive index detection range is much narrower in comparison to the present nanohole scheme. Furthermore, one major difference between our design and the traditional Kretschmann configuration is that the former detects the phase shift from the transmitted beam while the latter does this from the reflected beam. Our design therefore makes setting up of the Mach-Zehnder interferometer much easier. In light of the fact that the present scheme may offer advancement on important attributes including dynamic range, sensitivity threshold, data linearity and ease of practical implementation, there should be a good chance that this scheme will lead to practical SPR sensors, particularly for the multiple analyte biochip format.

3.3 Other design considerations

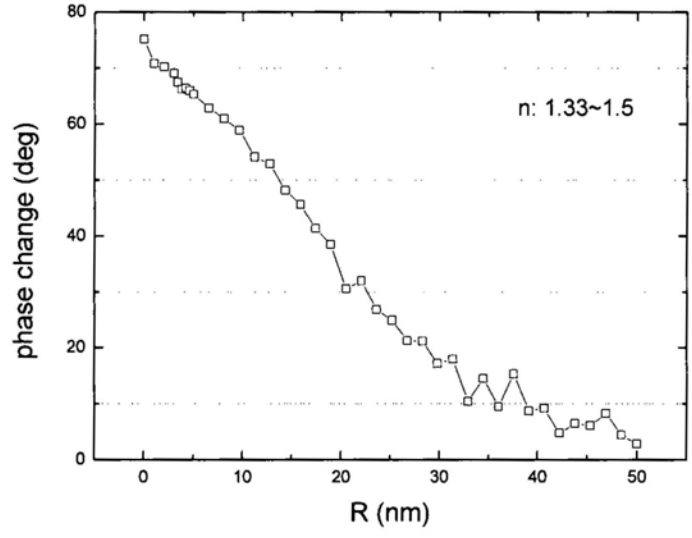
3.3.1 Curved corners at hole edges

It is likely that the actual device fabrication process would produce curved edges as shown in Fig. 3-8(a). Here we assume that the edge of the hole is round and the radius of curvature of the round edge is *R*. Fig. 3-8(b) reveals that curved edges will lead to smaller phase change. This is consistent with our expectation because of the

fact that as the edges are less abrupt it is less likely for reflection of the surface plasmon to take place at the top and bottom of the nanohole. The resonant behavior of the nanohole will become less prominent. The estimated sensitivity limit of using round cornered square holes having a 5nm radius of curvature is 5.03×10^{-9} RIU.



(a)



(b)

Fig. 3-8 Effect of curved corners at hole edges

(a) Schematic of curved corners. (b) Radius of curvature versus phase change.

3.3.2 Surface roughness of glass substrate

If one uses a focused ion beam to fabricate the nanohole array, it is inevitable that the surface of the glass substrate at the bottom of the hole becomes rough as depicted in Fig. 3-9. Here we assume that the average amplitude of the roughness is 1nm [88], the corresponding phase change for the case of shifting the refractive index from 1.33 to 1.50 in a device using silver as the sensor layer is 73.0°. The phase change is 75.1° if the substrate is perfectly smooth. This means that surface roughness can degrade device performance, but the effect is not too significant.

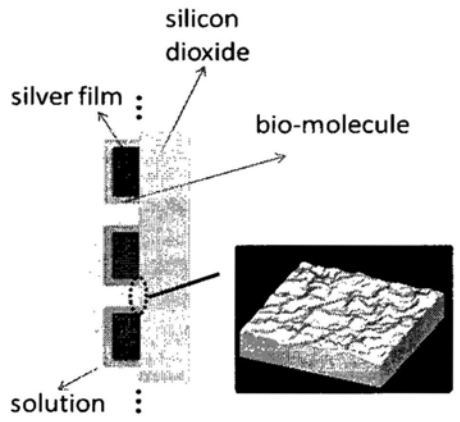
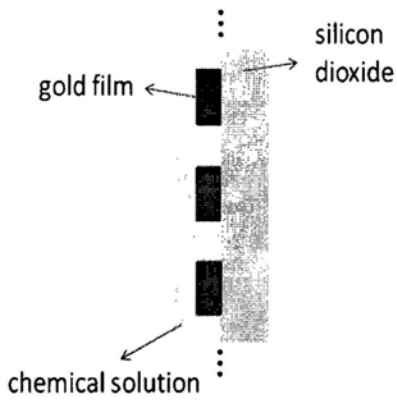


Fig. 3-9 Surface roughness on glass substrate

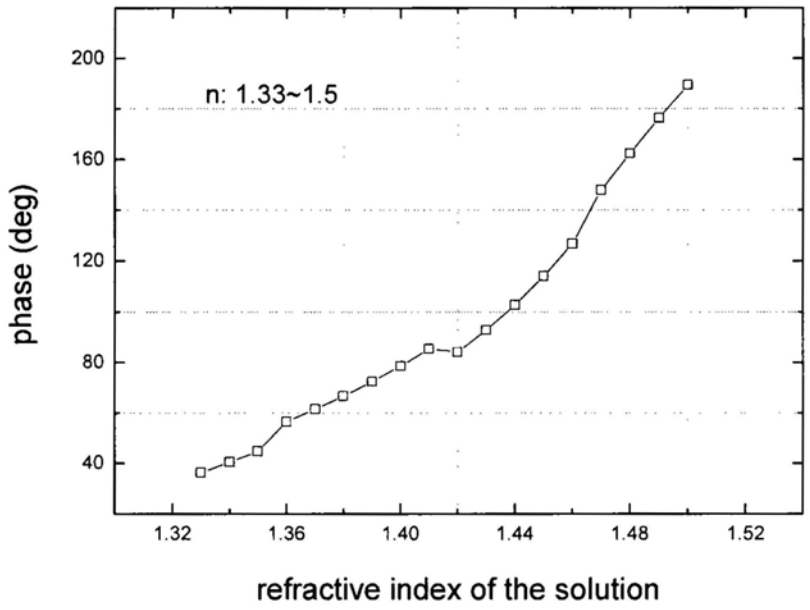
3.3.3 Chemical sensing

The nanohole array can also be used for detecting the concentration change of a chemical solution as demonstrated in Fig. 3-10(a). When the concentration of a chemical solution increases, the refractive index of the solution increases accordingly. The main difference of this case compared to the previous one, where one detects the immobilization of biomolecules within a thickness of 5nm, is that

the refractive index of the entire solution is being monitored. This means that the entire evanescent field is interacting with the analyte solution, thus leading to relatively larger response for a given change of refractive index value. The plot shown in Fig. 3-10(b) are obtained from a device having a hole width of 179nm, hole period of 364nm and gold film thickness of 108nm. For a refractive index range of 1.33 to 1.50, we get net a phase change of 153.1°. Assuming a phase detection resolution of 1.93×10^{-6} degree, as derived in Section 4.5, the sensitivity threshold is estimated to be 2.14×10^{-9} RIU, which is a very respectable value for SPR sensors.



(a)



(b)

Fig. 3-10 Chemical sensing

(a) Schematic of sensor structure; (b) Refractive index of the solution versus phase.

There is a small peak when refractive index of the solution is 1.41. This is because there is a resonant mode around this position as demonstrated in Section 3.2.6. The wavelength of SP at this position is 422.8nm. A quarter of this value is 106nm, which is very close to the thickness of the metal 108nm. Therefore, a sharp change of phase occurs at this position. We also notice that continuous changing of refractive index renders non- monotonic results. This may be because the system is in fact a multi-resonance system. Continuous variation of the refactive index leads to the fulfillment of difference resonance conditions.

3.3.4 Mesh size consideration

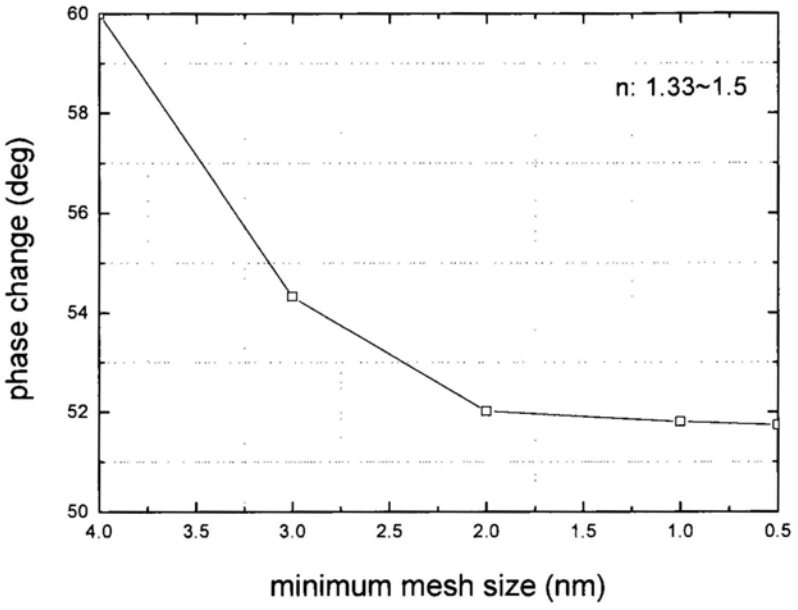


Fig. 3-11 the relation between mesh size and phase change

It is possible that we will get different results with different mesh size in the simulations. In theory, in order to get accurate simulation results, the mesh size should as small as possible. However, in three dimension simulations, if we double the resolution, the amount of memory increases by 8 and the amount of computational time increases by (at least) 16. Therefore, we should find suitable mesh sizes for our simulations. Normally, the minimum mesh size should be at least half of the minimum size of the structures. In Fig. 3-6, the maximum phase change is 51.8° for gold when the minimum mesh size is 1 nm. If we change the minimum mesh size, we can get different phase change as shown in Fig. 3-11. The simulation results prove that the 1 nm minimum mesh size is enough for simulation of this structure.

3.4 Summary

In summary, we have demonstrated that nanoholes fabricated in a metal film may lead to quite sizeable phase changes because of plasmonic effects, hence indicating the possibility of sensing devices. FDTD Solutions is used as the tool for the simulation experiments. FDTD is a time domain technique, meaning that the electromagnetic fields are solved as a function of time. Within our simulation experiments, space is divided into a discrete mesh and then the fields are evolved in time using discrete time steps. As the mesh and the time steps are made to have smaller dimensions, the numerical model becomes a close approximation of the analytical equations. The most important merit of time-domain methods is their ability to obtain system responses (or eigenfrequencies) in the entire frequency spectrum within a single simulation through the use of Fourier-transformation on the response of a short pulse. Finite-element methods (FEM) can also be used for time-evolving fields, but they have a serious disadvantage in terms of stability: one typically needs to use some form of implicit time-stepping, in which one must invert a matrix (i.e. solve a linear equation system) at every time step.

Our simulation results have shown that the 2D nanoholes are well suited for the sensor chip format in which high dense integration is readily achievable. While the sensor operates at normal illumination, practical implementation of the sensor is much easier in comparison to the traditional Kretschmann arrangement for SPR sensing. Various design parameters of the device have been studied by simulation.

Our results indicate that the scheme has a shot-noise limited sensitivity threshold of 4.37×10^{-9} RIU and a dynamic range of 0.17 RIU, which compare favorably with typical SPR sensors, particularly in terms of achieving high resolution and wide dynamic range sensor attributes. The phase change is also quite linear over the entire refractive index detection range. Curved hole edges and substrate roughness due to fabrication limitations have been quantitatively discussed. Application of the proposed device for chemical sensing is also demonstrated.

Chapter 4: Spectrometer Designs

4.1 Introduction

From this chapter, let's shift our concentration from bio-sensors to spectrometers, which are often used in astronomy, chemistry and some branches of biology. A spectrometer (spectrograph or spectroscope) is an optical instrument used to measure properties of light over a specific portion of the electromagnetic spectrum, typically used in spectroscopic analysis to identify materials. For example, when a material is heated to incandescence it emits light that is characteristic of the atomic makeup of the material. Particular light frequencies give rise to sharply defined bands on the scale which can be thought of as fingerprints. Therefore, information about the absorption or emission of radiation in a sample can be obtained by measuring the spectrum intensity ratio of the radiation "before" and "after" the sample [98-99].

Traditionally, dispersive instruments, available since the 1940s, were used to obtain infrared spectra. In recent decades, some very different methods of obtaining a spectrum have superceded the dispersive instrument. For example, Fourier transform (FT) spectrometers are now widely used and have improved the acquisition of spectra dramatically. However, Fourier transform spectrometers also suffer the disadvantages of big size and low resolution. In order to overcome the limitations of traditional devices and explore a cheap, compact and high

performance spectrometer, we propose a novel interferometer-based spectrometer. In this chapter, we introduce the principles of dispersive spectrometers, Fourier transform spectrometers and self-mixing interference spectrometers. The advantages and disadvantages of these spectrometers are also introduced.

4.2 Dispersive spectrometers

The first dispersive instruments employed prisms made of materials such as sodium chloride. The popularity of prism instruments fell away in the 1960s when the improved technology of grating construction enabled cheap, good quality gratings to be manufactured. Modern spectrometers, such as monochromators, generally use a diffraction grating, a movable slit, and some kind of photodetector, all automated and controlled by a computer. The spectrometer was invented by Joseph von Fraunhofer.

Figure 4-1 shows the optical path of a dispersive spectrometer which uses a grating monochromator [100]. Dispersion occurs when energy falling on the entrance slit is collimated onto the dispersive element and the dispersed radiation is then reflected back to the exit slit, beyond which lies the detector. The dispersed spectrum is scanned across the exit slit by rotating a suitable component within the monochromator. The widths of the entrance and exit slits may be varied and programmed to compensate for any variation of the source energy with wavenumber. In the absence of a sample, the detector then receives radiation of approximately constant energy as the spectrum is scanned. The detector must have adequate

sensitivity to the radiation arriving from the sample and monochromator over the entire spectral region required. In addition, the source must be sufficiently intense over the wavenumber range and transmittance range.

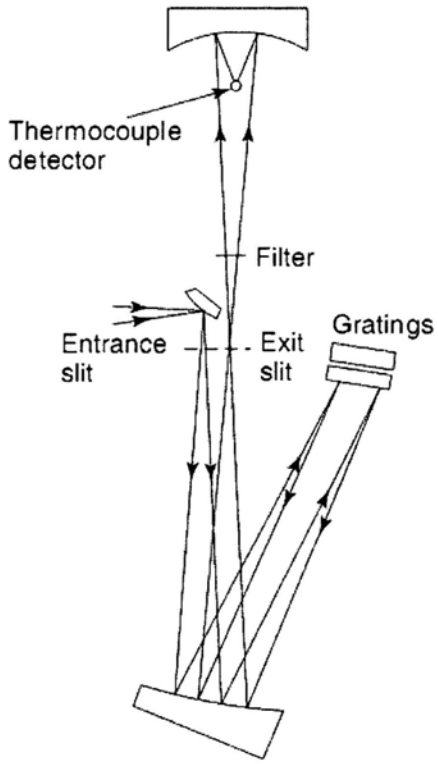


Fig. 4-1 Dispersive spectrometer

The essential problem of the dispersive spectrometer lies with its monochromator. This contains narrow slits at the entrance and exit which limit the wavenumber range of the radiation reaching the detector to one resolution width. Samples for which a very quick measurement is needed, for example, in the eluant from a chromatography column, cannot be studied with instruments of low sensitivity because they cannot scan at speed. However, these limitations may be overcome through the use of a Fourier-transform infrared spectrometer.

4.3 Fourier transform spectrometers

A Fourier transform spectrometer (FTS) is a Michelson interferometer with a movable mirror [101]. By scanning the movable mirror over some distance, an interference pattern is produced that encodes the spectrum of the incoming light. A Fourier transform spectrometer can be used to measure the spectra of a sample based on measurements of the coherence of a radiative source, using time-domain or space-domain measurements of the electromagnetic radiation or other type of radiation. Instead of recording the amount of energy absorbed when the frequency of the light is varied (monochromator), the light is guided through an interferometer. After passing through the sample, the measured signal is the interferogram. Performing a mathematical Fourier transform on this signal results in a spectrum identical to that from dispersive infrared spectroscopy.

Fourier transform spectrometers are cheaper than dispersive spectrometers because building of interferometers is easier than the fabrication of a monochromator. In addition, measurement of a single spectrum is faster for the Fourier transform technique because the information at all frequencies is collected simultaneously. This allows multiple samples to be collected and averaged together resulting in an improvement in sensitivity. Because of its various advantages, virtually all modern spectrometers are Fourier transform instruments. But Fourier transform instruments also suffer the disadvantages big size and low resolution.

4.3.1 Basic theory of Fourier transform spectrometer

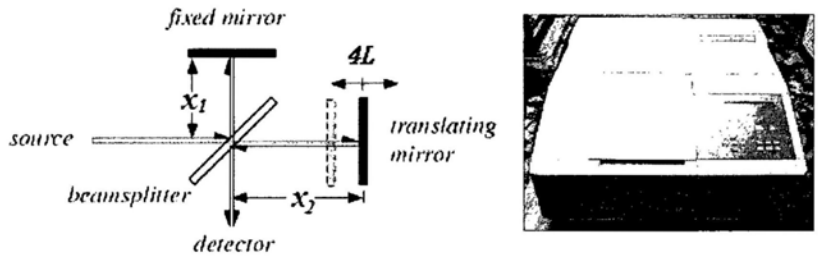


Fig. 4-2 Fourier transform spectrometer

As shown in Fig. 4-2, the heart of any Fourier transform spectrometer is a Michelson interferometer. In its simplest form, a Fourier transform spectrometer consists of two mirrors located at a right angle to each other and oriented perpendicularly, with a beamsplitter placed at the vertex of the right angle and oriented at a 45° angle relative to the two mirrors. Radiation incident on the beamsplitter from one of the two "ports" is then divided into two parts, each of which propagates down one of the two arms and is reflected off one of the mirrors. The two beams are then recombined and transmitted out the other port. When the position of one mirror is continuously varied along the axis of the corresponding arm, an interference pattern is swept out as the two phase-shifted beams interfere with each other.

If we assume the electronic field of the incoming light (single wavelength) is

$$\vec{E}=E \exp(\omega t-2\pi\nu x_0+\varphi_0) \qquad (5.1)$$

After passing through the interferometer, the electronic fields of the two phase-shifted beams are

$$\bar{E}_1=E_1 \exp(\omega t-2\pi\nu x_1+\varphi_0)=rtE\exp(\omega t-2\pi\nu x_1+\varphi_0) \quad (5.2)$$

$$\bar{E}_2=E_2 \exp(\omega t-2\pi\nu x_2+\varphi_0)=trE\exp(\omega t-2\pi\nu x_2+\varphi_0) \quad (5.3)$$

Here x_0, x_1, x_2 is the distance the light passing through, E, E_1, E_2 is the amplitude. t is transmittance coefficient of the beamsplitter, r is the reflection coefficient of the beamsplitter, φ_0 is the initial phase, ν is the wavenumber. Then the intensity measured by the detector is

$$\begin{aligned} I &= \left|\bar{E}_1+\bar{E}_2\right|^2 = (\bar{E}_1+\bar{E}_2)(\bar{E}_1^*+\bar{E}_2^*) \\ &= E_1^2+E_2^2+\bar{E}_1\bar{E}_2^*+\bar{E}_2\bar{E}_1^* \\ &= r^2t^2E^2\{2+\exp[2\pi\nu(x_1-x_2)]+\exp[-2\pi\nu(x_1-x_2)]\} \\ &= 2r^2t^2E^2\{1+\cos[2\pi\nu(x_1-x_2)]\} \\ &= 2r^2t^2E^2\{1+\cos[2\pi\nu x]\} \\ &= B[1+\cos(2\pi\nu x)] \end{aligned} \quad (5.4)$$

Here, we assume $x = x_1-x_2$ and $B=2r^2t^2E^2$. If we neglect DC component and consider the situation of multi-wavelength of the incoming beam, the detected intensity is

$$I(x) = \int_{-\infty}^{\infty} B(\nu)\cos(2\pi\nu x)d\nu \quad (5.5)$$

We can also write it as the following formula

$$I(x) = \int_{-\infty}^{\infty} B(\nu)\exp(2\pi\nu x)d\nu = F\{B(\nu)\} \qquad (5.6)$$

That means the intensity is the Fourier transform of the incoming spectrum. Then the spectrum can be obtained by inverse Fourier transform of the intensity detected by the detector. i.e.

$$B(\nu) = F^{-1}\{I(x)\} = \int_{-\infty}^{\infty} I(x)\exp(-i2\pi\nu x)dx \qquad (5.7)$$

or

$$B(\nu) = \int_{-\infty}^{\infty} I(x)\cos(2\pi\nu x)dx \qquad (5.8)$$

4.3.2 Spectrum resolution of Fourier transform spectrometer

In practice, Eq. (4.8) cannot be realized because there is a maximum value for pathlength difference $x = x_1\text{-}x_2$. We assume the maximum value for x is L. The feasible formula to calculate the spectrum is

$$S(\nu) = \int_{-L}^{L} I_t(x)\cos(2\pi\nu x)dx \qquad (5.9)$$

i.e.

$$S(\nu) = \int_{-\infty}^{\infty} I(x)T(x)\cos(2\pi\nu x)dx \qquad (5.10)$$

Here $T(x)$ is a rectangular function, that is

$$T(x) = rect(\frac{x}{2L}) \qquad (5.11)$$

When $|x| \leqslant L$, $T(x) = 1$; when $|x| > L$, $T(x) = 0$.

Therefore, the spectrum function is not the inverse Fourier transform of $I(x)$, but the convolution of the inverse Fourier transform of $I(x)$ and the inverse Fourier transform of $T(x)$, i.e.

$$S(\nu) = F^{-1}\{I(x)\} * F^{-1}\{T(x)\} = B(\nu) * t(\nu) \qquad (5.12)$$

Here $t(\nu)$ is an instrumental line shape (ILS) function, i.e.

$$t(\nu) = F^{-1}\{rect(\frac{x}{2L})\} = 2L\sin c(2\pi\nu L) \qquad (5.13)$$

Therefore, a monochromatic spectrum for a wavenumber ν_0 can be written as

$$S(\nu_0) = B(\nu_0) * 2L\sin c(2\pi\nu L) = 2B(\nu_0)\cdot\sin c[2\pi(\nu-\nu_0)L] \qquad (5.14)$$

Eq. (4.14) indicates the monochromatic spectrum is not a pulse function (δ function), but a sinc function with a center of ν_0. According to Rayleigh criterion, the images of two different points are regarded as just resolved when the principal maximum of one image coincides with the first minimum of the other. The first minimum point of sinc function is at $1/2L$ if the center is at 0. Therefore, the wavenumber resolution is

$$\delta\nu = \frac{1}{2L} \qquad (5.15)$$

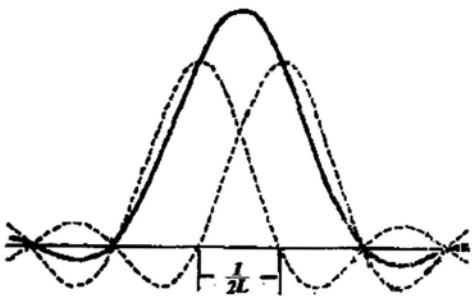


Fig. 4-3 Diagram of the spectrum resolution

Eq. (4.15) indicates that the resolution for a Fourier transform spectrometer is limited by the maximum path difference between the two beams. The limiting resolution in wavenumbers (cm^{-1}) is the reciprocal of two times of the pathlength difference (cm). For example, a pathlength difference of 10 cm is required to achieve a limiting resolution of 0.05 cm^{-1}. This simple calculation appears to show that it is easy to achieve high resolution. Unfortunately, this is not the case since the

precision of the optics and mirror movement mechanism become more difficult to achieve at longer displacements of pathlengths.

4.3.3 Sampling theory of Fourier transform spectrometer

We have to point out that we only can take the Fourier transform for some discrete points if we calculate the spectrum by a computer. Therefore, we need to sample the interference pattern. We assume the sampling interval is Δx, then the intensity pattern after the sampling should be

$$I_s(x) = I_t(x) \cdot \sum_{n=-\infty}^{\infty} \delta(x - n \cdot \Delta x) = I(x) \cdot T(x) \cdot comb(\frac{x}{\Delta x}) \qquad (5.16)$$

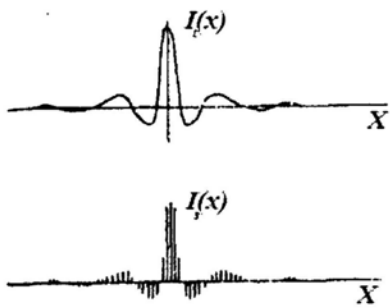


Fig. 4-4 Sampling of the interference pattern

Here $I_t(x)$ is the truncated intensity pattern. A comb function is used, which is a periodic Schwartz distribution constructed from Dirac delta functions. The definition of comb function is

$$\text{comb(x)}=\sum_{n=-\infty}^{\infty} \delta(\text{x-n}) \qquad (5.17)$$

According to the properties of comb function, the inverse Fourier transform of a comb function is still a comb function, i.e.

$$\text{F}^{-1}\{\frac{1}{\Delta x} comb(\frac{x}{\Delta x})\}=\text{comb}(\frac{\nu}{1/\Delta x}) \qquad (5.18)$$

Considering the inverse Fourier transform of Eq. (4.16), the sampling interval of the wavenumber spectrum is

$$\Delta\nu = \frac{1}{\Delta x} \qquad (5.19)$$

According to Nyquist–Shannon sampling theorem [102],

$$\Delta\nu \geq 2\nu_{max} \qquad (5.20)$$

Here ν_{max} is the largest wavenumber component of the incoming beam. From Eq. (4.19) and Eq. (4.20), we obtain the range of the sampling interval, i.e.

$$\Delta x \leq \frac{1}{2\nu_{max}} \qquad (5.21)$$

Eq. (4.21) indicates that as long as the sampling interval is less than $1/(2\ \nu_{max})$, the original spectrum can be reconstructed successfully. However, it is not necessary to make Δx far less than $1/(2\ \nu_{max})$ because it will lead to large calculation burden and will not generate more spectrum information.

Eq. (4.21) also indicates the Fourier transform spectrometer has a narrow detection range. Because the sampling interval should be less than half of the minimum wavelength of the source spectrum, the ordinary Fourier transform spectrometers can only be used to detect infrared beams due to their long wavelengths.

By the way, if the minimum wavenumber of the original spectrum ν_{min} is not 0, Eq. (4.21) can be changed into

$$\Delta x \leq \frac{1}{2(\nu_{max} - \nu_{min})} \qquad (5.22)$$

Then the sampling interval can be larger than that determined by Eq. (4.21).

4.3.4 Advantages of Fourier transform instruments

The Fourier transform instruments have several significant advantages over older dispersive instruments. Two of these are the Fellgett (or multiplex) advantage and the Jacquinot (or throughput) advantage. The Fellgett advantage is due to an improvement in the SNR per unit time or high speed. The time needed by the movable mirror for one scan cycle varies between 0.01 and 1 s, depending on the spectral resolution as well as the detector response. Typically, 20~200 scans are accumulated in one measurement to acquire a sufficient SNR. The number of accumulations depends on the experimental conditions and can be much higher if few spectral effects have to be studied. In addition, because Fourier transform spectrometry does not require the use of a slit or other restricting device, the total source output can be passed through the sample continuously. This results in a substantial gain in energy at the detector, hence translating to higher signals and improved SNRs. This is known as Jacquinot advantage.

4.4 Self-mixing interferometers

Although Fourier transform spectrometers have the Fellgett advantage and the Jacquinot advantage, they are quite bulky and sensitive to vibration due to use of Michelson interferometer. In order to remove big components such as beamsplitters and mirrors in spectrometers, we need to explore a compact interferometer instead

of the Michelson interferometer. Recently, a novel self-mixing interferometer has been proposed [103], which is the element of our new spectrometer.

The proposed self-mixing interferometer is simple and compact. It generates two beams by wavefront division. The beams are narrow and proximate to each other. Interference happens when the beams spread and merge by themselves within a short distance. Furthermore, it can be fabricated by the current semiconductor technology. The proposed interferometer is similar to the Rayleigh interferometer in that a beam is divided into side-by-side beams by wavefront division. It also likes the design of Young's experiment because interference occurs when beams mix together by themselves as they spread as a result of diffraction.

4.4.1 Principles of the self-mixing interferometer

Fig. 4-5 shows schematically Young's experiment, where a beam enters a slit S first and then reaches slits S1 and S2 respectively. Resulting beams from S1 and S2 spread while propagating and merge to cause interference in a shaded area in Fig. 4-5 [103]. The interference generates fringes on an observation screen. Pattern of the fringes depend upon spacing between and width of the slits and distance between the slits and screen. In a typical configuration, slit spacing is designed much larger than the wavelength of the beam and the screen is placed far away from the slits.

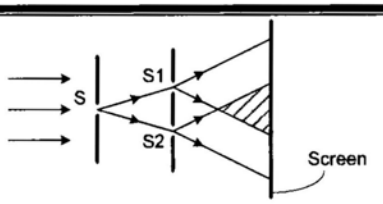


Fig. 4-5 Setup of classic Young's experiment

In order to create interference using fewer components and occupying less space, setup of the Young's experiment is modified as described in Fig. 4-6 [103]. First, width of slits is chosen to be similar to or smaller than the wavelength, such that beams spread due to diffraction at a larger angle. Second, spacing between the slits is reduced to make overlapping of the diffracted beams happen at a substantially short distance. Finally, a small screen is positioned near the slits.

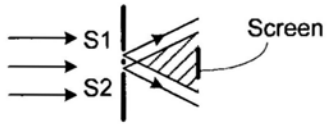


Fig. 4-6 Modified setup of Young's experiment

The configuration in Fig. 4-6 merely reduces overall dimensions, while contains similar quantity of components compared to the Young's experiment. But the new scheme can be utilized to create a self-mixing free space optical interferometer as shown in Fig. 4-7 [103].

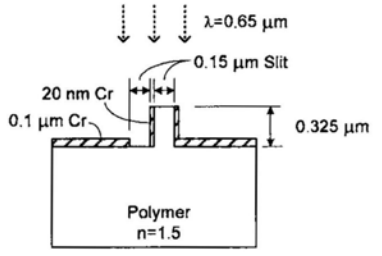


Fig. 4-7 Cross-sectional view of a self-mixing interferometer

Fig. 4-7 depicts graphically a cross-sectional view of an example of the self-mixing interferometer. A beam of wavelength λ = 0.65 μm is split into two

beams by wavefront division by two 0.15 μm wide slits. One slit is of a polymer stage protruded on a polymer surface. The stage is surrounded by 20 nm Cr films. The other slit is defined by one 20 nm Cr film and a 0.1 μm Cr film deposited on the polymer surface.

After passing through the slits, the beams are spaced apart by 20 nm and start spreading due to diffraction. Since the beams are smaller than the wavelength and proximate, they merge substantially, or interfere with each other, within a short distance. Simulation results confirm the phenomenon, which are shown in the next section.

Assume the stage height is 0.325 mm and refractive indexes of the polymer and the ambience are 1.5 and 1 respectively. Then, after passing through the slits, the beams have a phase difference of $\pi/2$.

The submicron self-mixing interferometer has a simple and compact structure. Since mirror and beam splitter are no longer needed, it has fewer components than the Michelson, Mach-Zehnder, or Rayleigh interferometer. The stage and substrate can also be made up of other materials such as silicon dioxide or silicon nitride. As a result of the uncomplicated structure and simple requirement, the interferometer can be built at a low cost using conventional semiconductor technology.

4.4.2 Simulation results of the self-mixing interferometer

To investigate the interference of the self-mixing interferometer, a series of simulation has been done by using FDTD Solutions [103]. Shown in Fig. 4-8 is a

setup of two-wave self-mixing interferometer in cross-sectional view. Two TE polarized sources with wavelength 0.65 μm are emitting along the x-axis direction, which are 0.14μm wide, and fed to the slits separately. The sources are arranged coherent to each other with the same power. The slits are 0.15 μm wide, 0.325 μm long, and made by 20 nm thick Cr films. A virtual power monitor measures energy flow in x-direction and is placed 1 μm away from the slits. Assume the refractive index is 1 everywhere except the Cr films. Thus, phase difference between the beams is just that between the sources.

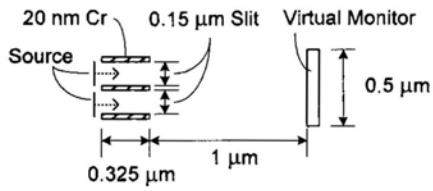


Fig. 4-8 Cross-sectional view of a self-mixing interferometer

FDTD simulation results of field intensity distribution of E_y are shown in Fig. 4-9 for three cases: in phase, out of phase (the lower source delayed by π), and the lower source delayed by□π/2 [103]. When the beams are in phase, as expected, the resulting beam is a single one propagating along *X*-direction. But unlike the classical two-wave interference, the beams don't cancel each other completely when they are out of phase. Instead of one beam, two weak beams are generated heading toward different directions. The central region experiences more extensive interference than the upper and lower part as it registers much lower field intensity of E_y, which is the reason why the virtual power monitor is placed there. Finally, when the phase difference is π/2, the resulting beam is single but "bent" downward.

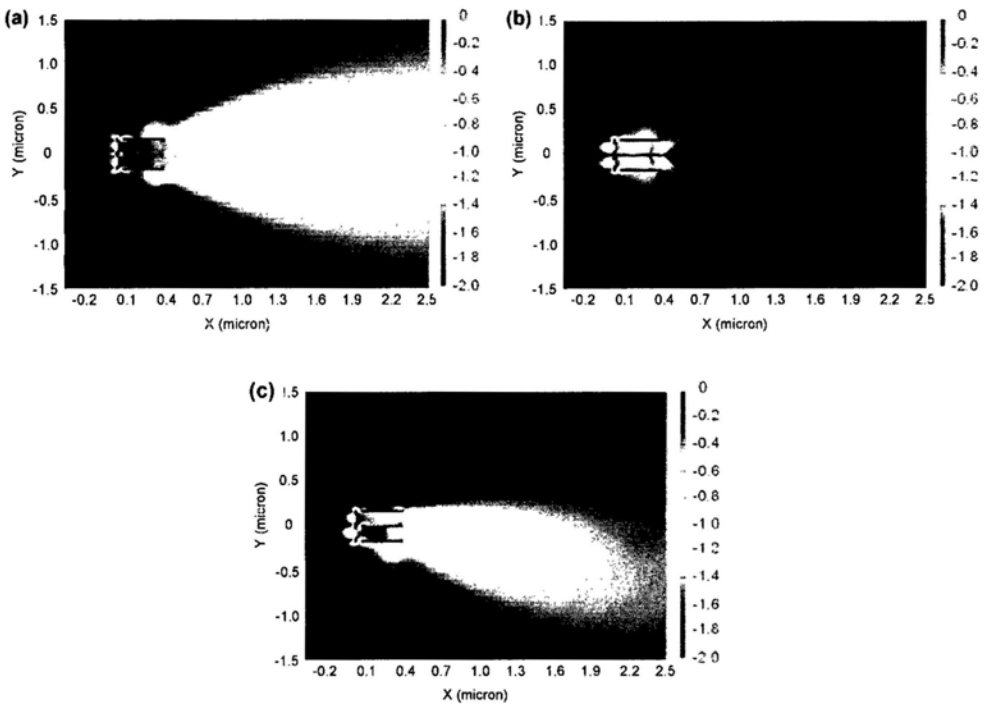


Fig. 4-9 Field intensity distribution of E_y in log scale

(a) Sources are in phase. (b) and (c) The lower source is phase delayed relative to the upper source by π and $\pi/2$.

Fig. 4-10 shows comparison between the self-mixing interferometer and a classical ideal one [103]. The black curve represents measurements from the power monitor after normalization, while the red curve follows the ideal Cosine function.

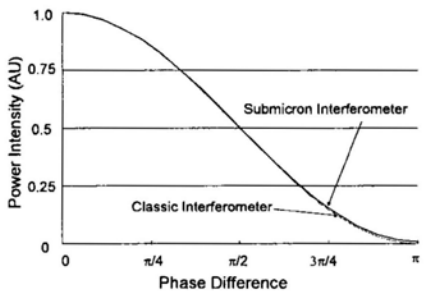


Fig. 4-10 Phase difference dependence of the interference intensity for the self-mixing interferometer (black line) and a classic ideal interferometer (red line)

4.4.3 Correlation induced spectral changes

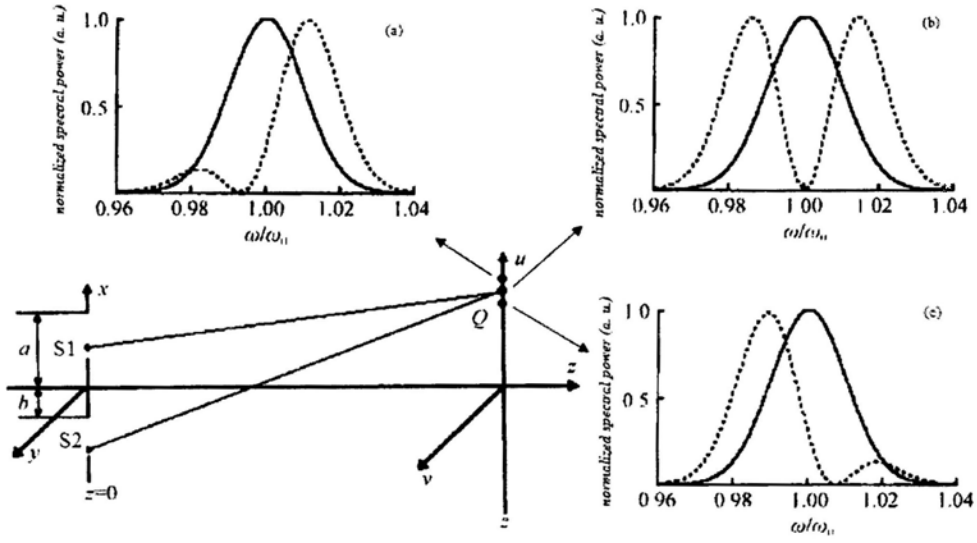


Fig. 4-11 The normalized spectra in the interference field

(The solid lines are the original spectrum and the dash lines are measured spectrum)

(a) u_N=10.6 (b) u_N=10.5253 (c) u_N=10.45

Spectral anomalies in Young's double slit interference experiment illuminated by completely coherent beam have been studied by Wolf and other researchers [104, 105]. As shown in Fig. 4-11 [105], assume the external and inner radius of the slit are a and b respectively. Then the central obstruction ratio is

$$\varepsilon = \frac{b}{a} \qquad (5.23)$$

Here $0 \leqslant \varepsilon < 1$. Assume the spectra of the incident light represent a Gaussian distribution with center angular frequency ω_0, i.e.

$$S^{(0)}(\omega) = S_0 \exp[-\frac{(\omega-\omega_0)^2}{2\Gamma_0^2}] \qquad (5.24)$$

Here S_0 is a constant; Γ_0 is the peak width of the spectra. Then the spectra in the uv plane can be expressed as [105]

$$S(u_N, z_N, \omega) = \frac{S^{(0)}(\omega)(\frac{\omega}{\omega_0})}{z_N}\left|\int_{\varepsilon}^{1}\exp[\frac{i[\pi(\frac{\omega}{\omega_0})]}{z_N}(x-u_N)^2]dx + \int_{-1}^{-\varepsilon}\exp[\frac{i[\pi(\frac{\omega}{\omega_0})]}{z_N}(x-u_N)^2]dx\right|^2 \quad (5.25)$$

Here $z_N = z/(a^2/\lambda_0), u_N = u/a$. From this equation, we realize that the spectra are related to the position of the monitor, the central obstruction ratio and the peak width.

Assume the parameters of the incident light are: the central angular frequency ω_0=3×10^{15}s^{-1}, peak width Γ_0=0.01 ω_0, z_N=8, the central obstruction ratio ε=0.9. Q is a point where the central angular wavelength is just out of phase. At this point, the power of the center wavelength is just 0 as shown in Fig. 4-11(b). Let's study the points just next to the point Q. For a point above the point Q, where u_N is increase and the phase delay is also increased, the corresponding spectrum is shown in Fig. 4-11(a). We can see the spectrum is blue-shift. Fig. 4-11(c) shows that the spectrum below the point Q, where phase delay is decreased and the spectrum is red-shift. Therefore, the phase delay has something to do with the spectrum. Some people use this property to design optical spectrum switches. But this is not the only application. Spectrometer design is also possible. As shown in Figs. 4-11, for different phase delay, the transmission coefficients of each frequency are different. It indicates that we can use phase delay to reconstruct the original spectrum. Considering the setup of the Young's double slit interference experiment is quite bulky and not easy to control, we use the above mentioned self-mixing interferometer to design a novel spectrometer. The self-mixing interferometer is different from Young's setup with ε=0.

4.5 Self-mixing interferometer spectrometers

In order to overcome the limitations of traditional devices and explore a low cost, small size and high performance spectrometer, we propose a novel interferometer-based spectrometer. The reconstruction process of the new method is just solving a linear system. The miniature spectrometer has the advantages of low cost, small size and high resolution simultaneously. Its polymer structure can be built on a chip using mature fabrication technologies such as moulding with the help of focused ion beam or electron beam lithography, so the device is easy to be fabricated as well.

4.5.1 Principles of self-mixing interferometer spectrometers

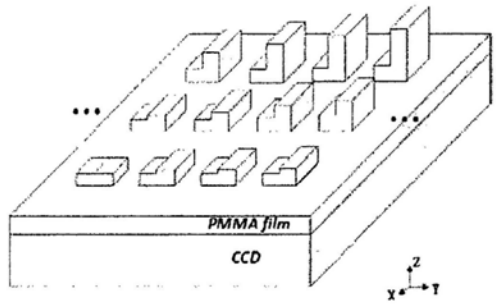


Fig. 4-12 Schematic of the optical spectrometer

As shown in Fig. 4-12, the design contains an interferometer array attached on a CCD chip. Each interferometer has two PMMA steps with a PMMA film as the substrate. The CCD pixel works as a detector beneath the PMMA substrate for each interferometer. Considering the signal to noise ratio (SNR) and the sensitivity, we use part of a CCD pixel as the detector for an interferometer by shading the other part.

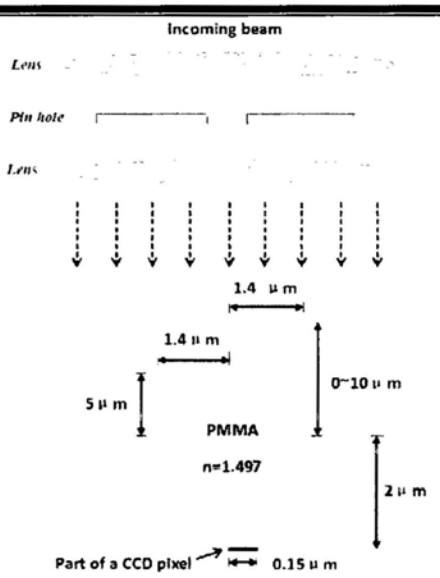


Fig. 4-13 Cross section of the optical interferometer

We enlarge the cross section of an interferometer in Fig. 4-13 to demonstrate the principles of the device. When a plane wave illuminates the surfaces of the interferometer, the incoming beam is divided into two parts by the two steps of the interferometer. Because the height of each step is different, the phase of each beam portion is separately delayed. When the two beam portions with different phase change merge again, interference occurs and the interference intensity is measured by a CCD pixel below.

The detected intensity contains the spectral information. Since each frequency component in the incoming beam corresponds to a unique phase difference of the two beam portions, the total intensity received by each CCD pixel, which is resulted from the addition of the interference signals from all the frequency components in the beam, should also be unique.

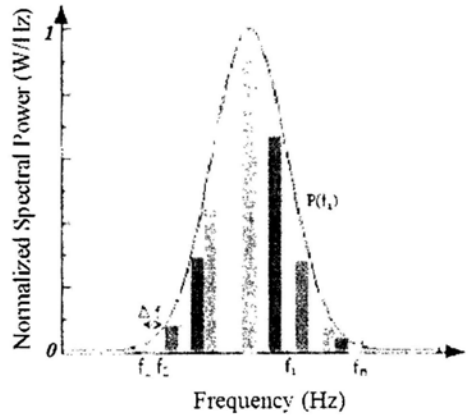


Fig. 4-14 Source spectrum used in simulation

Therefore, if the incoming beam is uniformly divided into n frequency components $f_1, f_2, \cdots f_n$ with width Δf as shown in Fig. 4-14, the total power of the incoming beam can be approximately calculated by using the theory of integral calculus when n is large enough, i.e.

$$P_0 = P(f_1)\,\Delta f + P(f_2)\,\Delta f + \ldots + P(f_n)\,\Delta f, \tag{5.26}$$

where $P(f_x)$ is the spectrum amplitude of f_x. After passing through the interferometer, the measured power can be represented by

$$P = C_1P(f_1)\,\Delta f + C_2P(f_2)\,\Delta f + \ldots + C_nP(f_n)\,\Delta f, \tag{5.27}$$

where $C_1, C_2, \cdots C_n$ are transmission coefficients of frequency parts f_1, f_2, $\cdots f_n$ respectively.

If the incoming beam illuminates n interferometers, a series of powers can be measured by the CCD array as follows

$$P_1 = C_{11}P(f_1)\,\Delta f + C_{12}P(f_2)\,\Delta f + \ldots + C_{1n}P(f_n)\,\Delta f, \tag{5.28}$$

$$P_2 = C_{21}P(f_1)\,\Delta f + C_{22}P(f_2)\,\Delta f + \ldots + C_{2n}P(f_n)\,\Delta f, \tag{5.29}$$

.....

$$P_n = C_{n1}P(f_1)\,\Delta f + C_{n2}P(f_2)\,\Delta f + \ldots + C_{nn}P(f_n)\,\Delta f, \qquad (5.30)$$

Therefore, once we know the transmission coefficients and the powers gotten from different CCD pixels, i.e.

$$C = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{pmatrix} \quad \text{and} \quad y = \begin{pmatrix} P_1 \\ P_2 \\ \vdots \\ P_n \end{pmatrix}, \qquad (5.31)\text{-}(5.32)$$

we can get a linear system

$$Cx = y \qquad (5.33)$$

where

$$x = \begin{pmatrix} P(f_1)\cdot\Delta f \\ P(f_2)\cdot\Delta f \\ \vdots \\ P(f_n)\cdot\Delta f \end{pmatrix}. \qquad (5.34)$$

Consequently, the spectrum of incoming beam can be obtained by fitting $P(f_1)$, $P(f_2)$, ... $P(f_n)$, which are the elements of the matrix

$$\tilde{x} = x/\Delta f = \begin{pmatrix} P(f_1) \\ P(f_2) \\ \vdots \\ P(f_n) \end{pmatrix}. \qquad (5.35)$$

Because the transmission coefficients can be calculated in advance by simulation or measurement, so the spectrum reconstruction is just a problem to solve Eq. (4.33). However, the data error of matrix A due to the limited signal to noise ratio makes this linear system become an ill-posed problem. It is difficult to solve such a large system of linear equations by using ordinary nonstationary iterative methods within the MATLAB environment due to failure in stability. The

unavoidable measurement noises and the round-off errors in the computing process may be amplified arbitrarily and therefore make the numerical solution completely useless. To make the numerical solution feasible, we use regularization method to solve Eq. (4.33). The main objective of regularization is to incorporate more information about the desired solution in order to stabilize the problem and find a useful and stable solution.

4.5.2 Reconstruction methods

Tikhonov regularization [106-112] is the most commonly used method of regularization of ill-posed problems named after Andrei Nikolaevich Tikhonov. In statistics, the method is also known as ridge regression. It is related to the Levenberg-Marquardt algorithm for non-linear least-squares problems. To obtain regularized solution to Eq. (4-33), we choose x to fit data y in least-squares sense, but penalize solutions of Euclidean norm. Solve minimization problem

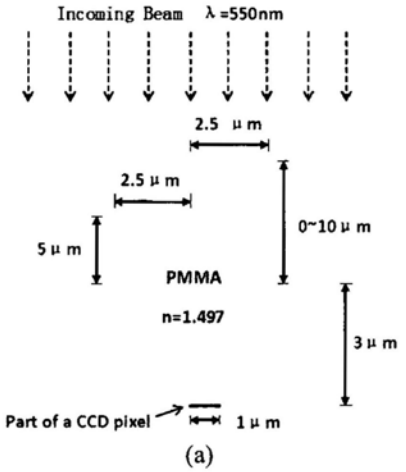
$$\begin{aligned} x_\alpha &= \arg\min_{x\in X} \|Cx-y\|_Y^2 + \alpha\|x\|_X^2 \\ &= (C^*C+\alpha I)^{-1}C^*y \end{aligned} \tag{5.36}$$

Here, x_α is the proximate solution of Eq. (4-33) by using Tikhonov regularization method. $\alpha>0$ is called the regularization parameter [113-116], which can be chosen by the linear model function method [117]. C^* is the adjoint matrix of C and I is the identity matrix. Therefore, the Tikhonov regularization is a method in which the regularized solution is sought as a minimizer of a weighted combination of the residual norm and a side constraint. The regularization

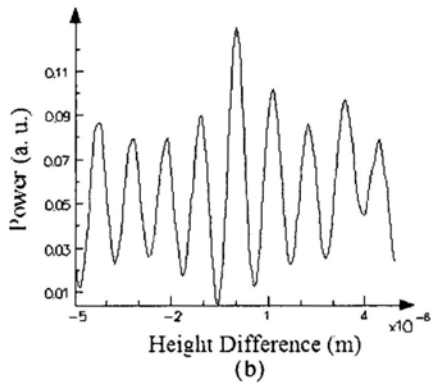
parameter controls the weight given to the minimization of the side constraint, which is determined by the data errors in C and y.

The reconstruction normally takes about several seconds in terms of the power of the computer, so it enables real-time measurement for many applications. Although fast Fourier transform (FFT) would cost less time, its conditions are difficult to satisfy because beam portions coming from the two steps of an interferometer are not uniform in intensity due to the confinement and absorption of the waveguide. At the same time, partial interference of the two beam portions due to the structure layout also make FFT infeasible.

We prove it by simulation experiments for the structures in Fig. 4-15(a) illuminated by a 550nm single wavelength source. As shown in Fig. 4-15(b), we can see the powers detected by the detectors for a series of interferometers with different optical length difference are not presented as a perfect Sine function. While perfect Sine function can be realized for a Fourier spectrometer. This difference makes FFT unpractical.



(a)



(b)

Fig. 4-15 Nonuniformity of maximum power for a single wavelength illumination with difference phase delay

Although FFT cannot be used, that is not a bad thing after all. Compared with Fourier transform spectrometer, the interferometer array spectrometer can realize high resolution. For Fourier transform spectrometer, the spectrum is given by

$$\begin{aligned} B(\nu)&=F^{-1}\{I(x)\} * F^{-1}\left\{rect(\frac{x}{2L})\right\} \\ &= B_0(\nu) * 2L\sin c(2\pi\nu L) \end{aligned} \qquad (5.37)$$

where ν is the wave number of the original spectrum, x is the distance of the moving mirror from the central position, $I(x)$ is the intensity measured at different position of the moving mirror, L is the maximum distance from the central position, $B(\nu)$ is the inverse Fourier transform of the intensity. Therefore, the wave number resolution of the Fourier transform spectrometer is limited by the Rayleigh criterion with a value of $1/2L$. However, for the interferometer array spectrometer, Fourier transform is not used at all, so high resolution can be realized with compact size.

4.5.3 Uniformity of the incoming light

In the previous analysis, we assume that the intensity of the incoming beam is uniform in XY plane. In reality, however, it is difficult to make the intensity absolutely uniform. Although there are many methods introduced in the book "laser beam shaping application" such as using kaleidoscopes, fly's eyes, phase modulation or other optical devices as shown in Fig. 4-16. [118], these methods need complex optical design or special optical design program such as ZEMAX. Furthermore, these uniformizers are only useful to a special source shape and

cannot be used to other kinds of sources. Therefore, when it is not necessary to realize a precise detection, we do not need to use these methods. In this thesis, we propose three choices for reference.

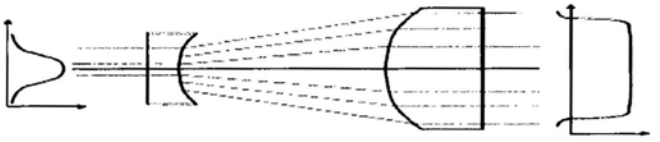


Fig. 4-16 Two lens system for large, uniform, collimated beam

First choice is to extend the original beam and only detect a small area in the center (Fig. 4-17) if the intensity of the original beam is large enough. However, the diffractions of the lens may change the spectrum of the original beam. Therefore, when we calculate (normalize) the transmission coefficients, we need to take the incoming beam from the entrance of the entire system or the emitter of original beam instead of the entrance of the spectrometer.

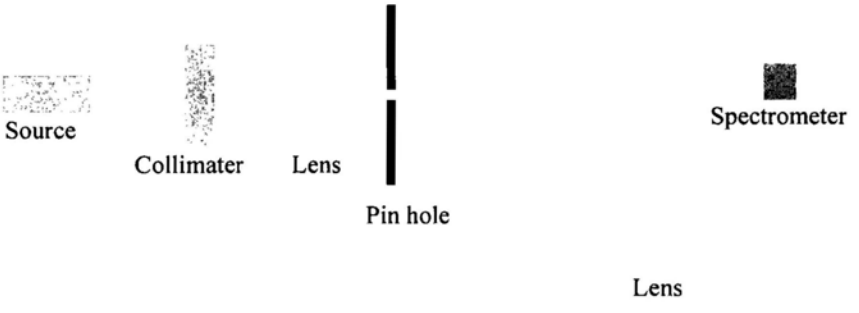


Fig. 4-17 The first design for uniform illumination

Second choice is using a waveguide with many scattering points in it as shown in Fig. 4-18. When the incoming lights are diffused by the scattering points, the

emitted scattered light may become somewhat uniform. However, the scattering points themselves may lead to the un-uniformity of the incoming lights.

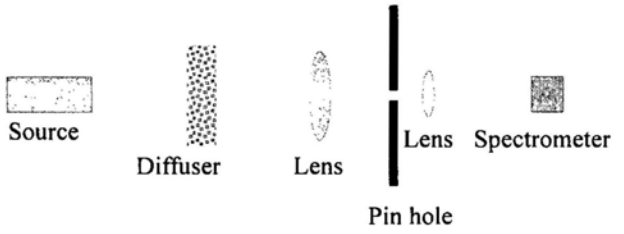


Fig. 4-18 The second design for uniform illumination

Third choice is inducing the incoming lights to a side-scattering waveguide (or fiber) [119-122]. When the lights propagate in the side-scattering waveguide, the light will be scattered by the scattering points and emitted from the side of the waveguide. If we design the scattering coefficients along the waveguide direction by changing the size or the density of the scattering points, we can get a uniform lights from the side of the waveguide. An equation for designing the scattering coefficients has been deduced, i.e.

$$k = \frac{k_0}{1 - k_0 z} \qquad (5.38)$$

Here we assume the lights propagate in the direction Z; k is the scattering coefficient at a position z; k_0 is scattering coefficient at z=0; z is the propagation length. The deduction process has been explained in detail in the reference [121]. The CCD we used here is one-dimensional array as shown in Fig. 4-19. The merit of this design is that it can also be used to detect the spectrum of waveguide lights. However, we need to make sure that the wavelengths of incoming lights must be larger than the cut-off wavelength of the side-scattering waveguide in case we lost

some frequencies during the process.

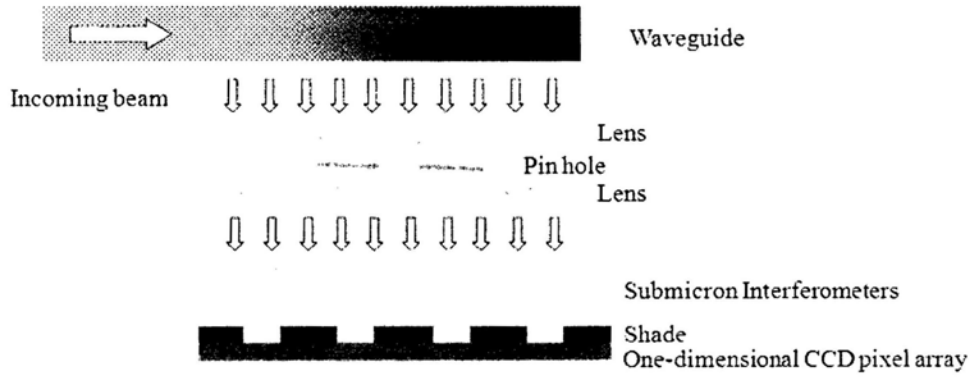


Fig. 4-19 The third design for uniform illumination

4.5.4 Correction method for the nonuniformity of CCD pixel response

Dark current and other factors may lead to the nonuniformity of CCD pixel response for same illumination intensity [123]. In order to remove the CCD pixel nonuniformity, a method based on least square method is introduced here. Firstly, we analyze the parameters related to the pixel nonuniformity

The relation between the charge obtained from each pixel and the illuminated power can be expressed as the following equation

$$Q = \eta \int_0^t P dt + \beta \qquad (5.39)$$

Here η is the conversion efficiency from photon to electron; t is the detection time. β is a parameter of the detection error due to dark current. Assume that the illumination power P does not change with time. The charge obtained from the i th pixel can be expressed as

$$Q_i = \alpha_i P + \beta_i \qquad (5.40)$$

Here α is a combined parameter of η and t. The subscript n means the parameter is for the ith pixel. Q_i can be measured by the CCD pixel. We assume that J_i is a value measured by a power meter. The detection error is

$$\omega_i = Q_i - \alpha_i J_i - \beta_i \qquad (5.41)$$

Corresponding to N input powers of the ith pixel, we can obtain N pixel gray levels. We can get the following matrix equation:

$$\omega = Q - MA \qquad (5.42)$$

Here $\omega = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_N \end{pmatrix}$, $Q = \begin{pmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_N \end{pmatrix}$, $M = \begin{pmatrix} J_1 & 1 \\ J_2 & 1 \\ \vdots & \\ J_N & 1 \end{pmatrix}$, $A = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}$, According to least square method,

$$J = \sum_{j=1}^{N} \omega_j^2 = (Q - MA)^T (Q - MA) \qquad (5.43)$$

$$\frac{\partial J}{\partial A} = 0 \qquad (5.44)$$

Here the superscript T means transpose of the matrix. Therefore, we can get the values of α and β by the equation:

$$A = (M^T M)^{-1} M^T Q \qquad (5.45)$$

4.5.5 Fabrication and measurement error

When we fabricate the device, it is possible that the step height of each interferometer is not what we design for it. It is also possible that the surface of the steps is rough. However, theoretical principles indicate that we can still reconstruct the original spectrum with such faulty devices. This is because we can calibrate the

transmission coefficients after we fabricate the device. As long as we can solve Eq. 4.33, the original spectrum can be successfully reconstructed. However, there is a limit for the roughness. If the roughness is too large, CCD would be not sensitive enough to separate the difference of two interferometers (roughness render the difference of the two interferometers is too small to be distinguished so that the transmission coefficients of the two interferometers are nearly same after calibration), which maybe leads to obvious distortions because we cannot solve a linear system with more unknown values but fewer equations. The problem can be solved by three methods: Firstly, we can increase the SNR of the CCD by using high sensitive device or design suitable sizes for the spectrometer components. Secondly, we can calculate a suitable regularization parameter α by using the method introduced in Ref. [117] to solve such an ill-posed problem due to the data errors in the transmission coefficient matrix and detected power matrix in Eq. 4.31 and Eq. 4.32. The roughness in fabrication and the noise in the CCD are the reason for the data errors. Regularization method can somewhat smooth over the distortions in the reconstructed spectrum if a suitable regularization parameter is used. Third, we can reduce the number of equations in Eq. 4.33 and choose some useful values detected by the CCD to calculate. That means the effective N number would be reduced. Although the resolution is reduced accordingly, it is also possible to reconstruct the original spectrum with small N value. The aim of the three methods is to solve the linear system. The first method is to increase the difference of the equation in the linear system. The second method is to solve the ill-posed

linear system. The third method is to reduce the unknown values in the linear system. Therefore, we can recontructe the original spectrum as long as we can solve the linear system. The fabrication errors only add some difficulty in solving the linear system.

4.5.6 Possible applications

There are some possible applications for the high resolution self-mixing interferometer spectrometer.

First, it can be used to determine a very narrow spectrum of a source or separate two very close spectral lines of a source.

Second, if the spectrum of a source is shifted with a change of temperature, we can use the spectrometer to detect the small change. What we need to do is reconstructing the spectrum before and after the spectral shift, then comparing the difference of the spectrums obtained from the two reconstructions.

Third, it can be used for optical communication system. Because modern optical communication system is based on dense wavelength division multiplexing (WEDM), we need to adopt more wavelengths in a narrow bandwidth in the future in order to increase the information channel. The spectrometer can be used to measure the signal with high wavelength division.

Fourth, because the spectrometer is quite compact, it is ideal for miniaturization and integration as the system in microfluidics architectures and lab-on-chip designs.

Last but not least, the spectrometer can realize real time measurement in

spectroscopic analysis to identify materials considering its small size, which is particularly useful for outdoor measurement and detection.

4.6 Summary

We present the structure and principles of an ultra compact and high resolution optical spectrometer and compare the advantages and disadvantages with dispersive spectrometers and Fourier transform spectrometers. The miniature interferometer based spectrometer is a series of submicron phase objects on a polymethyl methacrylate (PMMA) film with a CCD as the detector. The spectrum is obtained by solving a system of simultaneous linear equations. The Tikhonov regularization method is used to achieve high resolution. The methods for the uniformity of incoming beam intensity and CCD pixel response are discussed respectively. Compared with conventional spectrometers, the proposed device is low-cost and easy to fabricate for its simple structure. Furthermore, its compact feature renders the device ideal for miniaturization and integration as the system in microfluidics architectures and lab-on-chip designs.

Chapter 5: Simulation Experiments for Step Interferometer Based Spectrometers

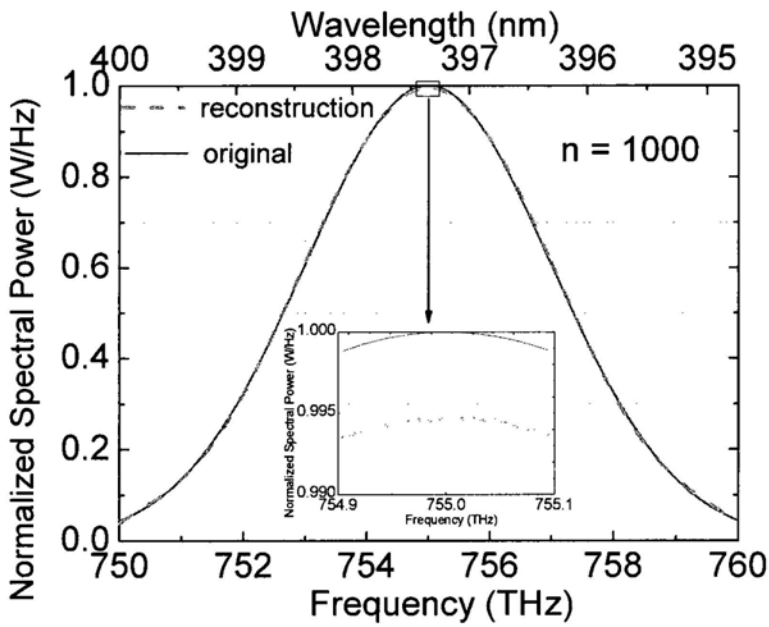
5.1 Introduction

The optical spectral analyzer is an indispensable tool for performing photonic measurements as many physical, chemical or biological phenomena may lead to some shifting of wavelength-dependent properties one way or the other. For example, absorption spectroscopy is an established technique for analyzing chemical and biological samples in a wide range of industrial and research applications [124-126]. Among all currently available approaches, Fourier transform spectroscopy is well-known for its high performance in terms of spectral discrimination and signal sensitivity [127]. However, current Fourier transform spectrometers, particularly those using the scanning mirror mechanisms, are quite bulky. Because such a compact and real-time operating analyzer could be used for monitoring the quality of gasoline at gas stations, the quality and consistency of products (e.g. food and drug industry), the safety in fermentation environment (CO_2), and countless other out-of-the lab applications, grating based spectrometers have more commercial market quota due to their small size [128]. But grating based spectrometers also suffer disadvantages of low spectral resolution and expensive price.

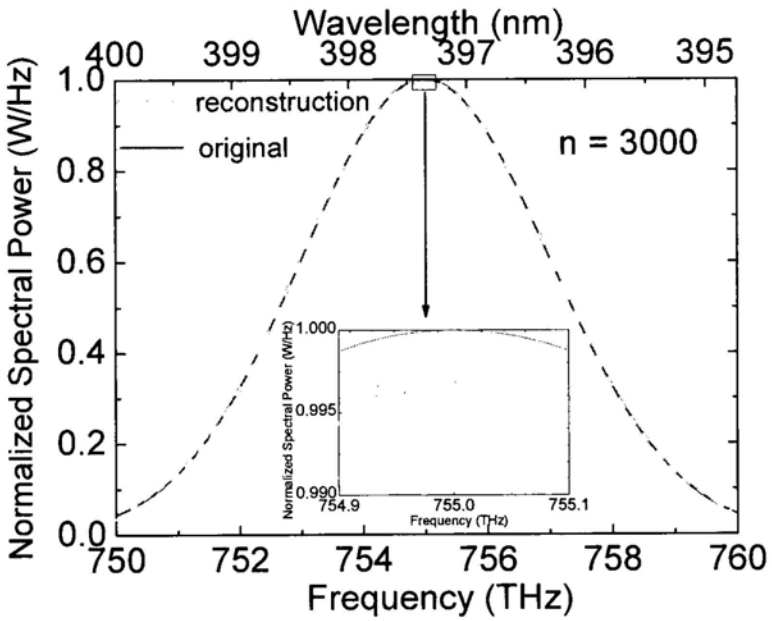
A chip-level integrated optical spectrometer and its broadband and high resolution performance are presented in the chapter by using FDTD simulations. In

this miniature interferometer-based spectrometer, a 2-dimensional array of micron-scale phase objects made from PMMA is mounted on a CCD imaging device. Our analysis shows that the 2-dimensional intensity distribution obtained by the CCD device can be described by a system of simultaneous linear equations when the phase delay within each interferometer element is pre-designed. One can readily recover the spectral contents of the input radiation with picometer resolution. The simulation results in this chapter prove the feasibility of producing miniature, low-cost, broadband and high performance optical spectrometers, which may find applications in a wide range of applications.

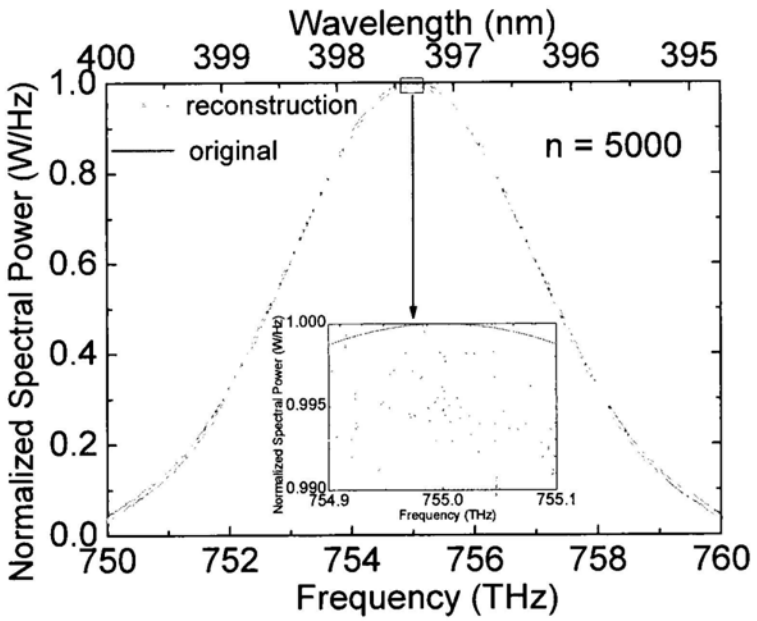
5.2 High resolution of self-mixing interferometer



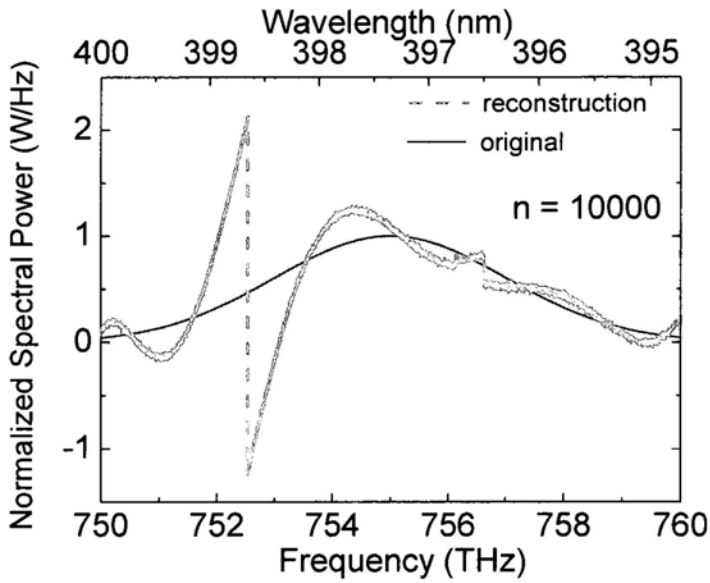
(a)


Wavelength (nm)
400 399 398 397 396 395
reconstruction
original
n = 3000
Normalized Spectral Power (W/Hz)
1.0 0.8 0.6 0.4 0.2 0.0
750 752 754 756 758 760
Frequency (THz)
1.000 0.995 0.990
754.9 755.0 755.1
Frequency (THz)

(b)


Wavelength (nm)
400 399 398 397 396 395
reconstruction
original
n = 5000
Normalized Spectral Power (W/Hz)
1.0 0.8 0.6 0.4 0.2 0.0
750 752 754 756 758 760
Frequency (THz)
1.000 0.995 0.990
754.9 755.0 755.1
Frequency (THz)

(c)



(d)

Fig. 5-1 Reconstruction in the ultra violet spectra range with different n

The spectrometer is anticipated to realize high resolution. We have performed a series of 2D simulation experiments by varying the height of one step from 0 to 10 μ m as demonstrated in Fig. 4-13. FDTD Solutions (Lumerical Solutions, Inc) with minimum 8nm mesh size is used to study the interferometer structure. In order to smooth out edge effects [129], we use a plane wave source acts as total- field scattered-field (TFSF) source with perfectly matched layer (PML) boundaries around. The values in matrix C and y can be obtained from a frequency domain power monitor by a series simulation experiments and the initial data can be analyzed by MATLAB.

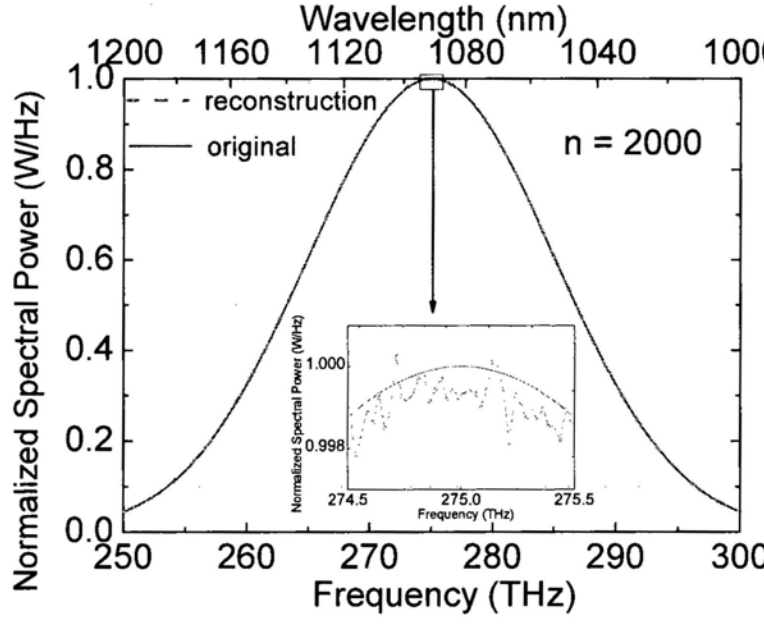
Fig. 5-1 shows the final results on the comparisons between the reconstructed and original spectra with different n. In the figures, the solid lines are the original source spectrum; the dash lines are the reconstruction spectrum. In the three figures,

the dash lines nearly cover the solid lines. It means that they are acceptable approximate reconstructions, although we can discover the dash lines are in fact biased from the solid lines in the enlarged figures. When n increases, the construction results become less convergent.

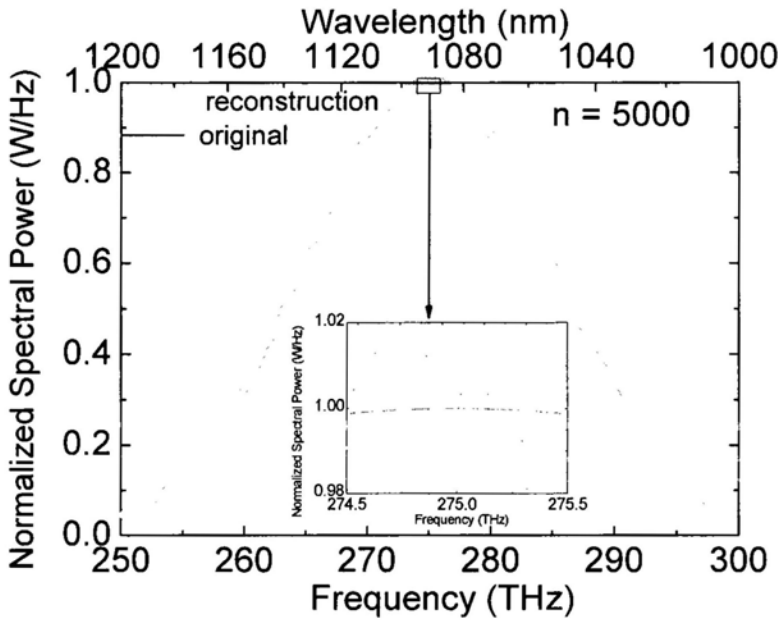
The simulation results also indicate that the interferometer array spectrometer has the potential to realize very high resolution. Fig. 5-1(a) represents 1000 frequency components are reconstructed in the range from 750THz to 760THz, so the resolution is 10GHz in frequency, which achieves 5.19pm in wavelength; Fig. 5-1(b) represents 3000 frequency components are reconstructed in the range from 750THz to 760THz, so the resolution is 3.33GHz in frequency, which reaches 1.78pm in wavelength; Fig. 5-1(c) represents 5000 frequency components are reconstructed in the range from 750THz to 760THz. The frequency interval is even narrower. The resolution is 2.00GHz in frequency, which reaches 1.07pm in wavelength. Therefore, it is no problem to solve 5000 or even more linear equations by using Tikhonov regularization method. However, if 10000 linear equations are used, major distortion occurs as shown in Fig. 5-1(d). It does not mean that we cannot solve 10000 linear equations by using Tikhonov regularization method. Because the spectrum range and the largest step height are not changed if the interferometers are increased to 10000, the difference of two values obtained from two close interferometers is very small. Therefore, it is even more difficult to solve the linear equations due to the similarity of the equations. In other word, we need to solve more unknown values with the fewer equations. In order to remove the

distortion, we need to check the values detected by the CCD in practice. If the values measured under similar interferometers are very close, we only take one value from these very similar values. That means we do not use all the data measured from the CCD pixels under the similar interferometers. Then the value of n can be decreased and the linear system can be solved so that no obvious distortion occurs.

The interferometer array spectrometer has a very wide detection range from ultra violet (UV) to infrared (IR). The original spectrum in Fig. 5-1 is in the range from 750THz to 760THz, which is in the ultra violet spectra range. Fig. 5-2 shows that the interferometer array spectrometer can also recover the spectrum in the infrared spectra range. Therefore, as long as the original spectrum is in the detection range of the CCD, it can be reconstructed by the interferometer array spectrometer.



(a)



(b)

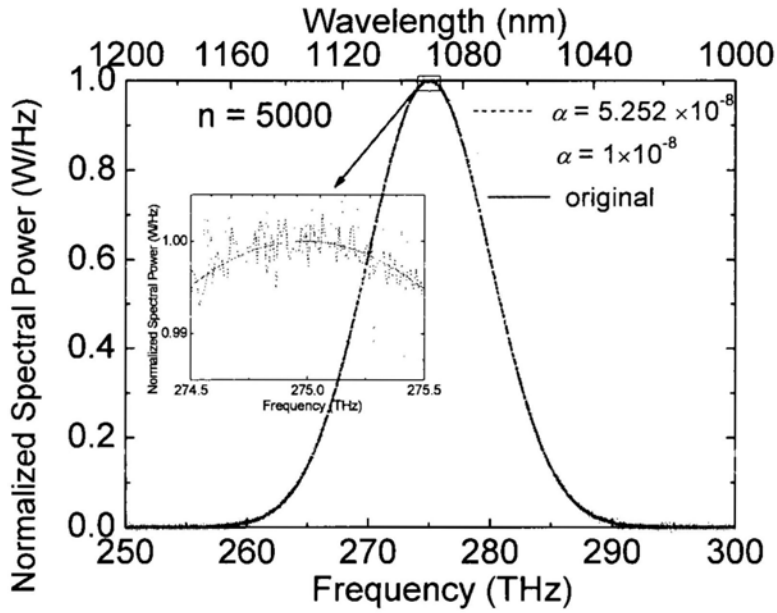
Fig. 5-2 Reconstruction in the infrared spectra range with different n

Successful reconstruction not only depends on physical factors such as the sensitivity and the SNR of the CCD pixels in experiments, but also depends on mathematical factors such as the significant digits kept for values in simulation and calculation, the width of the original spectrum peak, the spectrum range and the number of the frequency division. Among them, the width of the original spectrum peak is one of the factors because the effect of Tikhonov regularization method is making the curve smooth, which is difficult to be realized if the peak width is too narrow, therefore leading to obvious distortions. Corresponding simulation results are shown in Fig. 5-3. The frequency resolution in Fig. 5-3(a) is 10GHz, which is the same as that in Fig. 5-3(b). However, the peak width (the standard deviation of the Gauss distribution) in Fig. 5-3(a) is 5 THz, while that in Fig. 5-3(b) is only 1 THz. As a result, obvious distortion occurs in Fig. 5-3(b). The peak width in Fig.

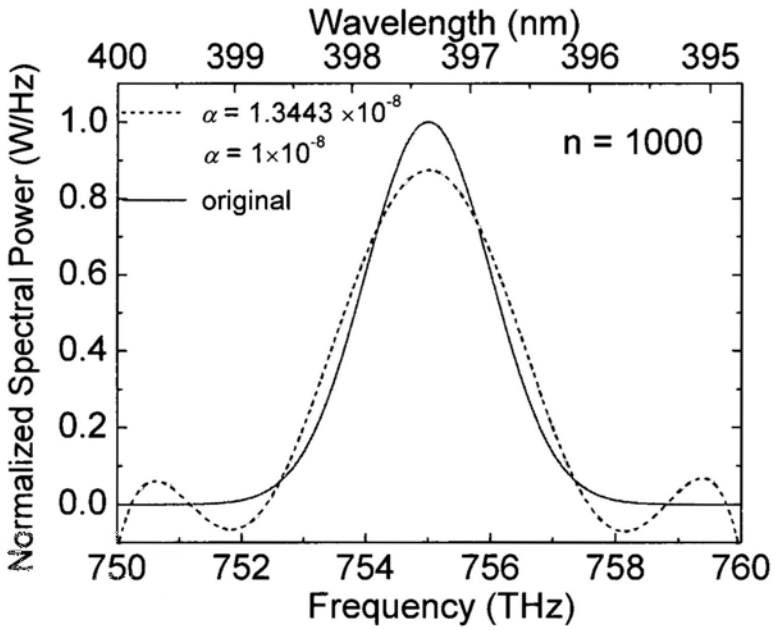
5-3(c) is also 1 THz, so obvious distortion still occurs. In contrast, Fig. 5-3(d) and Fig. 5-1 (a) have a peak width of 2 THz. We can see the distortions are much alleviated in these figures.

The regularization parameter α can also affect the reconstruction. The unavoidable measurement noises and the round-off errors in the computing process may be amplified due to the choice of an unreasonable regularization parameter and therefore make the numerical solution completely useless. So the effectiveness of a regularization method depends strongly on the choice of the regularization parameter. Because the regularization parameter mainly depends on the data errors in measurement and the significant digits kept in calculation, these factors are normally the same for different reconstructions for one device. As a result, we can use a fix regularization parameter for different reconstruction. The merit is saving time, which is valuable for real-time reconstruction application. If we calculate the regularization parameter every time for different reconstruction, the time costing is several seconds for each reconstruction. However, if we use a known regularization parameter for a reconstruction, the time costing is only one or two seconds. Therefore, we use 1×10^{8} as the regularization parameter in each reconstruction in the paper and the reconstruction results are marked as dash lines. In Fig. 5-3 we have also calculate the regularization parameters for these particular situations with a method introduced in Ref. [117]. The reconstruction results are the short dash lines as shown in Fig. 5-3 and the corresponding regularization parameters are also labeled in the figures. We can see the dash lines and the short dash lines are very

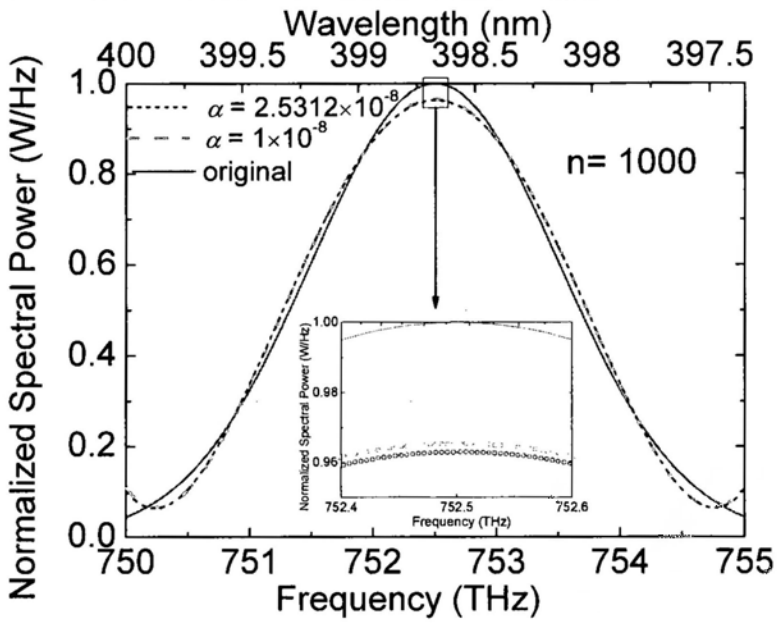
close, so the calculation of the regularization parameter for each reconstruction is not necessary.



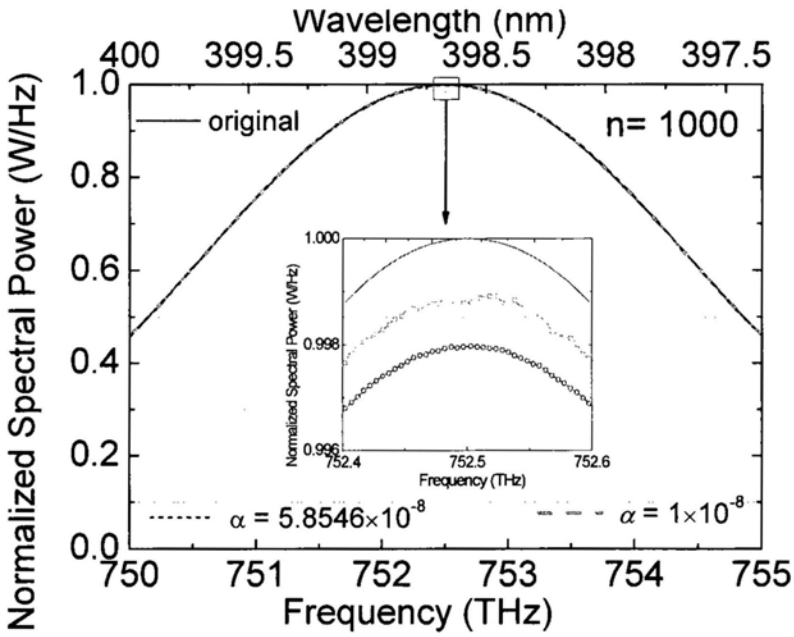
(a)



(b)



(c)



(d)

Fig. 5-3 Distortion factors comparison

5.3 Crosstalk consideration

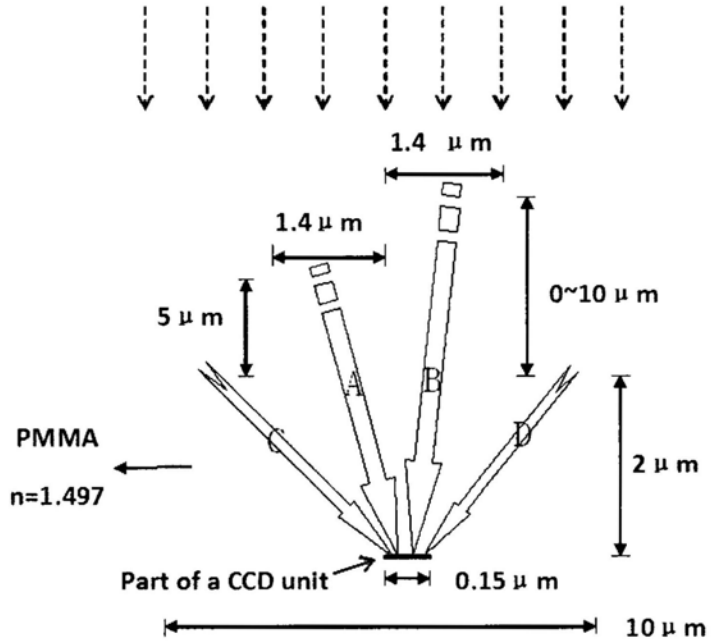
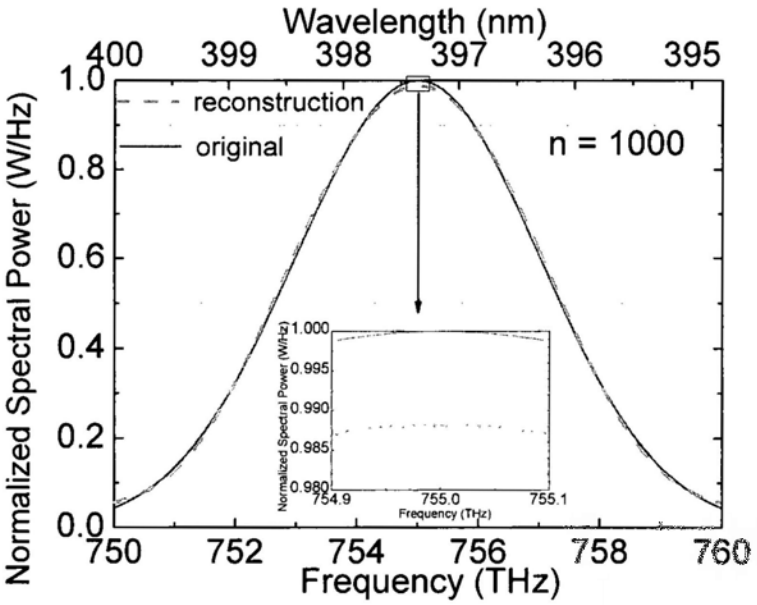


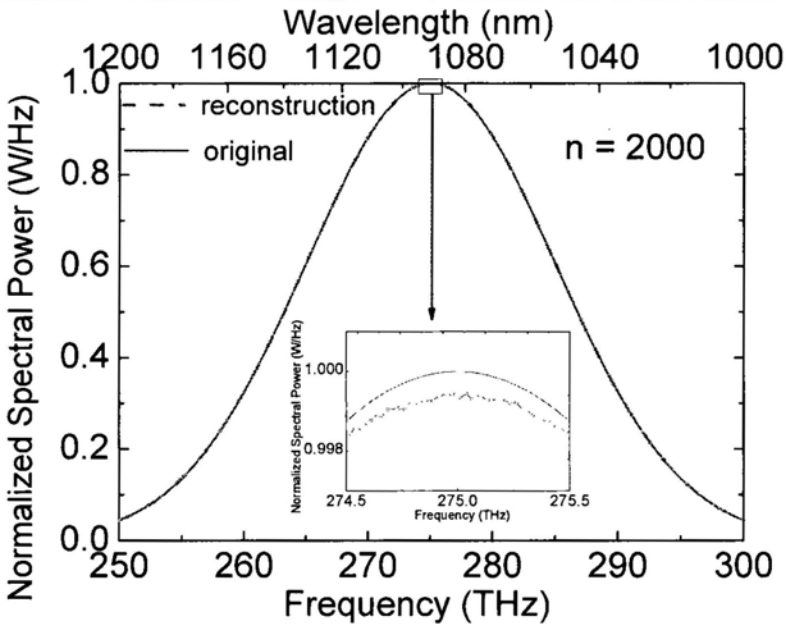
Fig. 5-4 Optical interference effect with side illumination

One may also think about the crosstalk between two interferometers. In the above simulation, the interferometers are studied one by one assuming that they are isolated. The illumination width of the incoming beam is only 3.5 μ m in the above simulation for this range is enough to cover the major structure of the interferometer. However, Fig. 4-12 indicates that these interferometers should be put together in an array configuration. In order to decrease the crosstalk between them, we need to separate the interferometers with enough space intervals. Considering even the distance between two interferometers is 10 μ m, a 1000×1000 array only occupies 1 cm^2. Therefore, the crosstalk problem can be solved by keeping large interval distance as shown in Fig. 5-4. In order to explore the effects of Beam *C* and Beam *D* in Fig. 5-4, we reconstruct the same spectrum as shown in Fig. 5-1(a) and Fig.

5-1(b) with 10 μ m interval and show the results in Fig. 5-5. By comparing Fig. 5-5(a) with Fig. 5-1(a) and Fig. 5-5(b) with Fig. 5-1(a), we find there is no obvious distortion. The original and reconstructed plots almost overlap with each other. It can be concluded that the side illumination has little influence on the success of reconstructions. The reason for that is that Beam *C* and Beam *D* are in fact scattering light of the incoming plane wave, which are very weak if the scatters such as the edge of interferometers are far away.



(a)



(b)

Fig. 5-5 Reconstruction with a large space interval (10 microns)

5.4 Reconstructions for nonlinear phase delay

In the above design, the phase delay of each interferometer element is pre-designed to follow a linear equation. We also consider the situation of following a parabola equation as shown in Fig. 5-6. Fig 5-7 (a) and (b) are the corresponding results. The simulation results indicate that we can still realize the reconstruction with small distortion even the phase delay of each interferometer is not uniform. That means as long as the transmission coefficients of each interferometer is different, even the actual device fabrication process may produce nonlinear phase delay for each interferometer due to the fabrication errors, we can realize the reconstruction with little distortion.

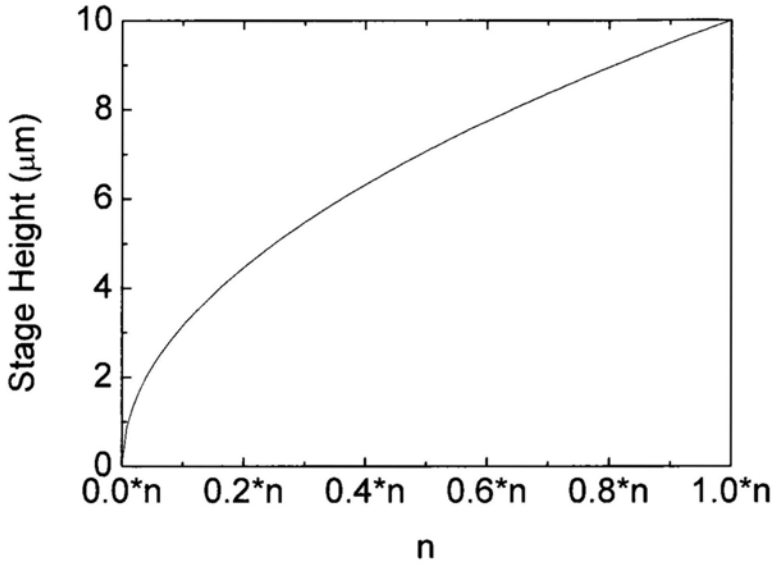
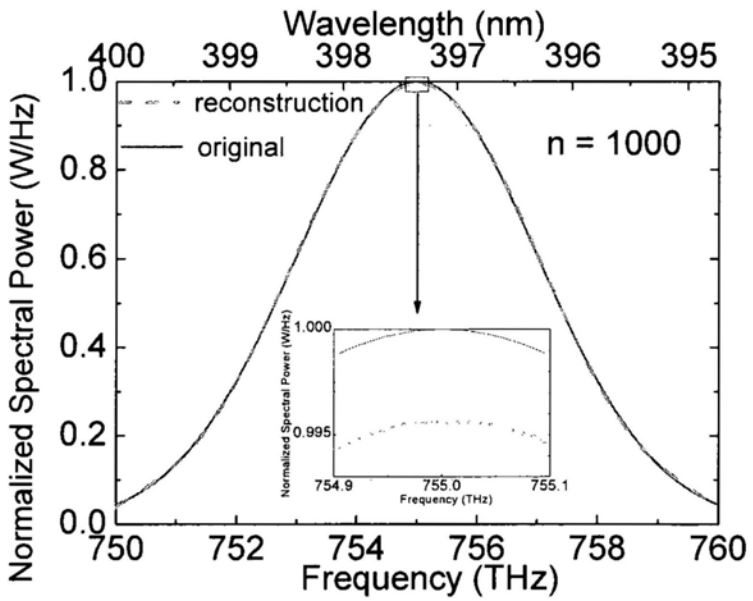
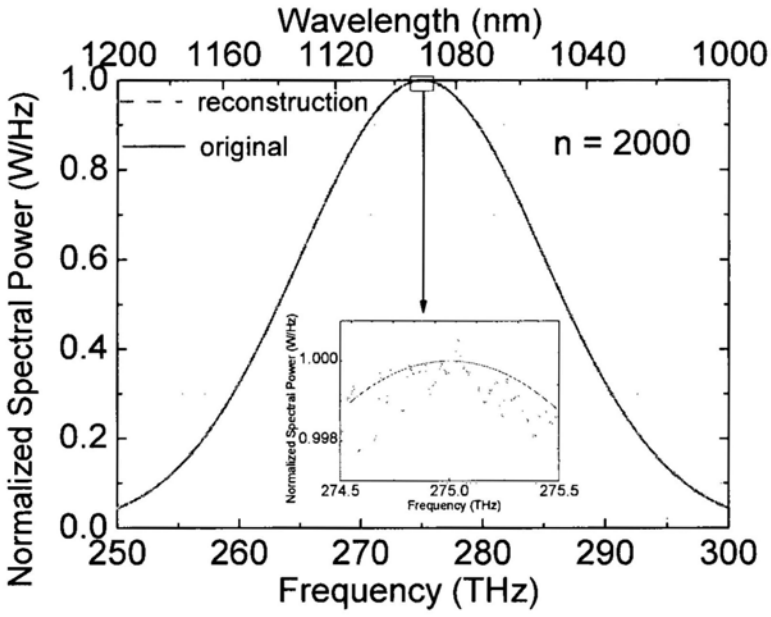


Fig. 5-6 Step height of each interferometer



(a)



(b)

Fig. 5-7 Reconstruction for nonlinear phase delay

5.5 Simulation results of pit structures

We also explore other structures of the interferometer. Considering the difficulty of fabricating the extruded arrays, it would be better to adopt the inverted pit structures as shown in Fig. 5-8. Both the extruded region and the pits can generate the required phase delay, so they can be used to realize the same function. The main advantage of the pit structure is that it is easier to fabricate with the mature optical storage technology, i.e. we can fabricate the pit array with a mould just like the fabrication of CD and DVD. In order to further simplify the structure, we can use one pit for an interferometer as shown in Fig. 5-9, because one pit is enough to achieve the phase difference. Fig. 5-10(a) and (b) are the reconstruction results of the simulation. In comparison with Fig. 5-1 (a) and Fig. 5-2(a), the distortion in

Fig. 5-10(b) is acceptable, but it becomes worse in Fig. 5-10(a). This means the resolution of the pit structure is worse than the extruded structure. The reason is attributed to the possibility that the positions of the monitors in the pit structure are further away from the polymer surface, so the detectors are more susceptible to influence due to side illumination.

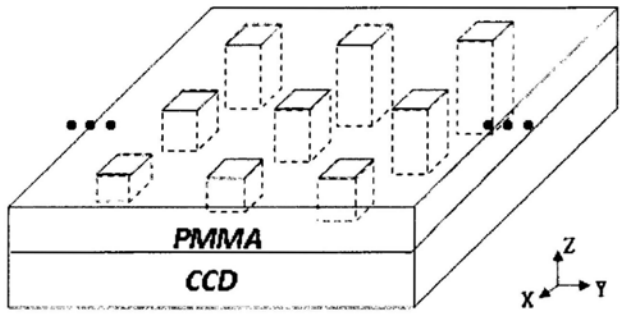


Fig. 5-8 Schematic of the pit structure spectrometer

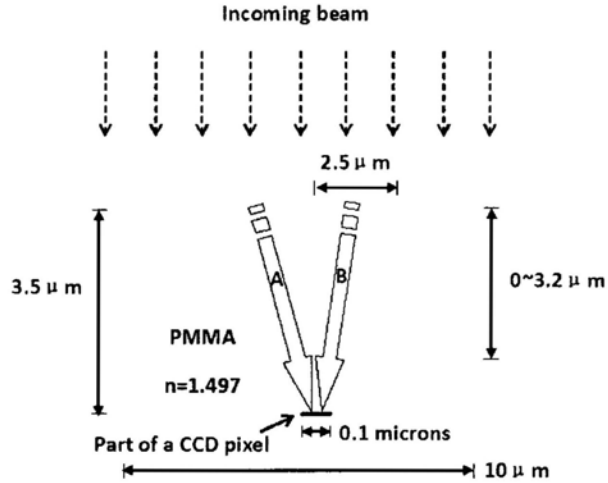
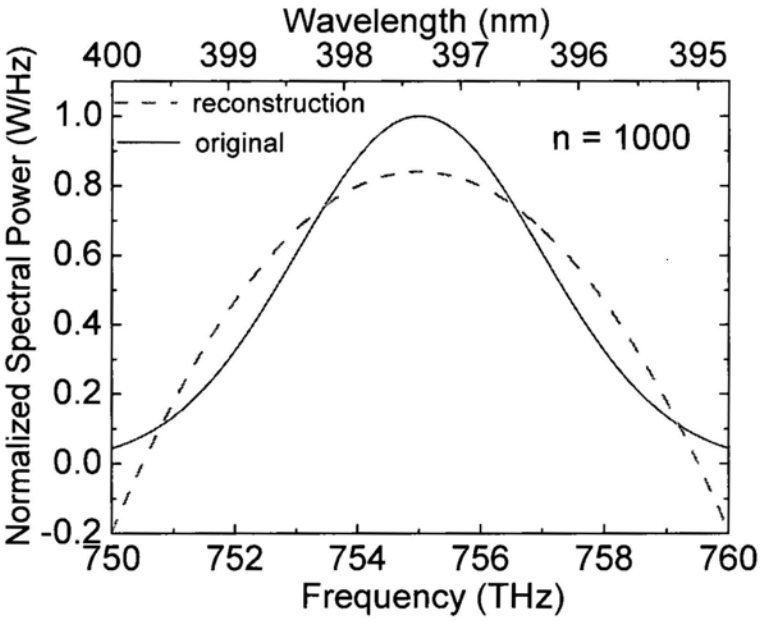
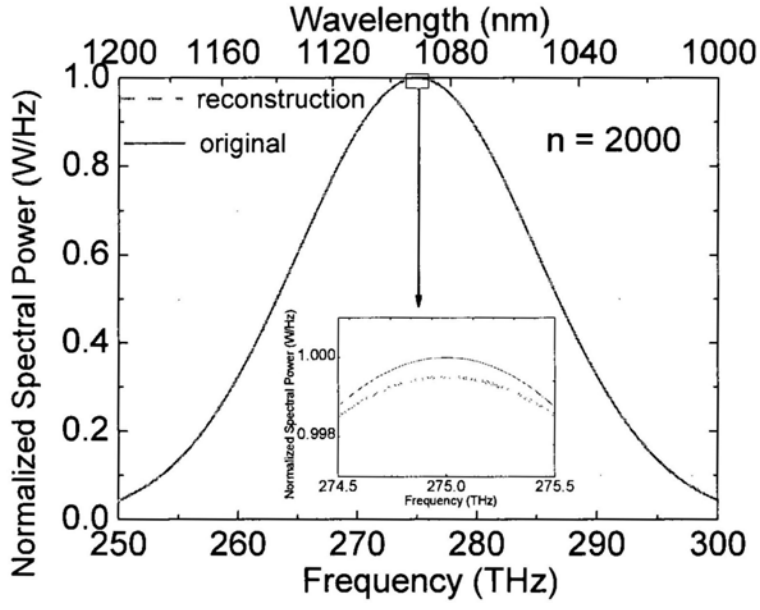


Fig. 5-9 Cross section of the pit structure interferometer



(a)



(b)

Fig. 5-10 Reconstruction for the pit structure

We also consider the situation when the shape of each pit is a torus from the top view. The cross section view and the corresponding reconstruction results are shown in Fig. 5-11 and Fig. 5-12 respectively. We find the distortion in Fig. 5-12(b)

is acceptable. In comparison with Fig. 5-10(a), the distortion in Fig. 5-12(a) is somewhat alleviated.

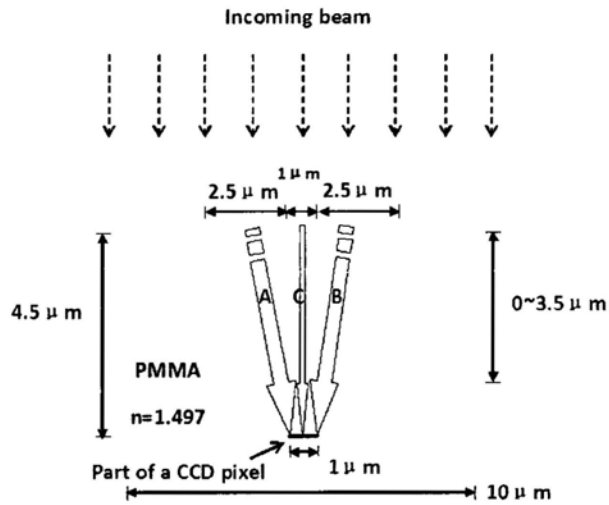
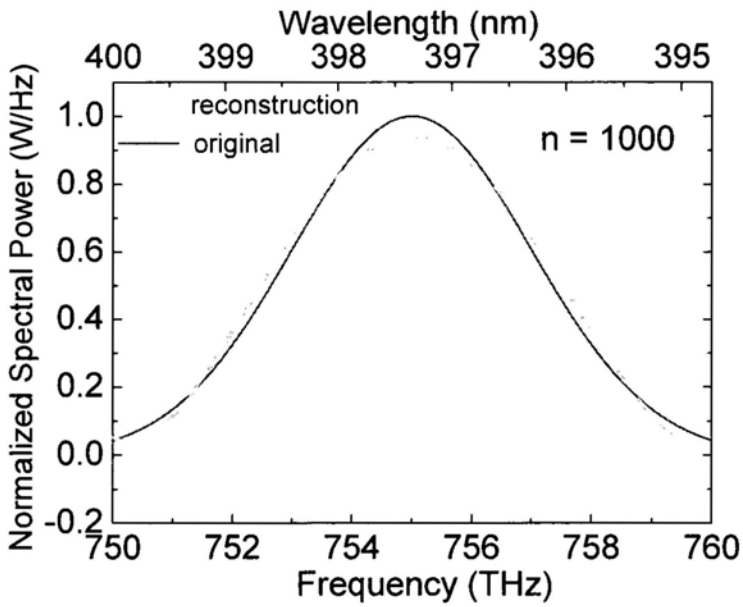
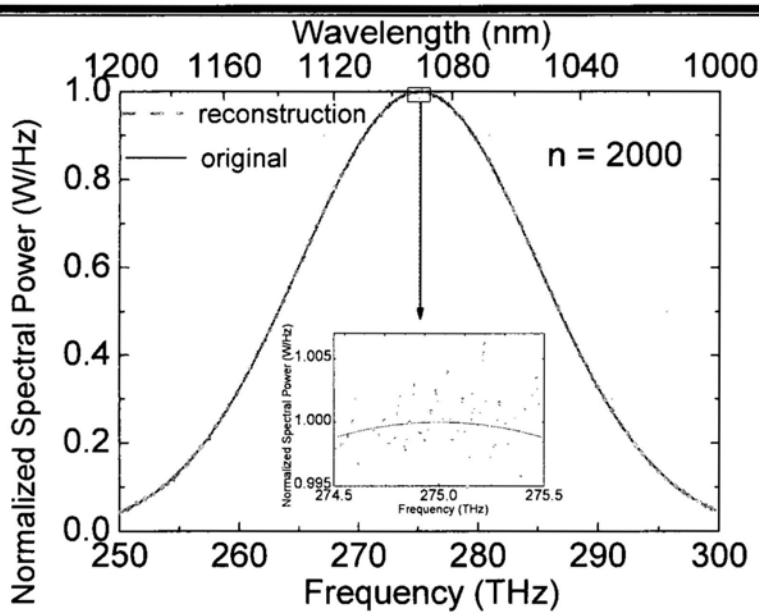


Fig. 5-11 The torus pit structure interferometer



(a)

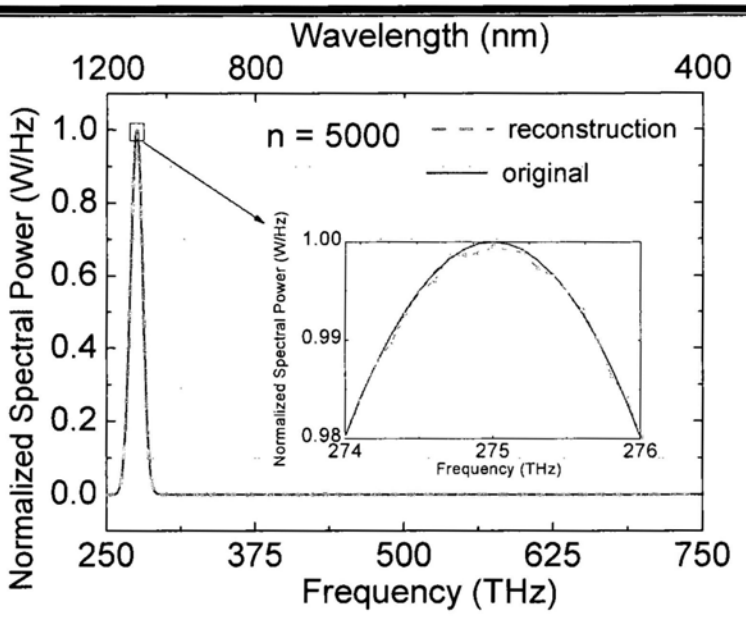


(b)

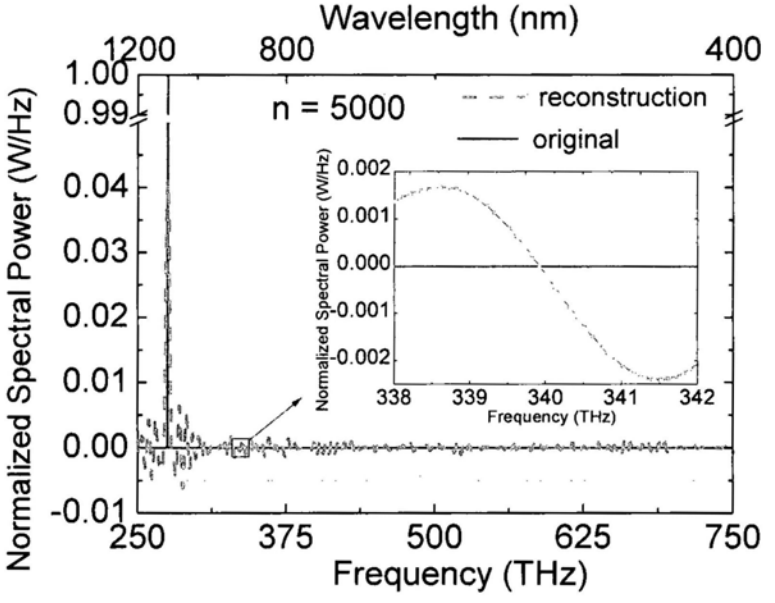
Fig. 5-12 Reconstruction for the torus pit structure

5.6 Broadband reconstruction

The wavelength range of optical spectroscopy extends from ultra-violet (UV) to infra-red (IR). However the traditional instruments only operate over a small portion of this total range because of the different techniques used to measure different portions of the spectrum. It is difficult to design one single spectral instrument capable of covering the whole region completely, providing high resolution and occupying small size at the same time. In this section, we prove that the novel interferometer-based spectrometers have the advantages of broad bandwidth, low cost, small size and high resolution simultaneously.



(a)



(b)

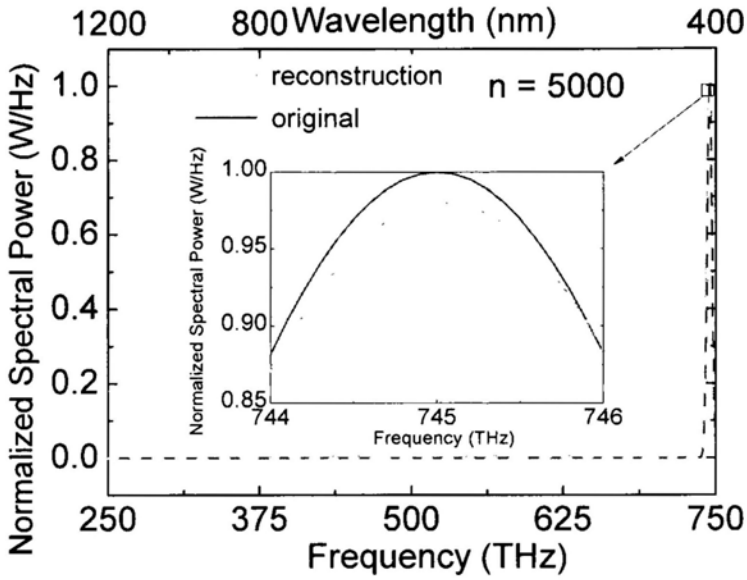
Fig. 5-13 Broadband reconstruction

Fig. 5-13(a) is the comparison of original spectrum and reconstruction spectrum. The anticipated detection range is from 250 THz to 750 THz, so the bandwidth is 500 THz. Because the standard deviation of the Gaussian distribution

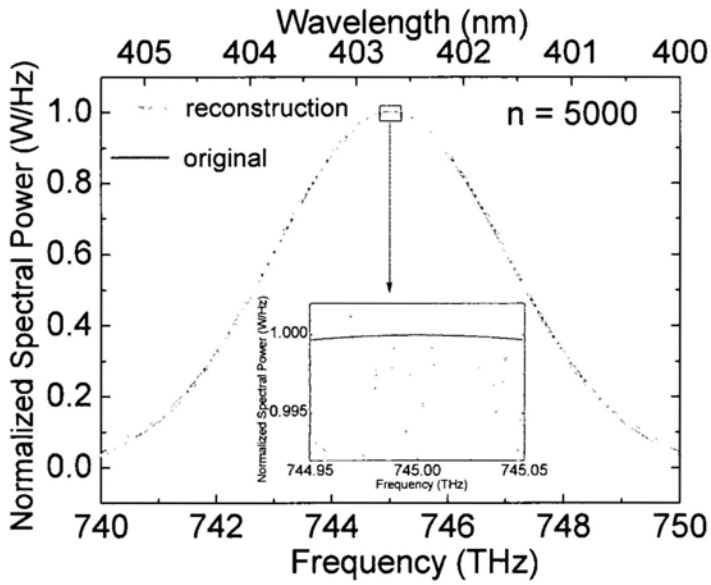
(a measure of the "width" of the bell) is 5THz, which is quite wider than the resolution 100 GHz, there is no obvious distortion after the reconstruction. In contrast, the "width" of the original spectrum in Fig. 5-13(b) is as narrow as 50 GHz, so the reconstruction spectrum cannot fully cover the original spectrum. But it still can represent the rough shape of the original spectrum.

In order to improve the resolution, if we know the probable range of the spectrum, we can shift to a fine coefficient efficiency matrix and try a second or third reconstruction. As shown in Fig. 5-14, if all the frequencies of the spectrum are in the range from 750THz to 760THz detected by the first reconstruction, we can do a second reconstruction in such a narrow range. Therefore, the frequency resolution is improved from 100GHz to 2GHz and the wavelength resolution is improved from 53.333pm to 1.067pm accordingly.

If there are several peaks in the original spectrum and we only need to research one part of the spectrum, we can use some filters to block other frequency components to do the second reconstruction. However, the transmission coefficients should be calibrated for the whole system, i.e. the power of incoming beam means the power before entering the filters.



(a)

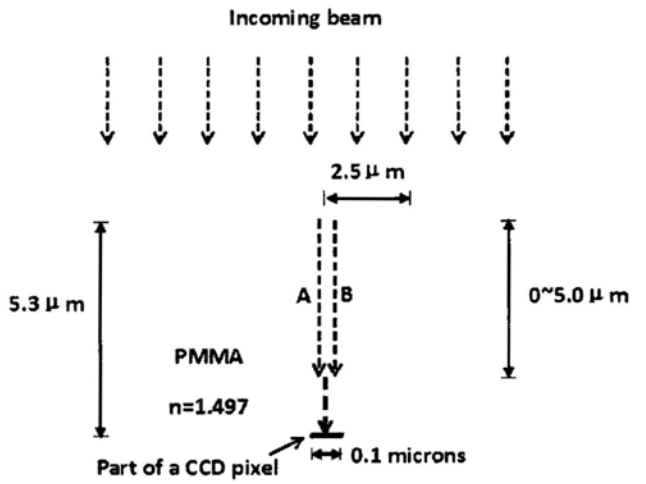


(b)

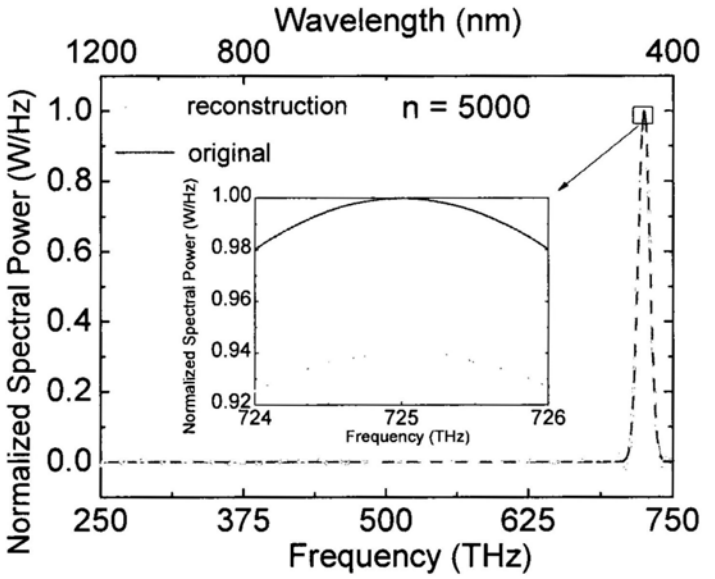
Fig. 5-14 Broadband reconstruction with high resolution

Considering the difficulty of fabricating the extruded arrays, it would be better to adopt the inverted pit structures as shown in Fig. 5-15 (a). Fig. 5-15(b) and (c)

are the reconstruction results, which also show little distortion.



(a)

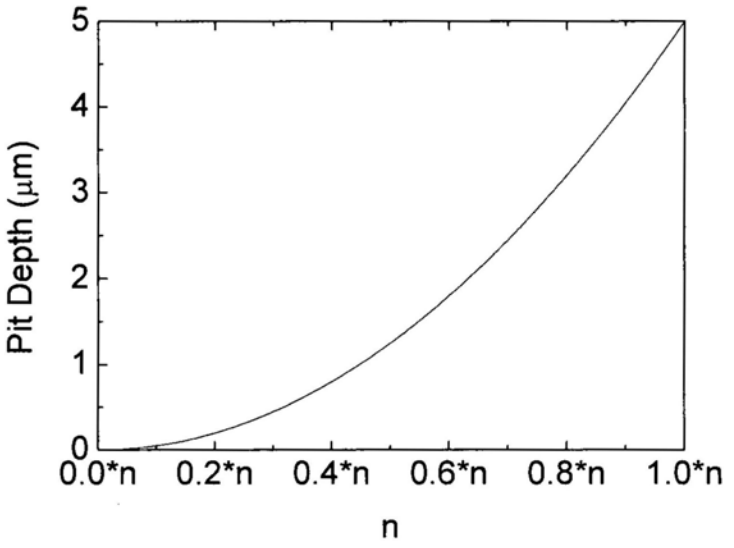


(b)

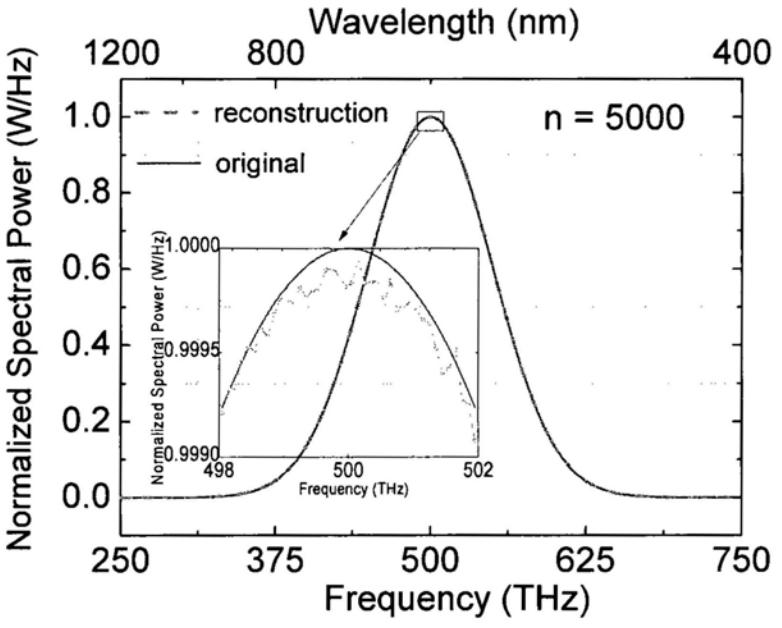
Fig. 5-15 Broadband reconstruction for pit structures

We also consider the situation that the phase delays of each interferometer element following a parabola equation. The depth of each pit is shown in Fig. 5-16(a). Fig 5-16(b) is the reconstruction result. We do this reconstruction because

the pit depth of real device maybe is not what we design for it. It is possible that the variation of the pit depth is not uniform. Simulation results below show that we can still get satisfactory results.



(a)

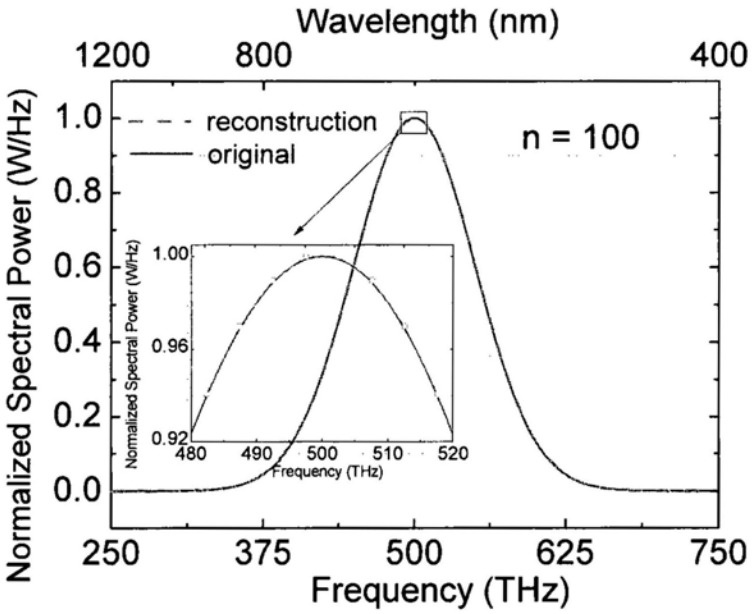


(b)

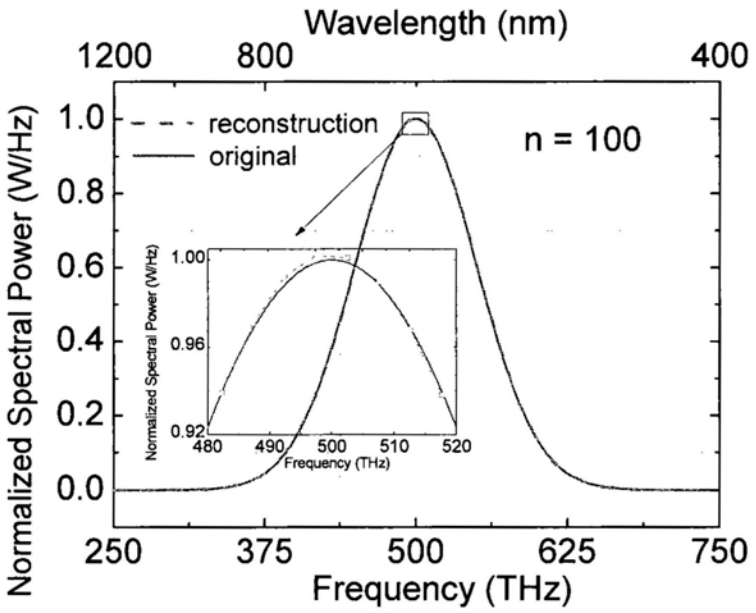
Fig. 5-16 Nonlinear phase delay reconstruction for pit structures

5.7 Few interferometer reconstruction

It is not necessary to use large number of the interferometers n to reconstruct the spectrum. Sometimes we need to reduce n because CCD is not sensitive enough to separate the difference of two interferometers if they are too similar, which maybe is caused by fabrication errors of the interferometers or low SNR of the CCD. If we still use large n, obvious distortions may occur as shown in Fig. 5-1(d). However, simulation results prove that the spectrometer can realize successful reconstruction even with much small n although the resolution is not high. We use 100 interferometers to reconstruct the original spectrum in Fig. 5-17(a) and Fig. 5-17(b). Fig. 5-17(a) is the simulation results for the structure as shown in Fig. 4-13; Fig. 5-17(b) is the simulation results for the structure as shown in Fig. 5-15(a). Therefore, both extruded and pit structures can realize acceptable reconstruction with few interferometers.



(a)

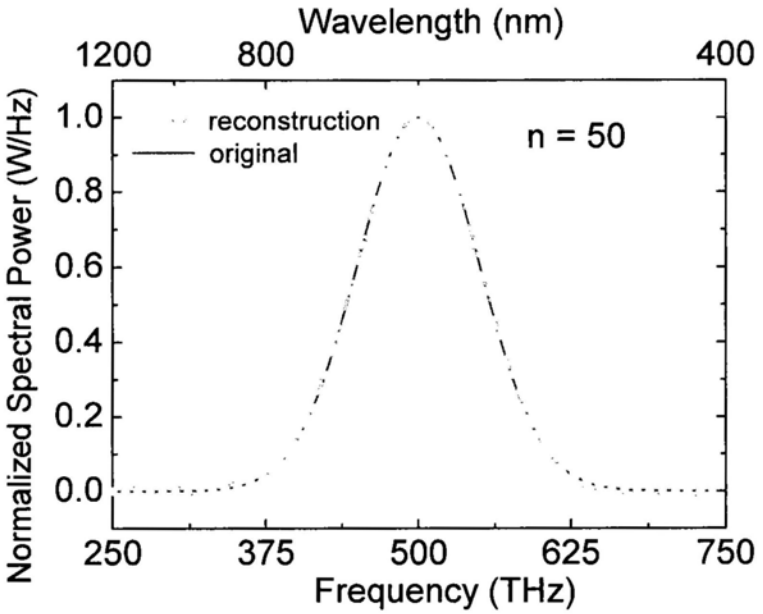


(b)

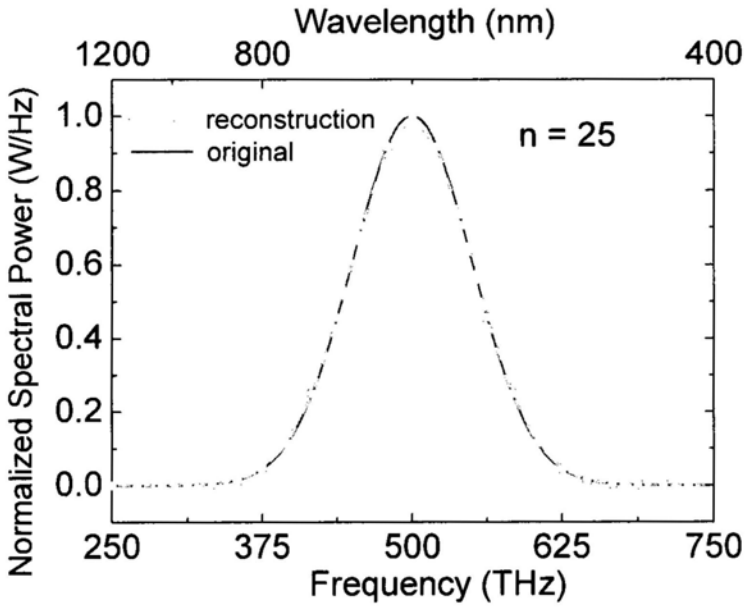
Fig. 5-17 Broadband reconstruction with 100 interferometers

(a) For extruded structure (b) For pit structure

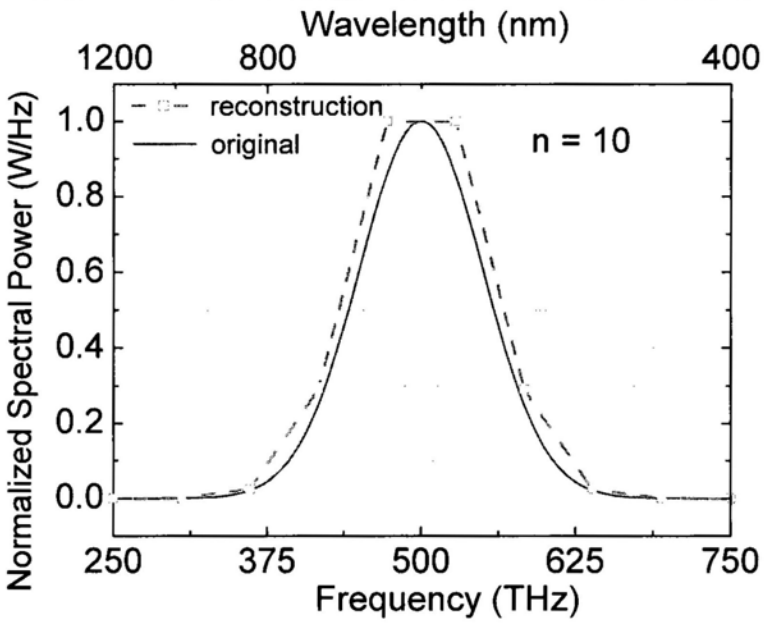
There is a limit for the number of the interferometers. If the interferometers are too few, obvious distortions may occur. As shown in Fig. 5-18(a), when the number of the interferometers is 50, the reconstruction spectrum (green line) still can cover the original spectrum (read line). But when the number is 25 as shown in Fig. 5-18(b), the top of the spectrum peak cannot be fully covered. The situation becomes even bad when the number of the interferometer decreases to 10 as shown in Fig. 5-18(c). However, we can still know the rough shape of the spectrum even 10 interferometers are used.


Wavelength (nm)
1200
800
400
reconstruction
original
n = 50
Normalized Spectral Power (W/Hz)
1.0
0.8
0.6
0.4
0.2
0.0
250
375
500
625
750
Frequency (THz)

(a)


Wavelength (nm)
1200
800
400
reconstruction
original
n = 25
Normalized Spectral Power (W/Hz)
1.0
0.8
0.6
0.4
0.2
0.0
250
375
500
625
750
Frequency (THz)

(b)



(c)

Fig. 5-18 Broadband reconstruction with large distortion for few interferometers

5.8 Reconstruction procedure

The spectrometers can be used to reconstruct a wide frequency range. However, the resolution is not very high due to limit n value. Here n is the effective number of interferometers. In order to improve the resolution, we need to do second or third reconstruction. The following is the steps we used to reconstruct an unknown spectrum:

1) Reshape the incoming beam to make the intensity uniform and ensure the direction perpendicular to the array surface.

2) Amend the data measured from CCD to remove the errors caused by the non-uniform response of the pixels due to dark current.

3) According to the spectrometer measure range $[f_a,f_b]$ (determined by the measure

range of the CCD) and the number of the self-mixing interferometers, we can get n frequency components. For each frequency components, the central frequency is used to represent the corresponding frequency component. The frequency width of each frequency component is $(f_b-f_a)/n$. In order to avoid obvious distortion, the value of n should be larger than 50. Here, the subscripts a, b, c, d... are used to separate the different frequencies.

4) Based on the transmission coefficients of each frequency components and each self-mixing interferometers as well as the data obtained from CCD, we can get a linear system. The transmission coefficients should be measured in advance. The measurement process is:

a) Put a tunable laser in front of the uniform and collimation system. The power of the emitting light with a certain frequency is measured.

b) Based on the data from each CCD pixels, the transmission coefficients for different interferometers with such frequency can be calculated.

c) Continually change the frequency of the tunable laser so as to get a series of transmission coefficients for the linear system.

5) The linear system is solved by Tikhonov regularization method. The regularization parameter is chosen by the linear model function method.

6) Calculate and calibrate the optical spectrum of the incoming beam.

7) If we want to further improve the resolution, we need to do the reconstruction again. We can use two methods as follows:

The first method is suitable to the reconstruction of an incoming beam with a

narrow peak width. The steps are:

a) According to the spectrum from the results of the first spectrum reconstruction (the resolution is poor) we can know the rough frequency range of the spectrum $[f_c,f_d]$. Here $[f_c,f_d]$ is one part of $[f_a,f_b]$. We divide the frequency range $[f_c,f_d]$ into n frequency components. For each frequency component, the central frequency is used to represent the corresponding frequency component. The frequency width of each frequency component is $(f_d-f_c)/n$. This width is the resolution of the second spectrum reconstruction.

b) We can measure the transmission coefficients of each frequency components and each self-mixing interferometers in advance (the transmission coefficients should be measured before the uniform and collimation system). By using the transmission coefficients as well as the data obtained from CCD, we can get a linear system.

c) The linear system is solved by Tikhonov regularization method. The regularization parameter is chosen by the linear model function method.

d) Calculate and calibrate the optical spectrum of the incoming beam.

e) If we want to further improve the resolution, we need to perform the third or fourth reconstruction.

The second method is suitable to the reconstruction of an incoming beam with a wide spectrum range (the reconstruction process needs to be done for several parts) or the reconstruction of an interesting part of the whole frequency range of the incoming beam.

a) Adopt a series of filters for the system. The transmission ranges of these filters

are different from each other. The position of these filters is between the collimator and interferometer array. If all the filters are used, the transmission range of these filters covers the whole range $[f_a,f_b]$. When we do the first reconstruction, we do not adopt any filter. The transmission coefficients are that do not count in any filters. When we do the reconstruction again, we can use one or more filters, which block other wavelengths and only transmit a small range $[f_e,f_g]$. Here $[f_e,f_g]$ is one part of $[f_a,f_b]$ or $[f_e,f_g]$ and $[f_a,f_b]$ have intersection. Assume that f_h is the large value of f_a and f_e; f_i is the small value of f_b and f_g. We divide the frequency range $[f_h,f_i]$ into n frequency components. For each frequency component, the central frequency is used to represent the corresponding frequency component. The frequency width of each frequency component is $(f_i-f_h)/n$. This width is the resolution of the second spectrum reconstruction.

b) We can measure the transmission coefficients of each frequency components and each self-mixing interferometers in advance (the transmission coefficients should be measured before the uniform and collimation system). By using the transmission coefficients as well as the data obtained from CCD, we can get a linear system.

c) The linear system is solved by Tikhonov regularization method. The regularization parameter is chosen by the linear model function method.

d) Calculate and calibrate the optical spectrum of the incoming beam.

e) If we want to research other frequency range except $[f_i,f_h]$, we need to perform the third or fourth reconstruction with other filters.

5.9 Factors affecting satisfactory reconstruction

Now we can summarize the impact factors for satisfactory reconstructions, which include:

physical factors such as

- ✧ the sensitivity of the CCD,
- ✧ the SNR of the CCD,
- ✧ size and material of the spectrometer components,
- ✧ precision of fabrication technology;

mathematical factors such as

- ✧ the significant digits kept for values in simulation or calculation,
- ✧ the width of the original spectrum peak,
- ✧ the spectrum range,
- ✧ the number of the frequency division,
- ✧ the regularization parameter α..

One may raise a question, i.e. whether is there a limit for each impact factor? In other word, whether does obvious distortion occur if an impact factor exceeds a limit? However, simulation results have shown that these factors are not isolated fromeach other. It is nearly impossible to state a range for the value of one factor because a satisfactory reconstruction depends on many parameters. For example, obvious distortion occurs when n equals to 1000 as shown in Fig. 5-3(c). It does not mean that 1000 is the limit for n because no obvious distortion occurs when n equals to 5000 as shown in Fig. 5-1(c). 5000 is also not the limit for n. We can take

an even large number when the spectrum range is increased. Then how to determine these parameters? In order to avoid obvious distortion and obtain satisfactory results, two rules must be satisfied:

1) The peak width of original spectrum should be at least larger than the reconstruction resolution. The resolution can be calculated by the spectrum range over the number of the frequency division. If the peak width is too narrow, it is too difficult to make the reconstruction curve smooth and continuous with Tikhonov regularization method and obvious distortion may occur.

2) The CCD pixels can separate the difference of two close interferometers. It is possible that two interferometers are different, but we get the same power from the CCD pixels with the two interferometers due to the poor SNR of CCD or small variation of step height or other reasons. If the CCD pixels obtain the same value from two expected different interferometers, it is difficult to solve the corresponding linear system hence obvious distortion may occur.

5.10 Fabrication error analysis

Before we use a spectrometer to measure the spectrum of an incoming beam, we must test whether the spectrometer can be directly used with all the fabricated interferometers. Due to limit fabrication precise, it is absolute that the real height (or the depth) of the step (or the pit) is not what we design for it. Then whether does it mean that the device cannot realize reconstruction function? The previous simulation results have shown that we can still obtain satisfactory results even with

such unperfect device. We can consider the following two cases:

Firstly, due to the fabrication error or noise, sometimes the height (or the depth) of the step (or the pit) is very similar or exactly same. Then the CCD pixels under these interferometers with small difference maybe obtain the same or very close value under the exposition of the illumination of the same incoming beam intensity. Therefore, the transmission coefficients of these similar interferometers are the same or very similar, which leads to obvious distortion because we try to solve the same number of unknown value in spite of the reduction of the effective number of linear equations (i.e. the effective number of the interferometers). To avoid the distortion, we have to count in the data with large variation and neglect the other values with small variation obtained from these CCD pixels. At the same time, we have to use few numbers of unknown frequencies in one calculation. Section 5.7 has proved that we can still realize reconstruction with just few interferometers although the resolution is not very good due to few unknown frequencies are used in the calculation. In order to improve resolution, we just need to do second or more reconstruction.

Secondly, it is possible that the variation of the height (or the depth) of the step (or the pit) is not uniform due to the error in the fabrication. However, Fig. 5-7 and Fig 5-16 have proven that we can still reconstruct the original spectrum with such unpefect devices.

In summary, as long as we can solve the linear system, we can reconstruct the spectrum. If the resolution is not fine enough due to small n value, we can improve

the resolution by doing a second reconstruction with a narrow spectrum range as shown in Fig. 5-14.

5.11 Summary

In this chapter, we have described an ultra compact and high resolution free space optical spectrometer and demonstrated it by using FDTD simulations. The spectrum can be reconstructed by solving a linear system. Tikhonov regularization method is used to make the resolution reach picometer level. Simulation results show that the spectrometer can realize a very wide measurement range from UV to IR. Reconstructions for the designs of a large space interval and a parabola phase delay indicate that the problems of crosstalk and fabrication error do not affect the results very much. Designs for extrude and the pit structures show that the proposed device is low-cost and easy to fabricate as well.

Real-time operation, high resolution, long and controllable measurement range, simple fabrication and easy integration are the major advantages of the proposed approach promising a great potential for development of compact spectrometer with good performance. Due to the flexibility of the actuation principle, the presented platforms can be actuated in different mechanical modes that can be used to implement other spectrometer configurations.

Chapter 6: Conclusions and Future Work

6.1 Conclusions

There are very few compact instruments currently available for the determination of refractive index change or spectral absorption of materials. The traditional instruments are normally quite large and expensive. To satisfy the miniaturization demand, compact optical sensors and optical spectrometers with high frequency resolution and wide dynamic range have been studied in this thesis.

In the first part of the thesis, we present a new high spatial resolution sensor for monitoring refractive index variations caused by binding of organic and biological molecules to the metallic surface containing arrays of nanoholes. Signal transduction is provided through detecting the optical phase change in the extraordinary optical transmission effected by surface plasmon resonance. These 2D nanoholes are well suited for the sensor chip format in which high dense integration is readily achievable. While the sensor operates under normal illumination, practical implementation of the sensor is much easier in comparison to the traditional Kretschmann arrangement for SPR sensing. Various design parameters of the device have been studied by simulation. Our results indicate that the scheme has a shot-noise limited sensitivity threshold of 4.37×10^{-9} RIU and a dynamic range of 0.17 RIU, which compare favorably with typical SPR sensors, particularly in terms of achieving high resolution and wide dynamic range sensor attributes. The phase

change is also quite linear over the entire refractive index detection range.

In the second part of the thesis, we present a chip-level integrated optical spectrometer and demonstrate its broadband and high resolution performance by using FDTD simulations. In this miniature interferometer-based spectrometer, a 2-dimensional array of micron-scale phase objects made from polymethyl methacrylate (PMMA) is mounted on a CCD imaging device. Our analysis shows that the 2-dimensional intensity distribution obtained by the CCD device can be described by a system of simultaneous linear equations when the phase delay within each interferometer element is pre-designed. One can readily recover the spectral contents of the input radiation with picometer resolution. The proposed design takes advantage of the well-established CCD and precision moulding technologies for producing miniature, low-cost and high performance optical spectrometers, which may find applications in a wide range of applications.

Any progress of current technology is based on new theory or new design. The exploration of novel sensors and spectrometers introduced here is on the fundament of optical transmission properties of dielectric aperture array. The theory studied in this thesis may also offers hints for the designs of other optical devices for miniaturization purpose.

6.2 Future work

6.2.1 Experimental test for the sensors and spectrometers

Based on above research results and findings, there are several issues that should be

investigated in the future. Most importantly, we need do some experiments to prove the feasibility of the optical sensors and optical spectrometers.

For the optical sensor design, we need to fabricate the nano-hole devices at first, which is somewhat difficult due to the limitation of fabrication precision using conventional molding techniques. However, with the advent of nano-imprinting and focused ion-beam milling technologies readily available nowadays, batch production of periodic nano-hole arrays should be quite possible. After the nano-hole devices have been fabricated, we can then test their performance. The device is illuminated with a broadband source and the position of the transmission peak may be checked according to our design parameters. Once this is done, the device can then able used for bio-molecular or chemical detection. The heterodyne interferometric system as shown in Fig. 2-15 should be well suited for measuring the phase.

For the optical spectrometer experiments, one can choose to start with using only a small number of interferometer. Our results have shown that 100 interferometers are enough for achieving reasonable spectral reconstruction, although the resolution may not be too high. As a proof-of-concept experiment, we propose to use a simple setup shown in Fig. 6-1.

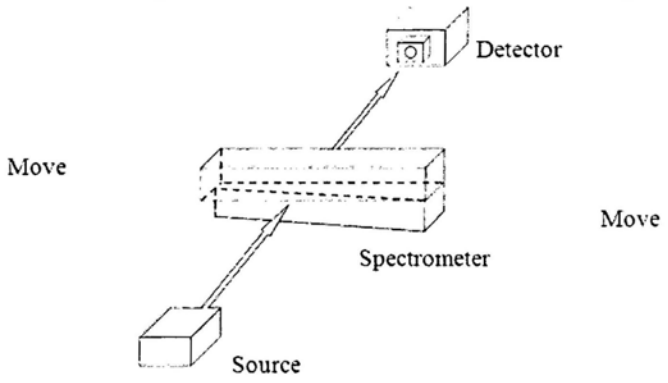


Fig. 6-1 Modeling experiment for spectrometer

The spectrometer contains two blocks of transparent plastics with one of them having a gradually increasing thickness to implement the step height variations. The transmitted interference signal will be detected by a single detector element as move the blocks along the horizontal direction. The source spectrum will then be reconstructed using the signal analysis algorithm described in this thesis. The detector can even be a commercial grating spectrometer, which will measure the transmission coefficients and total power at the same time. Once this step is completed, real 2-dimensional interferometer arrays with improved performances will be designed and fabricated using experience collected from the initial experiments.

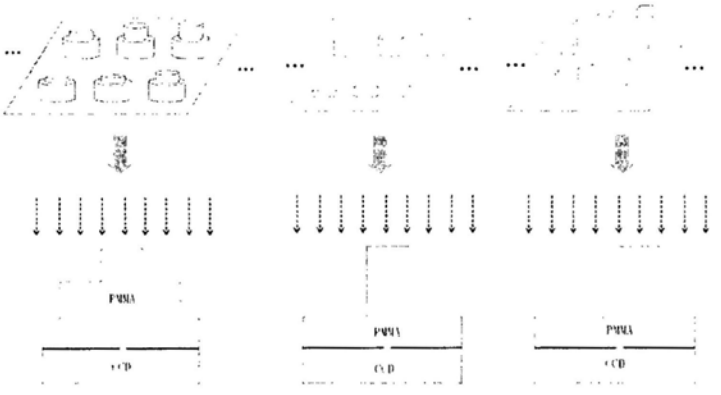


Fig. 6-2 Other step design for the interferometers

(The upper figures are 3D view and the nether figures are cross-section view)

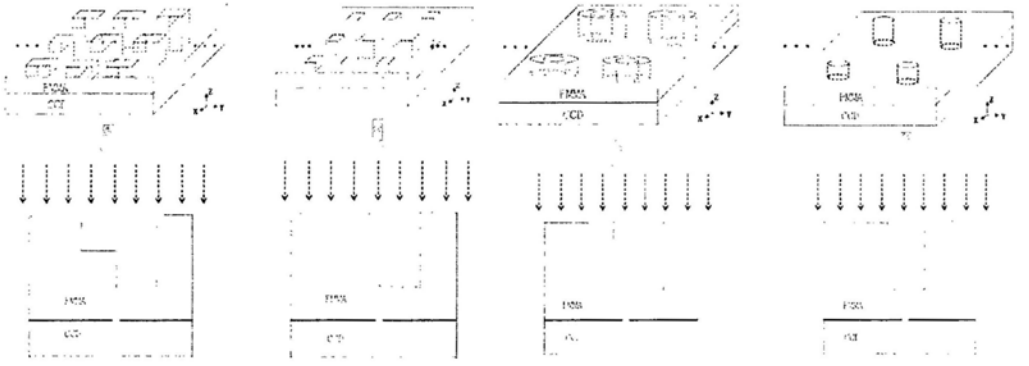


Fig. 6-3 Other pit design for the interferometers

(The upper figures are 3D view and the nether figures are cross-section view)

As for the interferometer design, besides the structure as shown in Fig. 4-12 and Fig. 4-13, we can also try other structures as shown in Fig. 6-2 and Fig. 6-3. Fig. 6-2 shows step structures and Fig. 6-3 shows pit structures. In these figures, the upper figures are 3D view and the nether figures are cross-section view. All these structures can generate the required phase delay, so they can be used to realize the same function. More simulations and experiments can be done for these structures to design a high performance device.

6.2.2 Other possible applications of the self-mixing interferometers

The operation principle of the self-mixing interferometers presented in the thesis may also find applications for the design of protein sensors, super-resolution lens, optical isolators or optical switches with the advantage of small size and high sensitivity. Some simulations and experiments for these applications should be done in future.

As shown in Fig. 6-4(a)-(d), we can design bio-sensors by using the self-mixing

interferometer. Just like the step and pit structure of spectrometers, there are two kinds of structures for the bio-sensor. Fig. 6-4(a) is a step structure bio-sensor and Fig. 6-4(b) is its cross-section view. In Fig. 6-4(b), the incoming beam is divided into two beams by the step. One reaches the detector through the step and substrate; another arrives at the detector through the cover, the bio-chemical solution and the substrate respectively. The solution fills the volume between the cover and the substrate. When the refractive index of the bio-chemical solution is changed, the power measured by the detector is accordingly changed because the phase difference of the two beams is changed. We can use this property to design a sensor. Our initial simulation results have shown its high sensitivity.

Fig. 6-4(c) is a pit structure bio-sensor and Fig. 6-4(d) is its cross-section view. In Fig. 6-4(d), the incoming beam is divided into two beams. One is through the substrate directly; another is through the cover, the bio-chemical solution and the substrate. The solution fills the volume between the cover and the substrate. When the refractive index of the bio-chemical solution is changed, the power measured by the detector should be changed accordingly. Both the step structure and pit structure can realize the function of a bio-chemical sensor.

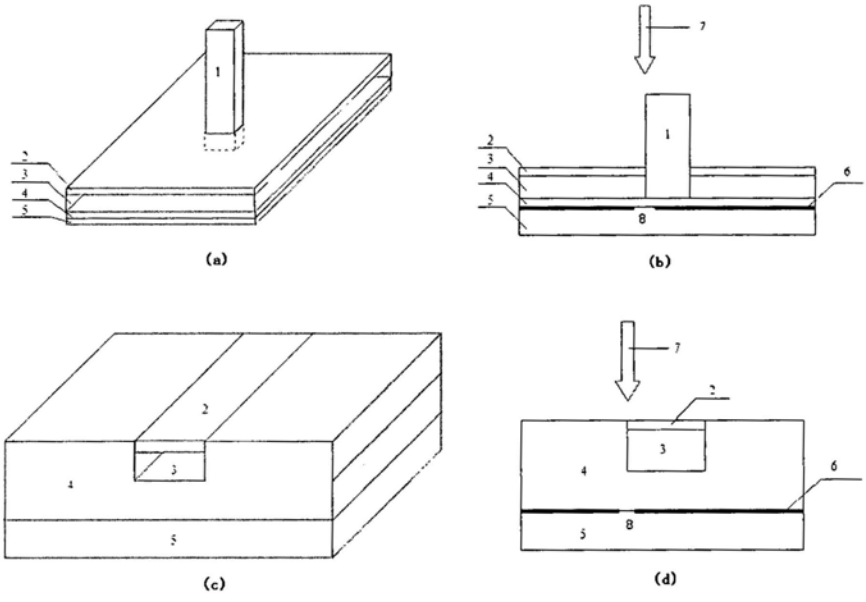


Fig. 6-4 Bio-chemical sensor designs with the self-mixing interferometer

(1 step, 2 cover, 3 bio-chemical solution, 4 substrate, 5 detector, 6 shade, 7 incoming beam, 8 hole)

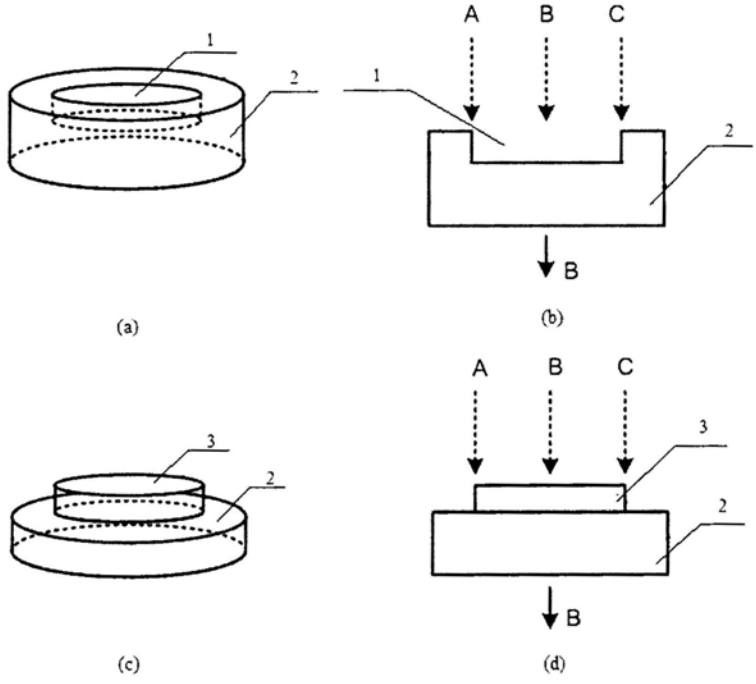


Fig. 6-5 Super-resolution imaging with the self-mixing interferometer

(1 pit, 2 spacer, 3 step)

The principles of the self-mixing interferometers can also be used to design super-resolution lenses. The step and pit structures of the super-resolution lenses are schematically shown in Fig. 6-5(a) and Fig. 6-5(c). Fig. 6-5(b) and Fig. 6-5(d) are cross section view of Fig. 6-5(a) and Fig. 6-5(c) respectively. The imaging system is not shown for simplicity reasons. Assume that three image points (not shown) are formed by imaging beams A, B, and C and the imaging points are too close to be distinguished. Now dispose a self-mixing interferometer at the image plane. The interferometer is arranged such that beams A and C are split by the edges of the pits or steps, preferably being evenly divided. On the other hand, beam B is affected by the phase objects by a much less extent.

Thus, beams A and C are split into two portions with a phase difference. Assume the phase object is designed to generate a pi phase difference between the portions. Then, when the portions mix by themselves, they interfere destructively with each other at a spacer region. Since beam B is much less influenced by the phase objects, its power is much less affected by the destructive interference than the other beams. Therefore, beam B dominates a surviving beam, or beam B is extracted from a mix of three beams. A detector (not shown) may be placed behind the spacer to measure the intensity of beam. Similarly, beams A and C can be retrieved by moving the interferometer setup to the left or right. Therefore, the resolving power is improved using the self-mixing interference principles. Assume the image system of Fig. 6-5(b) already reaches its diffraction limit. Then the improvement makes the system go beyond the diffraction barrier and achieve

super-resolution.

The above mentioned super-resolution embodiments are preferred to work in optical storage system. Compared to the traditional CD and DVD systems, we can fabricate high density disk and read the information with the super-resolution lens. The whole system must be much cheaper than the widely used blue-ray storage system.

The self-mixing interferometer must have other possible applications besides the above examples. We need to do more simulations and experiments to explore the potential applications in future.

Publication List

Conference Paper (Oral Presentation)

1. T. Yang, H. P. Ho, Phase Change Associated with Resonant Surface Plasmon Polariton-Assisted Transmission in Nanohole Arrays, International Conference on Metamaterials, Photonic crystals and Plasmonics (META'10), Cairo-Egypt, 22-25 February 2010.

2. T. Yang, C. C. Li, H. P. Ho, H. X. Zhang, Broadband Spectrometers Based on Nano-scale Difference Interferometers, 3nd IEEE International NanoElectronics Conference (INEC 2010) (accepted), Hong Kong SAR, 3-8 July 2010.

3. T. Yang, C. C. Li, H. P. Ho, Novel Ultra Compact and High Resolution Spectrometer, The 14th Opto-Electronics and Communications Conference (OECC 2009), Hong Kong SAR, 13-17 July 2009.

4. T. Yang, H. P. Ho, Simulation and Analysis of Phase-sensitive Surface Plasmon Resonance Sensor Based on the Enhanced Optical Transmission through Arrays of Nanoholes in Silver Films, OSA-COS Topical Meeting on Applications of Optical Metamaterials (AOM 2009), Tianjin, China, 22-25 June 2009.

5. T. Yang, C. C. Li, H. P. Ho, Novel Ultra Compact and High Resolution Spectrometer Based on the Optical Transmission through a Submicron Interferometer Array, OSA-COS Topical Meeting on Applications of Optical Metamaterials (AOM 2009), Tianjin, China, 22-25 June 2009.

6. T. Yang, H. P. Ho, A Phase-sensitive Surface Plasmon Resonance Sensor Based on the Enhanced Light Transmission through Arrays of Nanoholes in Gold Films, IEEE International Conference of Electron Devices and Solid-State Circuits(EDSSC 2008), Hong Kong SAR, 8-10 December 2008.

7. T. Yang, H. P. Ho, A Phase-sensitive Surface Plasmon Resonance Sensor with Sub-wavelength Apertures in Metallic Thin Film, The International Conference on Laser Applications in Life Sciences 2008 (LALS 2008), Taipei, Taiwan, 4-6 December 2008.

8. T. Yang, K. T. Chan, K. J. Chen, X. Y. Dong, H. Y. Tam, Generation of Continuous THz Wave by a Compact Dual-frequency Semiconductor Laser at Room Temperature, the 20th Annual Meeting of the IEEE Lasers and Electro-Optics Society (LEOS 2007), Lake Buena Vista, Florida, USA, 21-25 October 2007.

9. T. Yang, K. T. Chan, K. J. Chen, Z. C. Liang, A Novel Data Reading System of the WMM Based on the RW-DBR Laser and the AWG, The 7th IEEE Hong Kong LEOS Postgraduate Conference, Hong Kong SAR, 9 December 2006.

Journal Paper

1. T. Yang, H. P. Ho, Computational Investigation of Nanohole Array Based SPR Sensing Using Phase Shift, Optics Express, vol. 17, no. 13, pp. 11205~11216 (2009)

(The paper has been selected by the editor-in-chief, Gregory Faris, for publication in the Virtual Journal for Biomedical Optics (VJBO) Vol. 4, no. 8, 2009)

2. T. Yang, C. C. Li, Z. W. Wang, H. P. Ho, An Ultra Compact Spectrometer Based on the Optical Transmission Through a Micro Interferometer Array, Journal of Lightwave Technology (submitted).

3. T. Yang, H. P. Ho, Simulation and Analysis of Phase-sensitive Surface Plasmon Resonance Sensor Based on the Enhanced Optical Transmission through Arrays of Nanoholes in Silver Films, Chinese Journal of Optics and Applied Optics, vol. 3, no. 1, pp. 57~63 (2010).

4. T. Yang, C. C. Li, H. P. Ho, Novel Ultra Compact and High Resolution Spectrometer Based on the Optical Transmission through a Submicron Interferometer Array, Chinese Journal of Optics and Applied Optics, vol. 3, no. 1, pp. 38~44 (2010).

References

1. T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, P. A. Wolff, "Extraordinary optical transmission through sub-wavelength hole arrays," Nature 391, 667–669, 1998.
2. H. F. Ghaemi, T. Thio, D. E. Grupp, T. W. Ebbesen, H. J. Lezec, "Surface plasmons enhance optical transmission through subwavelength holes," Phys. Rev. B 58, 6779–6782, 1998.
3. L. Martin-Moreno, F. J. Garcia-Vidal, H. J. Lezec, K. M. Pellerin, T. Thio, J. B. Pendry, T. W. Ebbesen, "Theory of extraordinary optical transmission through subwavelength hole arrays," Phys. Rev. Lett. 86, 1114–1117, 2001.
4. L. Salomon, F. D. Grillot, A. V. Zayats, F. de Fornel, "Near-field distribution of optical transmission of periodic subwavelength holes in a metal film," Phys. Rev. Lett. 86, 1110–1113, 2001.
5. A. Krishnan, T. Thio, T. J. Kima, H. J. Lezec, T. W. Ebbesen, P. A. Wolff, J. Pendry, L. Martin-Moreno, F. J. Garcia-Vidal, "Evanescently coupled resonance in surface plasmon enhanced transmission, Opt. Commun. 200, 1–7, 2001.
6. A. Degiron, H. J. Lezec, W. L. Barnes, T. W. Ebbesen, "Effects of hole depth on enhanced light transmission through subwavelength hole arrays," Appl. Phys. Lett. 81, 4327–4329, 2002.

7. N. Bonod, S. Enoch, L. Li, E. Popov, M. Neviere, "Resonant optical transmission through thin metallic films with and without holes," Opt. Exp. 11, 482–490, 2003.

8. C. Genet, M. P. van Exter, J. P. Woerdman: Fano type interpretation of red shifts and red tails in hole array transmission spectra, Opt. Comm. 225, 331-336, 2003.

9. W. L. Barnes, W. A. Murray, J. Ditinger, E. Devaux, T. W. Ebbesen, "Surface plasmon polaritons and their role in the enhanced transmission of light through periodic arrays of subwavelength holes in a metal film," Phys. Rev. Lett. 92, 107401-1–107401-4, 2004.

10. R. Gordon, A. G. Brolo, A. McKinnon, A. Rajora, B. Leathem, K. L. Kavanagh, "Strong polarization in the optical transmission through elliptical nanohole arrays," Phys. Rev. Lett. 92, 37401-1–37401-4, 2004.

11. K. J. Klein Koerkamp, S. Enoch, F. B. Segerink, N. F. van Hulst, L Kuipers, "Strong influence of hole shape on extraordinary transmission through periodic arrays of subwavelength holes," Phys. Rev. Lett. 92, 183901-1–183901-4, 2004.

12. J. Gomez-Rivas, C. Schotsch, P. Haring Bolivar, H. Kurz, "Enhanced transmission of THz radiation through subwavelength holes," Phys. Rev. B 68(20), 201306-1–201306-4 (2003).

13. M. Beruete, M. Sorolla, M. Campillo, J. S. Dolado, L. Martin-Moreno, J. Bravo-Abad, F. J. Garcia-Vidal, "Enhanced millimeter-wave transmission

through subwavelength hole arrays," Opt. Lett. 29, 2500–2502, 2004.

14. H. J. Lezec, A. Degiron, E. Devaux, R. A. Linke, L. Martin-Moreno, F. J. Garcia-Vidal, T. W. Ebbesen, "Beaming light from a subwavelength aperture," Science 297, 820–822, 2002.
15. D. R. Jackson, A. A. Oliner, T. Zhao, J. T. Williams, "Beaming of light at broadside through a subwavelength hole: Leaky wave model and open stopband effect," Radio science 40, RS6S10.1-RS6S10.12,2005
16. D. F. V. James, E. Wolf, "Spectral changes produced in Young's interference experiment," Opt. Commun. 81, 150–154, 1991.
17. D. F. V. James, E. Wolf, "Some new aspects of Young's interference experiment," Phys. Lett. 157, 6-10, 1991.
18. A. J. Fresnel, Oeuvres completes d'Augustin Fresnel, eds. H. de Senarmont, E. Verdet and L. Fresnel, Imprimerie Imperiale, Paris, 1886.
19. F. Zernike, "The concept of degree of coherence and its application to optical problems," Phys. 5, 785-795 ,1938.
20. B. J. Thompson, E. Wolf, "Two-beam interference with partially coherent light," J. Opt. Soc. Am. 47, 895-902, 1957.
21. A. A. Michelson, "Studies in Optics", Univ. of Chicago Press, Chicago, Third impression, Chap. XI, 1962.
22. L. Mandel, E. Wolf, "Spectral coherence and the concept of cross-spectral purity," J. Opt. Soc. Am. 66, 529-*535*, 1976.
23. E. Wolf, "Young's interference fringes with narrow-band light," Opt. Lett. 8,

250-252, 1983.

24. L. Mandel, "Concept of Cross-Spectral Purity in Coherence Theory," J. Opt. Soc. Am. 51, 1342-*1350*, 1961.
25. E. Wolf, "Young's interference fringes with narrow-band light," Optics Lett. 8, 250-252, 1983.
26. E. C. Nice, B. Catimel, "Instrumental biosensors: new perspectives for the analysis of biomolecular interactions, " Bioessays 21, 339-352, 1999.
27. T. Vo-Dinh, B. M. Cullum, "Biosensors and. biochips, advances in biological and medical. diagnostics, Fresenius," J. Anal. Chem. 366, 540-551, 2000.
28. A. G. Brolo, R. Gordon et al, "Surface plasmon sensor based on the enhanced light transmission through arrays of nanoholes in gold films," Langmuir 20, 4813-4815, 2004.
29. S. Y. Wu, H. P. Ho et al, "Highly sensitive differential phase-sensitive surface plasmon resonance biosensor based on the Mach-Zehnder configuration," Opt. Lett. 29, 2378-2380, 2004.
30. X. L. Yu, D. X. Wang, Z. B. Yan, "Simulation and analysis of surface plasmon resonance biosensor based on phase detection," Sens. Actuators B 91, 285–290, 2003.
31. E. Altewischer, M. P. van Exter, and J. P. Woerdman, "Plasmon-assisted transmission of entangled photons," Nature 418, 304-306, 2002.
32. T. Vallius, K. Jefimovs et al, "Optical activity in subwavelength-period arrays of chiral metallic particles," Appl. Phys. Lett. 83, 234-236, 2003.

33. A. Nahata, R. A. Linke, T. Ishi,] K. Ohashi, "Enhanced nonlinear optical conversion from a periodically nanostructured metal film," Opt. Lett. 28, 423-425, 2003.

34. S. Shinada, J. Hashizume, and F. Koyama, "Surface plasmon resonance on microaperture vertical-cavity surface-emitting laser with metal grating," Appl. Phys. Lett. 83, 836-838, 2003.

35. A. Krishnan, T. Thio et al, "Evanescently coupled resonance in surface plasmon enhanced transmission," Opt. Commun. 197, 217-233, 2001.

36. T. J. Kim, T. Thio, T. W. Ebbesen et al, "Control of optical transmission through metals perforated with subwavelength hole arrays," Opt. Lett. 24, 256-258 1999.

37. S. M. Williams, A. D. Stafford et al, "Accessing surface plasmons with Ni microarrays for enhanced IR absorption by monolayers," J. Phys. Chem. B 107, 11871-11879, 2003.

38. T. Rindzevicius, Y. Alaverdyan, A. Dahlin, F. Hook, D. S. Sutherland, and M. Kall, "Plasmonic sensing characteristics of single nanometric holes," Nano Lett. 5, 2335-2339, 2005.

39. K. A. Tetz, L. Pang and Y. Fainman, "High-resolution surface plasmon resonance sensor based on linewidth-optimized nanohole array transmittance," Opt. Lett. 31, 1528-1530, 2006.

40. X. L. Yu, L. Q. Zhao, H. Jiang et al, "Immunosensor based on optical heterodyne phase detection," Sens. Actuators B 76, 199-202, 2001.

41. K. H. Chen, C. C. Hsu, D. C. Su, "Measurement of wavelength shift by using surface plasmon resonance heterodyne interferometry," Opt. Commun. 209, 167-172, 2002.

42. C. M. Wua, Z. C. Jian, S. F. Joec, L. B. Chang, "High-sensitivity sensor based on surface plasmon resonance and heterodyne interferometry," Sens. Actuators B 92, 133-136, 2003.

43. S. G. Nelson, K. S. Johnston, S. S. Yee, "High sensitivity surface plasmon resonance sensor based on phase detection," Sens. Actuators B 35, 187-191, 1996.

44. W. L. Barnes, A. Dereux, T. W. Ebbesen, "Surface plasmon subwavelength optics," Nature 424, 824-830, 2003.

45. J. Homola, "Surface plasmon resonance based sensors," Springer-Verlag Berlin Heidelberg, USA, 2006.

46. http://benjaminevans.net/files/surface_plasmons.pdf

47. V. M. Shalaev, S. Kawata, "Nanophotonics with surface plasmons," Elsevier B.V, The Netherlands, 2007.

48. S. A. Maier, "Plasmonics: Fundamentals and Applications," Springer Science Business Media LLC, United Kingdom, 2007.

49. M. L. Brongersma, P. G. Kik , "Surface plasmon nanophotonics," Elsevier, The Netherlands, 2007.

50. W. Yuan, "Surface plasmon enhanced effects in photonic biosensors," Ph. D. thesis, The Chinese University of Hong Kong, 2008

51. http://en.wikipedia.org/wiki/Surface_plasmon

52. H. Raether, "Surface plasmons with on smooth and rough surfaces and on gratings," Springer- Verlag Berlin Heidelberg, Hamburg, 1986.

53. H. P. Ho, W. W. Lam, "Application of differential phase measurement technique to surface plasmon resonance sensors," Sens. Actuators B 96, 554-559, 2003.

54. K. Matsubara, S. Kawata and S. Minami, "Optical chemical sensor based on surface plasmon measurement," Appl. Opt. 27, 1160-1163, 1988.

55. B. Liedberg, I. Lundstrom, E. Stenberg, "Principles of biosensing with an extended coupling matrix and surface plasmon resonance," Sens. Actuators B 11, 63–72, 1993.

56. L. M. Zhang, D. Uttamchandani, "Optical chemical sensing employing surface plasmon resonance," Elec. Lett. 23, 1469–1470, 1988.

57. C. Nylander, B. Liedberg and T. Lind, "Gas detection by means of surface plasmon resonance, Sens. Actuators 3, 79–88, 1982.

58. B. Liedberg, C. Nylander and I. Lundström, "Surface plasmon resonance for gas detection and biosensing," Sensors and Actuators 4, 299–304, 1983.

59. A. V. Kabashin, V. E. Kochergin, P. I. Nikitin, "Surface plasmon resonance bio- and chemical sensors with phase-polarisation contrast," Sens. Actuators B: Chemical 54, 51-56, 1999.

60. M. Born, E. Wolf, "Principle of optics," Pergamon, Oxford, 1980.

61. P. Yeh, "Optical waves in layered media," Wiley, New York, 1988.

62. B. Ran, S. G. Lipson, "Comparison between sensitivities of phase and intensity detection in surface plasmon resonance," Opt. Exp. 16, 5641-5650, 2006.

63. A. V. Kabashin , S. Patskovsky, A. N. Grigorenko, "Phase and amplitude sensitivities in surface plasmon resonance bio and chemical sensing," Opt. Exp. 17, 21191-21204, 2009.

64. http://www.capovani.com/dp/cat/107/63160/iinfo.cfm?LCl=986&TVTID=0&TItemNo=0&ItemNo=422&q=2.

65. R. B. M. Schasfoort, and A. J. Tudos, eds., "Handbook of Surface Plasmon Resonance," Royal Society of Chemistry, 2008.

66. M. Born, E. Wolf, "Principles of Optics," Cambridge University Press, Cambridge, UK, 2002.

67. A. Waksberg, and J. Wood, "Noise power spectrum characteristics for an HeNe laser operating under various discharge conditions," Rev. Sci. Instrum. 40, 1306–1313, 1969.

68. K. G. Baigent, D. A. Shaddock, M. B. Gray, D. E. McClelland, "Laser stabilisation for the measurement of thermal Noise," Gen. Relativ. Gravit. 32, 399–409, 2000.

69. G. Keiser, "Optical Communications Essentials," McGraw-Hill, 2003.

70. J. M. Liu, "Photonic devices," University press, Cambridge, 2005.

71. A. V. Kabashin, P. I. Nikitin, A. A. Beloglazov, "A method of examining biological, biochemical and chemical characteristics of a medium and apparatus for its embodiment," PCT Patent WO 98/57149,1998; US 6 628 376, 2003.

72. A. N. Grigorenko, P. I. Nikitin, A. V. Kabashin, "Phase Jumps and Interferometric Surface Plasmon Resonance Imaging," Appl. Phys. Lett. 75, 3917–3919, 1999.

73. A. G. Brolo, R. Gordon, B. Leathem, K. L. Kavanagh, "Surface plasmon sensor based on the enhanced light transmission through arrays of nanoholes in Gold Films," Langmuir 20, 4813-4815, 2004.

74. T. Rindzevicius, Y. Alaverdyan, A. Dahlin, F. Hook, D. S. Sutherland, M. Kall, "Plasmonic sensing characteristics of single nanometric holes," Nano Lett. 5, 2335-2339, 2005.

75. K. A. Tetz, L. Pang, Y. Fainman, "High-resolution surface plasmon resonance sensor based on linewidth-optimized nanohole array transmittance," Opt. Lett. 31, 1528-1530, 2006.

76. C. Genet, M. P. van Exter, and J. P. Woerdman, "Huygens description of resonance phenomena in subwavelength hole arrays," J. Opt. Soc. Am. 22, 998-1002, 2005.

77. L. S. Jung, C. T. Campbell et al, "Quantitative interpretation of the response of surface plasmon resonance sensors to adsorbed films," Langmuir 14, 5636-5648, 1998.

78. J. B. Pendry, L. Martin-Moreno, F. J. Garcia-Vidal, "Mimicking surface plasmons with structured surfaces," Science 305, 847-848, 2004.

79. L. Martín-Moreno, F. J. García-Vidal et al, "Theory of extraordinary optical transmission through subwavelength hole arrays," Phys. Rev. Lett. 86,

1114-1117, 2001

80. K. L. van der Molen, F. B. Segerink, and N. F. van Hulst, "Influence of hole size on the extraordinary transmission through subwavelength hole arrays," Appl. Phys. Lett. 85, 4316-4318, 2004.
81. K. J. Klein Koerkamp, S. Enoch, F. B. Segerink, N. F. van Hulst, L. Kuipers "Strong influence of hole shape on extraordinary transmission through periodic arrays of subwavelength holes," Phys. Rev. Lett. 92, 183901-1 - 183901-4, 2004.
82. X. L. Yu, L. Q. Zhao, H. Jiang et al, "Immunosensor based on optical heterodyne phase detection," Sens. Actuators B 76, 199-202, 2001.
83. K. H. Chen, C. C. Hsu, D. C. Su, "Measurement of wavelength shift by using surface plasmon resonance heterodyne interferometry," Opt. Commun. 209, 167-172, 2002.
84. C. M. Wua, Z. C. Jian, S. F. Joec, L. B. Chang, "High-sensitivity sensor based on surface plasmon resonance and heterodyne interferometry," Sens. Actuators B 92, 133-136, 2003.
85. S. G. Nelson, K. S. Johnston, S. S. Yee, "High sensitivity surface plasmon resonance sensor based on phase detection," Sens. Actuators B 35, 187-191, 1996.
86. http://ab-initio.mit.edu/wiki/index.php/Meep_Introduction
87. http://www.lumerical.com/fdtd_parallel_description.php
88. http://www.lumerical.com/fdtd_online_help

89. http://camfr.sourceforge.net/

90. L. Martın-Moreno, F. J. Garcıa-Vidal, "Optical transmission through circular hole arrays in optically thick metal films", Opt. Express 12, 3619-3628, 2004.

91. http://en.wikipedia.org/wiki/Finite-difference_time-domain_method

92. K. Yee, "Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media", Ante. and Propa. 14, 302 - 307, 1966.

93. A. Taflove and S. C. Hagness, "Computational Electrodynamics: The Finite-Difference Time-Domain Method (3rd ed.)," Artech House Publishers, 2005.

94. C. B. Scruby, L. E. Drain, "Laser ultrasonics techniques and applications (Chap. 3.)," LOP Publishing Ltd, 1990.

95. R. L. Rich, D. Myszka, "Survey of the year 2004 commercial optical biosensor literature," J. Mol. Recognit. 18, 457- 478, 2005.

96. D. K. Kambhampati, W. Knoll, "Surface-plasmon optical techniques, Current Opinion in Colloid & Interface," Science 4, 273-280, 1999.

97. J. Homola, "Present and future of surface plasmon resonance biosensors," Anal. Bioanal. Chem. 377, 528-539, 2003.

98. G. Gauglitz, T. Vo-Dinh, "Handbook of Spectroscopy," Wiley, 2003.

99. http://en.wikipedia.org/wiki/Spectrometer

100. B. Stuart, "Infrared spectroscopy – fundamentals and applications (Analytical techniques in the sciences (AnTs)," Wiley, 2004.

101. S. Shen, "Fourier transform spectroscopy introduction and progress," Progress

in Phys. 2, 275-322, 1982.

102.http://en.wikipedia.org/wiki/Nyquist-Shannon_sampling_theorem

103.C. C. Li, Oliver C. S. Choy, H. P. Ho, "Submicron free space optical interferometer," IEEE International Conference of Electron Devices and Solid-State Circuits(EDSSC 2008), Hong Kong SAR, 8-10 December 2008.

104.E. Wolf, "Invariance of the spectrum of light on propagation," Phys. Rev. Lett. 56, 1370-1372, 1986.

105.Z. Y. Chen, J. X. Pu, "Analytical characterization of spectral anomalies in Young's double-slit Interference Experiment," Acta Photonica Sinica 36, 733-737, 2007.

106.J. Honerkamp, J. Weese, Tikhonov, "Regularization method for ill-posed problems," Continuum Mech. Thermodyn. 2, 17-30, 1990.

107.A. N. Tikhonov, "On the stability of inverse problems," C. R. Acad. Sci. URSS (Nauk SSSR) 39, 195-198, 1943.

108.Y. F. Wang, Computational Methods for inverse problems and their applications, Beijing, China: Higher Education Press, 76-84, 2007.

109.A. E. Hoerl, "Application of ridge analysis to regression problems," Chemical Engineering Progress 58, 54-59, 1962.

110.M. Foster, "An application of the Wiener-Kolmogorov smoothing theory to matrix inversion," J. Soc. Indust. and Appl. Math. 9, 387-392, 1961.

111.D. L. Phillips, "A technique for the numerical solution of certain integral equations of the first kind," J. Assoc. Comput. Mach., 9, 84-97, 1961.

112.A. Tarantola, Inverse Problem Theory and Methods for Model Parameter Estimation, Society for Industrial and Applied Mathematics, 2004.

113.L. M. Wu, "A parameter choice method for Tikhonov regularization," Electronic Transactions on Numerical Analysis 16, 107-128, 2003.

114.J. L. Xie, J. Zou, "An improved model function method for choosing regularization parameters in linear inverse problems," Inverse Problems 18, 631-643, 2002.

115.J. Cheng, M. Yamamoto, "One new strategy for a priori choice of regularization parameters in Tikhonov's regularization," Inverse Problems 16, 31-38, 2000.

116.C. W. Groetsch, Inverse Problems in Mathematical Sciences, Vieweg, Braunschweig: Informatica International, Inc., 1993.

117.Z. W. Wang, J. Liu, New model function methods for determining regularization parameters in linear inverse problems, Applied Numerical Mathematics 59, 2489–2506, 2009.

118.F. M. Dickey, S. C. Holswade, D. L. Shealy, "Laser Beam shaping applications," CRC Press, Taylor & Francis Group, USA, 20-294, 2006.

119.J. Spigulis, D. Pfafrods, M. Stafeckis, and W. Jelinska-Platace, "The 'glowing' optical fibre designs and parameters," Proc. SPIE 2967, 226–236, 1996.

120.J. Spigulis and D. Pfafrods, "Clinical potential of the side-glowing optical fibers," Proc. SPIE 2977, 84–88, 1997.

121.T. Yang, Z. C. Liang, W. P. Qin, X. B. Cai, J. B. Chen, S. L. Zhuang, "Side-scattering properties of a novel waveguide multilayer memory," Optical

Engineering 45, 095201-1–095201-6, 2006.

122.T. Yang, Z. C. Liang, X. B. Cai, "Principles and experiments of uniform scattering waveguide multilayer memory," Chinese J. Lasers 33, 1168-1171, 2006.

123.W. S. Cheng, J. Zhao, H. G. Cai, "Correction method for pixel response nonuniformity of CCD," Optics and Precision Engineering 16, 314-318, 2008.

124.C. Ataman, H. Urey, "Compact Fourier transform spectrometers using FR4 platform," Sens. Actuators A 151, 9-16, 2009.

125.A. Schliesser, M. Brehm, F. Keilmann, "Frequency-comb infrared spectrometer for rapid, remote chemical sensing," Opt. Express 13, 9029-9039, 2005.

126.M. E. Gehm, R. John, D. J. Brady, R. M. Willett, T. J. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," Opt. Express 15, 14013-14027, 2007.

127.C. Ataman, H. Urey, A. Wolter, "A Fourier transform spectrometer using resonant vertical comb actuators," J. Micromechanics and Microengineering 16, 2517-2523, 2006.

128.J. C. Yang, H. Jung et al, "Micro-Electro-Mechanical- Systems-Based infrared spectrometer composed of multi-slit grating and bolometer array," Jap. J. of Appl. Phys. 47, 6943-6948, 2008.

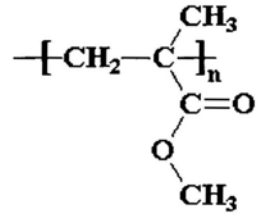
129.http://www.lumerical.com/fdtd_online_help/user_guide_planewave_edge.php

Appendix: Properties of Some Polymers

The material of the sensors and spectrometers are polymers such as PMMA or PC. Here we have a brief introduction of the polymers. The following information comes from the website: http://en.wikipedia.org

Appendix A: PMMA

1) Introduction:



Poly (methyl methacrylate) (PMMA) poly (methyl 2-methylpropenoate) is a transparent thermoplastic. Chemically, it is the synthetic polymer of methyl methacrylate. It is sold under many trade names, including Policril, Plexiglas, Gavrieli, Vitroflex, Limacryl, R-Cast, Per-Clax, Perspex, Plazcryl, Acrylex, Acrylite, Acrylplast, Altuglas, Polycast, Oroglass, Optix and Lucite and is commonly called acrylic glass, simply acrylic, perspex or plexiglas. Acrylic, or acrylic fiber, can also refer to polymers or copolymers containing polyacrylonitrile. The material was developed in 1928 in various laboratories and was brought to market in 1933 by Rohm and Haas Company.

2) Properties

PMMA is often used as an alternative to glass, and in competition with polycarbonate (PC). It is often preferred because of its moderate properties, easy handling and processing, and low cost, but behaves in a brittle manner when loaded, especially under an impact force. To produce 1 kg of PMMA, about 2 kg of petroleum is needed. PMMA ignites at 460 °C and burns, forming carbon dioxide, water, carbon monoxide and low molecular weight compounds, including formaldehyde. The elaborate introduction of properties is as follows:

PMMA:

* has a density of 1,150–1,190 kg/m^3. This is less than half the density of glass, and similar to that of other plastics.

* has a good impact strength higher than that of glass or polystyrene, but significantly lower than that of polycarbonate or engineering polymers. In the majority of applications, it will not shatter but instead breaks into large dull pieces.

* is softer and more easily scratched than glass. Scratch-resistant coatings (which may also have other functions) are often added to PMMA sheets.

* transmits up to 92% of visible light (3 mm thickness), and gives a reflection of about 4% from each of its surfaces on account of its refractive index of 1.4893 to 1.4899.

* filters ultraviolet (UV) light at wavelengths below about 300 nm. Some manufacturers add coatings or additives to PMMA to improve absorption in the 300–400 nm range.

* allows infrared light of up to 2800 nm wavelength to pass. IR of longer wavelengths, up to 25,000 nm, are essentially blocked. Special formulations of colored PMMA exist to allow specific IR wavelengths to pass while blocking visible light (for remote control or heat sensor applications, for example).

* has excellent environmental stability compared to other plastics such as polycarbonate, and is therefore often the material of choice for outdoors applications.

* has poor resistance to solvents, as it swells and dissolves easily. It also has poor resistance to many other chemicals on account of its easily hydrolyzed ester groups.

3) Modification of properties

Pure poly (methyl methacrylate) homopolymer is rarely sold as an end product, since it is not optimized for most applications. Rather, modified formulations with varying amounts of other comonomers, additives, and fillers are created for uses where specific properties are required. For example,

* A small amount of acrylate comonomers are routinely used in PMMA grades destined for heat-processing, since this stabilizes the polymer to depolymerization ("unzipping") during processing.

* Comonomers such as butyl acrylate are often added to improve impact strength.

* Comonomers such as methacrylic acid can be added to increase the glass

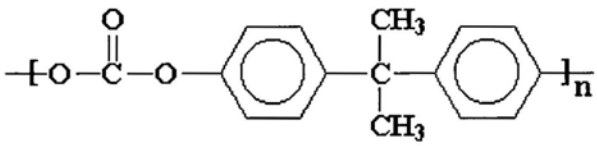
transition temperature of the polymer for higher temperature use such as in lighting applications.

* Plasticizers may be added to improve processing properties, lower the glass transition temperature, or improve impact properties.

* Dyes may be added to give color for decorative applications, or to protect against (or filter) UV light.

* Fillers may be added to improve cost-effectiveness.

Appendix B: PC



Polycarbonates (PC) received their name because they are polymers having functional groups linked together by carbonate groups (-O-(C=O)-O-) in a long molecular chain. Also carbon monoxide was used as a C1-synthon on an industrial scale to produce diphenyl carbonate, being later trans-esterified with a diphenolic derivative affording poly (aromatic carbonate)s.

Taking into consideration the C1-synthon polycarbonates can be divided into poly(aromatic carbonate)s and poly(aliphatic carbonate)s. The second one, poly(aliphatic carbonate)s are a product of the reaction of carbon dioxide with epoxides, which owing to the kinetic stability of carbon dioxide requires the use of a catalyst. The working systems are based on porphyrins, alkoxides, carboxylates, salens and beta-diiminates as organic, chelating ligands and aluminium, zinc, cobalt and chromium as the metal centres. Poly(aliphatic carbonate)s display promising characteristics, have a better biodegradability than the aromatic ones and could be employed to develop other specialty polymers.

The most common type of polycarbonate plastic is made from bisphenol A (BPA). This polycarbonate is a very durable material, and can be laminated to make bullet-proof "glass", though “bullet-resistant” would be more accurate. Although polycarbonate has high impact-resistance, it has low scratch-resistance and so a

hard coating is applied to polycarbonate eyewear lenses and polycarbonate exterior automotive components. The characteristics of polycarbonate are quite like those of polymethyl methacrylate (PMMA; acrylic), but polycarbonate is stronger, usable in a larger temperature range and more expensive. This polymer is highly transparent to visible light and has better light transmission characteristics than many kinds of glass. CR-39 is a specific polycarbonate material — although it is usually referred to as CR-39 plastic — with good optical and mechanical properties, frequently used for eyeglass lenses.

Polycarbonate has a variety of applications, from bulletproof windows to compact disks (CDs). The main advantage of polycarbonate over other types of plastic is unbeatable strength combined with light weight. While acrylic is 17% stronger than glass, polycarbonate is nearly unbreakable. Bulletproof windows and enclosures as seen inside banks or at drive-throughs are often made of polycarbonate. Add to this the advantage that polycarbonate is just one-third the weight of acrylic, or one-sixth as heavy as glass, and the obvious drawback is that it is more expensive than either acrylic or glass.