NOVEL OPTICAL DEVICES FOR INFORMATION PROCESSING

A Dissertation

by

ZHIJIE DENG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2006

Major Subject: Electrical Engineering

NOVEL OPTICAL DEVICES FOR INFORMATION PROCESSING

A Dissertation

by

ZHIJIE DENG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Philip R. Hemmer
Committee Members,	Henry F. Taylor
	Kai Chang
	Richard Arnowitt
Head of Department,	Costas N. Georghiades

May 2006

Major Subject: Electrical Engineering

ABSTRACT

Novel Optical Devices for Information Processing.

(May 2006)

Zhijie Deng, B. Eng., Tsinghua University, Beijing, China

Chair of Advisory Committee: Dr. Philip R. Hemmer

Optics has the inherent advantages of parallelism and wide bandwidths in processing information. However, the need to interface with electronics creates a bottleneck that eliminates many of these advantages. The proposed research explores novel optical devices and techniques to overcome some of these bottlenecks. To address parallelism issues we take a specific example of a content-addressable memory that can recognize images. Image recognition is an important task that in principle can be done rapidly using the natural parallelism of optics. However in practice, when presented with incomplete or erroneous information, image recognition often fails to give the correct answer. To address this problem we examine a scheme based on free-space interconnects implemented with diffractive optics. For bandwidth issues, we study possible ways to eliminate the electronic conversion bottleneck by exploring all-optical buffer memories and all-optical processing elements. For buffer memories we examine the specific example of slow light delay lines. Although this is currently a popular research topic, there are fundamental issues of the delay-time-bandwidth product that must be solved before slow light delay lines can find practical applications. For all-optical processing we examine the feasibility of constructing circuit elements that operate directly at optical frequencies to perform simple processing tasks. Here we concentrate on the simplest element, a sub-wavelength optical wire, along with a grating coupler to interface with conventional optical elements such as lenses and fibers. Even such a simple element as a wire has numerous potential applications.

In conclusion, information processing by all-optical devices are demonstrated with an associative memory using diffractive optics, an all-optical delay line using room temperature slow light in photorefractive crystals, and a subwavelength optical circuit by surface plasmon effects.

To my parents

and

beloved late grandfather

Shenghua Chen.

ACKNOWLEDGMENTS

Any creations are products of cooperations. This dissertation would not be possible without my family, mentors, colleagues, friends or even strangers I barely know. The list is too long for a complete acknowledgement here. I would take this opportunity to give special thanks to my advisor, Dr. Philip Hemmer, for his support, guidance, encouragement and patience. In addtion, I want to thank my committee members, Dr. Henry Taylor, Dr. Kai Chang and Dr. Richard Arnowitt, for their tutelage and instructions. It was a pleasure to work with Dr. M. Suhail Zubairy on the associative memory and slow light projects. I have also learned a lot from my colleagues HoNam Yum, Chang-Seok Shin, Mughees Khan, Dr. De-Kui Qing, Dr. Elizabeth Trajkov, Huiliang Zhang and Raheel Zubairy. Dr. Winfried Teizer and Arlene Ford have been tremendous help on the electron beam lithography. Yulia Vasilyeva and Shyamashree Chatterjee were always ready to deliver assistance in the experiments at the TAMU Materials Characterization Facility (MCF). Robert Atkins has contributed to our lab in many ways. Dr. Andreas Holzenburg, Dr. Mike Pendleton and Tom Stephens gave me quite some advice on my experiments at the TAMU Microscopy and Imaging Center (MIC). Dr. Andreas Kronenberg was very generous in helping us on the measurements at the Department of Geology & Geophysics, TAMU. I would also like to thank Dr. Konrad Bussman at the Naval Research Laboratory, Dr. Robert Armstrong at New Mexico State University, Dr. Mikhail Lukin and Darrick Chang at Harvard University, for their helpful discussions during my visits. And last, I want to thank my parents for their unconditional love and support all these years of long term separation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

A.   Overview

Optics has been reshaping the landscape of information processing with its inherent parallelism and wide bandwidths. Optical fibers and dense wavelength division multiplexing (DWDM) are the corner stones of the information world we are living in. Active research in optical devices is underway to address the information processing bottlenecks. One example is the optical interconnects between microprocessors. The computation power of modern processors has been growing exponentially, until recently, following the famous Moore's law. But the data input/output (I/O) interface cannot keep up with it. It is projected that the computing power growth over a decade will be three times of the I/O capacity growth in the same period [1]. In the near-term, inter-chip communication will become the computer performance bottleneck. People are turning to optical fibers when electrical wires cannot handle the communication task. Agilent Technologies demonstrated a multiple-fiber, multiple-wavelength prototype that has a 10 Gbps (giga bits per second) capacity applicable to inter-chip data exchange in servers [2]. More aggressive research has been reported by Intel Corporation, with the final goal to integrate the whole photonic interconnect system onto silicon substrates [3–5]. Another example is the optical wireless local area network (LAN), where free space optical links may replace the radio frequency (RF) links in the Wi-Fi (wireless fidelity) LAN under specific environments. A data rate of 155 Mbps was reported in 2001 [6], which is almost three times of the aggregate

bandwidth (54 Mbps) in the current Wi-Fi standard—IEEE 802.11g. Texas Instruments (TI) demonstrated the multiple-link Optical Wireless Solutions^{TM}, where each dedicated connection has a capacity of 100 Mbps, with more than 10 Gbps bandwidth in aggregate. In comparison, the next generation Wi-Fi standard IEEE 802.11n will feature a bandwidth of 600 Mbps shared among multiple connections.

Above are examples of largely serial optical networking in communications, which is a well developed part of information processing. It is also known that optics is well suited to parallel processing, especially for tasks involving images such as feature recognition or extraction. However, in real world situations feature recognition has often proved unreliable due to its emphasis on precise pattern matching. Any deviation in the input due to incomplete or distorted information rapidly degrades the selectivity of pattern matching. To address this problem we examine technologies that are far from fully developed—artificial intelligence (AI). Among the many AI algorithms, neural networks (NN) are perhaps the best known ones. In neural networks models, information is decomposed into pieces, i.e., neurons, which are linked to each other by some weights. The weights specify the strengths by which the information pieces relate to each other. In case some pieces are missing or unclear, they are hoped to be recovered from the remaining ones through the links. It is immediately seen that the weighted links can be implemented by properly designed optical interconnects. A diffractive optical element (DOE) can make such interconnections in free space. Since the interconnections can be established in parallel, it becomes a great advantage over software implementations in conventional computers, where two-dimensional input data must be converted and handled serially by each central processing unit (CPU).

The neural networks can be regarded as memory that associate fractions of information with each others. Memories, or buffers are also desired in controlling information flow over the optical communication networks. One example is the congestion

control. Since data is transmitted in the form of optical pulses over fiber optic networks, problems arise when several pulse trains simultaneously request right of the same route. Certain mechanisms will be helpful if they can save the pulse trains of low priorities to yield for the high priority one. Currently, no good optical buffer memories exist to deal with this need, and so the optical signal must first be converted to an electronic signal, which severely limits the overall network speed. In 1999, Hau *et al.* reported that the group velocity of light was slowed down to the unusual 17 m/s in an ultra cold atomic gas [7]. This has spurred ultraslow light research for applications in optical buffers, delay lines and quantum information storage. For the sake of commercial competitiveness, ultraslow light has to be realized at room temperature by solid-state devices. However, there is an inherent conflict between the delay time and bandwidth, which limits the device capacities to an unacceptable value. The capacity limit is represented by the delay-time-bandwidth product which is related to the length of an optical pulse stream that can be buffered by this technique. Here, we will study the problem of delay-time-bandwidth product in slow light and a discuss a possible solution.

Ultraslow light is of interest because it can greatly reduce the dimensions and latency time of optical buffers based on fiber optic delay lines. Otherwise, a delay line for 1 ms can easily take hundreds of kilometers (km) of optical fibers. Applying this approach to optical processing elements, more ambitious miniaturizations will eventually shrink photonic component sizes below the corresponding wavelength. By confining light in subwavelength space, not only can we pack more information processing power, but also discover novel photonic devices where light interacts with materials in an unprecedentedly strong manner. These novel devices may serve as key components, such as light sources and detectors, in the information infrastructures tomorrow. One fundamental element of such infrastructures will be the subwave-

length waveguide, or optical wire. Accordingly, we will investigate the properties of subwavelength waveguides that utilize the surface plasmon effect.

The three examples discussed above will be elaborated in the following sections.

B.   Associative memory by diffractive optics

Content-addressable associative memory find broad applications such as image recognitions for counterfeit and access control purposes. The most efficient machines are human brains up to now, which can associate partial information with its original version stored previously. In 1982, Hopfield presented a neural networks model that emulates the operations of biological intelligence [8]. Digitized information is represented by a group of neurons. They are binarized memory cells, which are close to the bits in a computer. The basic operations in the algorithm are weighted interconnections and thresholding (see Fig. 1). The input neurons are first interlinked to form "intermediate neurons", which refer to the interconnect outputs hereafter. The interconnections are weighted by predefined values to emulate the association strengths between the biological neurons in human brains. Non-binary numbers are thus stored in the intermediate neurons. They are then converted into the binary output neurons by comparison to a threshold value. Fig. 1 illustrates the process, where $\{V^{in}\}$ are the input neurons, $\{V^{s}\}$ are the intermediate neurons, $\{V^{out}\}$ are the output neurons, $\{W\}$ are the interconnection weights and T stands for the thresholding operation. The output neurons can be fed back into the input for further iterations if necessary.

The interconnection operation is described in mathematics as a multiplication of a matrix (with elements $\{W\}$) to a vector (formed by $\{V^{in}\}$). The CPU in a conventional computer is designed to take care of scaler multiplications rather than matrix multiplications. It can only make one interconnection in each computation
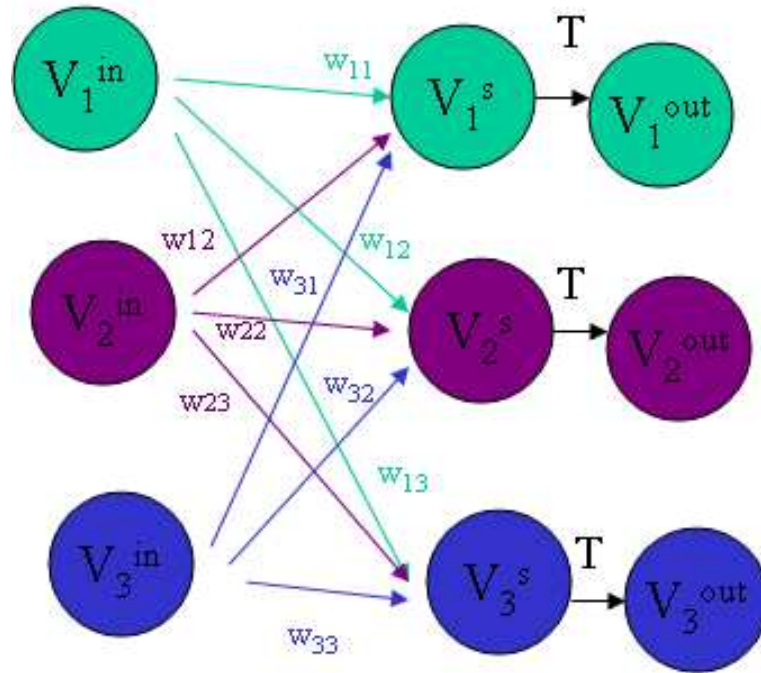
Fig. 1. Operations in the Hopfield model. $\{V^{in}\}$: input neurons; $\{V^s\}$: intermediate neurons; $\{V^{out}\}$: output neurons; $\{W\}$: interconnection weights; T: thresholding operation.

cycle. The interconnection time grows by the square power of the number of neurons. Consequently, conventional computers become extremely inefficient for high volume associative memory, as in the general case of image recognition. The sluggishness caused by the serial processing within CPUs can be overcome with the parallelism of optics in establishing multiple interconnections [9–13]. The parallelism is inherent in free space optical interconnects. Photorefractive holograms [10, 12] and lenslet arrays [11, 13] have been used for this purpose, but they are relatively expensive components. It will be desirable to replace them with components that can be mass produced with mainstream technologies to lower the cost. Diffractive optical elements fall into this category. They have been demonstrated for inter-chip optical intercon-

nects, for instance, as reported by Fey *et al.* [14]. Computer-generated-holograms (CGHs) are one sort of DOEs, with a history traced back to the 1960s [15, 16]. The Lohmann type CGHs can be fabricated by coating metals on a glass substrate and then opening apertures in the metal with lithography and etching. These procedures are commonplace in the semiconductor industry. Once the apertures to open are designed with Fourier optics simulations, CGHs can be massively produced by flat panel display foundries without much difficulties. CGHs manufactured this way will greatly reduce the system price of the associative memory.

Complexities and costs of the system can also come from the thresholding part in the associative memory. Hopfield's original model utilizes negative interconnection weights. Despite the advantage of a constant threshold, it introduces subtractions that are much harder to implement in optics than additions. Psaltis *et al.* addressed this issue by space multiplexing, which doubles the required interconnections [9]. Gao *et al.* followed similar strategy with time-multiplexed interconnections, which doubles the required processing time [13]. The multiplexing overheads can be eliminated by adding an offset to Hopfield's interconnection weights, so that they are all non-negative to avoid subtractions [12, 17]. In the proposed associative memory system, we will choose the model modified in this way by Rizvi and Zubairy [17].

A problem remains that the Rizvi-Zubairy-like models create input dependent thresholds rather than a constant one. Song *et al.* handled this problem by incorporating additional detectors to measure the threshold. To complete the thresholding, a comparison is necessary between outputs from the intermediate-neuron-measurement system and the output from the threshold-measurement system. Such comparisons will become more difficult if there are discrepancies in dynamic ranges and linearities of the two systems. It is noted that plenty of detectors are available on a charge-couple-device (CCD) array in a camera. Therefore, we propose a proper design of the

optics, so that the optical interconnect outputs will fall on one part of the CCD array while the other part can be reserved for measuring the threshold. This will further streamline the system and lower the cost. The thresholding is also expected to be more accurate, since pixel performances are more uniform on the same CCD chip.
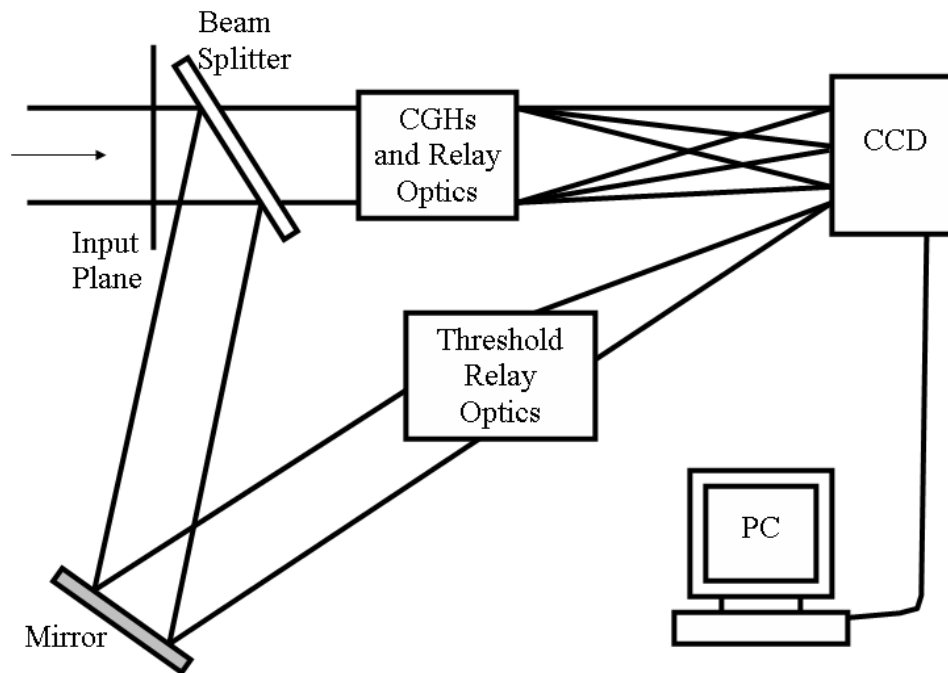
Fig. 2. Proposed architecture for the associative memory.

Briefly summarized, work is proposed to implement the Rizvi-Zubairy associative memory model with computer-generated-holograms, which is economical for mass production. A thresholding scheme is proposed to streamline the system and further reduce the cost, by merging the detection systems for the interconnect outputs and the threshold. Fig. 2 shows the architecture to test the associative memory.

C.   Delay-time-bandwidth problem in room temperature slow Light

The group velocity slowdown of light has been known to people for quite some time, for example, the pulse delay in optical waveguides due to their dispersions. But the delay was too small for realistic optical buffers and delay lines [18]. In 1999, the interests were renewed with Hau's report on the ultra slow group velocity of 17 m/s [7]. It is attributed to the ultra narrow transmission bandwidth for the light going through the cold atoms. According to the Kramers-Kronig theory, it is accompanied by an extremely strong dispersion in the refractive index, which greatly delays light pulses in the medium. Ultraslow light opens opportunities for potential applications in optical communications (all optical buffers, ultra-sensitive switches), radar beam-steering (true time delay), and quantum information storage.

A major challenge to current slow light technologies is the substantial distortion or short delay times of the pulses. It is a dilemma inherent in the Kramers-Kronig analysis, where the narrow bandwidth of a system implies a large group delay and the loss of high frequency components simultaneously. This problem is often characterized by the delay-time-bandwidth product. There may be illusions that stopped light can solve the problem. For example, in photorefractive crystals, stopped light has been demonstrated [19] and theoretically can have a storage time up to months, but only one pulse can be saved in the crystal. A figure of merit would be the number of pulses that can be slowed (or stopped) in the material simultaneously, which is equal to the delay-time-bandwidth product in ultraslow light. The delay-time-bandwidth product indicates the capacity or word length of optical buffers and delay lines.

Key applications generally require delay-time-bandwidth products of 100-1000 or more. The best result so far is about five in lead vapor [20], that is to say, the delay time is five times of the pulse width. Electromagnetic induced transparency (EIT) [7]

can overcome this limit in principle with a highly absorptive material and applying enough pump laser power to saturate the EIT transition. It is possible in vapors, because the transmission window is very narrow, and the material can withstand a high laser power. This scheme is also applicable to EIT in certain solids under cryogenic conditions, such as Pr:YSO [21]. However, in room temperature solids, which are strongly preferred for commercially competitive applications, the windows are significantly broadened by phonons or other mechanisms. The saturation requires a higher pump power, which would easily exceed the damage threshold of the material well before 100 pulses could be slowed. This analysis applies to most if not all slow light and stopped light techniques demonstrated so far.

To increase the delay-time-bandwidth, the pulse distortions can be suppressed by simultaneously slowing down all the input harmonic components. It can be done by piecing together multiple neighboring slow-light transmission windows into a big one, while maintaining the dispersion characteristics within each narrow window. This can improve the line shapes of the homogeneously broadened windows in room temperature ultraslow light, which typically assume the Lorentzian profile otherwise. It is as if an inhomogeneously broadened window had been created.

Fig. 3 illustrates the general system proposed for increasing the delay-time-bandwidth. The input pulses are decomposed into different spectral components (sub-pulses or wavelets), which can be done with a prism or grating as in femtosecond pulse shaping. The sub-pulses have slightly different central frequencies from each other, so that adequate number of harmonics can be covered. The duration of every sub-pulse is matched to the inverse of the slow-light element bandwidth, which guarantees minimum distortion that will be imposed on it. The transmission peak of each slow-light element is tuned to the center frequency of the respective sub-pulse, so that all sub-pulses receive the same delay. The delayed sub-pulses are then recom-

bined with another prism or grating to produce the output pulses in good fidelity.
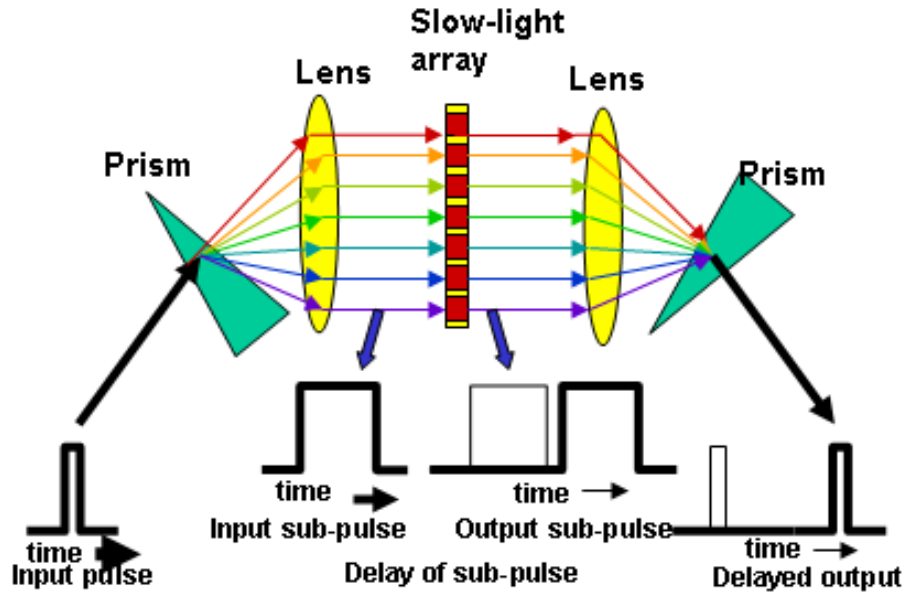
Fig. 3. Proposed architecture to address pulse distortions in ultraslow light.

Since each slow-light element handles a more manageable signal than when it works alone, sharper pulses can be delivered through their cooperations. Suppose an individual array element has a delay-time-bandwidth product of one, the overall system will have a delay-time-bandwidth product equal to the number of channels. A theoretical analysis of inhomogeneous broadening in EIT systems has also been carried out recently [22] and shows improved pulse fidelity.

The proposed approach is suitable in general for any slow light material that has a narrow-band frequency-adjustable gain or transmittance, such as EIT, optical resonators [23], optical spectral hole burning [24], resonance Raman [21], coherent

population oscillation [25], photorefractive two-wave mixing [19, 26, 27], narrow-band fibers and vertical cavity surface-emitting laser amplifier [28]. Due to the available budget, a proof-of-principle experiment will be demonstrated with photorefractive two-wave mixing in Ce:BaTiO$_3$. The crystal has a narrow-band gain of two-wave mixing effect [29]. It also acts as a holographic beam combiner, which can eliminate the lens and prism at the output end in Fig. 3. Hence the channels are spatially overlapped inside the crystal at different angles. Ideally, the channels should be spatially separated to avoid cross-talks (mutual interferences). In this case, the cross-talks are suppressed by the angular Bragg selectivity of the thick holograms in the crystal [30]. The demonstration system will have three channels. Although it is far from a complete one for practical applications, efforts will be made to show the basic concepts of simultaneous slowdown of multiple spectral components and their recombination. There are two ways to manifest the delay-time-bandwidth enhancement, by improving the shape of a highly distorted output pulse at the same delay time, or by increasing the delay time of an originally clean output pulse. The first strategy will be adopted here.

In summary, a general architecture is proposed to make ultraslow light more realistic for solid-state optical buffers and delay lines working at room temperature. The experiment will be based on photorefractive two-wave mixing in Ce:BaTiO$_3$. The delay-time-bandwidth enhancement will be demonstrated with shape improvements in a highly distorted delayed pulse.

D. Subwavelength surface plasmon waveguides and couplers

The surface plasmon wave is the coupled oscillation of electrons and electromagnetic (EM) wave at the interface between metal and dielectrics. Its quantization is called

surface plasmon (SP). SP waves can be excited by electron beams impinging into metallic thin films [31]. Incident light can also induce such oscillations. Some literatures refer to the latter case as surface plasmon polariton (SPP) oscillation. This differentiation is not strictly followed, as the term "surface plasmon" refers to the optically excited case in most publications. For convenience of discussion, surface plasmon is presumed to be generated by incident photons hereafter without further distinctions.

The SP wave in the microwave regime is more commonly known to engineers as surface waves. It has received more interests in the optical society, probably due to its strong localization of light below the corresponding wavelength. The tight focus of the electromagnetic field can be down to 100 nanometers and below in the visible spectrum, which enables strong interactions with substances of interests. To make things more intriguing, the interactions can be further enhanced by the SP resonance. Such features present surface plasmon as a competitive candidate for subwavelength optics, ultra-sensitive detectors and novel optical sources.

Research on surface plasmon can be traced back to the light scattered by small particles of noble metals in the cluster form, which are signified by the vibrant colors unknown with their bulk counterparts. In 1908, Mie analytically solved the Maxwell equations where a plane wave is scattered by a sphere of arbitrary radius and dielectric constant. For an extremely small sphere in terms of the wavelength, Mie's solution leads to the condition for the resonant maximum scattering. It states that the sphere dielectric constant must be twice and opposite to the ambient, which is satisfied by metals across the visible spectrum. The scattering cross sections can be an order higher than their physical cross sections at resonance. The distinctive colors stem from the dispersiveness of scattering capabilities around the resonance. A closer study found that these phenomena are related to the surface waves on the spheres [32].

Around the same time as Mie's dissertation, Wood discovered in 1902 that at a particular angle, a minimum showed up in the light intensity diffracted by metal-coated gratings. It was later dubbed "Wood's anomaly". The disappearance of the optical energy can be attributed to its conversion into SP waves propagating on the coating surface.

Mechanisms similar to Wood's anomaly can explain the so-called fluorescence "quenching" for dye molecules adsorbed on a smooth metal film. The SP wave supported by the film is evanescent in the sense that the field strength decays exponentially in the surface normal direction. Thus it is tightly bound to the surface. And the field magnitude is remarkably high at the metal-dielectric interface, which introduces a strong electrical coupling to the adsorbed molecules according to Fermi's golden rule [33]. Consequently, the excited molecules relax dominantly into SP waves traveling within the plane, which is gradually converted into loss in the metal due to its finite conductivity. The failure to observe fluorescent light going out of the plane appears as if the molecular fluorescence has been suppressed or quenched by the metal.

The SP-molecule interaction is actually much stronger than that between a molecule and a free space photon (spontaneous emission/fluorescence is the photon radiation into free space during the decay of an excited atom/molecule). Enhanced fluorescent signals are anticipated if the SP wave can be extracted out of the surface. Ritchie *et al.* were able to do so with a rough metallic surface instead of a smooth one [34]. The SP wave was scattered out of the plane by irregularities on the evaporated semi-continuous film. The fluorescence was amplified by almost an order of magnitude.

The metal islands on the semi-continuous thin films can be treated as small particles, which comes back to the case of metal clusters in two dimensions (2D).

Numerous literatures have reported magnifying the fluorescence or Raman scattering signals by attaching target molecules to noble metal particles, which were tuned to their SP resonance. For instance, Nie *et al.* successfully detected surface enhanced Raman scattering (SERS) signals from single molecules [35]. Furthermore, if the SP-molecule interaction is viewed as a first stage amplifier of light emissions, a second stage can be added by coupling the SP into a resonator, which is well known for augmenting the spontaneous emission rate by the Purcell factor [36]. A gigantic multiplicative enhancement ($10^9 \sim 10^{11}$) was estimated by Kim *et al.* in an experiment, where metal clusters were adhered to the inner wall of a quartz micro-cylinder (a whisper gallery mode resonator) [37].

The extensive research on SP reveals its potential for subwavelength optics and novel optoelectronic devices, by exploiting the strong field localization and enhanced interactions with substances [38, 39]. The subwavelength concentration of SP makes its behavior highly dependent on structure sizes of the same order. As a result, performance control of prospective SP devices will not be possible without precise fabrication and design methodologies. Locations of the SP maxima (hot-spots) are unpredictable both in the metal clusters prepared with chemical methods, and in the semi-continuous films prepared with evaporation. Recently, people began fabricating SP devices with tools in semiconductor research in pursuit of more engineerable devices. For instance, extraordinary transmissions have been observed in hole arrays punched on metal thin films with focused ion beams (FIBs) [40]. Gain improvement was reported by combining a metal layer with photonic crystal structures in quantum cascaded lasers [41]. Superlens with sub-diffraction limit resolution was demonstrated using tools like FIB, evaporator, spincoator and atomic force microscope (AFM) [42].

Our interests are mainly on SP waveguides and couplers. Optical dielectric waveguides have bounded mode sizes on the order of the wavelength or above. In

contrast, SP waveguides allow significantly smaller modes sizes. They are potential building blocks of the highly integrated photonic chips in the future (see [39] and references therein). Here the major concerns are the guiding mode size which is their main advantage over the dielectric counterparts, and the loss, which determines their usefulness in various scenarios. A general observation is that these two are contradictory. Insights will be gained by the proposed research on these issues.

Another issue stands out when connecting the SP waveguide to the outside world. Their subwavelength mode sizes are mismatched to the characteristic dimensions of both guided and free space optical links, which are on the order of the wavelength or above. Couplers must be designed to bring in a smooth mode size transition as well as a momentum match between the SP and the incoming/outgoing photons. Prisms were widely used in previous SP experiments as couplers, where the total internal reflection (TIR) generates evanescent waves that can be readily converted into SP waves [31]. Since planar fabrications are preferred for the waveguide, grating coupler concepts from integrated optics are favored over the prism scheme. It is an area yet to be explored by the science society.

At last, the devices have to be simulated before they are made, otherwise the accuracy provided by the advanced fabrication techniques will be offset by the unpredictability in design. It is a task as challenging as making them. Finite difference in time domain (FDTD) is a well tested method in simulating electromagnetic problems. It is a rigorous simulation of the Maxwell equations. Moreover, it is versatile to handle arbitrary geometries and substances, as long as the material model is available. Metal dielectric constants can be described by the Drude model, which is safe to use for features greater than 20 nm. When the size is below that, the dielectric constant will be dependent on the geometry due to quantum effects [43]. Because the device sizes are not expected to be smaller than 20 nm, the Drude model will suffice in our

case. To save the development time, a commercial FDTD package, CST Microwave Studio$^{\circledR}$ combined with the Drude model, is selected to simulate the waveguide and couplers. The software has been used to simulate the focus of superlens by Liu *et al.* [44]. An optional choice of solving the waveguide problem is the eigenmode analysis based on the mapped Galerkin method (MGM). It has been applied to the analysis on dielectric optical waveguides [45] and may be extended to the SP case. But the MGM can only manage waveguides, not the couplers.

In regard to the discussions above, SP waveguides and couplers working in the visible spectrum are proposed as the third and last topic for the Ph.D. dissertation. Silver will be the chosen metal considering its relatively low loss in the visible. Simulations will be carried out with a FDTD software along with the Drude model. Attempts may also be made to extend the MGM to the SP waveguide analysis. The device will be fabricated with electron beam lithography (EBL), where the planar pattern is generated by exposing the "photoresist" to a directed electron beam in a scanning electron microscope (SEM). It is a mature technology in making planner features on the sub 100 nanometer scale with proven precisions [46]. The exposed patterns will be transferred to silver by the lift-off technique [47]. Because of their small sizes, an oil immersion microscope will be built to characterize the optical responses of the SP devices.

E. Outline of contents

In chapter II, we will discuss the CGHs implementation of the Rizvi-Zubairy model for associative memory. In chapter III, the delay-time-bandwidth problem in slow light will be addressed with a solution of artificial inhomogeneous broadening. Chapter IV will present the research on nano waveguides and coupling antennas by surface

plasmon effects. Chapter V concludes this dissertation.

CHAPTER II

ASSOCIATIVE MEMORY IMPLEMENTED WITH
COMPUTER-GENERATED-HOLOGRAMS*

A. Overview

Rizvi and Zubariy proposed a variation of the Hopfield model for associative memory
of binary images [17]. Every stored image is an array of binary neurons assuming
0s and 1s. The memory is supposed to retrieve a stored image closest to the input.
The operations consist of the multiple interconnections between neurons and the
subsequent thresholding of the interconnection outputs.

In mathematics, a binary image can be represented by a vector with elements of 0s
and 1s. The interconnection is then a matrix operator applied to the input vector. The
product vector of the multiplication is binarized into 0 and 1 elements again according
to an input-dependent threshold. Rizvi and Zubairy derived the matrix operator and
the threshold value in an intuitive manner. Since conventional computers are not
efficient in doing matrix multiplications, they suggested parallel interconnections by
diffractive optical elements (DOEs) for experimental demonstrations instead.

In the optical realization, the binary input image is a simple transmission mask
modulating an illumination beam with opaque and transparent pixels. It will be
transformed into an output image by a DOE, which is a computer-generated-hologram
(CGH). By properly designing the CGH, its optical transformation can perform the

interconnection matrix multiplication to the input image. A CCD camera detects the multiplication product as well as the threshold. The captured image is then binarized to generate the associative memory output.

In section B, we will briefly review how Rizvi *et al.* arrived at their associative memory model. The mathematical formulation will also be introduced. In section C, we will discuss the optical architecture to implement the model and the design details of the CGH. In section D, experimental results will be presented and compared to simulations. Section E will summarize our work on the associative memory.

B.   The Rizvi-Zubairy model of associative memory

The philosophy of the Rizvi-Zubairy model is simple and straight forward [17]. An image is composed of "on" and "off" pixels. Each "on" pixel in an input image recalls for once all the stored images that has the same pixel "on". This procedure is repeated for every "on" pixels in the input image. All the recalled images are added up to get a grayscale image, where the grayscale reading counts the time a pixel has been recalled. The higher the grayscale reading, the stronger this pixel is related to the input image. An arbitrator mechanism is required here to determine to what extend this relation is strong enough to keep the pixel in the final output. This can be completed by binarizing the grayscale pixels. An intuitive threshold value will be the total number of "on" pixels in the input image, because it is equal to the frequency of the recall operation. If the grayscale reading is higher or equal to this number, it implies on average the corresponding pixel has been requested at least once in each recall operation, which is a sign of strong correlation to the input image.

Fig. 4 illustrates it with a simple example, which is a memory of three binary images "+", "T" and "X". Each image has nine pixels. The "on"(transparent) pixel

is 1 and the "off"(opaque) pixel is 0. The three "on" pixels in the input image of "-"
recalls "+" three times, "T" and "X" once each. They are summed up as a grayscale
image, which is binarized by the threshold value "3" into an output image of "+". It
is the closest one to the "-" among the three stored images.
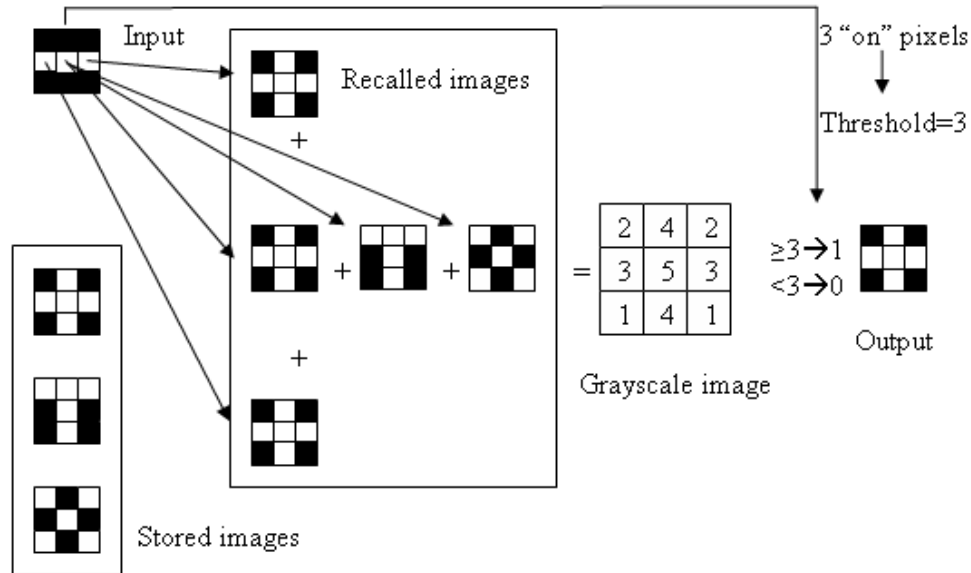
Fig. 4.   Illustration of the Rizvi-Zubairy model. It handles binary images. The trans-
parent pixel is 1 and the opaque pixel is 0. The example has a memory of three
images. Each image has nine pixels. The output is the closest one to the input
among the three stored images.

As evident from the operations described above, the memory associating process
boils down to the correlation of the input image to the stored ones. Thus the Rizvi-
Zubairy model can be interpreted in an equivalent perspective of correlations. The
grayscale image is obtained by combining the stored images, where the contribution
from each image is weighted by its correlation with the input image. The correlation
will be minimum if the two images are orthogonal, i.e., the contribution turns out to
be zero. If the two images are identical, the correlation will be maximum, which is

equal to the number of "on" pixels in the input. When a set of orthogonal images are stored in the memory and one of them is presented at the input, a reasonable threshold value should allow it to reproduce itself. In that case, the threshold value should be the autocorrelation of the input image, which is the total number of "on" pixels in the input.

The correlation interpretation is convenient in deriving the mathematical formulation. Assume that there are M images stored in the memory, each image having $N \times N$ pixels (neurons). The mth image is represented by a two-dimensional (2-D) array $\{V_{ij}^m, 1 \leq i \leq N, 1 \leq j \leq N\}$, where V_{ij} is either 0 or 1. Suppose the input image is a 2-D array $\{V_h l^{in}, 1 \leq h \leq N, 1 \leq l \leq N\}$, the grayscale image $\{V_{ij}^s, 1 \leq i \leq N, 1 \leq j \leq N\}$ is

$$V_{ij}^s = \sum_{m=1}^{M} V_{ij}^m \sum_{h=1}^{N} \sum_{l=1}^{N} V_{hl}^{in} V_{hl}^m, \tag{2.1}$$

where $\sum_{h=1}^{N} \sum_{l=1}^{N} V_{hl}^{in} V_{hl}^m$ is the combining weight for the mth stored image determined by correlation. The order of summations in (2.1) can be changed

$$V_{ij}^s = \sum_{h=1}^{N} \sum_{l=1}^{N} V_{hl}^{in} \sum_{m=1}^{M} V_{ij}^m V_{hl}^m = \sum_{h=1}^{N} \sum_{l=1}^{N} V_{hl}^{in} W_{ij}^{hl}, \tag{2.2}$$

where

$$W_{ij}^{hl} = \sum_{m=1}^{M} V_{ij}^m V_{hl}^m. \tag{2.3}$$

The 2-D input array can be treated as a $N^2 \times 1$ column vector

$$V^{in} = \left[V_{11}^{in}, V_{12}^{in}, \ldots, V_{1N}^{in}, \ldots, V_{N1}^{in}, \ldots, V_{NN}^{in} \right]^T. \tag{2.4}$$

Similarly, the grayscale image is rearranged into a $N^2 \times 1$ column vector

$$V^s = \left[V_{11}^s, V_{12}^s, \ldots, V_{1N}^s, \ldots, V_{N1}^s, \ldots, V_{NN}^s \right]^T. \tag{2.5}$$

A $N^2 \times N^2$ matrix W can be filled as

$$W = \begin{bmatrix} W_{11}^{11} & W_{11}^{12} & \cdots & W_{11}^{1N} & \cdots & W_{11}^{N1} & \cdots & W_{11}^{NN} \\ W_{12}^{11} & W_{12}^{12} & \cdots & W_{12}^{1N} & \cdots & W_{12}^{N1} & \cdots & W_{12}^{NN} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ W_{1N}^{11} & W_{1N}^{12} & \cdots & W_{1N}^{1N} & \cdots & W_{1N}^{N1} & \cdots & W_{1N}^{NN} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ W_{N1}^{11} & W_{N1}^{12} & \cdots & W_{1N}^{1N} & \cdots & W_{N1}^{N1} & \cdots & W_{N1}^{NN} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ W_{NN}^{11} & W_{NN}^{12} & \cdots & W_{NN}^{1N} & \cdots & W_{NN}^{N1} & \cdots & W_{NN}^{NN} \end{bmatrix}. \tag{2.6}$$

Now the grayscale image is available by a simple multiplication of a matrix to a vector

$$V^s = WV^{in}. \tag{2.7}$$

The $N^2 \times 1$ binary output vector (image)

$$V^{out} = \left[V_{11}^{out}, V_{12}^{out}, \ldots, V_{1N}^{out}, \ldots, V_{N1}^{out}, \ldots, V_{NN}^{out} \right]^T, \tag{2.8}$$

is obtained after the thresholding operation

$$V_{ij}^{out} = \begin{cases} 1 & if V_{ij}^s \geq \sum_{h=1}^{N} \sum_{l=1}^{N} V_{hl}^{in} \\ 0 & otherwise \end{cases} \tag{2.9}$$

C.   Optical implementation with computer-generated-holograms

As suggested by (2.2), each input neuron (pixel) V_{hl}^{in} contributes to the $N \times N$ intermediate neurons $\{V_{ij}^s\}$ through the $N \times N$ weights $\{W_{ij}^{hl}, 1 \leq i \leq N, 1 \leq j \leq N\}$. The accumulated contributions are binarized according to (2.9). Fig. 5(a) shows the optical implementation using CGHs. The ten numbers in Fig. 5(b) are used to demonstrate the memory with the proposed optical architecture. Every stored image has

(a)

(b)
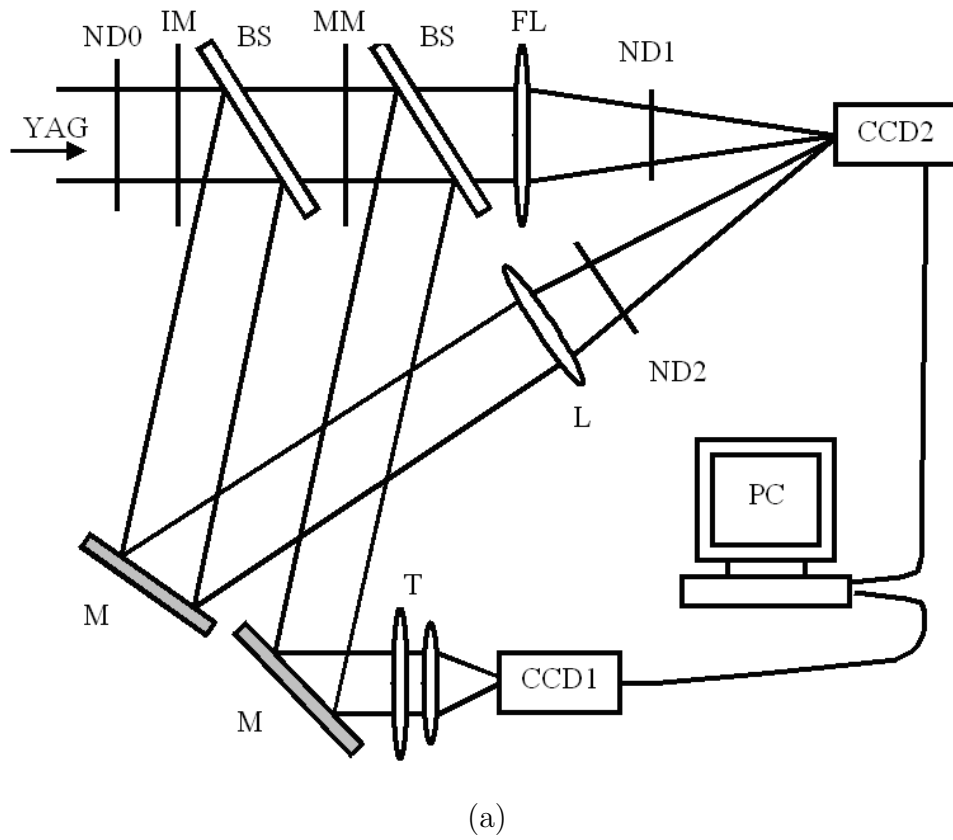
Fig. 5. Associative memory experimental setup.

12×12 pixels that are either opaque or transparent, corresponding to 0 and 1. The pixel size is $0.675mm \times 0.675mm$. Input images should be in the same format accordingly. They are introduced by the input masks (IMs), the pixels of which either block or transmit the incoming laser light. The memory is encoded in the 12×12 CGHs on the memory mask (MM), where the size of each CGH is also $0.675mm \times 0.675mm$, same as a pixel. Each input pixel is always lined up with a CGH on the optical path, which is monitored through the telescope T and camera CCD1. Therefore, the CGHs also form a 12×12 array on the MM. Since the input pixels and CGHs are assembled in arrays of the same dimensions, we can use h as the row index and l as the column index in both arrays. If an input pixel V_{hl}^{in} is opaque, it turns off the corresponding CGH behind it and has no effects to the final output. If V_{hl}^{in} is transparent, it turns on the CGH indexed by h and l in the array, which creates diffraction spots at 12×12 fixed locations on the detection plane of the camera CCD2. The diffraction intensities are proportional to the weights $\{W_{ij}^{hl}, 1 \leq i \leq 12, 1 \leq j \leq 12\}$, where i and j indexes the spots on the camera plane. We will discuss the CGHs design later on. As every transparent input pixel cast its own set of diffraction beams onto the fixed locations, their contributions to the light intensities at the fixed spots are accumulated due to the superposition of optical waves. The parallel interconnections in (2.2) are completed. The intermediate neurons $\{V_{ij}^s\}$ are obtained through the CCD readings at the diffraction locations.

The threshold in (2.9) is equal to the total number of transparent input pixels. Utilizing this fact, we propose a simple binarizing scheme where the total light intensity transmitted through IM serve as the threshold light intensity I_{th}. In Fig. 5, the lens L collects I_{th} and focuses it onto a pixel of CCD2 for measurement. This measuring pixel on CCD2 is spatially separated from the detection spots of the intermediate neurons. A calibration in necessary to compensate for the different attenuations along

the two optical paths through lens L and FL. A test image was first chosen as the input in the calibration, which by the Rizvi-Zubairy model is known to generate some intermediate neurons equal to the threshold intensity. By adjusting neutral density attenuators ND1 and ND2, the I_{th} was set equal to these neuron intensities. Once calibrated, ND1 and ND2 remain unchanged in experiments thereafter. In the memory association process, a computer program compares the detected intensities of $\{V_{ij}^s\}$ individually with I_{th}. The binarized pixel (neuron) will be 1 if the former is greater or equal to the latter, otherwise it will be 0. Such a thresholding method cancels out the detection problems caused by non-linearity and limited dynamic ranges of the sensors, since such effects are identical to the threshold measurement and the intermediate neuron measurements using the same CCD chip.

A key issue remains in designing the aforementioned CGHs that can diffract light beams with intensities $\{W_{ij}^{hl}\}$. In Fig. 5(a), the lens FL right behind MM is the Fourier transform lens whose focal plane is the same as the detection plane of CCD2. Consequently, the transmittance $\{H_{pq}^{hl}, 1 \le p \le N, 1 \le q \le N\}$ of a CGH is linked to the diffraction beam intensities $\{W_{ij}^{hl}\}$ through the well-known Fourier transform property of a lens [48]

$$W^{hl}(i\delta, j\delta) = \frac{A \exp\left[\widetilde{j}\frac{\pi}{\lambda f}(i^2\delta^2 + j^2\delta^2)\right]}{\widetilde{j}\lambda f} \sum_{p=1}^{N}\sum_{q=1}^{N} H^{hl}(p\Delta, q\Delta) \exp\left[-\widetilde{j}\frac{2\pi\Delta\delta}{\lambda f}(pi + qj)\right].$$
(2.10)

Here, h and l indexes the CGH on the MM, $\{W_{ij}^{hl}\}$ is given by (2.3), $\widetilde{j}$ is used to represent the square root of -1 because the conventional symbol for this purpose j has been assigned already, A is an insignificant constant, λ is the wavelength of light ($532nm$ in our case), f is the focal length of the lens ($75mm$ for FL in Fig. 5(a)), δ is a constant dimension of $591\mu m$ on the CCD2 plane, and Δ is a constant dimension of $27\mu m$ on the MM plane. In (2.10), terms before the double summations
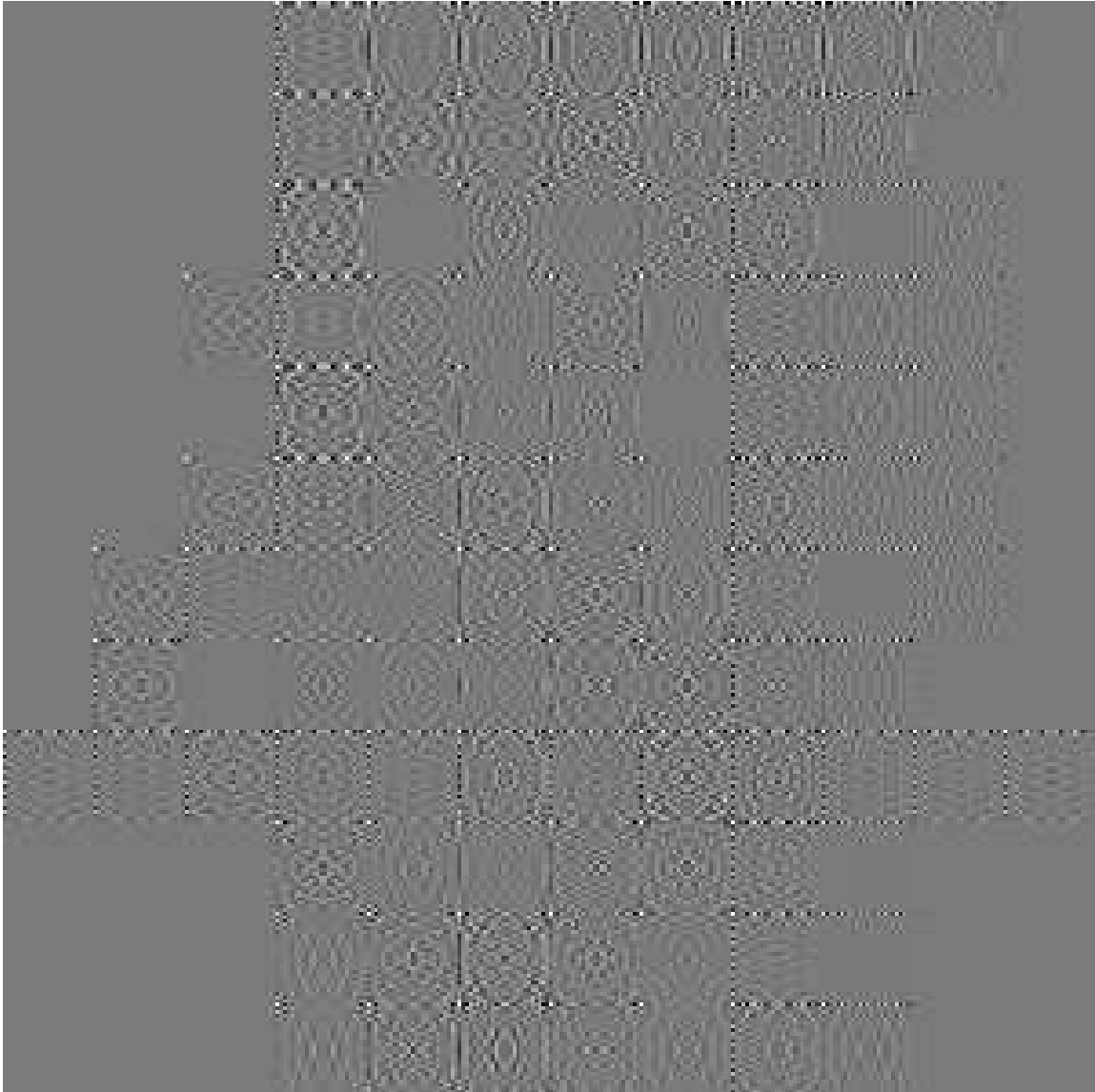
Fig. 6. Normalized transmittances between 0 and 1 for the computer-generated-holograms. The white pixels represent transmittance of 1 and the black pixels stand for 0 transmittance. Gray pixels have transmittances in between.

introduce an additional optical phase distributions on the CCD2 plane. They can be ignored since the camera only detects magnitudes. Now the CGH transmittances are available by doing a inverse fast Fourier transform (IFFT) of the interconnection weights $\{W_{ij}^{hl}\}$. Nevertheless, complex transmittances requires the more challenging optical phase modulations in addition to the amplitude modulations. To ensure real transmittances, $\{W_{ij}^{hl}\}$ are duplicated and extended into four quadrants in an origin-symmetric manner before doing the IFFT. The duplication introduces redundancy in the final CGH, but it relaxes the design and fabrication requirements.

27um
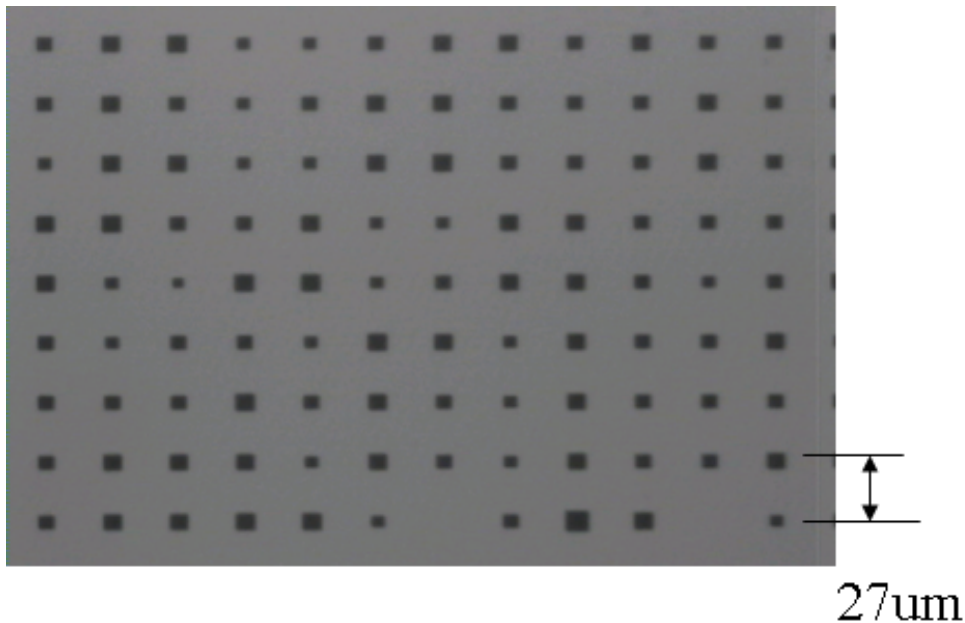
Fig. 7. Computer-generated-holograms fabricated by the laser writer technique for making photomasks. It was inspected with a Olympus BH2 microscope. The bright areas are open apertures, the sizes of which determine the transmissions of the holograms.

The computed transmittances were normalized to the range $[0, 1]$. They are continuous numbers rather than binary ones. Fig. 6 shows the transmittance distribution over the MM plane, where the brightest spot represent transmittance of 1 and the
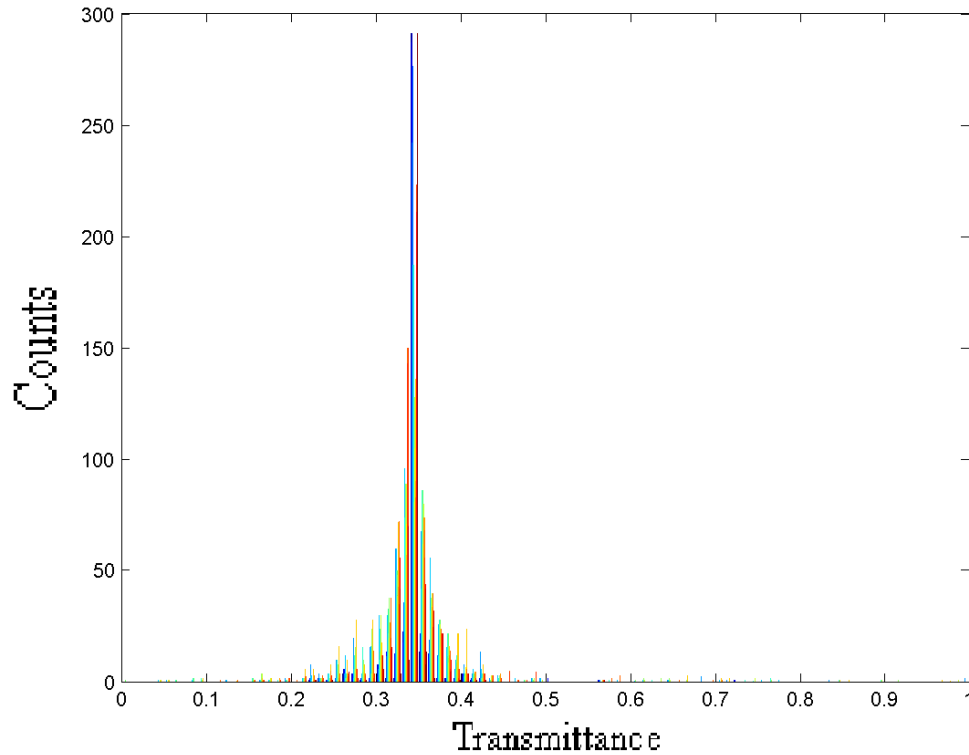
Fig. 8. Histogram of the transmittances in the computer-generated-holograms. Most of the transmittance values fall in the region of [0.2,0.5].

darkest spot stands for 0 transmittance. It is clear that multiple levels of transmission attenuations are demanded on the MM. The MM is produced by opening square apertures of various sizes on an otherwise opaque Chromium-coated glass substrate. The CGH is basically Lohmann type [15, 16]. However, the centers of the apertures are now fixed to a square lattice with a $27\mu m$ spacings, because there is no need to do phase modulations. The design scheme was verified with FFT and IFFT in Matlab simulations. The memory mask design was laid out in GDSII format and fabricated with the laser writing technique in photolithography mask production. Fig. 7 is the memory mask inspected by an optical microscope in epi-illumination (reflection) mode. The dark regions in the figure are the apertures. They have ten different

sizes as shown in Tab. I. The mapping of the transmittance to the aperture sizes

Table I. The square aperture sizes on the computer-generated-holograms. The sizes encode the transmittances $\{H_{pq}^{hl}\}$ as calculated by the IFFT.

H_{pq}^{hl}	[0, 0.23)	[0.23, 0.26)	[0.26, 0.29)	[0.29, 0.32)	[0.32, 0.35)	[0.35, 0.38)	[0.38, 0.41)	[0.41, 0.44)	[0.44, 0.47)	[0.47, 1]
μm^2	0	9	16	25	36	49	56.25	64	72.25	81

is linear in the region $0.2 \leq H_{pq}^{hl} \leq 0.5$, where most of the transmittance values are concentrated according to their histogram in Fig 8.

D. Experimental results

As a test of the system, we tried to retrieve the numeric number "4" from the memory. When an input image as in Fig. 9(a) showed up, the intermediate neuron intensities were detected by CCD2 as in Fig. 9(c), which agrees with the FFT simulation (Fig. 9(e)). Fig 9(g) shows the binarized output image, indicating a successful retrieval. Similarly, Fig. 9(b,d,f,h) demonstrate a successful recall of the numeric number "9".

Moreover, the Rizvi-Zubairy model can also recall images from partial inputs. Fig. 10(a) is a partially blocked input image of numerical number "9". Fig. 10(b) is the interconnection outputs, which is binarized into Fig. 10(c). The above results show that the associative memory based on Rizvi-Zubairy model and the proposed thresholding scheme are effective in pattern recognition and data retrieval.
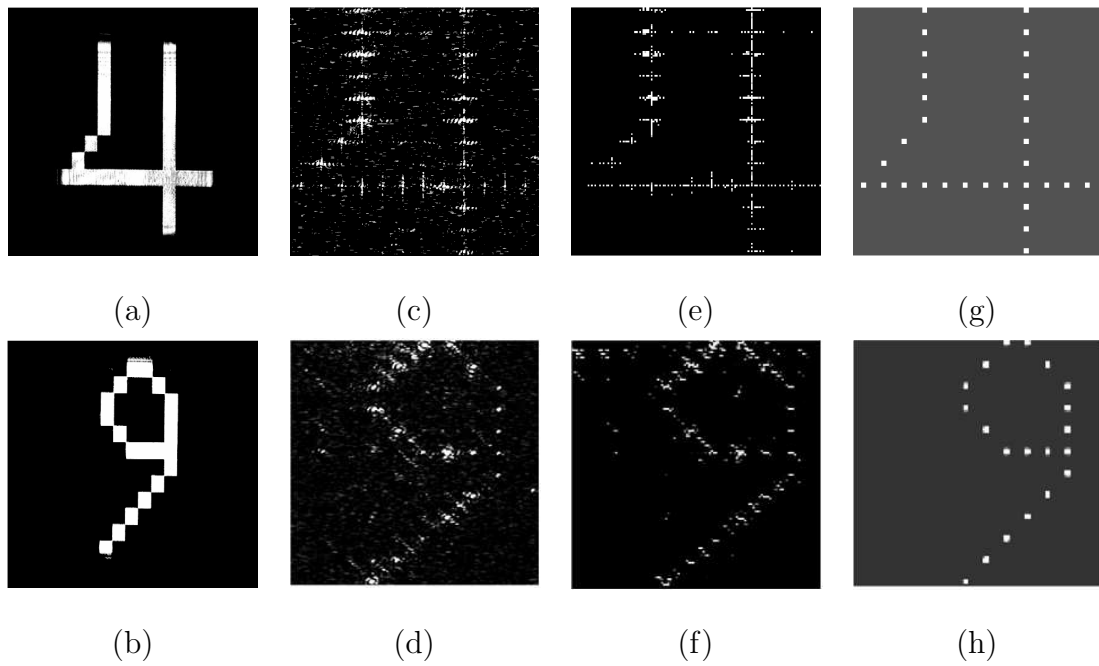
Fig. 9. Experimental tests of the memory recall function. Images of (a)(b) input images captured by CCD1, (c)(d) intermediate neuron intensities and threshold intensities (the bright spots on the right edge) detected by CCD2, (e)(f) FFT simulation of the intermediate neuron intensities, and (g)(h) binarized outputs.

(a)

(b)

(c)
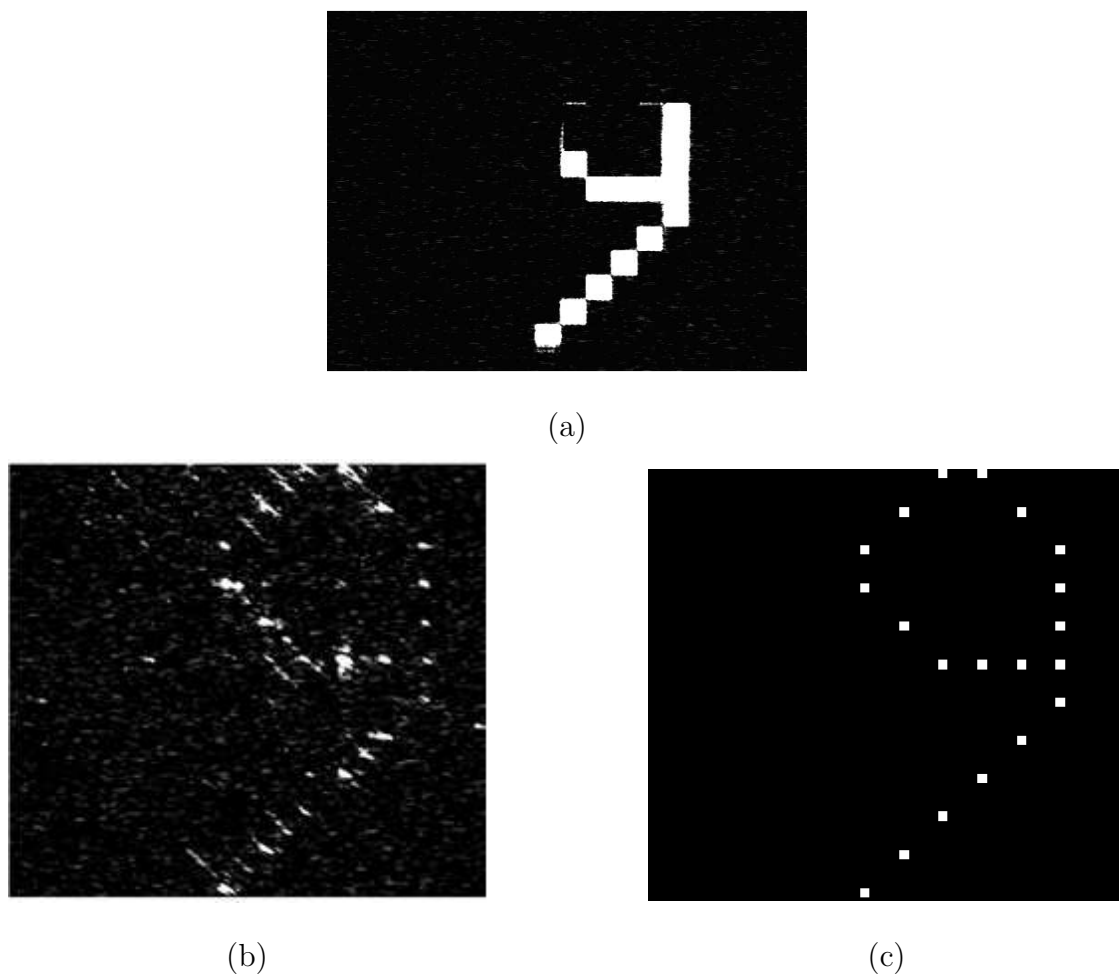
Fig. 10. Experimental tests of the memory recover function. Images of (a) partial inputs for "9", (b) the intermediate neuron intensities and (c) the binarized output.

E.   Summary of the associative memory

In summary, the Rizvi-Zubairy asscociative memory model is very suitable for optical implementation due to the absence of subtractions. A simple real-time thresholding scheme streamlines the system and suppresses noise. An experiment with computer-generated-holograms was demonstrated. The stored images were successfully retrieved complete or partial inputs through the content-addressable mechanism.

CHAPTER III

DELAY-TIME-BANDWIDTH PROBLEM IN ROOM TEMPERATURE

ULTRASLOW LIGHT[*]

A.   Overview

In 1999, ultraslow light was first reported in the ultra-cold atomic gas, which can be explained by the extremely narrow frequency window for light to transmit through the gas. The narrow transmission band implies a sharp dip in the imaginary part of the complex refractive index

$$\mathsf{N}(\omega) = \mathsf{n}(\omega) + i\mathsf{k}(\omega), \tag{3.1}$$

where ω is the angular frequency of light. The real part and the imaginary part are linked to each other by the Kramers-Kronig relations

$$\mathsf{n}(\omega) \;=\; 1 + \frac{2}{\pi}P\int_0^\infty \frac{\Omega\mathsf{k}(\Omega)}{\Omega^2 - \omega^2}\mathrm{d}\Omega, \tag{3.2}$$

$$\mathsf{k}(\omega) \;=\; \frac{-2\omega}{\pi}P\int_0^\infty \frac{\mathsf{n}(\Omega)}{\Omega^2 - \omega^2}\mathrm{d}\Omega, \tag{3.3}$$

where P is a constant [32]. The sharp dip in $\mathsf{k}(\omega)$ leads to a strong dispersion of $\mathsf{n}(\omega)$ around the pass-band, as evident from (3.2). The large $\mathrm{d}\mathsf{n}(\omega)/\mathrm{d}\omega$ creates an

ultraslow group velocity v_g as given by

$$v_g = \frac{c}{\mathsf{n}(\omega) + \omega[\mathsf{dn}(\omega)/\mathsf{d}\omega]}. \tag{3.4}$$

Consequently, an optical pulse was slowed down to 17 m/s by the ultra-cold atomic gas [7]. The first ultraslow light in solid state material was reported in 2002 [21]. One year later, ultraslow light was realized in room temperature solid [25]. The explosion in ultraslow light research is driven by the prospects of optical delay lines, optical buffers and quantum information storage.

The rapid progress gives a delusion that the commercialization is just around the corner. But an obstacle remains that the narrow pass-band introduce strong distortions to the pulse shapes at high data rates. It seems the long delay time and wide bandwidth of slow light devices are incompatible with each other, otherwise the Kramers-Kronig relations will be violated. This problem can be characterized by the delay-time-bandwidth product of the slowlight device. It describes the capacity or word length of the optical delay line or buffers, which is the number of bits (pulses) that can be slowed down simultaneously. The best delay-time-bandwidth product observed so far is five in lead vapors [20].

To increase the delay-time-bandwidth product, a general architecture will be introduced in section B. A proof-of-principle experiment will be presented in section D. Section E will conclude this chapter.

B. Augmented delay-time-bandwidth product by artificial inhomogeneous broadening

The basic idea of increasing the delay-time-bandwidth is illustrated in Fig. 11. As shown in the left box, a single ultraslow light element has a narrow transmission

window, where the absorption component $k(\omega)$ in the complex refractive index is small. Instead of passing light through a single slow light element, the optical signal is simultaneously fed to an array of slow light elements, the central frequencies of which are slightly separated from each other. Outputs from the array are then combined to produce the final signal. In effect, the transmission window of the system is artificially broadened, while the dispersiveness of $n(\omega)$ is preserved. The widened transmission window allows higher data rate or less distortions in the pulse shape. The corresponding optical architecture can be found in Fig. 3 in Chapter I.

The broadening of a single dip into a wide window is similar to that from the Lorentzian lineshape in homogeneous optical absorption to the Gaussian lineshape in inhomogeneous optical absorption by solid state materials [33]. However, the broadening proposed here is created artificially and can be controlled with more freedom and accuracy. Interestingly, the structure in Fig. 11 also resembles the subband filter banks in signal processing research.

The general architecture can be implemented with devices that have narrowband frequency-adjustable gain or transmittance, such as EIT, optical resonators [23], optical spectral hole burning [24], resonance Raman [21], coherent population oscillation [25], photorefractive two-wave mixing [19,26,27], narrow-band fibers and vertical cavity surface-emitting laser amplifier [28]. A proof-of-principle experiment will be presented next with photorefractive two-wave mixing in Ce:BaTiO$_3$. This material has a very narrow band in the two-wave mixing gain and is ideal for demonstrating room temperature ultraslow light. Meanwhile, it is a thick holographic material, which can provide multiple two-wave mixing channels by angular multiplexing, and also serve as a beam combiner for outputs from different channels. The Bragg selectivity of thick holograms [30] guarantees the inter-channels isolations are acceptable for demonstration purpose. Although in the real-world applications a dedicated solid
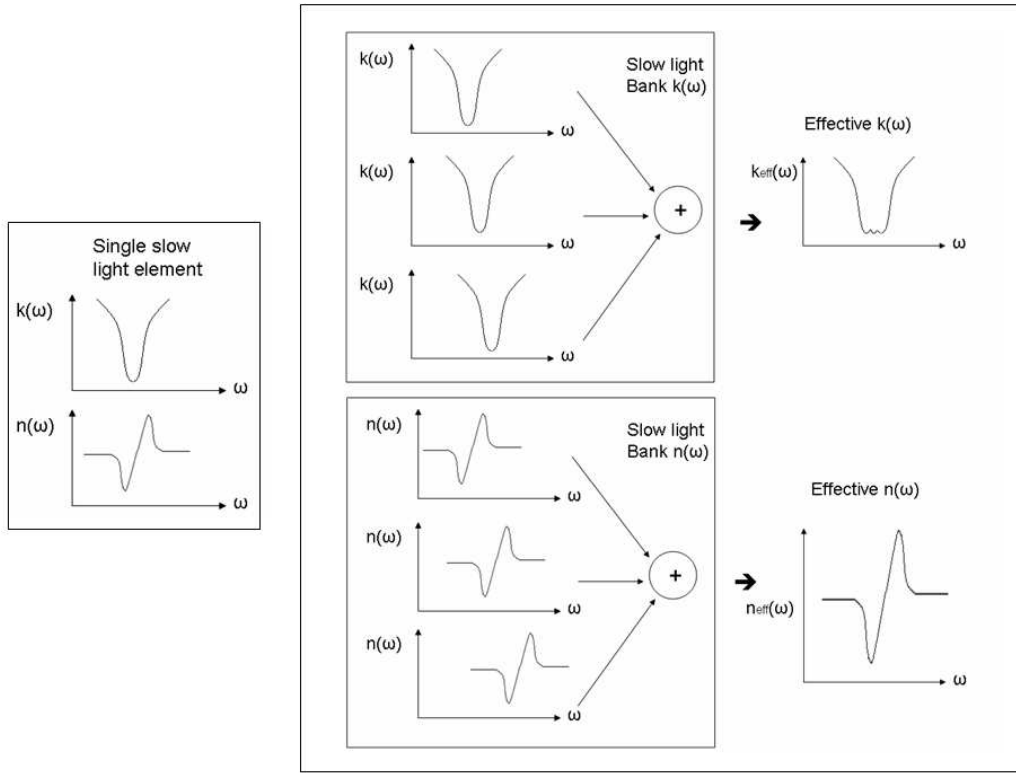
Fig. 11. Slow light bank to achieve artificial inhomogeneous broadening. The slow
light elements have center frequencies offset incrementally. The effective com-
plex refractive index features a wider transmission window and keeps the
strong dispersiveness.

state device will be desirable for each channel to achieve more superior performance,
the all-in-one characteristics of Ce:BaTiO$_3$ makes it a economical choice for demon-
strations.

C.   Photorefractive two-wave mixing

Photorefractive materials feature refractive indexes that can be changed by illumi-
nation. The electron states in photorefractive crystals are depicted in Fig. 12 [49].

They are composed of a valance band where it is hard for electrons to move around and a conduction band where electrons are mobile. Dopants in the crystal also create donor centers and acceptor centers at discrete sites across the crystal. Electrons at donor centers have energies higher than the valence band electrons, and thus are more ready to be excited by an incident photon into the conduction band. The photoelectrons may wander around and run into the acceptor centers, where the energy level is lower than the conduction band. This energy trap makes it difficult for the captured electrons to go back to the conduction band to regain their mobility.
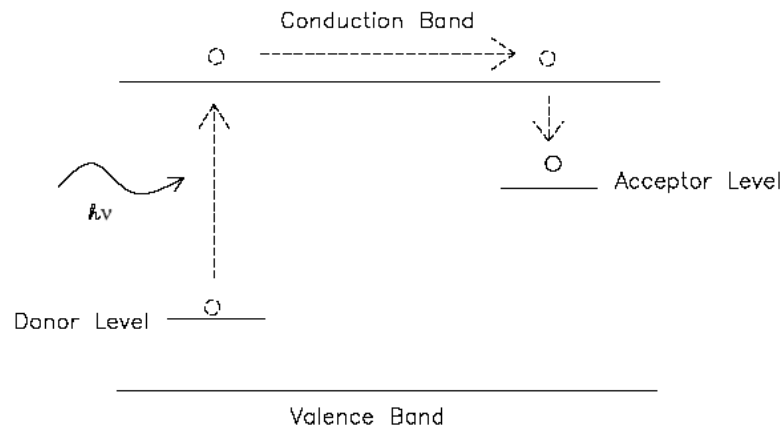
Fig. 12.   Photo-electron model of photorefractive crystals. Incident photons to photorefractives can excite electrons from the donor level to the conduction band, where they are possible to move and fall into acceptor traps.

Now consider the two-wave mixing case in Fig. 13, where two coherent laser beams, with very close angular frequencies ω_p and ω_s, interfere inside a photorefractive crystal. In the dark stripes of the interference pattern, exciting light intensity is weaker than in the bright stripes. Hence, more photoelectrons are excited in the bright region. This spatial density differential causes diffusion of the photoelectrons to the dark regions. In the steady state of the diffusion process, more photoelectrons

migrate to and get trapped in the dark regions, which creates an electric field (space charge field) to balance the diffusion potential. The space charge field changes the refractive index through the electrooptic effect [50]. A holographic grating of modulated refractive index is recorded in the crystal. At proper incident angles, the grating transfers energy from one beam (the pump) into the other (the probe).
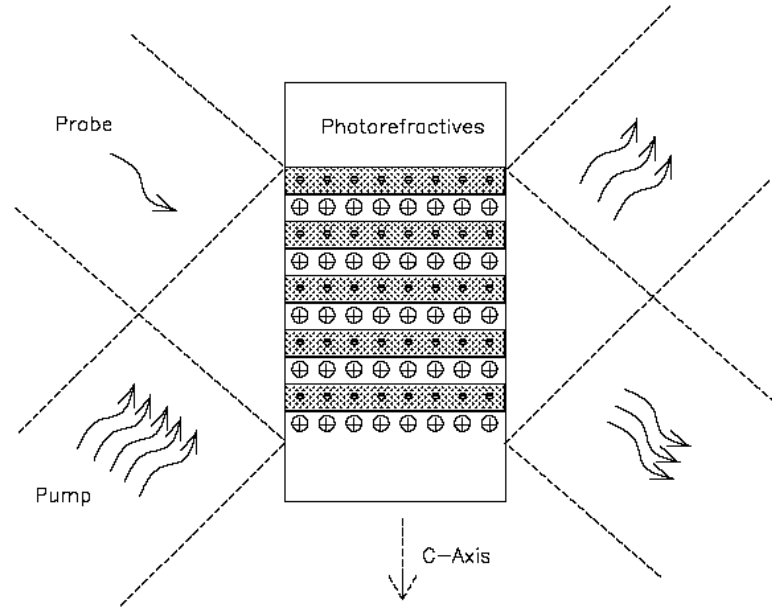
Fig. 13. Two-wave mixing in photorefractive crystals. Photoeclectrons concentrate in the dark stripes (shaded region in the figure) of the interference pattern, creating a space charge field to modify refractive index. The refractive grating diffract pump light into the probe output.

If the pump beam is much stronger than the probe, the amplified probe field is given by [19, 51]

$$A_s(\omega_s, d) = A_s(\omega_s, 0) \exp\left[i(\frac{2\pi n}{\lambda})d\right] \exp\left[\frac{\Gamma(\Delta\omega)d}{2}\right] \exp\left[i(\frac{2\pi Q(\Delta\omega)}{\lambda})d\right], \quad (3.5)$$

where $A_s(\omega_s, 0)$ is the probe field at the entrance into the crystal, d is the path length of probe beam inside the crystal, n is the refractive index of the material, λ

is laser wavelength, and $\Delta\omega$ is the frequency detune between the pump and probe $\Delta\omega = \omega_s - \omega_p$. $\Gamma(\Delta\omega)$ is the two-wave mixing gain coefficient

$$\Gamma(\Delta\omega) = \frac{\Gamma_0}{1 + \Delta\omega^2\tau^2}, \tag{3.6}$$

with Γ_0 as a constant and τ as the rising time of the space charge field. $Q(\Delta\omega)$ is a unit length phase shift

$$Q(\Delta\omega) = \frac{\Gamma_0\lambda}{4\pi}\left[\frac{\Delta\omega\tau}{1 + \Delta\omega^2\tau^2}\right]. \tag{3.7}$$

For light propagating in a medium of thickness d and with a complex refractive index $\mathsf{n}(\omega) + i\mathsf{k}(\omega)$, the emerging signal is given by [32]

$$A(\omega, d) = A(\omega_s, 0)\exp\left[\frac{-2\pi\mathsf{k}(\omega)d}{\lambda}\right]\exp\left[i(\frac{2\pi\mathsf{n}(\omega)}{\lambda})d\right]. \tag{3.8}$$

Similarity is immediately observed between (3.5) and (3.8), where the photorefractive gain coefficient $\Gamma(\Delta\omega)$ corresponds to $-\mathsf{k}(\omega)$ and the unit length phase shift $Q(\Delta\omega)$ corresponds to $\mathsf{n}(\omega)$. Therefore, the previous discussions on ultraslow light also apply to the photorefractive two-wave mixing case. Compared to other ultraslow light demonstrations, the only difference in the photorefractive implementation is the delayed signal is amplified rather than attenuated. Following the same procedures in [27], the group velocity of the probe beam in photorefractives is derived

$$v_g(\Delta\omega) = \frac{c}{n + c\ \mathrm{d}Q/\mathrm{d}\omega_s} \approx \frac{2[1 + \Delta\omega^2\tau^2]^2}{\Gamma_0\tau[1 - \Delta\omega^2\tau^2]}. \tag{3.9}$$

Fig. 14 shows the normalized gain coefficient, the unit length phase shift and the group velocity of the probe beam as functions of $\Delta\omega$.

A small group velocity is evident in Fig. 14 for the probe beam within the passband of the two-wave mixing gain. It can be estimated at the zero frequency detune
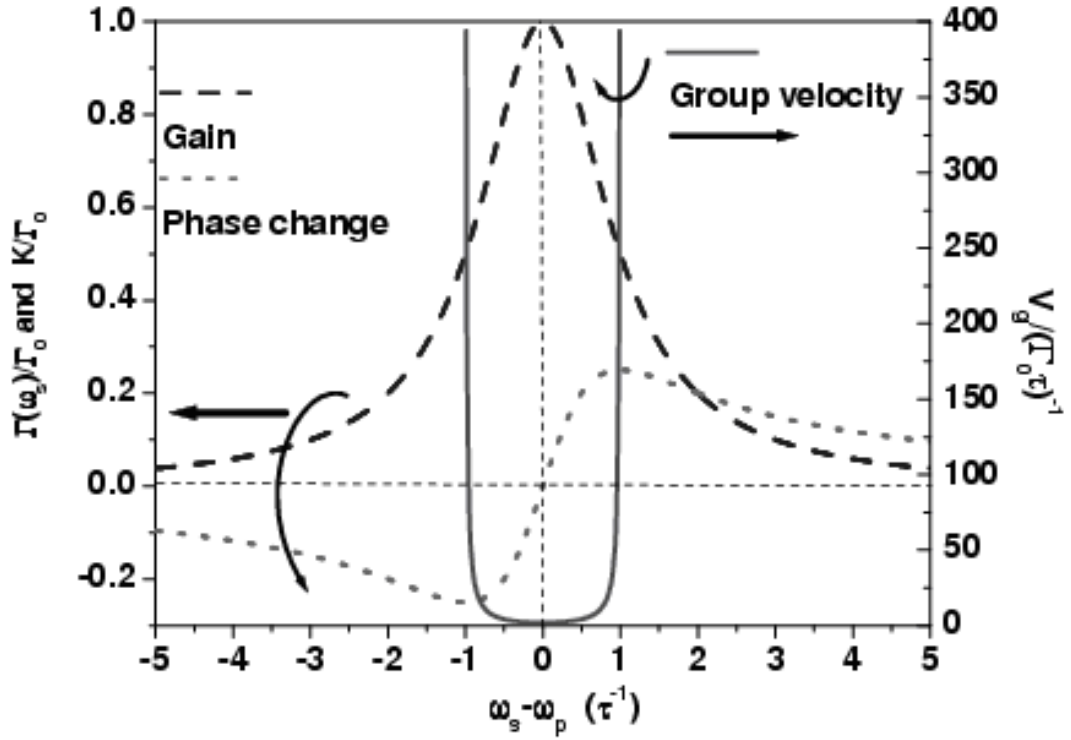
Fig. 14. Characteristic performances of the photorefractive two-wave mixing. Normalized group velocity, unit length phase shift, and two-wave mixing gain coefficient of the probe beam as functions of its frequency detune from the pump beam.

in (3.9)

$$v_g(0) = \frac{2}{\Gamma_0 \tau}, \tag{3.10}$$

which gives the delay-time T_d of the probe beam through the crystal

$$T_d \approx \frac{d}{v_g(0)} = \frac{\Gamma_0 \tau d}{2}. \tag{3.11}$$

The signal gain G is available through (3.5) and (3.6)

$$G(\Delta\omega) \propto \exp\left[\frac{\Gamma_0 d}{1 + \Delta\omega^2 \tau^2}\right], \tag{3.12}$$

the 3dB bandwidth of which is

$$BW = \frac{1}{\pi\tau}\sqrt{\frac{\ln 2}{\Gamma_0 d - \ln 2}} \ (Hz). \tag{3.13}$$

Equations (3.11) and (3.13) predict the delay-time-bandwidth product is indeed a constant in the slow light effects by photorefractive two-wave mixing

$$T_d BW = \frac{\Gamma_0 d}{2\pi}\sqrt{\frac{\ln 2}{\Gamma_0 d - \ln 2}}. \tag{3.14}$$

It once again confirms that there is a physical limit for the capacity (number of pulses to be handled) of all optical delay lines and buffers. High data rates and long delay time cannot be achieved simultaneously. A fast-changing signal forced through such devices will experience severe distortions at the output. A deceptive solution may be posed by (3.13) to increase the delay-time-bandwidth product by increasing $\Gamma_0 d$. But it would destroy the slow light device before there are any significant changes in the delay-time-bandwidth product, since the optical power grows exponentially according to $\Gamma_0 d$. The limitations in a single slow light element incite us to look into solutions by multiple slow light elements. In the next section, the proposed architecture of artificial inhomogeneous broadening will be applied to a photorefractive crystal to improve the distorted output, i.e., the delay-time-bandwidth product.

D.   Photorefractive demonstration of artificial inhomogeneous broadening

For photorefractive crystals, the pulse-delay effect is well known and has been applied to optical delay lines [52]. Recently, room temperature slow light were also demonstrated by photorefractive two-wave mixing [19, 26, 27]. However, no solutions were attempted to address the pulse distortion problems (limited delay-time-bandwidth product). In our experiment, Ce:BaTiO$_3$ [29] was used to demonstrate the pulse

delay and moreover the proposed scheme of artificial inhomogeneous broadening to correct the pulse distortions. The concept will be illustrated with two channel and three channel systems respectively.

Fig. 15 shows the experimental setup of a two channel system. A collimated 532 nm Nd:YAG laser (Compass 315M-100, Coherent Co.) was split into a probe beam (S) and two pump beams (P_1 and P_2). The beams intersected in a Ce:BaTiO$_3$ crystal (7.73 x 7.04 x 4.8 mm, Photox Opt. Sys.) with an angle about $40°$ between S and P_2, $15°$ between P_2 and P_1, $30°$ between P_1 and the C-axis of the crystal, respectively. The photorefractive crystal also worked as a holographic beam combiner. Hence, a single outgoing beam was formed by joining the two-wave mixing output from S and P_1, and an independent one from S and P_2. The independence between these two processes are guaranteed by the Bragg selectivity due to the angular separation between P_1 and P_2 [30]. A photodiode (PDA400, Thorlab Inc.) was used to detect the output signal, and the time domain waveforms were recorded by a digital oscilloscope (TDS640A, Tektronix). Photos of the apparatus are shown at the end of this chapter.

All the incident waves were p-polarized (polarizations parallel to the optical table). The typical beam intensity were about $6.5 mW/cm^2$ for the S beam, $637 mW/cm^2$ for the P_2 beam, and $716 mW/cm^2$ for the P_1 beam, respectively. Under this conditions, preliminary measurements gave the rising time $\tau = 3.5s$ and $\Gamma_0 = 6.2 cm^{-1}$. These two measurements allowed theoretical calculations of the group velocity by (3.10), the delay time by (3.11), the 3dB bandwidth by (3.13) and the delay-time-bandwidth product by (3.14). They are listed in Tab. II. It is interesting to note that the predicted delay-time-bandwidth product is only 0.261, much less than 1. That means any number of delayed pulses, even if just a single delayed pulse, would suffer from significant distortions. It made the crystal an ideal candidate for demonstrating
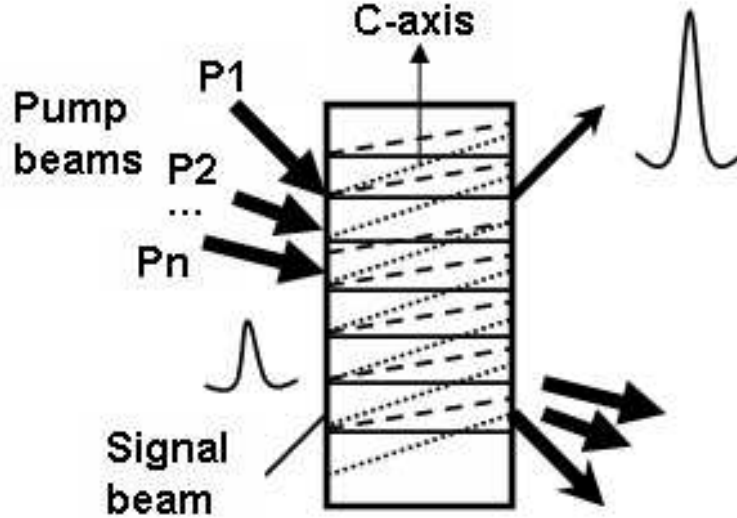
Fig. 15. The experiment layout of slow light in photorefractives. $P_1 \sim P_n$ are pump beams. Each one of them will involve in a two-wave mixing with the signal beam respectively. The multiple two-wave mixing processes serve as the multiple slow light channels as proposed in the artificial inhomogeneous broadening architecture.

the pulse distortion problems in slow light.

In the experiment, the S beam was gated by an acoustic-optics-modulator (AOM, 1206C, Isomet Co.) to generate a rectangular pulse with 1 second duration and a 0.025 Hz repetition rate. The optical frequency detunes Δf_1 and Δf_2 from the S beam carrier, are introduced respectively by the other two AOMs in the paths of P_1 and P_2 beams.

First of all, the ultraslow light was demonstrated by applying the S beam and P_2 beam only, with $\Delta f_2 = 0$. The top trace of Fig. 16(a) shows the output signal has a peak about 3 seconds lagging behind the input. The group velocity is estimated to be around 1.6 mm/s, which is in agreement with the predicted ballpark number in Tab. II. The delayed signal is badly distorted, as predicted by the small delay-time-

Table II. Predictions of the group velocity, the delay time, the 3dB bandwidth and the delay-time-bandwidth product for the slow light experiment using the Ce:BaTiO$_3$ crystal, based on preliminary measurements of $\tau = 3.5s$ and $\Gamma_0 = 6.2cm^{-1}$.

Group velocity (mm/s)	Delay time (s)	3dB bandwidth (Hz)	Delay-time-bandwidth product
0.992	5.208	0.0501	0.261

bandwidth product in Tab. II.

Next, we show the simultaneous slow down of two frequency components in the input. This was done by applying the s, P_1 and P_2 beams simultaneously, where $\Delta f_1 \equiv 0$ and Δf_2 was varied among 0.19 Hz, 0.3 Hz, 0.54 Hz, 0.66 Hz and 0.78 Hz. Now two accentuated harmonics were recovered in the delayed output. A beating between them was supposed to follow. This was indeed observed in experiment, as shown by the traces below the top one in Fig. 16(a). The beating frequencies matched the aforementioned Δf_2, as expected.

If more harmonics are recovered in the delayed output, the pulse shape can be furthered improved. This was demonstrated by adding a third pump beam P_3 with frequency detune Δf_3 in the setup. The frequency detunes were $\Delta f_1 = 0$, $\Delta f_2 = 0.4Hz$, and $\Delta f_3 = 0.54Hz$. The pump beams P_1, P_2 and P_3 were adjusted to come in 54°, 49°, and 42° in regard to the c-axis, respectively. The experimental results are presented in Fig. 16(b). The top three traces are the output signals when any two of the three pump beams were on, and are similar to the beating delayed outputs in Fig. 16(a), except that the signal to noise is degraded because each available pump power was further diluted by introducing P_3. The next trace below is the output signal when all three pumps were on. It has a peak delayed by

almost the same time, and is beginning to show a sharper peak. In comparison, the bottom trace with circles shows the theoretically predicated output $I_{123s}(t)$ for the three-pump-beam case, which is synthesized by

$$I_{123s}(t) = I_{12s}(t) + I_{23s}(t) + I_{13s}(t) - I_{1s}(t) - I_{2s}(t) - I_{3s}(t). \qquad (3.15)$$

Here the subscripts identify by which beams the output was produced, for example, $I_{12s}(t)$ is the output intensity when the S, P_1 and P_2 beams are present. $I_{123s}(t)$ can be computed, once all the terms on the right hand side of (3.15) are measured with procedures described previously. As Fig. 16(b) shows, the calculated waveform is in approximate agreement with the experimental one.

Fig. 16 shows experimental results from two-channel and three-channel systems of artificial inhomogeneous broadening to correct the pulse distortion. It suggests that the output pulse shape can be improved further if more detuned pump beams of adequate powers are available. Unfortunately, we are unable to demonstrate that in experiments due to limited budgets. A simulation of six-pump-beam case is presented in Fig. 17 to support our argument. The calculation is similar to (3.15). This is done by noting that the beating terms $I_{12s}(t)$, $I_{23s}(t)$ and $I_{13s}(t)$ in (3.15), rather than measured directly, can be calculated from $I_{1s}(t)$, $I_{2s}(t)$ and $I_{3s}(t)$ by translation operations and summations in the Fourier transform domain. Thus, $I_{123s}(t)$ can be calculated solely from $I_{1s}(t)$, $I_{2s}(t)$ and $I_{3s}(t)$. This procedure can be extended to the six-pump-beam case, where $I_{1s}(t) \sim I_{6s}(t)$ were obtained at $\Delta f_1 = 0Hz$, $\Delta f_2 = 0.19Hz$, $\Delta f_3 = 0.3Hz$, $\Delta f_4 = 0.54Hz$, $\Delta f_5 = 0.66Hz$ and $\Delta f_6 = 0.78Hz$, respectively. The simulated output in Fig. 17 has a pulse width about 1 s and a delay time about 4 seconds. The improved pulse shape implies a system bandwidth of at least 2 Hz, which is a significant improvement over the 0.05 Hz in Tab. II. The delay-time-bandwidth product also increases from 0.26 to around 8.

E.   Summary of the slow light delay-time-bandwidth product

In conclusion, we have proposed a general technique to solve the delay-time-bandwidth limitation in solid-state room-temperature slow light. To demonstrate proof of principle, we have performed an experiment in a Ce:BaTiO$_3$ crystal (photos of the setup in Fig. 18 and Fig. 19), wherein two to three spectral components of an input optical pulse were simultaneously slowed. A theoretical projection for multiple pump beams shows that the effective delay-time-bandwidth product can be significantly extended with this approach.
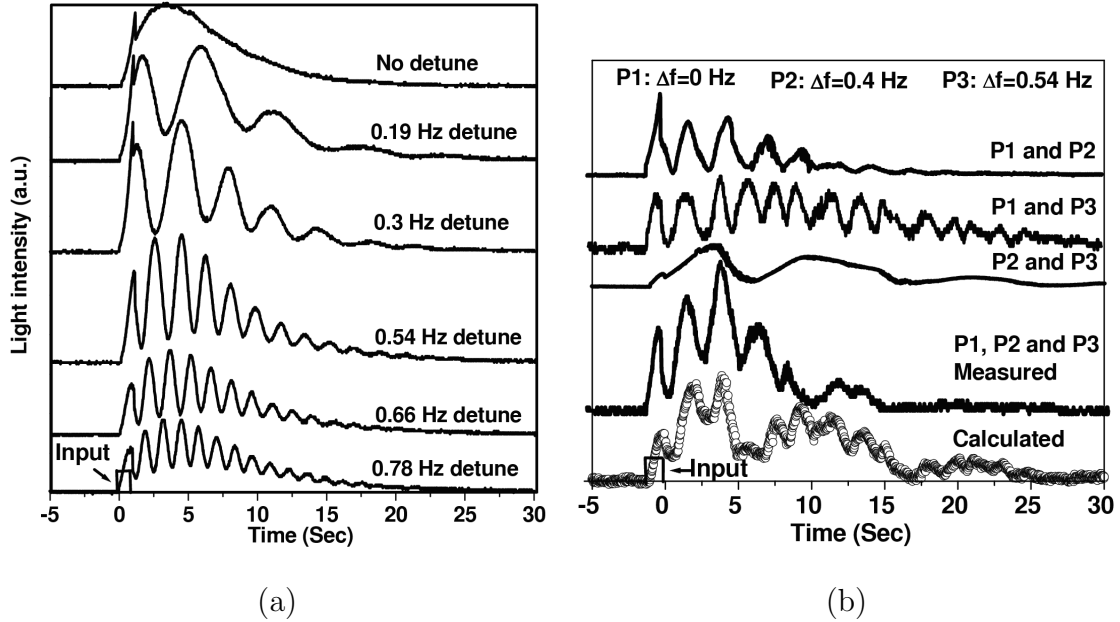
Fig. 16. Experimental demonstration of artificial inhomogeneous broadening in ultra-slow light. Demonstrated is the simultaneous slowing of up to three Fourier components of signal beam S, which is a rectangle pulse with a duration of 1 second and a repetition time of 40 seconds. (a) Output pulses when two pump frequencies are used. From top to bottom, P2 is detuned away from P1 by 0, 0.19, 0.3, 0.54, 0.66 and 0.78 Hz, respectively. (b) Detailed analysis of the case when three pump beams are used. P2 and P3 are detuned away from P1 by 0.4 Hz and 0.54 Hz, respectively. The top three curves are the output pulses when two of the three pump beams are applied, respectively. The next trace is the output when all the three pump beams are turned on. The bottom trace, with circles, is a theoretical prediction corresponding to the three frequency data.
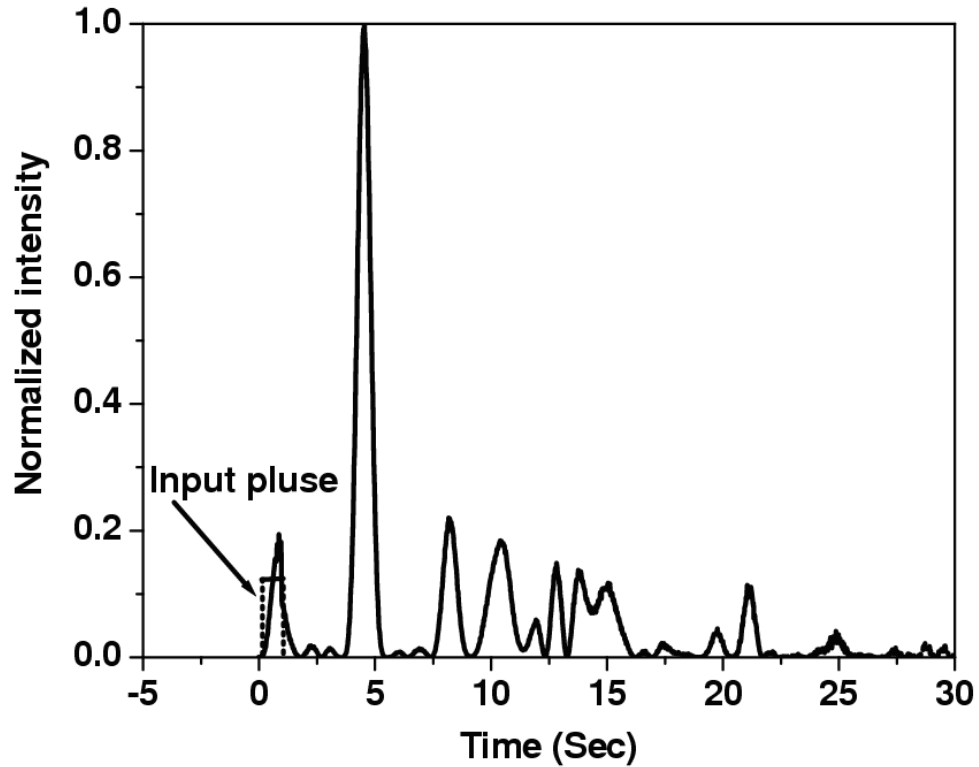
Fig. 17. Simulated improvement of the pulse distortion in ultraslow light. The simulation is based on six repeated measurements of two-wave mixing between a single probe and a single pump. The frequency detunes of the pump beams in these six measurements are $\Delta f_1 = 0Hz$, $\Delta f_2 = 0.19Hz$, $\Delta f_3 = 0.3Hz$, $\Delta f_4 = 0.54Hz$, $\Delta f_5 = 0.66Hz$ and $\Delta f_6 = 0.78Hz$, respectively.
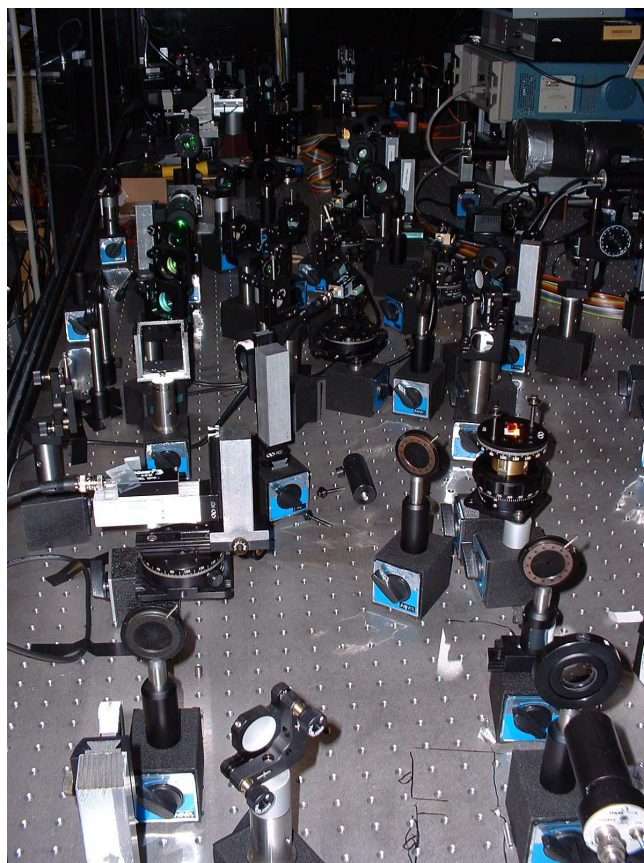
(a)
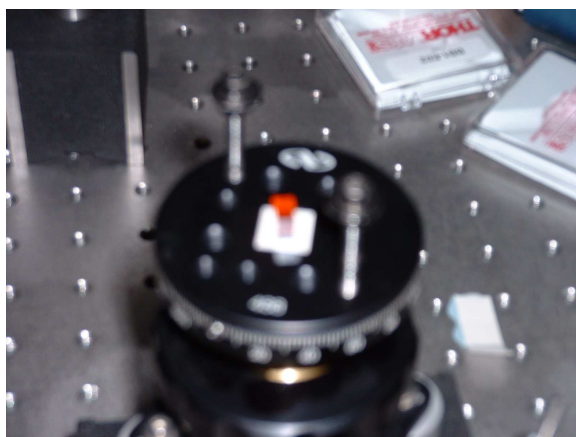
(b)

Fig. 18. (a) Optics setup for the ultraslow light demonstration. (b) Close-up of the Ce:BaTiO$_3$ photorefractive crystal mounted on a rotation stage.
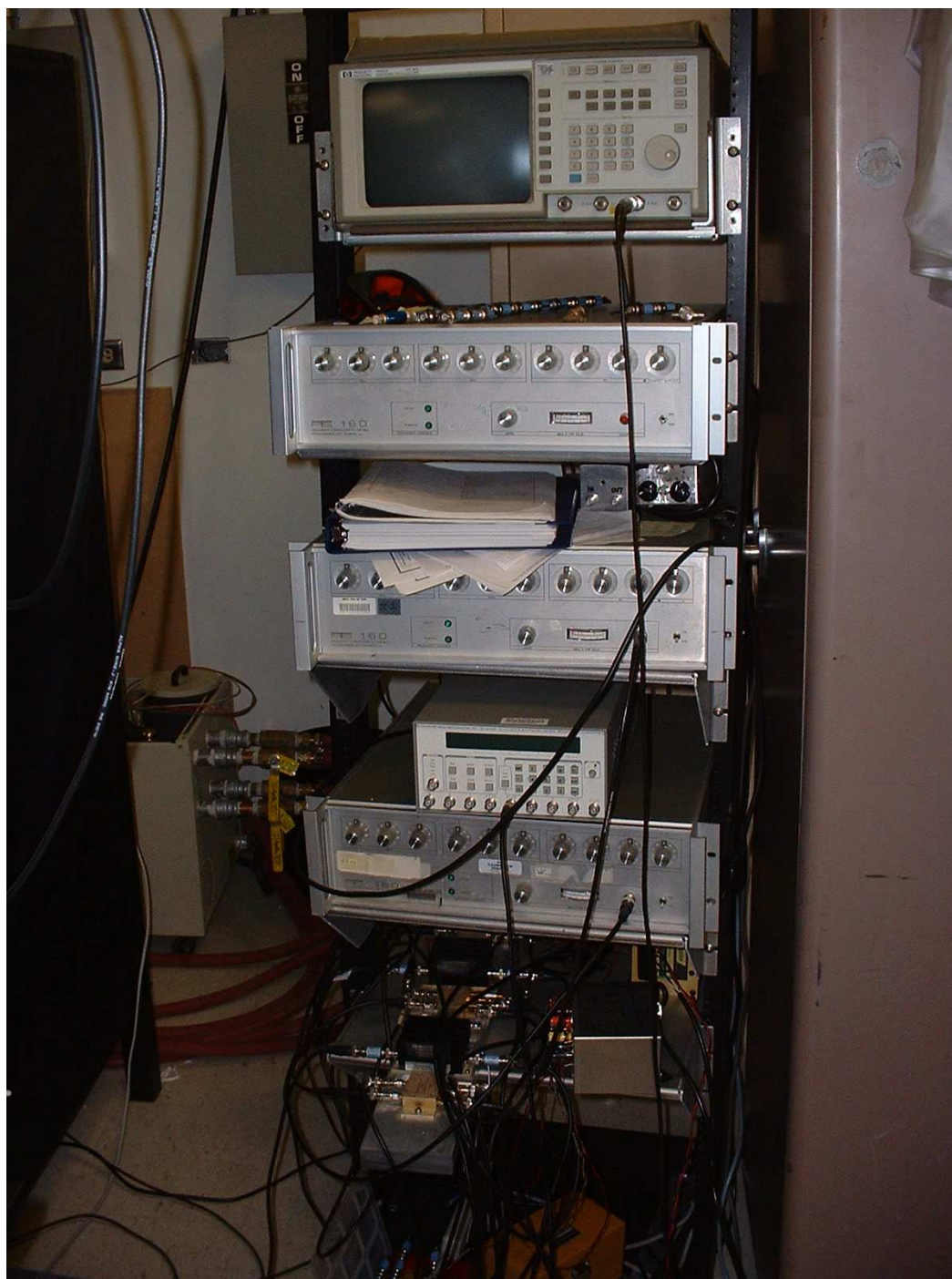
Fig. 19. Electronics in the ultraslow experiment for light modulation and detection.

CHAPTER IV

SURFACE PLASMONIC NANO WAVEGUIDES AND COUPLING ANTENNAS[*]

A.   Overview

Surface plasmon (SP) wave is the coupled oscillation of electrons and electromagnetic waves that are tightly bound to metal surfaces. The tight focus of the field introduces many novel optical phenomena which promises applications for the information age to come [38, 39]. For example, superlens has been demonstrated using SP effect that could resolve details below the diffraction limit [42], which may be used for near-field nano-lithography as well as microscopy. SP structures have also been used to enhanced light emission efficiencies in light emitting diodes (LEDs) by 32-fold [53], organic LEDs (OLEDs) [54] and lasers [41], which may serve as superior light sources for displays, communications and remote sensing. Highly directional beam forming has also been reported by utilizing the SP waves in hole arrays [55], which can boost the performance of optical sources and detectors.

One fundamental device of special interests is the SP waveguide. Their geometries may be simple, but they will be the building blocks of the future SP subwavelength circuits. Active research have been going on to understand their operations and identify potential applications. For example, Takahara *et al.* proposed SP guides based on circular metal fibers of nano scale (around 100 nm or below) diameters [56]. Such wires were suggested for making single photon sources [57].

Major concerns for SP waveguide performances are the mode size and propaga-

tion loss. SP waveguides can achieve mode sizes well below the wavelength, which are probably the most significant advantage over their dielectric counterparts. However, SP waveguides suffer more from the propagation attenuation, because metals are lossier than dielectrics. These two qualities are hard to optimize simultaneously. To make things worse, it is still a challenging task to characterize the mode sizes and losses for nano SP waveguides, which are important for many applications that have been proposed. This is attributed to the difficulties in the designing, fabricating and testing of such tiny waveguides. The design problems are due to the lack of mature computation methods for arbitrary SP structures that are accurate, fast and easy to implement. It is another problem to fabricate devices at the nano scale in a precise and repeatable manner. Testing of a single nano device is also in question due to the weak signal, which is in turn caused by the coupling issue of the small device to the much larger outside world. Tab. III compiles the recent progress on SP waveguides.

Table III. Summary of recent research on surface plasmon waveguides. $\star$: depending on metals and exact dimensions; $*$: depending on coupling alignments; $\dagger$: experimental; $\ddagger$: theoretical.

Ref.	Size (μm)	Range (μm)	λ (nm)	Notes
[58–60]	$\sim 8 \times 0.02$	$\sim 1000^\star$	1.55	Simulated by the method of lines. Made by photolithography. End-fire coupled to fibers. Poor localization.
[61, 62]	$\sim 1 \times 0.07$	< 10	0.633	No rigorous calculation. Asymmetric cladding. Made by photolithography or EBL. Prism coupled.
[63, 64]	0.09×0.03	0.2	0.633	Simulation by FDTD.

Table III. (continued)

Ref.	Size (μm)	Range (μm)	λ (nm)	Notes
	$\times 0.03$ chain			Asymmetric cladding. Made by EBL. Excited by NSOM.
[65, 66]	0.2×0.05	2.5	0.800	No rigorous calculation. Asymmetric cladding. Made by EBL. Prism coupled to a tapered strip.
[67]	0.6×1 V groove	$90 \sim 250^*$	~ 1.5	Calculation by effective index method. No cladding. Made by focused ion beam. End-fire coupled to fibers.
[68]	0.73×1 wedge	$1.5^\dagger$ $2.25^\ddagger$	0.633	Simulation by FDTD. No cladding. Made by focused ion beam. Coupled to in-plane apertures.
[69, 70]	6.2×0.3	$50^\dagger$ $320^\ddagger$	~ 1.6	Simulation by FDTD. Si/Au/air. Made by lithography. Excited as a whole by a tapered fiber in parallel.
[71]	0.23×0.12 dot chain	4.0	0.785	Simulation by FDTD. Made by focused ion beam. Excited at the focus of a SP condenser.
[72]	0.25×0.05	2.0	0.532	Calculation with dipole models. Made by focused ion beam. Excited at the focus of a SP condenser.
[73]	0.02	N/A	0.532	No rigorous calculation. Made

Table III. (continued)

Ref.	Size (μm)	Range (μm)	λ (nm)	Notes
			0.820	by template-directed electrosynthesis. Excited by a total internal reflection objective.

If a tightly bound mode is supported by a SP waveguide, a general observation is that the propagation loss will increase as the cross-section shrinks. As shown in Tab. III, the reported propagation lengths do not exceed two microns so far for subwavelength waveguides in the visible spectrum [72]. The waveguide width was 250 nm, about half of the wavelength. However, the propagation length is much longer in the infrared (IR) region. It is due to the fact that the optical properties of metals are highly dispersive across visible and IR domain. As the wavelength goes up, the field is more displaced into the low-loss dielectric cladding rather than the lossy metal. Extremely long travel range was reported with a strip that has a width five times of the telecommunication wavelength in the test [60].

For applications in detectors and light sources, it is desirable to have a field concentration as high as possible. We will look into a nano SP waveguide working in the visible with critical dimensions in the 100 nm regime or bellow, where few attempts have ever been made. Its propagation loss will be characterized, as the importance has been emphasized already. This is possible by introducing a novel coupling mechanism between the nano waveguide and the conventional diffraction-limited optics. Tab. III evidently shows the commonplace end-fire coupling with fibers is challenged by its large mode size mismatch to the ever decreasing SP waveguides [67]. Special coupling

techniques have been used, such as hole/dot array condensors [71, 72] and apertures [68]. We will apply the grating coupler concept to design the couplers.

Since the field is strongly focused in SP devices in general, variations in the structure dimensions can affect the performance significantly. Accurate simulations and precise fabrications are required to guarantee predictable optical responses. Numerical techniques are widely used to analyze the devices, such as the finite difference in time domain (FDTD) method. Tools in semiconductor research are strongly favored for making them, such as electron beam lithography (EBL) and focused-ion-beam (FIB) milling. They are mature technologies for making nano devices, and more importantly, prototypes fabricated this way can be easily transferred to manufacturing processes in integrated circuit (IC) industry. In regard of these issues, we will use FDTD and the Galerkin method to design the devices, and use EBL to fabricate them.

Here is the organization of this chapter. In section B, theories of surface plasmon will be reviewed. In section C, experiment details will be covered. Discussions on the measurement results will also be presented. In section D, a study will be presented on the interaction between an atom and the SP waveguide. The last section will conclude this chapter.

B. Design of the surface plasmon waveguide and couplers

The dielectric constants of metals can be explained classically by the Drude model of free electron dipolar oscillations under an external field

$$\epsilon = \epsilon' + i\epsilon'' = 1 - \frac{\omega_p}{\omega^2 + i\Gamma\omega} \ , \tag{4.1}$$

where ω is the angular frequency of the external field, ω_p is the plasma frequency representing the ensemble inertia of the free electron gas, and Γ is the collision frequency representing the loss caused by the inter-electron collisions [74]. Experimentally measured values approximately follow (4.1). But there are noticeable deviations in the visible region for copper and gold. Quantum mechanically, it is due to the interband absorptions by these metals, which give the reddish and golden colors [43]. Silver is believed to be the best metal for surface plasmon experiments in the visible spectrum, not only because it fits the Drude model to a high degree, but also because it has the lowest loss among metals in this wavelength range. For geometries bigger than 20 nm, the Drude model packages the optical properties of metals in the convenient form of dielectric constants. Quantum effects become prominent for smaller structures, where the dielectric constant becomes case dependent [43]. In our research, the device size will be bigger than 20 nm. The Drude model is still a convenient and good choice for the design purpose. At the target wavelength of 632 nm, the measured dielectric constants of silver are not consistent in literatures [75, 76]. We will use -15.78+1.07i in the calculations [75]. The plasmas frequency 1.2175×10^{16} rad/s and collision frequency 1.8869×10^{14} Hz were inferred to fit the Ag permittivity at the wavelength of $632 nm$.

The simplest SP waveguide is probably the interface between a semi-infinite dielectric medium and a semi-infinite bulk metal, along which the SP wave propagates. A straightforward treatment with Maxwell equations reveals the field decays exponentially in both directions that are perpendicular to the interface [31]. Thus the field is highly concentrated around the surface. The propagation constant of the surface wave is

$$k = k' + ik'' = \frac{\omega}{c}\sqrt{\frac{\epsilon_m \epsilon_d}{\epsilon_m + \epsilon_d}} \approx \frac{\omega}{c}\sqrt{\frac{\epsilon'_m \epsilon_d}{\epsilon'_m + \epsilon_d}} + i\frac{\omega}{2c}\left(\frac{\epsilon'_m \epsilon_d}{\epsilon'_m + \epsilon_d}\right)^{3/2}\frac{\epsilon''_m}{(\epsilon'_m)^2}, \qquad (4.2)$$

where $\epsilon_m = \epsilon_m' + i\epsilon_m''$ is the dielectric constant of the metal, and ϵ_d is the dielectric constant of the surrounding medium. In the visible and IR, $\epsilon_m' < \epsilon_m' + \epsilon_d < 0$ in general. Therefore,

$$k' > \frac{\omega\sqrt{\epsilon_d}}{c}, \tag{4.3}$$

where the right hand side is the propagation constant of plane waves in the dielectric medium. The momentum mismatch in (4.3) prevents the SP from radiative loss into the open space. The propagation length, as given by the inverse of k'', not only depends on the loss property of the metal, but also on the permittivity contrast between the metal and the dielectric. Moreover, it can qualitatively explain the pronounced growth of traveling range in IR. Metal permittivities are generally highly dispersive across visible and IR, where $|\epsilon_m'|$ and $|\epsilon_m''|$ both increase rapidly as the wavelength grows. Accordingly, the k'' drops hard due to the nonlinearly diminishing term $\epsilon_m''/(\epsilon_m')^2$ in (4.2). In the physical picture, the rapid increase in $|\epsilon_m'|$ and $|\epsilon_m''|$ leads to a field more displaced to the loss free dielectric side rather than the lossy metal, reducing the dissipations. Although we arrive at the above conclusions in a very simple SP waveguide, they repeat themselves in more complicated guiding structures.

Among the planar circuits that can be fabricated by EBL, the most simple SP waveguide is probably a rectangular silver wire embedded in a dielectric medium. Although it sounds simple, no demonstration has been made for such nano waveguides working in the visible spectrum. It is the goal of our project to accomplish that. To couple the waveguide to conventional diffraction-limited optics such as a microscope objective, a tapering structure is first designed at both ends of the wire to enable a smooth transition from the subwavelength waveguide mode size to the focus size of an objective that is on the order of the wavelength. The tapering structures are

then connected to grating couplers (antennas), which convert the SP wave into out-of-plane optical waves that can be collected by the objective, and vise verse. In order to do this, the grating period Λ has to be matched to the the SP wavelength λ_{sp}

$$\Lambda = \lambda_{sp} \ . \tag{4.4}$$

Detailed designs and performance optimizations of the device relies on solving the Maxwell equations.

Unfortunately, the close form solutions of the Maxwell equations are only available in rare cases. The mathematical solutions are usually too complicated to handle without the help of computers. Previously when SP research was more focused on noble metal particles and their clusters, the transfer matrix method [32] and the discrete dipole approximation method [77] were two popular algorithms for simulations . Now our interests move from particles to waveguide and couplers. It makes more sense to follow the paths of integrated optics and microwave simulations, where similar devices are well-studied [78, 79]. Promising numerical techniques are the finite difference in time domain (FDTD) method and the Galerkin method. The FDTD method is a very popular technique which is a direct discrete emulation of the Maxwell equations in the time domain. It is computational intensive, but versatile to address almost any problems including waveguides and grating couplers. The Galerkin method was designed to analyze optical dielectric waveguides in the harmonics domain by expanding the field into Fourier series and reducing it to a standard matrix eigenvalue problem. It is fast and simple in handling waveguides. A comparison of these two methods will give more insights into the simulations of SP devices. The next two subsections will describe the waveguide and couplers design with FDTD and the Galerkin method, respectively.

(a)   (b)
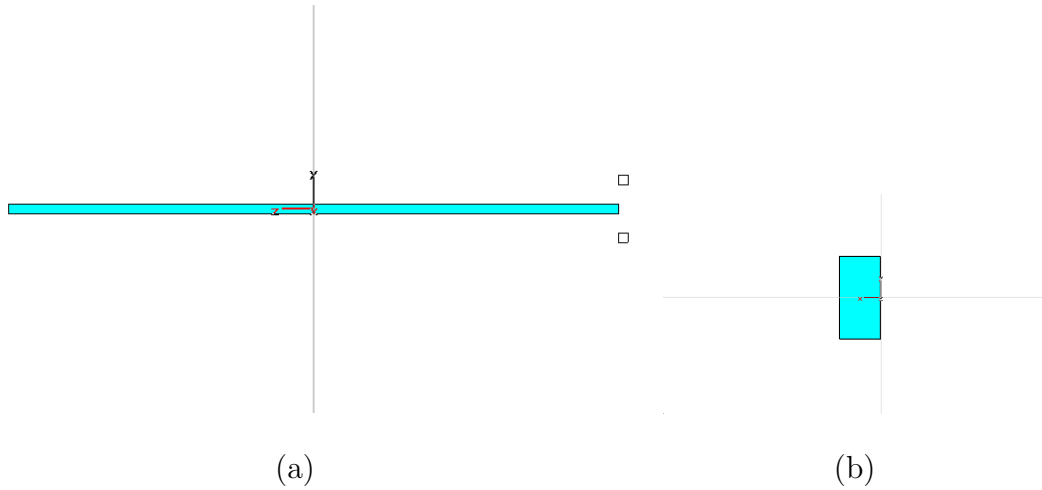
Fig. 20. Schematic of the SP waveguide in FDTD simulation: (a) top view; (b) back
view of the $100nm \times 50nm$ cross-section. The $6.3\mu m$ long silver core is em-
bedded in MgF_2.

1.   FDTD simulations of a rectangular SP waveguide and grating antennas

Finite difference in time domain method was first proposed by Yee in 1966 to directly
emulate the evolution of the electromagnetic field in a leap-frog manner. It can
simulate almost any geometries as long as the material model is provided. FDTD has
been applied to electromagnetic problems in microwave and photonics [80]. A popular
commercial tool in the microwave society is CST Microwave Studio, which implements
FDTD in three dimensional (3D) space. The Drude model was incorporated in the
simulator recently. It is a widely used model for metallic structures greater than 20
nm. Therefore, it is chosen to simulate the waveguide and grating couplers. Fig. 20
shows the waveguide in simulation. The rectangular cross-section is $100nm \times 50nm$.
In the simulation, the strip length is $6.3\mu m$. The surrounding medium is MgF$_2$, which
has a refractive index of 1.38, or a dielectric constant of 1.9044 [75].

The resulting lowest loss mode is dominantly polarized in the plane and perpen-

```
V/m
2.14e8
8.98e7
3.97e7
0
-2.39e7
-6.1_e7
-1.29e8
-2.14e8
```

```
Type        = E-Field (peak)
Monitor     = e-field (f=474683) [1]
Component   = y
Plane at x  = 0.025
Frequency   = 474683
Phase       = 225 degrees
Maximum-2d  = 2.13807e+008 V/m at 0.025 / 0.05 / -3.185
```
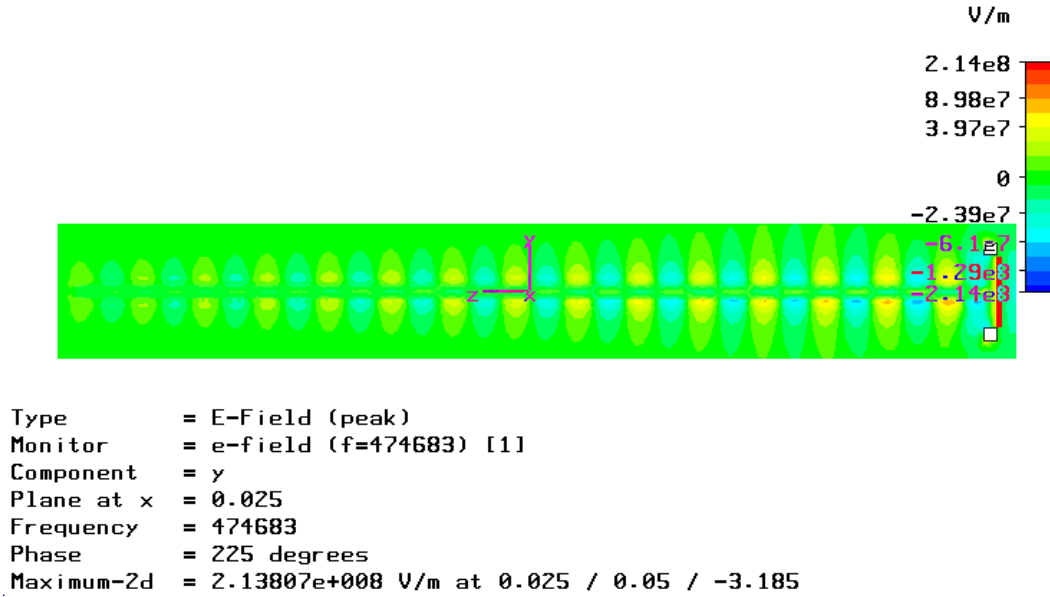
Fig. 21. SP wave propagation along the Ag wire in the top view. Only the dominant polarization is shown here. The electric field is primarily in the plane and perpendicular to the axis.

dicular to the axis. Fig. 21 shows the SP wave propagation along the axial direction in the top view. The attenuation is clearly seen due to the metallic loss. It is determined to be $1.8dB/\mu m$. Fig. 22 shows the grating that couples the SP wave to free space radiations. The SP wave is first spread out by coupling to neighboring strips (with a spacing of 200 nm) parallel to the waveguide. It is then fed to a grating area to be converted into outgoing waves in the direction normal to the plane. The grating vector must be equal to the SP propagation vector to do this, which means the periodicity of the grating in the axial direction is equal to the SP wavelength of 436 nm in the simulated mode. Thus the wavelet scattered by each island in the gratings area will add up constructively in the surface normal direction. Fig. 22(c)(d) shows polarization sensitiveness of the grating radiation. There is a main lobe along the out-of-plane direction in the radiation pattern of the dominant polarization, while

the cross polarization is suppressed in radiation. It implies the grating coupler may function as a polarizing mode-selector.
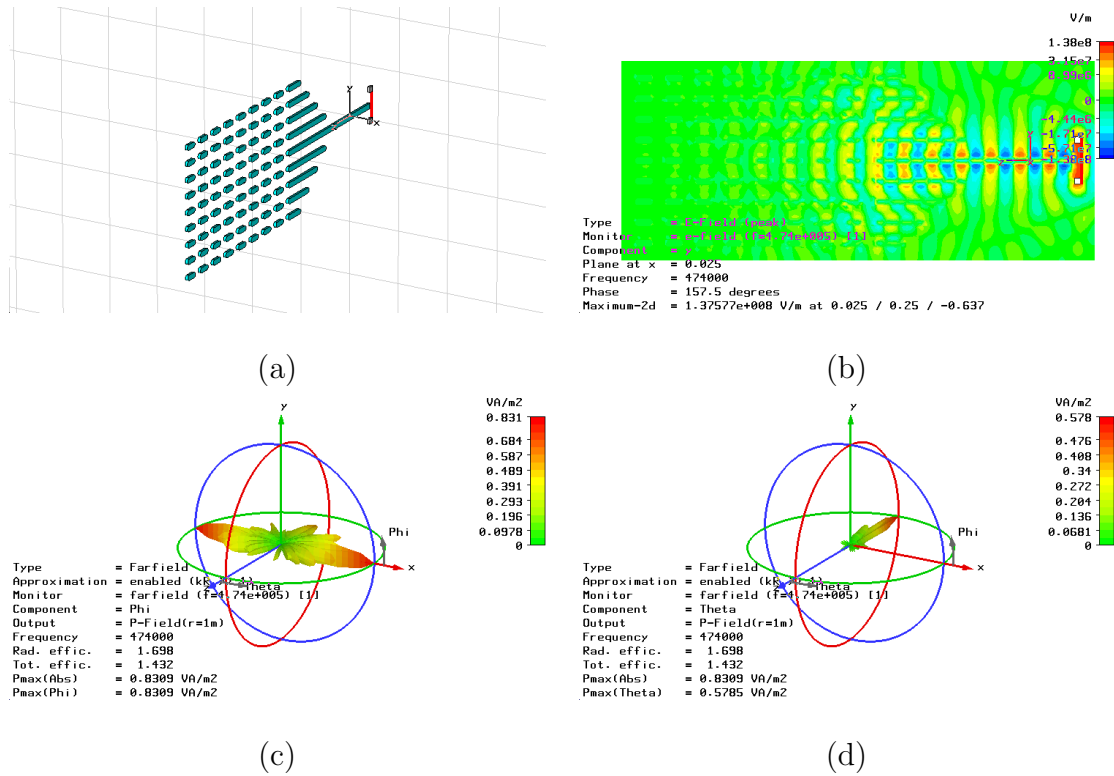
Fig. 22.   Radiation far field of the grating coupler. (a) perspective overview of the grating coupler; (b) propagation of the dominant polarization in the grating area, top view; (c) the radiating far field of dominant polarization in the perspective overview; (d) the radiating far field of the component cross to the dominant polarization in the perspective overview.

2.   Galerkin analysis of rectangular SP waveguides

Although FDTD is versatile and able to obtain simulation results at different frequencies simultaneously, it has well-known drawbacks of huge memory demands and slow computation speed, especially when the simulation domain is large in terms of wavelength. Unfortunately, the waveguide simulation, which is a long strip running

multiple wavelengths, falls into one of such worst categories. The problem will be aggravated when more than one modes have to be calculated. It would be desirable to look into some other simulation tools that have advantages on waveguide problems. Comparison would also be beneficial to clarify uncertainties about the CST Microwave Studio, as it has not been extensively tested for SP simulations. Among the many algorithms targeting waveguides, the mapped Galerkin method is attractive due to the relative simplicity in implementation and fast computation speed. It is also capable of solving multiple modes in a single run.

Henry and Verbeek proposed the scalar wave solutions for dielectric waveguides in 1989 using the Galerkin method. The scalar wave assumption will be violated if the permittivity contrast is high between the core and the cladding. Marcuse addressed this problem by deriving the vectorial form of the method [81]. In both the scalar and vector scenarios, the field is decomposed by Fourier basis functions and the Maxwell equations are subject to projections onto each basis. The projections lead to a matrix eigenvalue problem, where the eigenvector to be solved contains the decomposition coefficients and the unknown eigenvalue will give out the propagation constant of the guided mode. The aforementioned projections were calculated by integrals over a large but limited domain as suggested by Marcuse, which is at odds with the open space that surrounds the waveguide in reality. This question was answered by Hewlett and Ladouceur, though in the scalar case only. The open space including the waveguide is first mapped to a unit square in a transformed domain. The Galerkin method is then applied to solve the problem in the mapped domain [82]. Recently, Xiao *et al.* extended the mapped Galerkin method to the vectorial case using an electric (E) field formulation, to analyze guided modes in optical dielectric waveguides including rectangular ones [45].

The SP waveguide has a dielectric constant with a different sign from that of the

cladding. It seems Xiao's formulation is closest to what we need. A caveat has to be pointed out here, that the E field component normal to the waveguide/cladding interface will change sign across it, as enforced by Gauss's law. Such a disruptive change will result in a slow convergence of the Fourier series decomposition, which has not been encountered in the previous research on dielectric waveguides. Therefore, we will derive our own vectorial formulation of the mapped Galerkin method based on the magnetic (H) field, which is continuous across the waveguide/cladding interface.

a.   *H-field formulation of the mapped Galerkin method*

Fig. 23(i) shows the cross-section of the rectangular SP waveguide. The metal core size is $2a_0 \times 2b_0$ and has a permittivity of ϵ_m. The surrounding dielectric medium has a permittivity of ϵ_d. The time harmonic (assuming $\exp^{-j\omega t}$) form of Maxwell equations are

$$\nabla \times \mathbf{E} \ = \ j\omega\mu\mathbf{H} \ , \tag{4.5}$$

$$\nabla \times \mathbf{H} \ = \ -j\omega\epsilon\mathbf{E} \ , \tag{4.6}$$

where μ is the permeability, and

$$\epsilon = \epsilon(x,y) = \begin{cases} \epsilon_m & -a_0 \leq x \leq a_0 \ and \ -b_0 \leq y \leq b_0 \\ \epsilon_d & otherwise \end{cases} \tag{4.7}$$

is the spatial distribution of the permittivity that is axial independent.

Equations (4.5) and (4.6) can be combined to yield

$$\nabla^2\mathbf{H} + \omega^2\mu\epsilon\mathbf{H} - j\omega\nabla\epsilon \times \mathbf{E} = 0 \ . \tag{4.8}$$
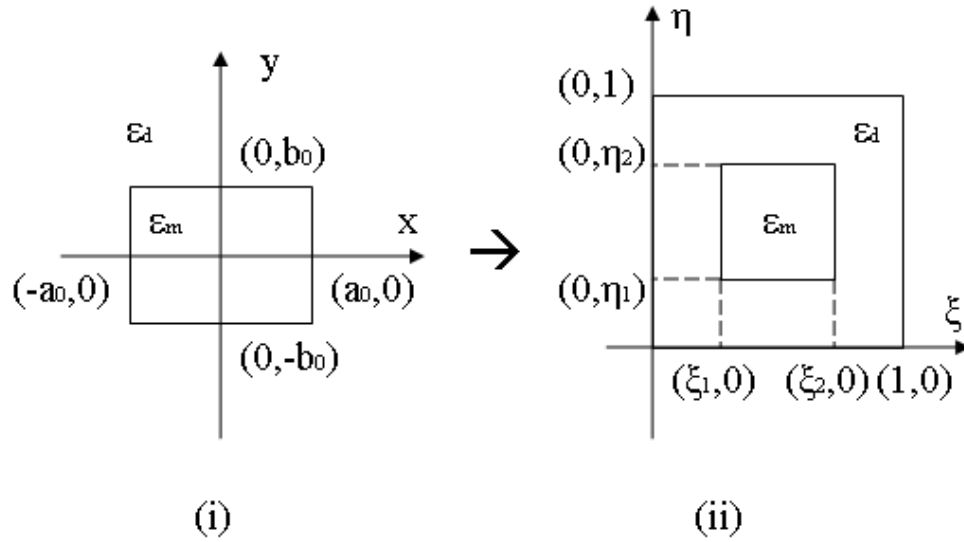
Fig. 23. Transforming the rectangular cross-section of the SP waveguide in the mapped Galerkin method: (i) in the original (x, y) domain; (ii) in the transformed (ξ, η) domain. The waveguide core region $\{-a_0 \leq x \leq a_0, -b_0 \leq y \leq b_0\}$ is metal with a permittivity of ϵ_m. It is embedded in a dielectric medium of permittivity ϵ_d. The transformation mapped $\{-\infty \leq x \leq \infty, -\infty \leq y \leq \infty\}$ to $\{0 \leq \xi \leq 1, 0 \leq \eta \leq 1\}$, and $\{-a_0 \leq x \leq a_0, -b_0 \leq y \leq b_0\}$ to $\{\xi_1 \leq \xi \leq \xi_2, \eta_1 \leq \eta \leq \eta 2\}$.

Note that the guided mode of the waveguide assumes the general form of

$$\mathbf{E} = \mathbf{U}(x, y) \exp^{j\gamma z}, \tag{4.9}$$

$$\mathbf{H} = \mathbf{V}(x, y) \exp^{j\gamma z}, \tag{4.10}$$

where $\gamma = \alpha + j\beta$ is the propagation constant.

It can be shown that any two of the six components in $\mathbf{U}$ and $\mathbf{V}$ can derive the

other four, for example,

$$U_x = \frac{\gamma}{\omega\epsilon}V_y - \frac{1}{\omega\epsilon\gamma}\left(\frac{\partial^2 V_x}{\partial x \partial y} + \frac{\partial^2 V_y}{\partial y^2}\right) , \tag{4.11}$$

$$U_y = -\frac{\gamma}{\omega\epsilon}V_x + \frac{1}{\omega\epsilon\gamma}\left(\frac{\partial^2 V_x}{\partial x^2} + \frac{\partial^2 V_y}{\partial x \partial y}\right) , \tag{4.12}$$

$$U_z = \frac{-j}{\omega\epsilon}\left(\frac{\partial V_x}{\partial y} - \frac{\partial V_y}{\partial x}\right) , \tag{4.13}$$

$$V_z = \frac{j}{\gamma}\left(\frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y}\right) . \tag{4.14}$$

Thus (4.8) becomes

$$\frac{\partial^2 V_x}{\partial x^2} + \frac{\partial^2 V_x}{\partial y^2} + (\omega^2\mu\epsilon - \gamma^2)V_x - \frac{1}{\epsilon}\frac{\partial\epsilon}{\partial y}\left(\frac{\partial V_x}{\partial y} - \frac{\partial V_y}{\partial x}\right) = 0 , \tag{4.15}$$

$$\frac{\partial^2 V_y}{\partial x^2} + \frac{\partial^2 V_y}{\partial y^2} + (\omega^2\mu\epsilon - \gamma^2)V_y + \frac{1}{\epsilon}\frac{\partial\epsilon}{\partial x}\left(\frac{\partial V_x}{\partial y} - \frac{\partial V_y}{\partial x}\right) = 0 . \tag{4.16}$$

Next the infinite $x - y$ plane is mapped to a unit square $\{0 \le \xi \le 1, 0 \le \eta \le 1\}$ on the $\xi - \eta$ plane by the following transformation

$$x = a\tan[\pi(\xi - 1/2)] , \tag{4.17}$$

$$y = b\tan[\pi(\eta - 1/2)] , \tag{4.18}$$

where a and b are mapping parameters that can affect the convergence of the solution and will be discussed later. The rectangular metal core preserves its shape on the $\xi - \eta$ plane as shown in Fig. 23(ii), where

$$\xi_1 = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\frac{a_0}{a} , \tag{4.19}$$

$$\xi_2 = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\frac{a_0}{a} , \tag{4.20}$$

$$\eta_1 = \frac{1}{2} - \frac{1}{\pi}\tan^{-1}\frac{b_0}{b} , \tag{4.21}$$

$$\eta_2 = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\frac{b_0}{b} . \tag{4.22}$$

The mapping also changes (4.15) and (4.16) into

$$\frac{\partial^2 V_x}{\partial \xi^2} \left(\frac{d\xi}{dx}\right)^2 + \frac{\partial V_x}{\partial \xi}\frac{d^2\xi}{d^2x} + \frac{\partial^2 V_x}{\partial \eta^2}\left(\frac{d\eta}{dy}\right)^2 + \frac{\partial V_x}{\partial \eta}\frac{d^2\eta}{d^2y}$$

$$+ (\omega^2\mu\epsilon - \gamma^2)V_x - \frac{1}{\epsilon}\frac{\partial \epsilon}{\partial \eta}\frac{d\eta}{dy}\left(\frac{\partial V_x}{\partial \eta}\frac{d\eta}{dy} - \frac{\partial V_y}{\partial \xi}\frac{d\xi}{dx}\right) = 0 , \qquad (4.23)$$

$$\frac{\partial^2 V_y}{\partial \xi^2} \left(\frac{d\xi}{dx}\right)^2 + \frac{\partial V_y}{\partial \xi}\frac{d^2\xi}{d^2x} + \frac{\partial^2 V_y}{\partial \eta^2}\left(\frac{d\eta}{dy}\right)^2 + \frac{\partial V_y}{\partial \eta}\frac{d^2\eta}{d^2y}$$

$$+ (\omega^2\mu\epsilon - \gamma^2)V_y + \frac{1}{\epsilon}\frac{\partial \epsilon}{\partial \xi}\frac{d\xi}{dx}\left(\frac{\partial V_x}{\partial \eta}\frac{d\eta}{dy} - \frac{\partial V_y}{\partial \xi}\frac{d\xi}{dx}\right) = 0 , \qquad (4.24)$$

where

$$\frac{d\xi}{dx} = \frac{1}{2\pi a}\left[1 - \cos(2\pi\xi)\right] , \qquad (4.25)$$

$$\frac{d^2\xi}{dx^2} = \frac{1}{2\pi a^2}\sin(2\pi\xi)\left[1 - \cos(2\pi\xi)\right] , \qquad (4.26)$$

$$\frac{d\eta}{dy} = \frac{1}{2\pi b}\left[1 - \cos(2\pi\eta)\right] , \qquad (4.27)$$

$$\frac{d^2\eta}{dy^2} = \frac{1}{2\pi b^2}\sin(2\pi\eta)\left[1 - \cos(2\pi\eta)\right] . \qquad (4.28)$$

If a vector V is defined

$$V = [V_x, V_y]^T , \qquad (4.29)$$

(4.23) and (4.24) can be rearranged into

$$\Phi V = \gamma^2 V , \qquad (4.30)$$

where

$$\Phi = \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{bmatrix}, \tag{4.31}$$

$$\mathcal{A} = \left(\frac{\mathrm{d}\xi}{\mathrm{d}x}\right)^2 \frac{\partial^2}{\partial \xi^2} + \frac{\mathrm{d}^2\xi}{\mathrm{d}^2 x} \frac{\partial}{\partial \xi} + \left(\frac{\mathrm{d}\eta}{\mathrm{d}y}\right)^2 \frac{\partial^2}{\partial \eta^2} + \frac{\mathrm{d}^2\eta}{\mathrm{d}^2 y} \frac{\partial}{\partial \eta}$$
$$+ \omega^2 \mu \epsilon - \frac{1}{\epsilon} \frac{\partial \epsilon}{\partial \eta} \frac{\mathrm{d}\eta}{\mathrm{d}y} \frac{\mathrm{d}\eta}{\mathrm{d}y} \frac{\partial}{\partial \eta}, \tag{4.32}$$

$$\mathcal{B} = \frac{1}{\epsilon} \frac{\partial \epsilon}{\partial \eta} \frac{\mathrm{d}\eta}{\mathrm{d}y} \frac{\mathrm{d}\xi}{\mathrm{d}x} \frac{\partial}{\partial \xi}, \tag{4.33}$$

$$\mathcal{C} = \left(\frac{\mathrm{d}\xi}{\mathrm{d}x}\right)^2 \frac{\partial^2}{\partial \xi^2} + \frac{\mathrm{d}^2\xi}{\mathrm{d}^2 x} \frac{\partial}{\partial \xi} + \left(\frac{\mathrm{d}\eta}{\mathrm{d}y}\right)^2 \frac{\partial^2}{\partial \eta^2} + \frac{\mathrm{d}^2\eta}{\mathrm{d}^2 y} \frac{\partial}{\partial \eta}$$
$$+ \omega^2 \mu \epsilon - \frac{1}{\epsilon} \frac{\partial \epsilon}{\partial \xi} \frac{\mathrm{d}\xi}{\mathrm{d}x} \frac{\mathrm{d}\eta}{\mathrm{d}y} \frac{\partial}{\partial \eta}, \tag{4.34}$$

$$\mathcal{D} = -\frac{1}{\epsilon} \frac{\partial \epsilon}{\partial \xi} \frac{\mathrm{d}\xi}{\mathrm{d}x} \frac{\mathrm{d}\xi}{\mathrm{d}x} \frac{\partial}{\partial \xi}. \tag{4.35}$$

The guided mode solutions in the transformed domain are available by solving (4.30) with the boundary condition that V must be zero at the edges of the outer square in Fig. 23(ii), which implies the guided modes must be bounded and be zero at infinity in the original domain.

The unknown V_x and V_y can be represented by a series of orthonomal sinusoidal functions

$$V_x = \sum_{m=1}^{M} \sum_{l=1}^{L} 2\sin(m\pi\xi)\sin(l\pi\eta)c_{ml}^x, \tag{4.36}$$

$$V_y = \sum_{p=1}^{P} \sum_{q=1}^{Q} 2\sin(p\pi\xi)\sin(q\pi\eta)c_{pq}^y. \tag{4.37}$$

All the sinusoidal basis functions satisfy the boundary condition of being zero at the perimeter of the outer square in Fig. 23(ii). Thus a solution to (4.30) in the form of (4.36) and (4.37) is implicitly guaranteed to be a guided mode. For convenience in

formulation, (4.29), (4.36) and (4.37) are combined in the matrix form

$$V = \begin{pmatrix} V_x \\ V_y \end{pmatrix} = \begin{bmatrix} \mathcal{F} & 0 \\ 0 & \mathcal{G} \end{bmatrix} \begin{pmatrix} C^x \\ C^y \end{pmatrix} = \Psi C \ , \tag{4.38}$$

where

$$\mathcal{F} = [f_{11}, f_{12}, \ldots, f_{1L}, \ldots, f_{ml}, \ldots, f_{M1}, f_{M2}, \ldots, f_{ML}] \ , \tag{4.39}$$

$$f_{ml} = 2 \sin(m\pi\xi) \sin(l\pi\eta) \ , \tag{4.40}$$

$$\mathcal{G} = [g_{11}, g_{12}, \ldots, g_{1Q}, \ldots, g_{pq}, \ldots, g_{P1}, g_{P2}, \ldots, g_{PQ}] \ , \tag{4.41}$$

$$g_{pq} = 2 \sin(p\pi\xi) \sin(q\pi\eta) \ , \tag{4.42}$$

$$C^x = [c_{11}^x, c_{12}^x, \ldots, c_{1L}^x, \ldots, c_{ml}^x, \ldots, c_{M1}^x, c_{M2}^x, \ldots, c_{ML}^x]^T \ , \tag{4.43}$$

$$C^y = \left[c_{11}^y, c_{12}^y, \ldots, c_{1Q}^y, \ldots, c_{pq}^y, \ldots, c_{P1}^y, c_{P2}^y, \ldots, c_{PQ}^y\right]^T \ , \tag{4.44}$$

$$\Psi = \begin{bmatrix} \mathcal{F} & 0 \\ 0 & \mathcal{G} \end{bmatrix} \ , \tag{4.45}$$

$$C = [C^x, C^y]^T \ . \tag{4.46}$$

Next, the sinusoidal expansion of the magnetic field V (4.38) is applied to the field equation (4.30), so that

$$\Phi\Psi C = \gamma^2 \Psi C \ . \tag{4.47}$$

Ψ actually forms a series of extended basis functions for vector V. Based on these extended basis functions, a projection operator can be defined as $\int_0^1 d\xi \int_0^1 d\eta \Psi^T$. It can be shown that the projection operator and the extended basis Ψ still satisfy the orthonormal condition

$$\int_0^1 d\xi \int_0^1 d\eta \Psi^T \Psi = I \ , \tag{4.48}$$

where I is an identity matrix of dimension $(ML + PQ) \times (ML + PQ)$. Applying the

projection operator to both sides of (4.47), we obtain

$$\Theta C = \gamma^2 C \; , \tag{4.49}$$

where

$$\Theta = \int_0^1 d\xi \int_0^1 d\eta \Psi^T \Phi \Psi \tag{4.50}$$

is a matrix of dimension $(ML + PQ) \times (ML + PQ)$.

Equation (4.49) is a standard eigenvalue problem of matrix Θ, where the eigenvalue γ^2 is the square of the propagation constant, and the eigenvector C is the expansion coefficients of the unknown magnetic field $V_x(\xi, \eta)$ and $V_y(\xi, \eta)$ in the transformed domain. Before the eigenvalue problem can be solved, the matrix Θ must be evaluated first using (4.31)$\sim$(4.35), (4.39)$\sim$(4.42) and (4.45). This can be done numerically, but it turns to be too time consuming. Since the rectangular waveguide has a simple shape, the evaluation of Θ can also be completed analytically. The process is straightforward but too tedious to list here. After Θ is evaluated, the eigenvalue problem can be solved with a routine in Matlab. Once $V_x(\xi, \eta)$ and $V_y(\xi, \eta)$ are available, the other four components of the field can be computed with (4.11)$\sim$(4.14), (4.25) and (4.27).

b.   Numerical results

The H-field mapped Galerkin method was first calibrated with the E-field mapped Galerkin method which Xiao $et\ al.$ used to analyze rectangular dielectric waveguides. The example in Fig. 2 of [45] is considered, where $a_0 = 2\mu m$, $b_0 = 1\mu m$, $\epsilon_m = 2.25\epsilon_0$ in the waveguide core, $\epsilon_d = 2.1025\epsilon_0$ in the cladding region, and $\lambda = 1.15\mu m$. Here ϵ_0 is the permittivity of vacuum. It is a bench mark example for several numerical methods (e.g. [83]). The effective index n_{eff} $(= \gamma\lambda/2\pi)$ is inferred from [45] to be
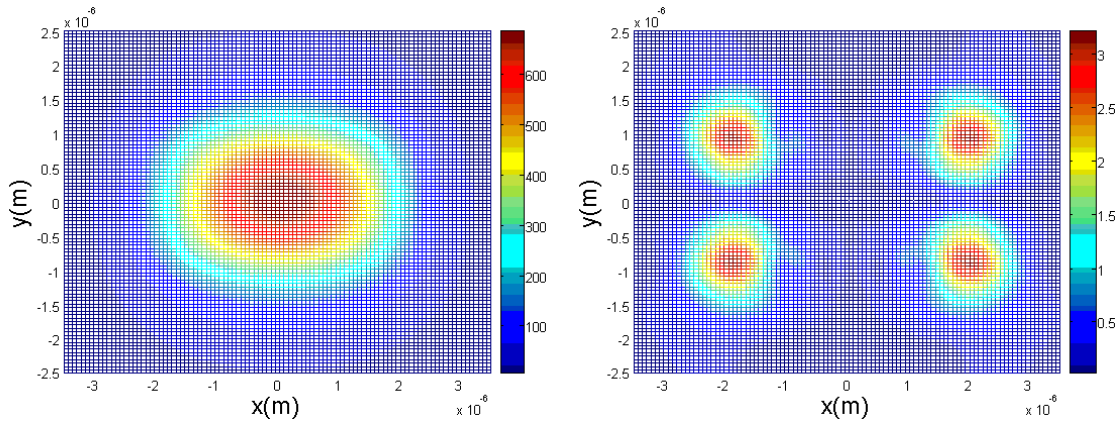
Fig. 24. Calibration of the H-field mapped Galerkin method. Field components of the fundamental mode are calculated for a dielectric waveguide: (a) x polarization; (b) y polarization. They match the results in Fig. 2 in [45].

1.4757 for the fundamental mode polarized along the long edge. The H-field mapped Galerkin method was coded in Matlab to solve the same problem. It should be noted here that the transformation constant a in (4.17) and b in (4.18) will affect the convergence. Optimal values by rule of thumb are $a = a_0$ and $b = b_0$ for rectangular dielectric waveguides. The expansion includes 800 terms of harmonics, i.e., $M = L = P = Q = 20$. The computed effective index is 1.4832 with the H-field formulation. The calculated mode is shown in Fig. 24, which matches Fig. 2(a)(b) in [45].

The tested program is now ready to be applied to the SP waveguide problem. Because the SP waveguide has more confined mode profiles than the dielectric waveguide, the transformation constants have to be tuned specifically to achieve convergence. They were found to be $a = 0.6a_0$ and $b = 1.2b_0$ in our case. The matrix Φ has dominant diagonal elements as shown in Fig. 25, which allows eigenvalue solutions by fast iterative methods like the Arnoldi algorithm. An extra advantage of these algorithms is that multiple eigenvalues (modes) can be computed in a single run. Thus
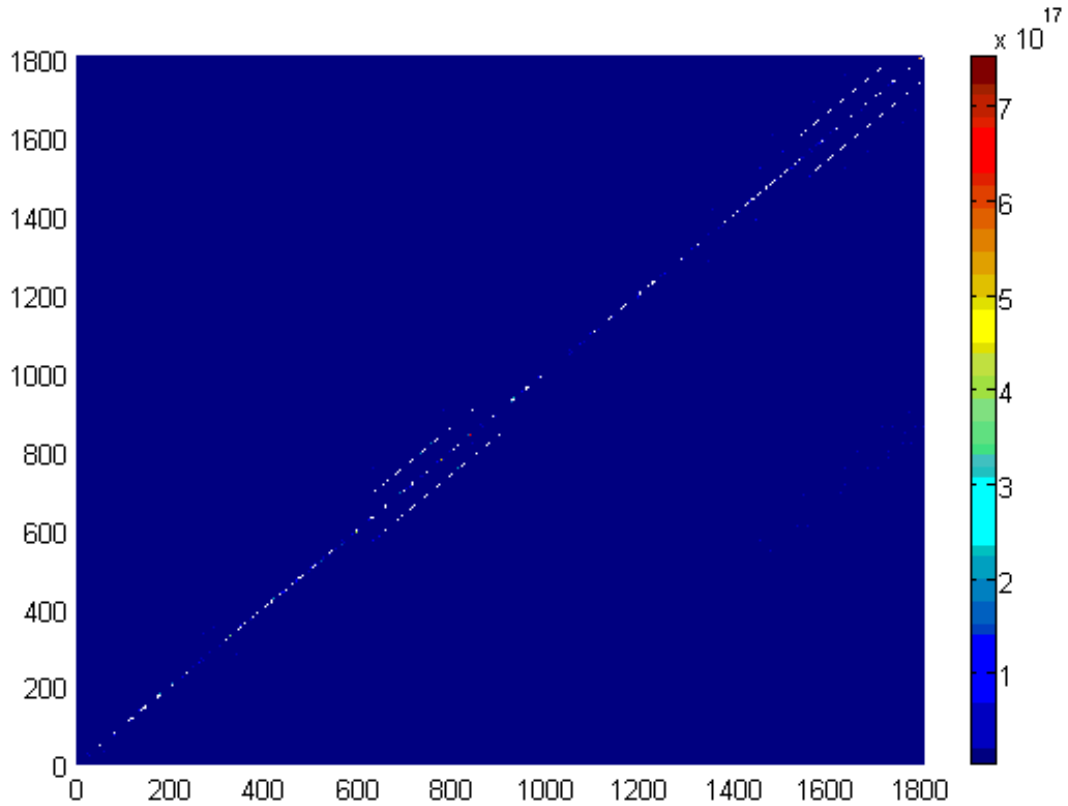
Fig. 25. Amplitudes of the matrix elements produced by the Galerkin method. The x and y axis indexes the row and column of an element in the matrix Φ produced by the Galerkin method. Amplitude of the element is encoded by color. It is a diagonal dominant matrix, which allows fast iterative algorithms to solve the eigenvalue problem.

the lowest loss mode can be identified.

The dispersion curve and attenuation of the computed mode are shown in Fig. 26. They were obtained with 1800 terms of harmonics, which means $M = L = P = Q = 30$. With an exciting laser of 632 nm wavelength, the SP wavelength of the guiding mode is $431nm$, slightly different from the $436nm$ obtained by FDTD. The propagation loss is $0.46dB/\mu m$ at this wavelength (see Fig. 26(b)) as compared to the $1.8dB/\mu m$ given by the FDTD. The computation time is shorter with the mapped
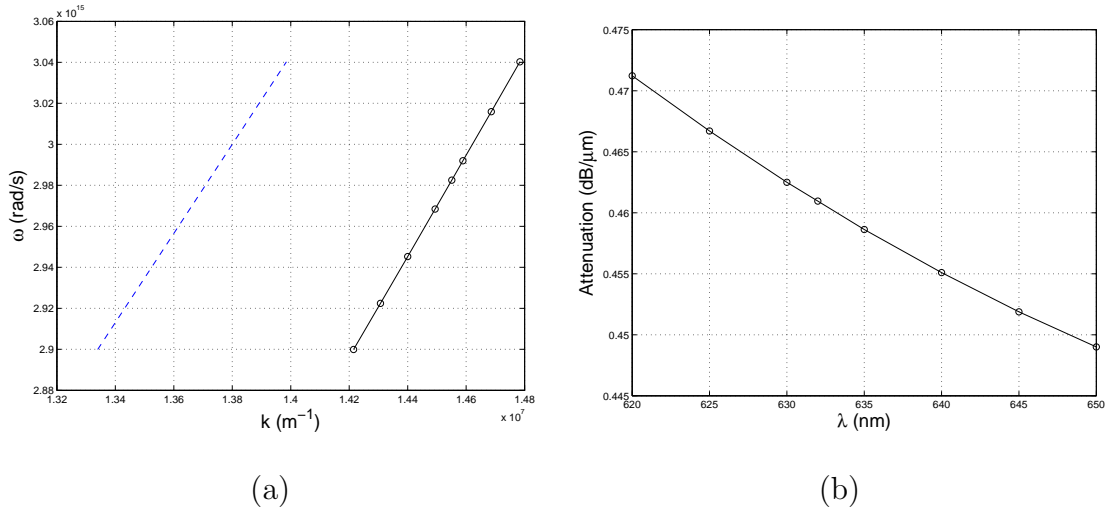
Fig. 26.   Dispersion curve and losses of the SP waveguide. (a) Dispersion curve of the
SP waveguide mode as computed by the mapped H-field Galerkin method.
It was calculated at 620 nm,625 nm, 630 nm, 632 nm, 635 nm, 640 nm, 645
nm, and 650 nm. The blue trace is the light line in MgF_2. (b) The waveguide
attenuation vs. the optical wavelength.

Galerkin method, probably due to its semi-analytical nature. It took 212 seconds, in
contrast to the 42 minutes by the FDTD package. Fig. 26(a) shows the waveguide is
operating in a moderately dispersive region. The group velocity is inferred to be 0.78
of the light speed in vacuum.

Fig. 27 shows the fields of the SP waveguide mode simulated with the H-field
mapped Galerkin method. The SP waveguide mode obtained by FDTD is also plottd
in Fig. 28 for comparison. Both methods give out electric fields polarized in the x-
direction. The distribution of $|E_x|$, $|E_y|$, $|E_z|$, $|H_x|$ and $|H_y|$ in the FDTD solution are
somehow compressed or stretched in the Galerkin solution. The dominant field E_x is
more loosely bound to the metal core in the Galerkin case, probably accounting for its
lower attenuation as mentioned. A significant difference arises in the H_z component,
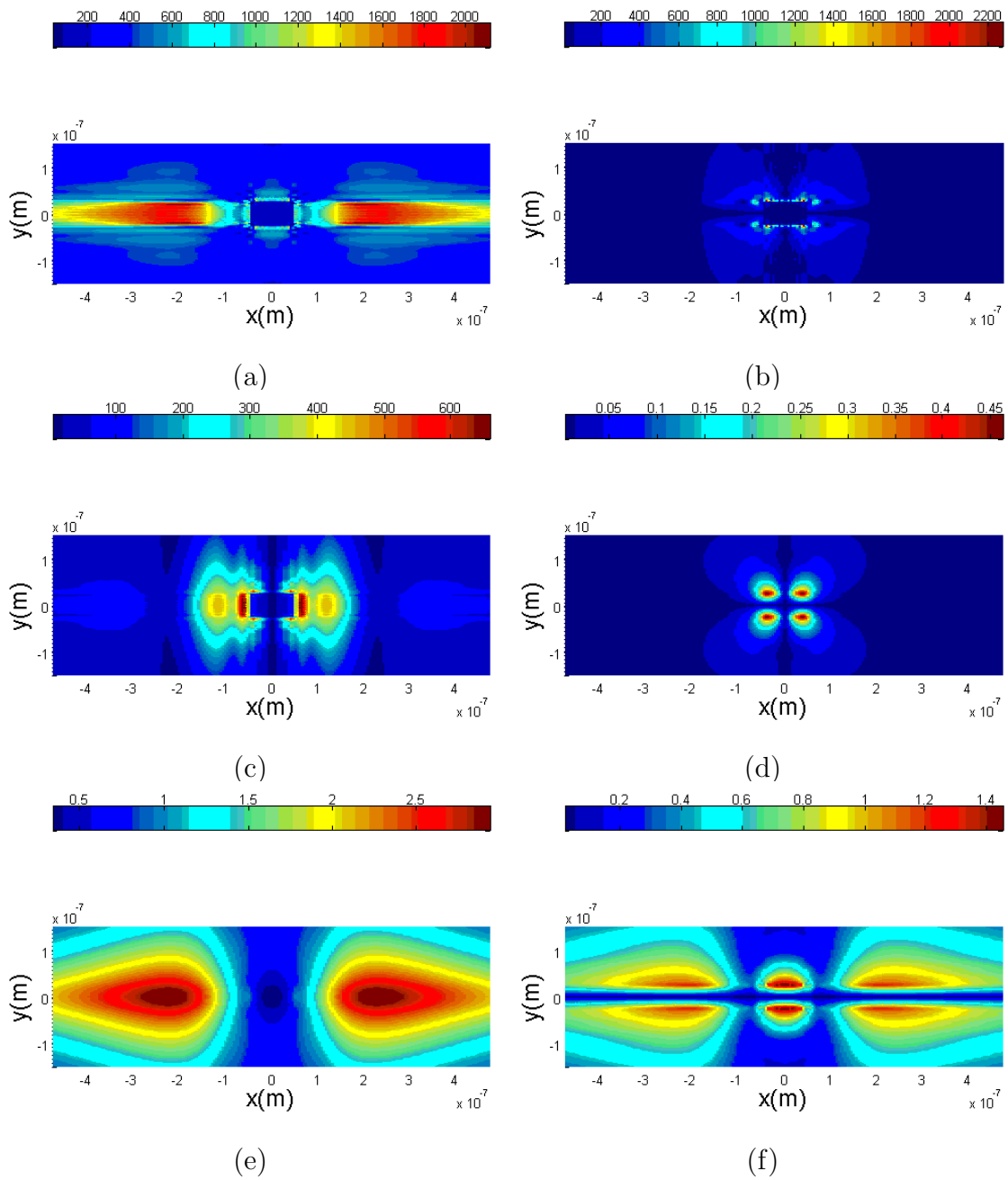where additional sidebands are present away from the core.

Fig. 27. The SP waveguide mode computed by the H-field mapped Galerkin method: (a) $|E_x|$; (b)$|E_y|$; (c) $|E_z|$; (d)$|H_x|$; (e) $|H_y|$; (f)$|H_z|$.
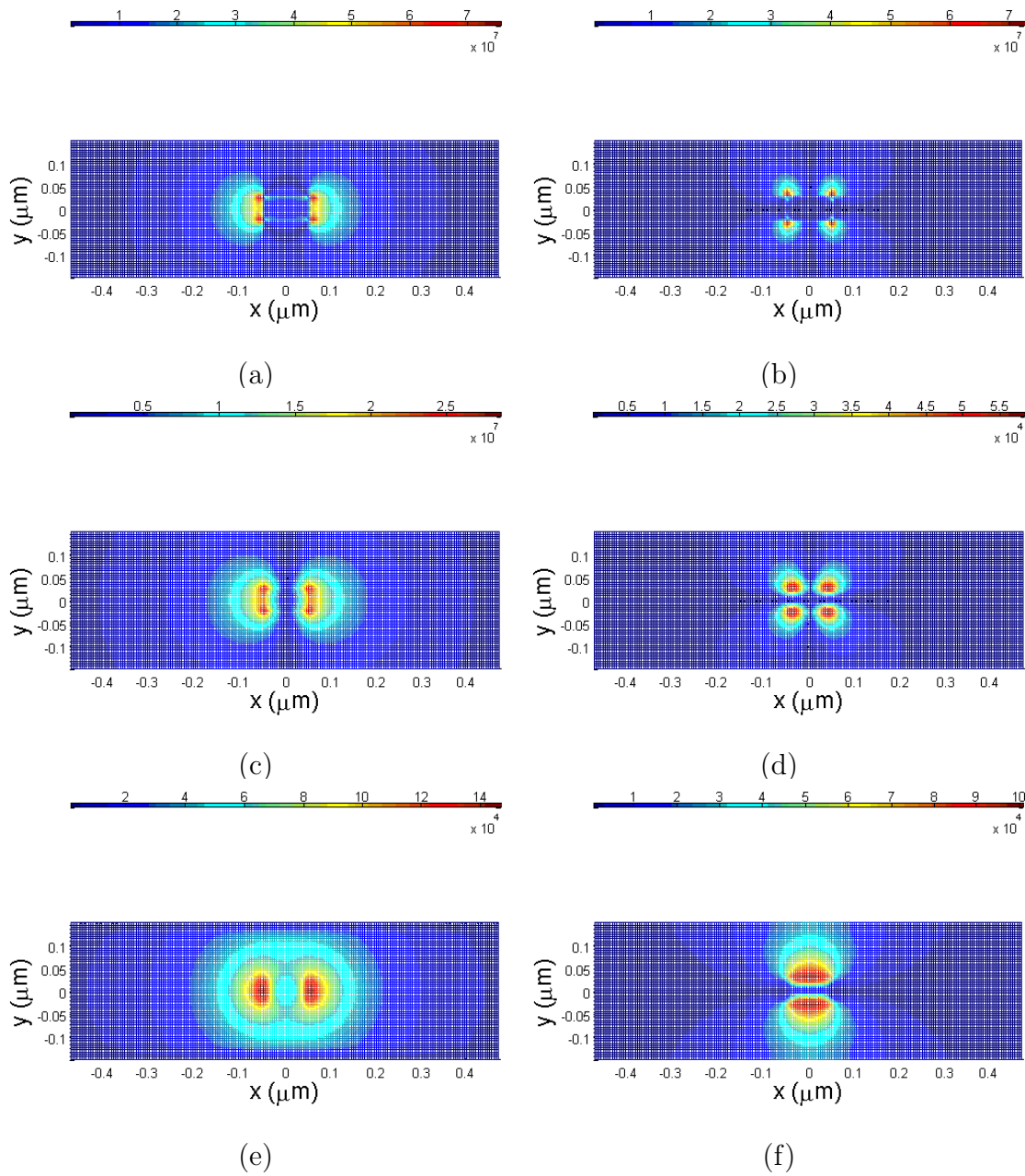
Fig. 28. The SP waveguide mode computed by the FDTD package: (a) $|E_x|$; (b)$|E_y|$; (c) $|E_z|$; (d)$|H_x|$; (e) $|H_y|$; (f)$|H_z|$.
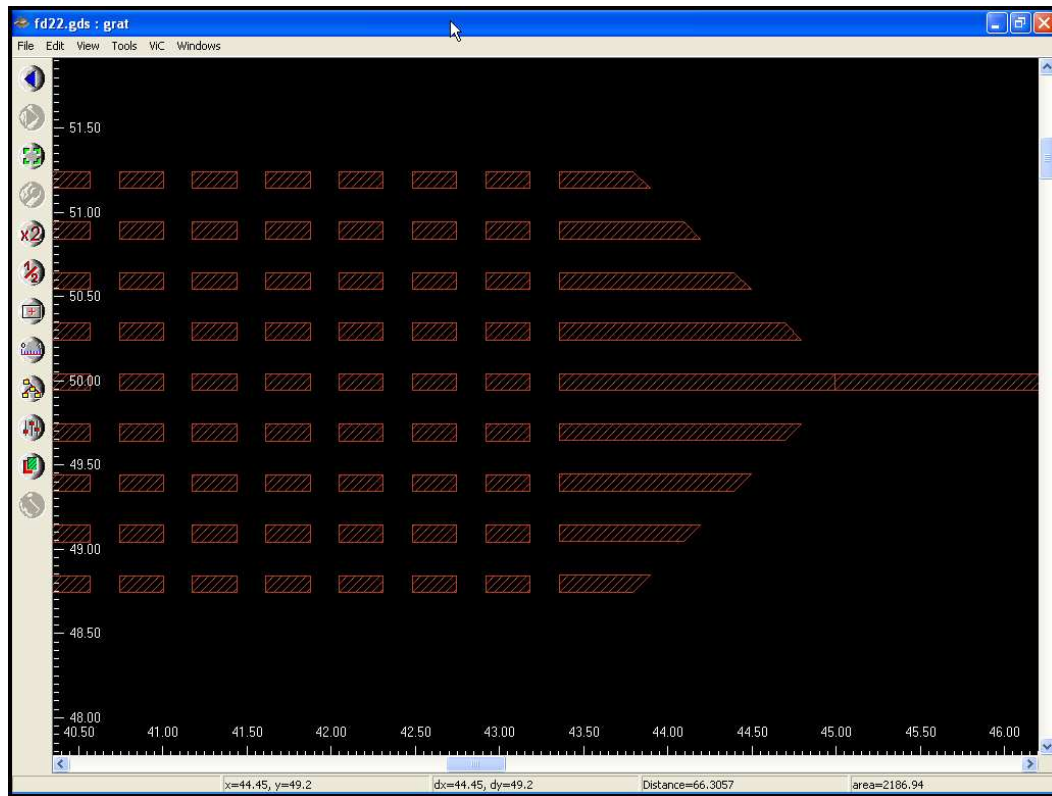
Fig. 29. Devices laid out in GDSII format.

C.   Fabrication and testing of the surface plasmon devices

The designed devices were laid out in GDSII format, as shown in Fig. 29. Electron beam lithography was used to fabricate them. It is a well developed plannar fabrication technique for making features below 100 nm [84]. Structures as small as 20 nm are possible with state-of-the-art EBL facilities [46]. Fig. 30 illustrates the EBL process. The MgF_2 substrate (MTI Corp.) was spin-coated with a thin ($\sim$150 nm) layer of EBL resist Polymethylmethacrylate (PMMA, 950K molecular weight, 3% in monochlorobenzene from Brewer Science). Patterns were written on the PMMA in the scanning electron microscope (SEM) with a 30 keV focused electron beam. They
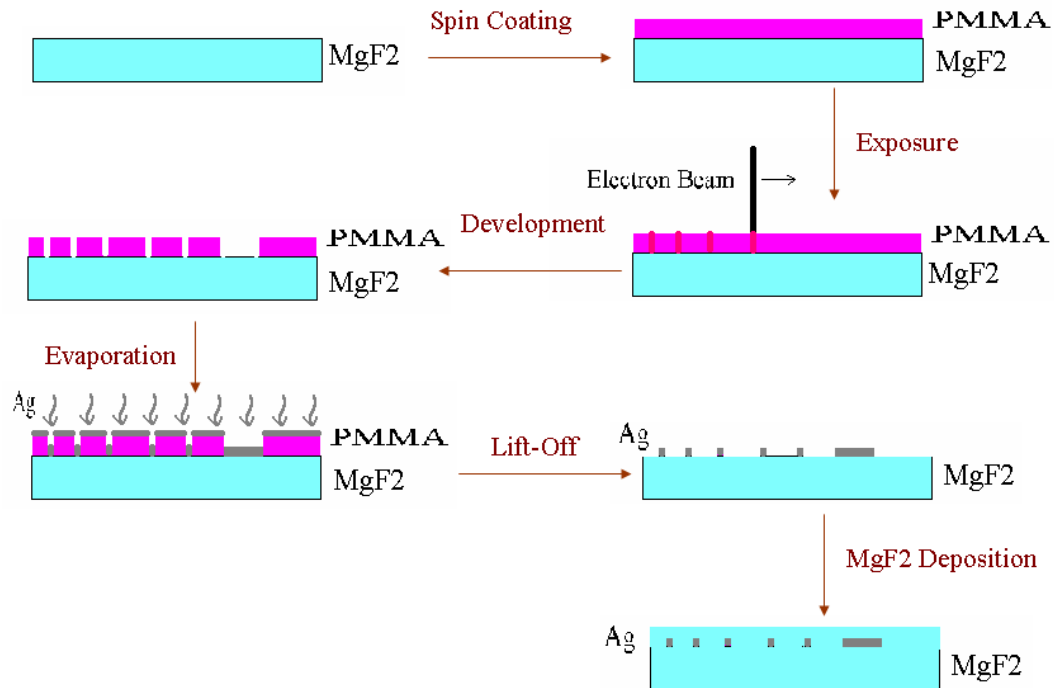
Fig. 30. Procedures of electron beam lithography.

were then developed in MIBK:IPA (methyl isobutyl ketone:isopropyl alcohol). Since the PMMA is a positive tone resist, the exposed regions were removed. Next, silver was deposited on the sample by thermal evaporation. To get rid of the undesired PMMA capped by Ag, a lift-off step was introduced by immersing the sample in organic solvents, which dissolved the PMMA and washed it away from the substrate. At last, 500 nm MgF_2 was overcoated by thermal evaporations.

Fig. 31 is a picture of a JSM-6460 SEM from JEOL (Japan Electron Optics Laboratory). It is the model at the Center for Nano Science and Technology (CNST), Texas A&M University. Fig. 32 shows the patterns developed by the EBL process on campus. Features as small as 50 nm were successfully made.

Fig. 33 shows the test setup to characterize the optical performance of the grating

Fig. 31. EBL facility at CNST, TAMU. It is a JSM-6460 SEM from JEOL, controlled by NPGS.

couplers and the waveguide. An s-polarized HeNe laser beam (632 nm) was passed through a quarter wave plate to become circularly polarized. A polarization cube was used to select the desired linear polarization. The beam was then collimated by a telescope. A lens of 75 cm focal length was used to focus the collimated beam at the back focal plane of an oil-immersion microscope objective (numerical aperture 1.4, Olympus), which finally delivered a tiny laser spot ($\sim 1.5\mu m$) on the sample surface. The small laser spot selectively excited one of the grating couplers connected to the waveguide. At this moment, illumination of the sample surface by the halogen light would help in aligning the laser spot with the target gratings. Once aligned, the halogen light was shut off. The incident laser was converted to SP wave by the
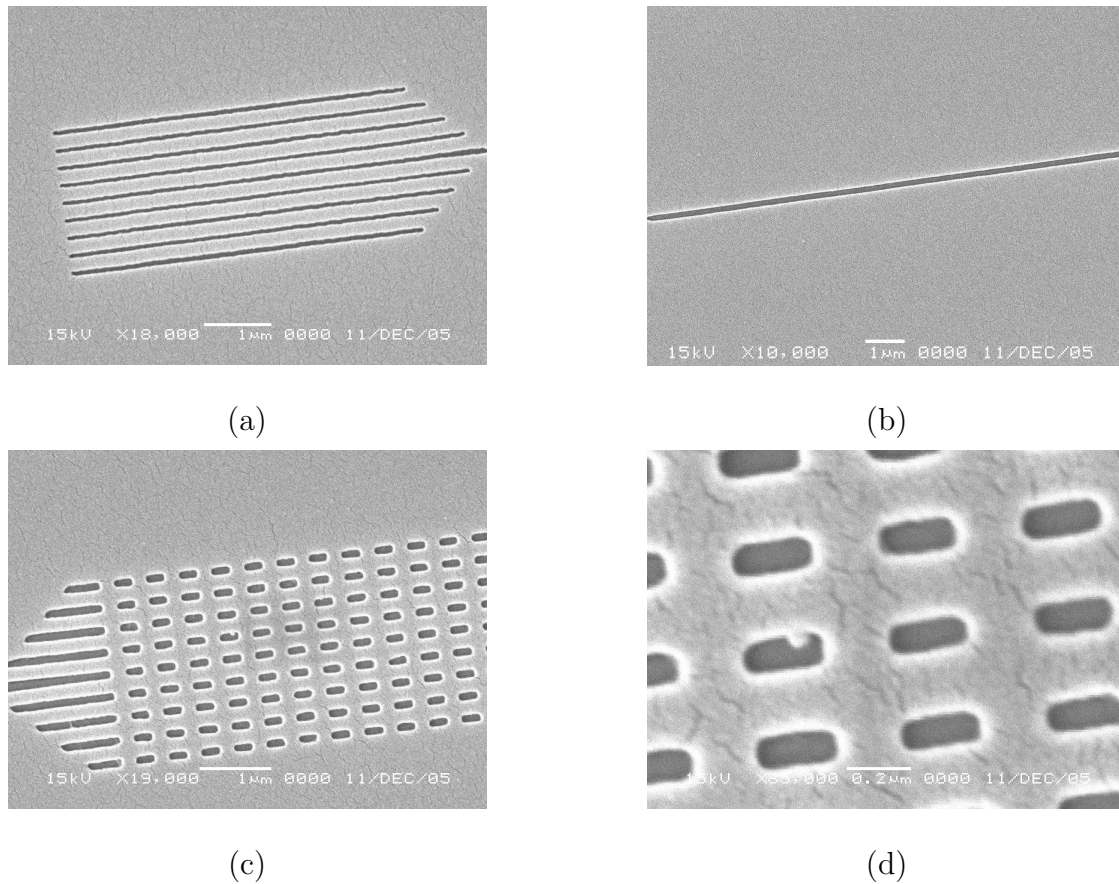
Fig. 32. SEM micrographs of developed patterns: (a) wire arrays of 50 nm linewidths; (b) a 100 nm wire running for more than $10\mu m$; (c) grating couplers with the smallest dimension of 100 nm; (d) zoom-in of (c).

gratings, which would be transmitted along the wire and re-emitted by the grating coupler on the other end. The same objective collected the re-emitted light to cast an image at the CCD camera (SXV-H9, Starlight Xpress Ltd.) through the beam splitter cube and the relay lens (f = 50.2 mm). Photos of the test apparatus are shown at the end of this chapter.

Fig. 34 shows the focus size of the exciting laser with a power around $245\mu W$. The spot size was about $1.5\mu m$, to have a good size match with the grating coupler.
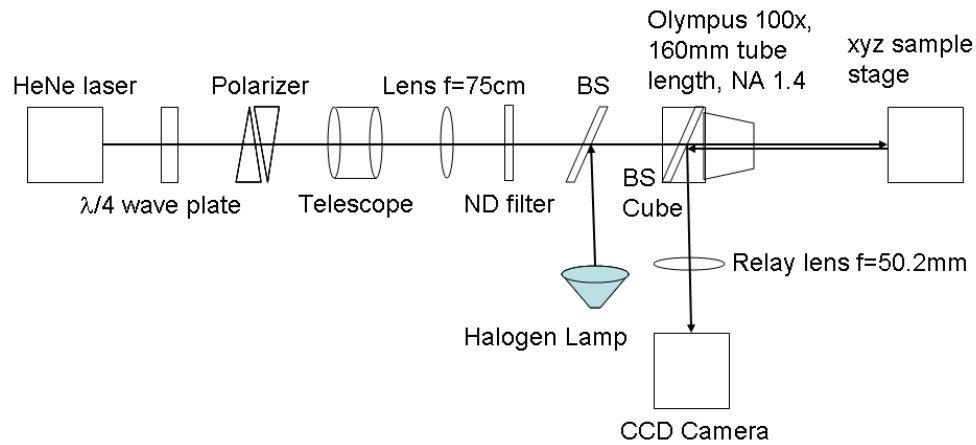
Fig. 33. Optical setup to characterize the fabricated surface plasmon devices. BS: beam splitter; ND: neutral density attenuator; NA: numerical aperture.

Fig. 34 also shows satisfactory device fabrications by the EBL and lift-off process. Fig. 35 illustrates the laser alignment process with the illumination from the halogen light. Fig. 35(b) shows a control device where the waveguide is missing. It was used to determine the scattered background light when calculating the waveguide loss, as will be discussed later.

Fig. 36(a) demonstrates the light transmission through the device. The white arrow in the figure designates the polarization of the laser incident on the left grating coupler. The incident laser light was converted into SP wave propagating along the wire. The output light became visible again when the SP wave hit the right grating. The dark region in between implies that light is tightly-bounded to the wire rather than scattered out of the plane. Scattered light was present in the wire region of the center image, which was caused by the waveguide roughness in that particular device. The waveguide lengths are $5\mu m$, $10\mu m$ and $15\mu m$ from top to bottom. The incident light might also be scattered by the left gratings directly onto the right gratings without going through the wire, accounting for an in-plane wireless transmission. It
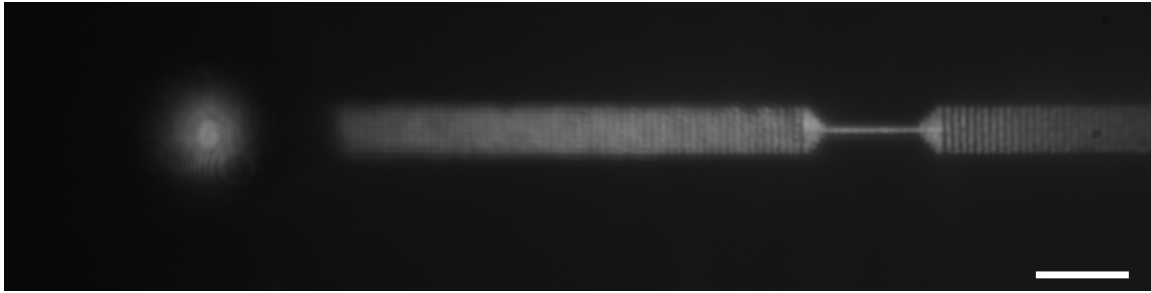
(a)

(b)

Fig. 34.   Laser alignment and inspection of the SP devices. (a) The fabricated devices were inspected with the halogen light illumination. The EBL and lift-off quality is satisfactory. The HeNe laser spot has a diameter about $1.5\mu m$. The camera integration time is 100 ms. (b) The laser focus with the halogen light off. Almost no stray light from the laser hit the neighboring device. The camera integration time is 300 ms. The scale bar is $5\mu m$.

would affect our characterization of the waveguide loss. To find out their contributions in the output, control patterns were used where the wire is missing. Fig 36(b) shows the "wireless transmission" effect. It is clear that it becomes negligible at the $15\mu m$ range. The output in the bottom image of Fig. 36(a) is mainly due to the "wired transmission". The images were captured with an integration time of 300 ms. The waveguide loss characterization is illustrated in Fig. 37. Light intensities coming out of the output grating region (indicated with the white box) were integrated in
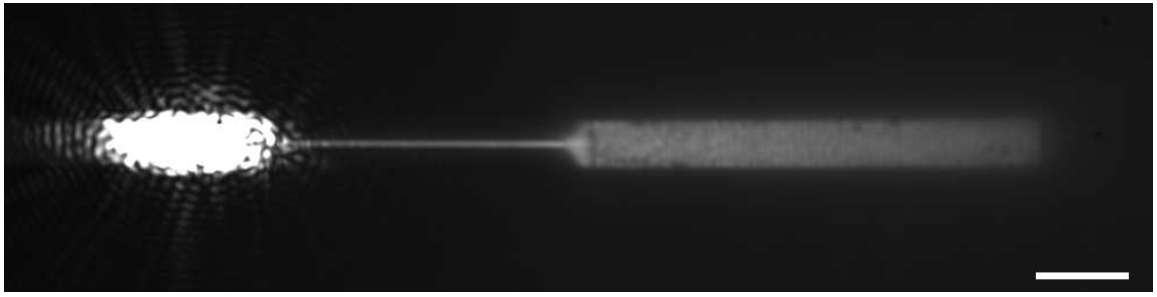
(a)
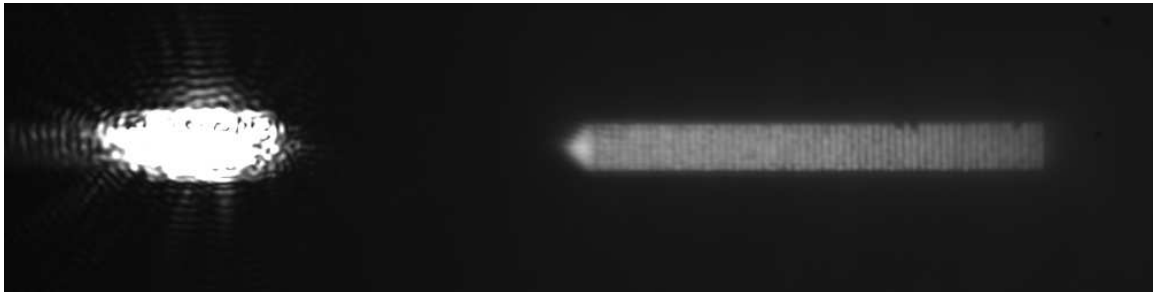
(b)

Fig. 35. Laser alignment for the devices under test: (a) the input and output grating couplers connected by the designed waveguide; (b) the control pattern with the waveguide missing. The pictures were taken under the illumination of the halogen light. It helps aligning the focuses of the exciting HeNe laser with the input grating couplers, which are on the left. The bright regions over the input couplers were due to the reflected laser light by the gratings. The wire is $15\mu m$ long. The scale bar is $5\mu m$.

both the wired and wireless cases. After subtracting the wireless contribution, the attenuations by waveguides of different lengths are shown by the "□" in Fig. 38. They are all normalized to the $5\mu m$ wire case. An attenuation curve was fitted out of these data points, as shown by the red trace in Fig. 38. The same procedures were repeated with the maximum light intensity in the output region. It shows similar trends, as described by the blue trace in Fig. 38. The waveguide loss is estimated to be $0.4 \sim 0.5 dB/\mu m$. It is close to the $0.46 dB/\mu m$ predicted by the Galerkin method,

(a)

(b)
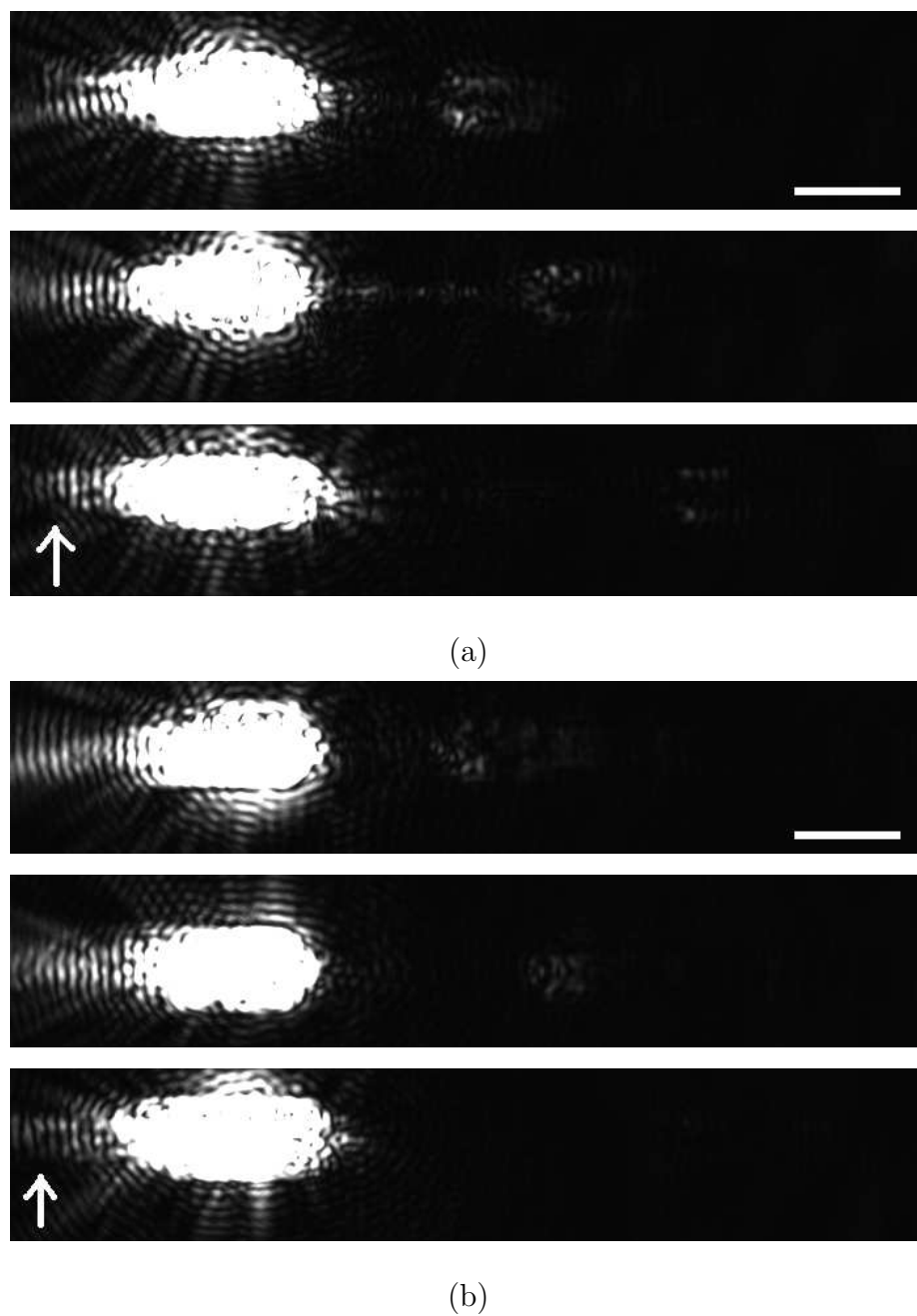
Fig. 36. Light transmission test of the SP waveguides: (a) complete devices; (b) control patterns without the wires. The exciting laser was focused on the left gratings. Light came out of the gratings on the right. The output light in (a) is stronger than the counterpart in (b), with the difference coming from the light transmitted by the wires. The scale bar is $5\mu m$.

(a)

(b)
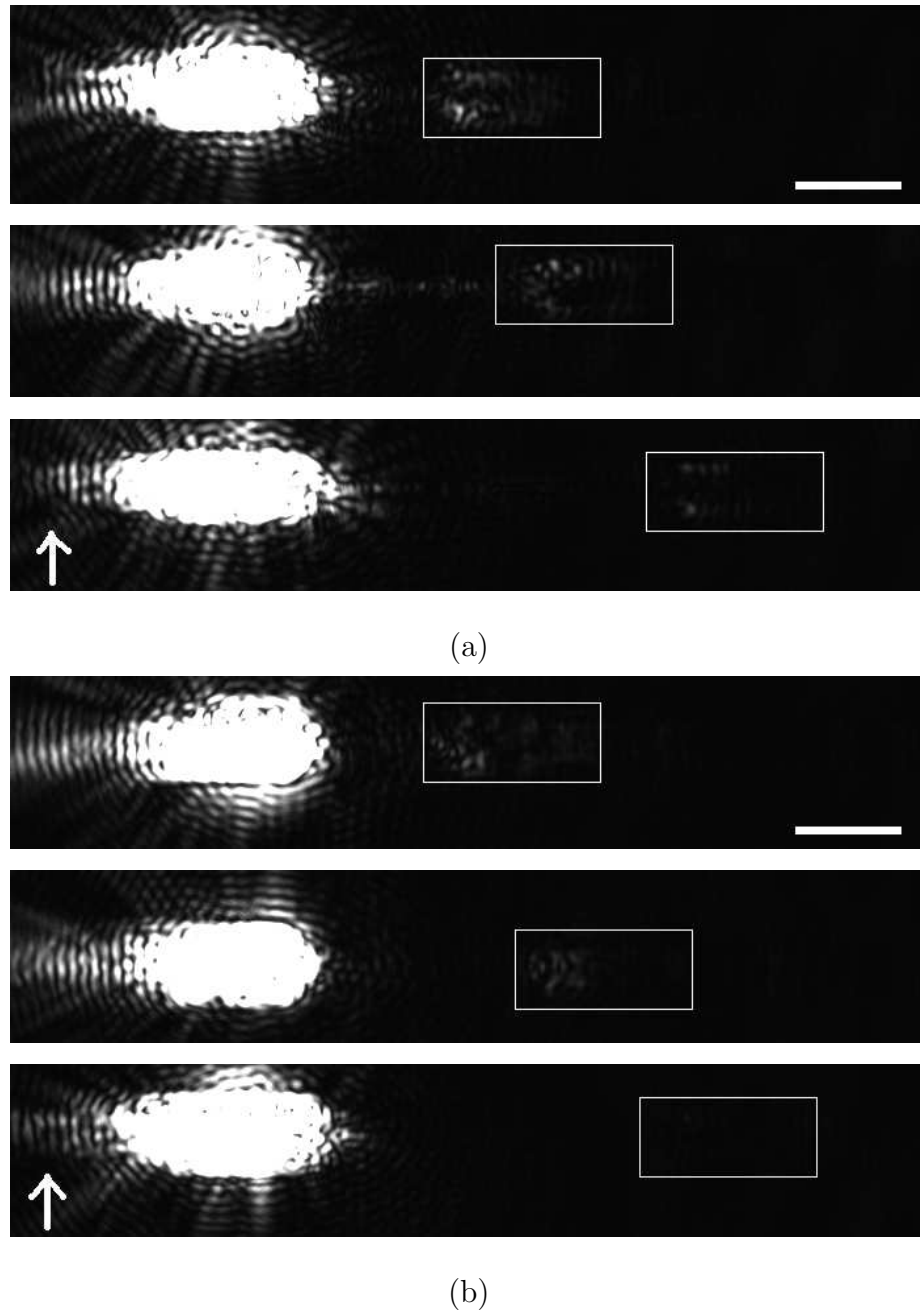
Fig. 37. Attenuation measurements of the SP waveguides. The white boxes in the pictures define the output region of interests, where the light intensities were processed for (a) complete devices, and (b) control pattterns without the wires. The processed data were used to calculate the propagation loss presented in Fig. 38. The scale bar is $5\mu m$.

but different from the $1.8dB/\mu m$ in the FDTD simulation. The propagation length L_p can be calculated from the attenuation $dB/\mu m$

$$L_p = \frac{10 \log e}{dB/\mu m} \ (\mu m) \ . \tag{4.51}$$

That means a measured attenuation of $0.5dB/\mu m$ corresponds to a traveling range of $8.7\mu m$ on our waveguide of $100nm \times 50nm$ cross-section. In terms of the wire size and traveling range, the closest result in the visible spectrum is $2\mu m$ traveling range with a $250nm \times 50nm$ waveguide ([72],see Tab. III also), which translates to a loss of $2.17dB/\mu m$.

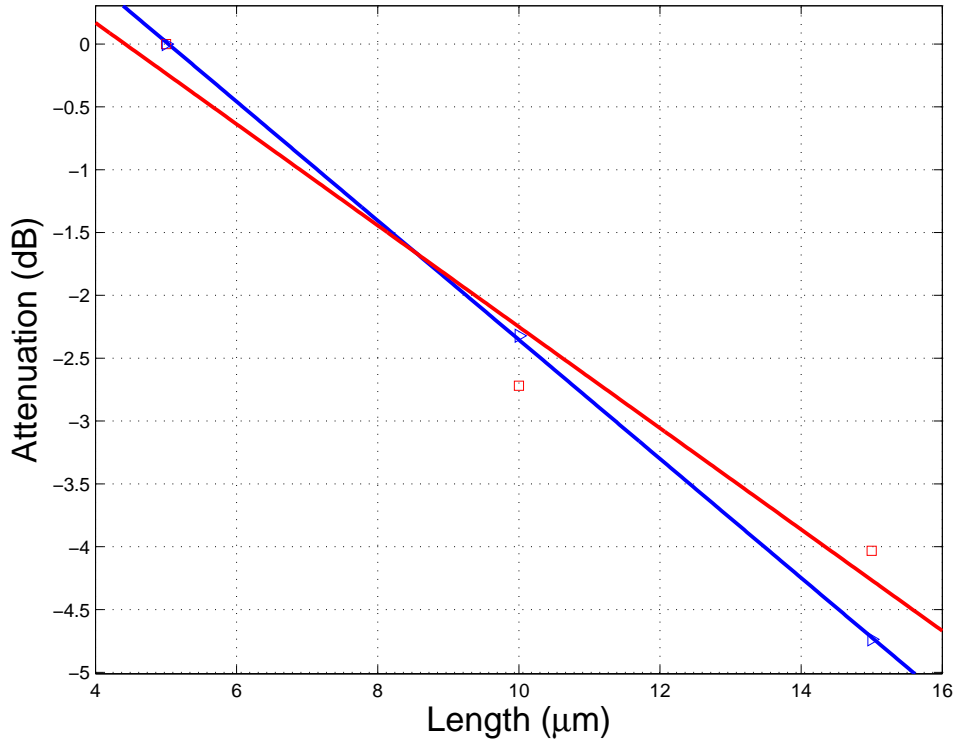Fig. 38.   Attenuations of the output signals by waveguides of different lengths. All data are normalized to values at $5\mu m$. □: total intensity of the output; ▷: maximum intensity of the output. The red trace is the curve fitting of the total intensity. The blue trace is the curve fitting of the maximum intensity.

Fig. 22(c)(d) implies the polarization sensitive behaviors of the grating couplers.

This is confirmed by an experiment as presented in Fig. 39, where the white arrows designate the two perpendicular polarizations in respective cases. Light was seen coming out in Fig. 39(a) with the s-polarization, while the p-polarization failed to produce the same output in Fig. 39(b). This supports the simulations by FDTD. The images were captured with an integration time of 300 ms.

(a)

(b)

Fig. 39.   The polarization property of the SP devices. (a) S-polarized light at 632 nm was coupled to the gratings and transmitted by a $15\mu m$ wire. (b) Almost no light comes out when the polarization is rotated by 90°. The camera integration time is 300 ms in both cases. The polarizing property match simulation predictions. The scale bar is $5\mu m$.

The grating couplers were designed to work with normal incident/outgoing waves, as shown in Fig. 22(c)(d). In reality, it was found that they work best with incident wave eight degrees from the surface normal direction. To further understand functions of the grating couplers, it is also desirable to estimate its coupling efficiency. It can be done with Fig. 34(a) and Fig. 37. The integrated light intensity in the region ($6.8\mu m \times 6.6\mu m$) around the focus spot is 321701 arbitrary units (AUs) in

Fig. 34(a) after subtracting the background. Since it was caused by the reflection at the MgF_2/oil interface, the total incident intensity is inferred to be 1.3804×10^8 AUs after compensating for the reflection coefficient [74]

$$R = \left(\frac{n_{MgF_2} - n_{oil}}{n_{MgF_2} - n_{oil}} \right)^2 , \tag{4.52}$$

where $n_{MgF_2} = 1.38$ and $n_{oil} = 1.52$ are the refractive indexes for the MgF_2 and the immersion oil, respectively. After subtracting the background, the integrated output intensity is 11385 within the white box in Fig. 37 for the $15\mu m$ long device. By taking into account of the three times difference in integration times and the $0.42dB/\mu m \times 15\mu m$ attenuation by the waveguide, the combined transmission is 1.214×10^{-4} through the input and output grating couplers. The coupling efficiency of a single grating is approximated as $\sqrt{1.214 \times 10^{-4}} \approx 1.1\%$. The linear compensations in the above estimations are valid because no saturations occur in the integration regions in either image.

The 1.1% coupling efficiency is comparable to the best efficiency (1.6%) reported so far for near-field scanning optical microscope (NSOM) [85]. It is achieved in our case with an objective in the far field ($> 100\mu m$ from the device surface), whereas the NSOM delivers similar efficiencies with a taper fiber in the near-field ($< 100nm$ to the device surface). The enhanced coupling efficiency by the grating is also demonstrated with Fig. 40, where image (a) shows the alignment of the laser spot with a bare waveguide of $15\mu m$ long under the halogen illumination. Image (b) shows little light comes out of the other end of the wire when only the laser was present. Image (c), which is the same picture in Fig. 36(a), shows more light was emitted from the output grating compared to the bare wire case.
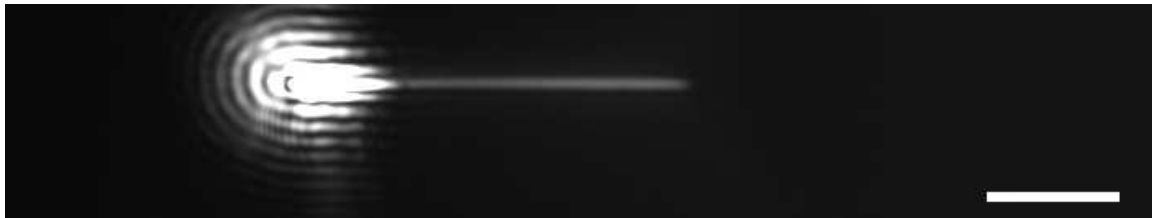
(a)

(b)

(c)

Fig. 40. Coupling efficiency enhancement by the gratings. (a) The HeNe laser spot is aligned with one end of a $15\mu m$ wire. The camera integration time is 100 ms. (b) Almost no light comes out of the other end of the wire. The camera integration time is 300 ms. (c) Stronger light was recorded when the wire is connected to an input and an output gratings at both ends. The coupling efficiency is enhanced with the grating couplers. The scale bar is $5\mu m$.

D.   Enhanced atomic decay rates by a surface plasmon waveguide

The FDTD and mapped Galerkin solutions both confirm strong field focusing effects by the described SP waveguide. Theoretical research also suggest strong dispersiveness are possible in SP waveguides, which stands for high densities of guiding modes. According to Fermic's Golden Rule, these two factors will both enhance the emission rate of an atom nearby. It may found important applications in LEDs, OLEDs, quantum information systems [57], and micro-lasers [86]. A plasmonic-waveguide-coupled-atom system will be analyzed next. Enhanced emission rate will be compared with the spontaneous emission rate in vacuum. Insights may be gained by examining the special case of an atom coupled to the rectangular SP waveguide mode discussed previously.

Suppose an atom is in the neighborhood of a single mode waveguide oriented along the z-axis. The atom in the excited state $|a\rangle$ decays to the ground state $|b\rangle$ by emitting a surface plasmon of energy $\hbar\omega_{ab}$ into the waveguide mode. The photon emission rate can be estimated with Fermi's Golden Rule [33]

$$\Gamma_{wgd} = 2\pi \frac{|\mathcal{D}_{ab} \cdot \widetilde{\mathbf{E}}(\mathbf{r}_{atm})|^2}{\hbar^2} \rho(\omega_{ab}) \ . \tag{4.53}$$

Here, $\mathcal{D}_{ab} = e\langle a|\mathbf{r}'|b\rangle$ is the vectorial electric dipole moment of the transition and is calculated under the nucleus-centered coordinates $\mathbf{r}'$, $\rho(\omega_{ab})$ is the density of SP waveguide modes at the frequency of interests, and $\widetilde{\mathbf{E}}(\mathbf{r}_{atm})$ is the normalized electric field of the waveguide mode at the atom.

We will examine the waveguide mode density first. For the time being, the waveguide mode is approximated as lossless. The loss free assumption allows us to follow the similar procedures in deriving the vacuum mode density [87], except that a simpler case of one dimensional space is considered instead of the three dimensional

one. For the SP wave traveling along a wire of a virtual length L, the propagation vector γ

$$\gamma = \frac{\pi n_z}{\mathsf{L}} \ , \tag{4.54}$$

where n_z is an integer $(0, \pm 1, \pm 2, \dots)$ identifying the mode. The number of modes in the γ space is given by

$$N = \frac{\mathsf{L}}{\pi} \int \mathrm{d}\gamma \ . \tag{4.55}$$

Note that $\mathrm{d}\gamma$ is related to the group velocity by

$$\mathrm{d}\gamma = \frac{1}{v_g(\omega)} \mathrm{d}\omega \tag{4.56}$$

where $v_g(\omega)$ is derived from the effective refractive index $n_{eff}(\omega) = c\gamma/\omega$ for the guided wave (c is the light speed in vacuum)

$$v_g(\omega) = \frac{c}{[n_{eff}(\omega) + \omega n'_{eff}(\omega)]} \ . \tag{4.57}$$

Therefore, (4.55) is transformed into

$$N = \frac{\mathsf{L}}{\pi} \int \frac{1}{v_g(\omega)} \mathrm{d}\omega \ , \tag{4.58}$$

which gives out the mode density by

$$\rho(\omega) = \frac{\mathrm{d}N}{\mathrm{d}\omega} = \frac{\mathsf{L}}{\pi v_g(\omega)} \ . \tag{4.59}$$

Next, the field normalization will be undertaken. Suppose the field distribution $\{\mathbf{E}(x,y), \mathbf{H}(x,y)\}$ is available by solving Maxwell equations in the SP waveguide problem, the normalization factor χ satisfies

$$\begin{aligned}
\hbar\omega_{ab} &= \frac{1}{2} \int \epsilon(\mathbf{r}) |\chi \mathbf{E}(\mathbf{r})|^2 + \mu(\mathbf{r}) |\chi \mathbf{H}(\mathbf{r})|^2 \mathrm{d}^3 r \\
&= \frac{\chi^2 \mathsf{L}}{2} \int \int \epsilon(x,y) |\mathbf{E}(\mathbf{x},\mathbf{y})|^2 + \mu |\mathbf{H}(x,y)|^2 \mathrm{d}x \mathrm{d}y \ ,
\end{aligned} \tag{4.60}$$

where the axial independence of the waveguide has been utilized. Thus the normalization factor χ is

$$\chi = \sqrt{\frac{\hbar\omega_{ab}}{\mathsf{L}I}} \ , \tag{4.61}$$

where

$$I = \frac{1}{2} \int \int \epsilon(x,y)|\mathbf{E}(x,y)|^2 + \mu(x,y)|\mathbf{H}(x,y)|^2 \mathrm{d}x\mathrm{d}y \ . \tag{4.62}$$

Without loss of generality, the electric dipole moment is assumed to be aligned with the local electric field at the atom. By applying the SP mode density (4.59) and the normalizing factor (4.61), the emission rate (4.53) becomes

$$\Gamma_{wgd} = \frac{2\omega_{ab}|\mathcal{D}_{ab}|^2}{\hbar v_g} \frac{|\mathbf{E}(\mathbf{r}_{atm})|^2}{I} \ . \tag{4.63}$$

Recall that the emission rate to vacuum modes is [87]

$$\Gamma_{vac} = \frac{\omega_{ab}^3|\mathcal{D}_{ab}|^2}{3\pi\hbar\epsilon_0 c^3} \ , \tag{4.64}$$

the emission enhancement factor is

$$\frac{\Gamma_{wgd}}{\Gamma_{vac}} = \frac{3}{2\pi} \frac{c}{v_g} \frac{\epsilon_0|\mathbf{E}(\mathbf{r}_{atm})|^2}{I/\lambda_0^2} \ , \tag{4.65}$$

where $\lambda_0 = 2\pi c/\omega_{ab}$ is the corresponding wavelength in vacuum. By applying solutions from the FDTD method, the emission enhancement is plotted in Fig. 41 for locations across the waveguide cross-section. The maximum enhancement factor is 59 around the four corners. The figure also tells how well confined the waveguide mode is. It suggests the propagating power is mainly confined in two regions of $100nm \times 100nm$ on both sides of the waveguide.
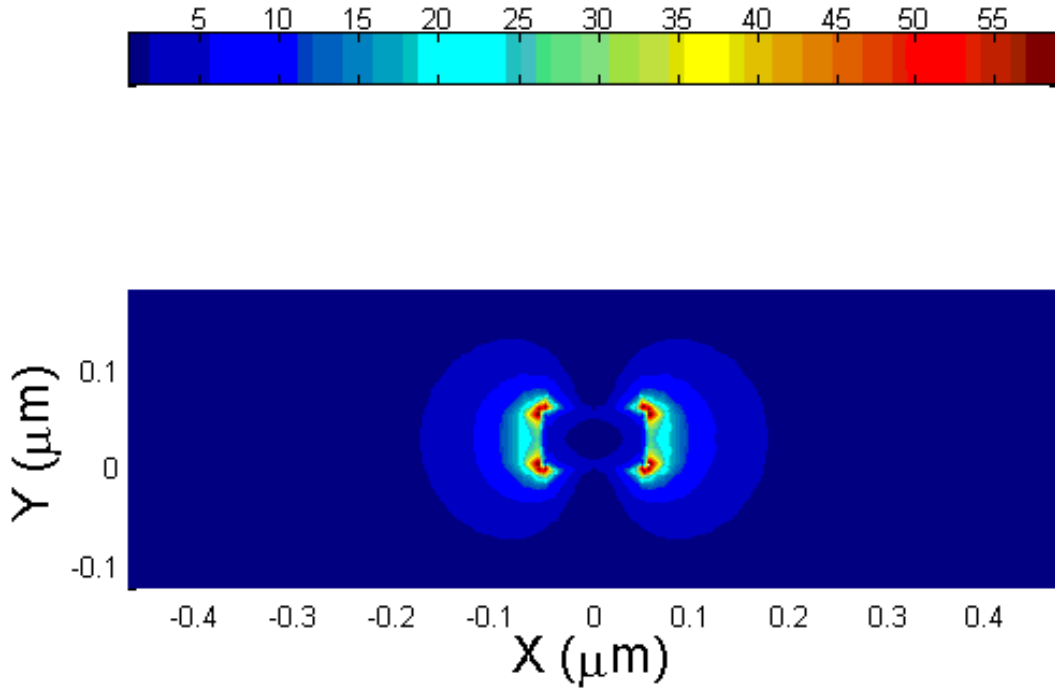
Fig. 41. Map of emission enhancement sites over the waveguide cross-section. The pixel color indicates the emission enhancement factor by the SP waveguide for an atom at the site.

E.   Summary of the SP waveguide and grating couplers

In summary, a surface plasmon waveguide was simulated with FDTD. A grating coupler was also designed with the same software package to couple the nano wire to the diffraction-limited optics,without which the mode size mismatch would otherwise forbid efficient excitation and collections. A vectorial H-field formulation was derived for the mapped Galerkin method for the first time. It was applied to analyze surface plasmon waveguides. The mapping constants were found for the special case of SP rectangular waveguides. Due to the diagonal-dominant nature of the resulting matrix, the subsequent eigenvalue problem can be solved with a fast iterative method. Speedup by a factor of ten was observed for the waveguide analysis by using the H-

field mapped Galerkin method instead of FDTD. Discrepancies were present in the simulation results obtained by the two methods. The designed devices were laid out in GDSII format, and fabricated with electron beam lithography. Feature sizes as small as 100 nm or below were successfully fabricated. An oil-immersion microscope (refer to Fig. 42 and Fig. 43 for the photos) was built to characterize the optical performance of the device. With this microscope, we were able to observe the conversion of optical wave into SP wave by the grating antennas, or vice verse. Also observed was the SP wave transmission by the nano wire with a cross-section of $100nm \times 50nm$. The SP waveguide loss was measured to be $0.4 \sim 0.5dB/\mu m$, as compared to the predicted values of $1.8dB/\mu m$ by the FDTD method and $0.46dB/\mu m$ by the Galerkin method. That corresponds to a traveling range greater than $8.7\mu m$. It is four times longer, and on a wire a half narrower, than the best result reported so far in the visible spectrum. The polarization selectiveness of the coupler was both confirmed in simulation and experiment. The grating coupler achieved an estimated efficiency of 1%, comparable to the best efficiency 1.6% by NSOM, but with a much longer distance between the source and the device. The waveguide can potentially enhance the emission rate 59 folds for an atom.
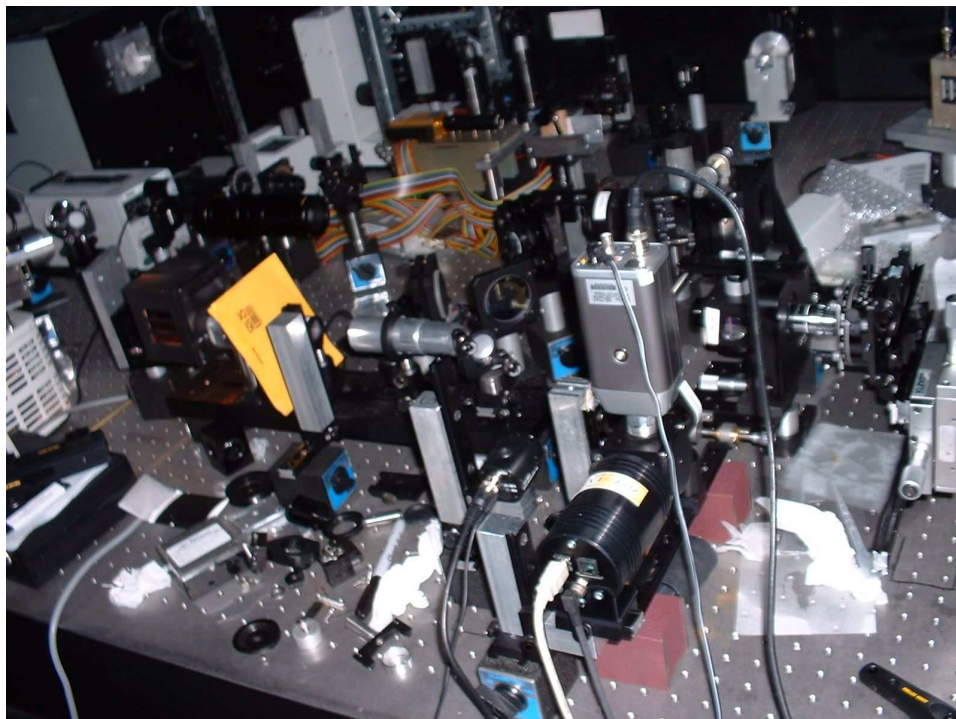
(a)
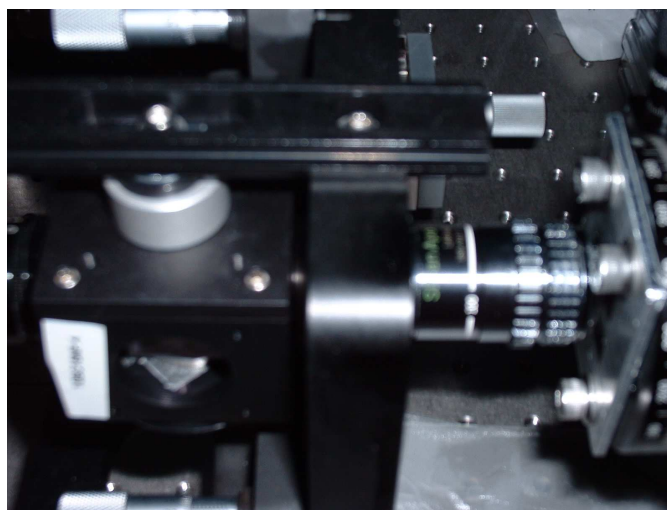
(b)

Fig. 42. (a) Optics setup for testing the SP devices; (b) Close-up of the oil-immersion
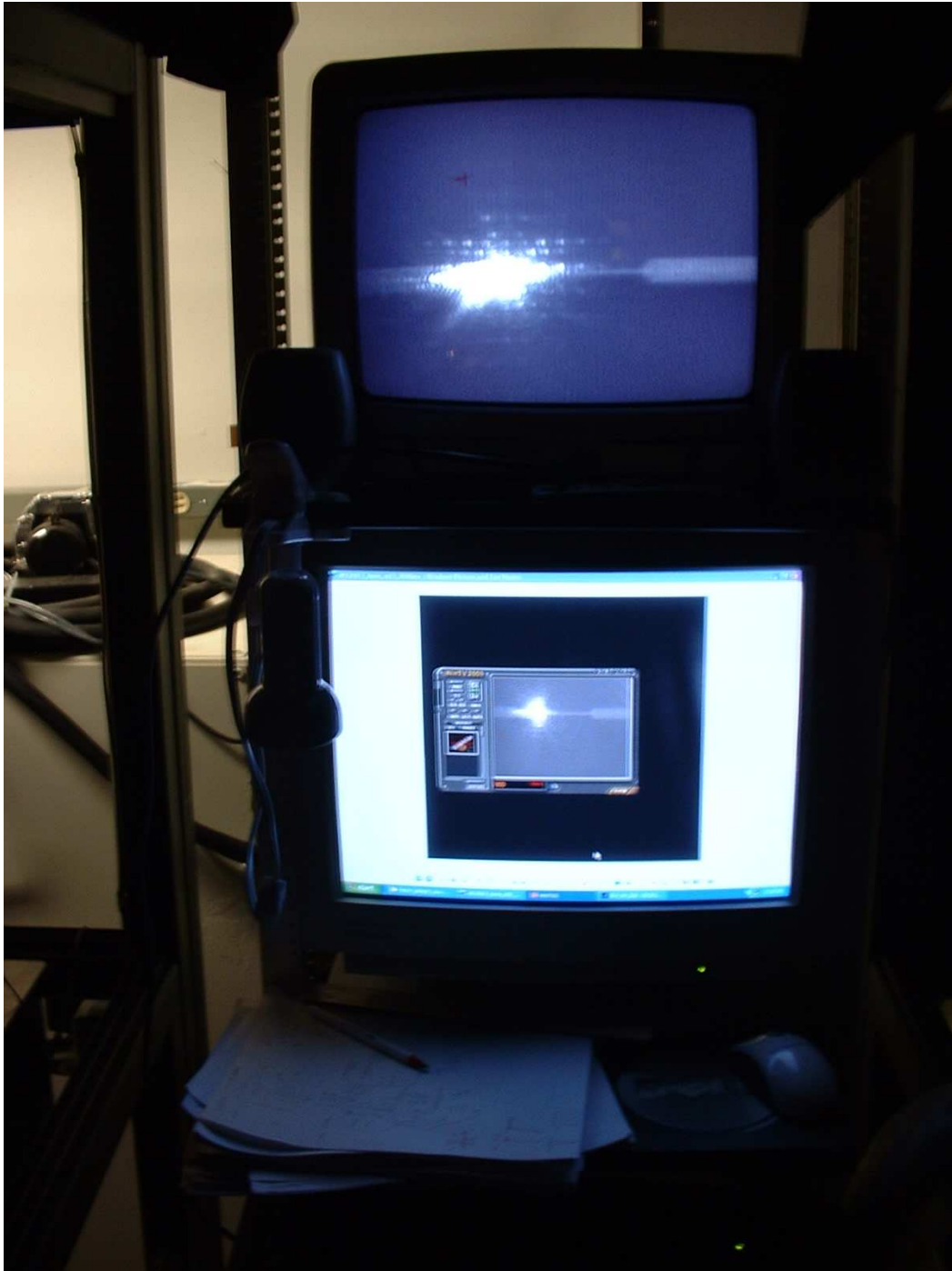objective and the sample mounting stage in the setup.

Fig. 43. Electronics for testing the SP devices. The device under test is displayed on both the computer and TV monitors.

CHAPTER V

CONCLUSIONS

In conclusion, novel optical devices for information processing have been studied in this dissertation. First, computer-generated-holograms were proven for parallelism in making multiple free space interconnects, which enabled the optical implementation of the Rizvi-Zubairy associative memory model. Recalls of complete images were demonstrated as well as recoveries from the incomplete inputs. Second, the delay-time-bandwidth product was analyzed in room temperature slow light, which must be solved to make all-optical delay lines and buffers applicable in reality. Optical pulses were slowed down to 1 mm/s inside a $Ce:BaTiO_3$ crystal using photorefractive two-wave mixing. The pulse distortion was improved by the proposed architecture of artificial inhomogeneous broadening. Finally, a surface plasmon waveguide was realized with a nano silver wire working in the visible spectrum. The wire had a cross section comparable to the smallest reported. Its guiding loss was measured to be the lowest, only one fourth of the best reported so far. The designed antennas achieved an efficiency comparable to the state of the art, with a distinct advantage of much greater working distance.

REFERENCES

[1] A. F. Benner, M. Ignatowski, J. A. Kash, D. M. Kuchta, and M. B. Ritter, "Exploitation of optical interconnects in future server architectures," *IBM Journal of Research and Development*, vol. 49, pp. 755–775, 2005.

[2] B. E. Lemoff, M. E. Ali, G. Panotopoulos, G. M. Flower, B. Madhavan, A. F. J. Levi, and D. W. Dolfi, "MAUI: Enabling fiber-to-the-processor with parallel multiwavelength optical interconnects," *IEEE Journal of Lightwave Technology*, vol. 22, pp. 2043–2054, 2004.

[3] A. Liu, R. Jones, L. Liao, D. Samara-Rubio, D. Rubin, O. Cohen, R. Nicolaescu, and M. Paniccia, "A high-speed silicon optical modulator based on a metal-oxide-semiconductor capacitor," *Nature*, vol. 427, pp. 615–618, 2004.

[4] H. Rong, A. Liu, R. Jones, O. Cohen, D. Hak, R. Nicolaescu, A. Fang, and M. Paniccia, "An all-silicon Raman laser," *Nature*, vol. 433, pp. 292–294, 2005.

[5] H. Rong, R. Jones, A. Liu, O. Cohen, D. Hak, A. Fang, and M. Paniccia, "A continuous-wave Raman silicon laser," *Nature*, vol. 433, pp. 725–728, 2005.

[6] V. Jungnickel, T. Hauslein, A. Forck, and C. von Helmolt, "155 Mbit/s wireless transmission with imaging infrared receiver," *Electronics Letters*, vol. 37, pp. 314–315, 2001.

[7] L. V. Hau, S. E. Harris, Z. Dutton, and C. H. Behroozi, "Light speed reduction to 17 metres per second in an ultracold atomic gas," *Nature*, vol. 397, pp. 594–598, 1999.

[8] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences*, vol. 79, pp. 2554–2558, 1982.

[9] D. Psaltis and N. Farhat, "Optical information processing based on an associative-memory model of neural nets with thresholding and feedback," *Optics Letters*, vol. 10, pp. 98–100, 1985.

[10] X. Wang, G. Mu, and Y. Zhang, "Optical associative memory using an orthogonalized hologram," *Optics Letters*, vol. 16, pp. 100–102, 1991.

[11] F. T. S. Yu, X. Yang, and T. Lu, "Space-time-sharing optical neural network," *Optics Letters*, vol. 16, pp. 247–249, 1991.

[12] S. H. Song, S.-C. Park, and S. S. Lee, "Optical implementation of a quadratic associative memory by using the polarization-encoding process," *Optics Letters*, vol. 15, pp. 1389–1391, 1991.

[13] S. Gao, J. Yang, Z. Feng, and Y. Zhang, "Implementation of a large-scale optical neural network by use of a coaxial lenslet array for interconnection," *Applied Optics*, vol. 36, pp. 4779–4781, 1997.

[14] D. Fey, W. Erhard, M. Gruber, J. Jahns, H. Bartelt, G. Grimm, L. Hoppe, and S. Sinzinger, "Optical interconnects for neural and reconfigurable VLSI architectures," *Proceedings of the IEEE*, vol. 88, pp. 838–848, 2000.

[15] B. R. Brown and A. W. Lohmann, "Complex spatial filtering with binary masks," *Applied Optics*, vol. 5, pp. 967–969, 1966.

[16] A. W. Lohmann and D. P. Paris, "Binary fraunhofer holograms, generated by computer," *Applied Optics*, vol. 6, pp. 1739–1748, 1967.

[17] A. A. Rizvi and M. S. Zubairy, "Implementation of associative memory using grating structures," *Applied Optics*, vol. 33, pp. 3642–3646, 1994.

[18] G. Lenz, B. J. Eggleton, C. K. Madsen, and R. E. Slusher, "Optical delay lines based on optical filters," *IEEE Journal of Quantum Electronics*, vol. 37, pp. 525–532, 2001.

[19] E. Podivilov, B. Sturman, A. Shumelyuk, and S. Odoulov, "Light pulse slowing down up to 0.025 cm/s by photorefactive two-wave coupling," *Physical Review Letters*, vol. 91, pp. 083902, 2003.

[20] A. Kasapi, M. Jain, G. Y. Yin, and S. E. Harris, "Electromagnetically induced transparency: propagation dynamics," *Physical Review Letters*, vol. 74, pp. 2447–2450, 1995.

[21] A. V. Turukhin, V. S. Sudarshanam, M. S. Sharhriar, J. A. Musser, B. S. Ham, and P. R. Hemmer, "Observation of ultraslow and stored light pulses in a solid," *Physical Review Letters*, vol. 88, pp. 023602, 2002.

[22] Q. Sun, Y. V. Rostovtsev, J. P. Dowling, M. O. Scully, and M. S. Zubairy, "Optically controlled delays for broadband pulses," *Physical Review A*, vol. 72, pp. 031802, 2005.

[23] J. E. Heebner, R. W. Boyd, and Q. H. Park, "Slow light, induced dispersion, enhanced nonlinearity, and optical solitons in a resonator-array waveguide," *Physical Review E*, vol. 65, pp. 036619, 2002.

[24] M. S. Bigelow, N. N Lepeshkin, and R. W. Boyd, "Observation of ultraslow light propagation in a ruby crystal at room temperature," *Physical Review Letters*, vol. 90, pp. 113903, 2003.

[25] M. S. Bigelow, N. N. Lepeshkin, and R. W. Boyd, "Superluminal and slow light propagation in a room-temperature solid," *Science*, vol. 301, pp. 200–202, 2003.

[26] G. Zhang, R. Dong, F. Bo, and J. Xu, "Slowdown of group velocity of light by means of phase coupling in photorefractive two-wave mixing," *Applied Optics*, vol. 43, pp. 1167–1173, 2004.

[27] G. Zhang, F. Bo, R. Dong, and J. Xu, "Phase-coupling-induced ultraslow light propagation in solids at room temperature," *Physical Review Letters*, vol. 93, pp. 133903, 2004.

[28] X. Zhao, P. Palinginis, B. Pesala, C. J. Chang-Hasnain, and P. R. Hemmer, "Tunable ultraslow light in vertical-cavity surface-emitting laser amplifier," *Optics Express*, vol. 13, pp. 7899, 2005.

[29] C. Yang, Y. Zhang, P. Yeh, Y. Zhu, and X. Wu, "Photorefractive properties of Ce:BaTiO$_3$ crystals," *Optics Communications*, vol. 113, pp. 416–420, 1995.

[30] H.J. Coufal, D. Psaltis, and G. Sincerbox, *Holographic Data Storage*, Springer, Berlin, 2000.

[31] H. Raether, *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*, Springer, Berlin, 1988.

[32] C. F. Bohren and D. R. Huffman, *Absorption and Scattering of Light by Small Particles*, Wiley, New York, 1983.

[33] A. Yariv, *Quantum Electronics*, Wiley, New York, 1989.

[34] G. Ritchie and E. Burstein, "Luminescence of dye molecules absorbed at a Ag surface," *Physical Review B*, vol. 24, pp. 4843–4846, 1981.

[35] S. Nie and S. R. Emory, "Probing single molecules and single nanoparticles by surface-enhanced raman scattering," *Science*, vol. 275, pp. 1102–1106, 1997.

[36] E. M. Purcell, "Spontaneous emission probabilities at radio frequencies," *Physical Review*, vol. 69, pp. 681, 1946.

[37] W. Kim, V. P. Safonov, V. M. Shalaev, and R. L. Armstrong, "Fractals in microcavities: giant coupled, multiplicative enhancement of optical responses," *Physical Review Letters*, vol. 82, pp. 4811–4814, 1999.

[38] W. L. Barnes, A. Dereux, and T. W. Ebbesen, "Surface plasmon subwavelength optics," *Nature*, vol. 424, pp. 824–830, 2003.

[39] E. Ozbay, "Plasmonics: merging photonics and electronics at nanoscale dimensions," *Science*, vol. 311, pp. 189–193, 2006.

[40] T. W. Ebbesen, H. J. Lezec, H. F. Ghaemi, T. Thio, and P. A. Wolff, "Extraordinary optical transmission through sub-wavelength hole arrays," *Nature*, vol. 391, pp. 667–669, 1998.

[41] R. Colombelli, K. Srinivasan, M. Troccoli, O. Painter, C. F. Gmachl, D. M. Tennant, A. M. Sergent, D. L. Sivco, A. Y. Cho, and F. Capasso, "Quantum sascade surface-emitting photonic crystal laser," *Science*, vol. 302, pp. 1374–1377, 2003.

[42] N. Fang, H. Lee, C. Sun, and X. Zhang, "Sub-diffraction-limited optical imaging with a silver superlens," *Science*, vol. 308, pp. 534–537, 2005.

[43] U. Kreibig and M. Vollmer, *Optical Properties of Metal Clusters*, Springer, Berlin, 1995.

[44] Z. Liu, J. M. Steele, W. Srituravanich, Y. Pikus, C. Sun, and X. Zhang, "Focusing surface plasmons with a plasmonic lens," *Nano Letters*, vol. 5, pp. 1726–1729, 2005.

[45] J. Xiao, X. Sun, and M. Zhang, "Vectorial analysis of optical waveguides by the mapped Galerkin method based on E fields," *Journal of the Optical Society of America B*, vol. 21, pp. 798–805, 2004.

[46] J. G. Goodberlet, J. T. Hastings, and H. I. Smith, "Performance of the Raith 150 electron-beam lithography system," *Journal of Vacuum Science and Technology B*, vol. 19, pp. 2499–2503, 2001.

[47] S. A. Campbell, *The Science and Engineering of Microelectronic Fabrication*, Oxford University Press, New York, 2001.

[48] J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, San Francisco, 1968.

[49] N. V. Kukhtarev, V. B. Markov, S. G. Odulov, M. S. Soskin, and V. L. Vinetskii, "Holographic storage in electrooptic crystals. I. steady state," *Ferroelectrics*, vol. 22, pp. 949–960, 1979.

[50] A. Yariv, *Optical Electronics in Modern Communications*, Oxford University, New York, 1997.

[51] P. Yeh, "Two-wave mixing in non-linear media," *IEEE Journal of Quantum Electronics*, vol. 25, pp. 484–519, 1989.

[52] G. Zhou and D. Z. Anderson, "Photorefractive delay line for the visualization and processing of time-dependent signals," *Optics Letters*, vol. 18, pp. 167, 1993.

[53] K. Okamoto, I. Niki, A. Scherer, Y. Narukawa, T. Mukai, and Y. Kawakami, "Surface plasmon enhanced spontaneous emission rate of InGaN/GaN quantum wells probed by time-resolved photoluminescence spectroscopy," *Applied Physics Letters*, vol. 87, pp. 071102, 2005.

[54] C. Liu, V. Kamaev, and Z. V. Vardeny, "Efficiency enhancement of an organic light-emitting diode with a cathode forming two-dimensional periodic hole array," *Applied Physics Letters*, vol. 86, pp. 143501, 2005.

[55] H. J. Lezec, A. Degiron, E. Devaux, R. A. Linke, L. Martin-Moreno, F. J. Garcia-Vidal, and T. W. Ebbesen, "Beaming light from a subwavelength aperture," *Science*, vol. 297, pp. 820–822, 2002.

[56] J. Takahara, S. Yamagishi, H. Taki, A. Morimoto, and T. Kobayashi, "Guiding of a one-dimensional optical beam with nanometer diameter," *Optics Letters*, vol. 22, pp. 475–477, 1997.

[57] D. E. Chang, "Cavity quantum electrodynamics in the optical domain via surface plasmons," personal communications, 2004.

[58] R. Charbonneau, P. Berini, E. Berolo, and E. Lisicka-Shrzek, "Experimental observation of plasmonpolariton waves supported by a thin metal film of finite width," *Optics Letters*, vol. 25, pp. 844–846, 2000.

[59] P. Berini, "Plasmon-polariton waves guided by thin lossy metal films of finite width: Bounded modes of symmetric structures," *Physical Review B*, vol. 61, pp. 10484–10503, 2000.

[60] P. Berini, R. Charbonneau, N. Lahoud, and G. Mattiussi, "Characterization of long-range surface-plasmon-polariton waveguides," *Journal of Applied Physics*,

vol. 98, pp. 043109, 2005.

[61] B. Lamprecht, J. R. Krenn, G. Schider, H. Ditlbacher, M. Salerno, N. Felidj, A. Leitner, and F. R. Aussenegg, "Surface plasmon propagation in microscale metal stripes," *Applied Physics Letters*, vol. 79, pp. 51, 2001.

[62] J.-C. Weeber, J. R. Krenn, A. Dereux, B. Lamprecht, Y. Lacroute, and J. P. Goudonnet, "Near-field observation of surface plasmon polariton propagation on thin metal stripes," *Physical Review B*, vol. 64, pp. 045411, 2001.

[63] S. A. Maier, P. G. Kik, H. A. Atwater, S. Meltzer, E. Harel, B. E. Koel, and A. A.G. Requicha, "Local detection of electromagnetic energy transport below the diffraction limit in metal nanoparticle plasmon waveguides," *Nature Materials*, vol. 2, pp. 229–232, 2003.

[64] S. A. Maier, P. G. Kik, and H. A. Atwater, "Observation of coupled plasmon-polariton modes in Au nanoparticle chain waveguides of different lengths: Estimation of waveguide loss," *Applied Physics Letters*, vol. 81, pp. 1714, 2002.

[65] J. R. Krenn and J.-C. Weeber, "Surface plasmon polaritons in metal stripes and wires," *Phylosophical Transactions of the Royal Society A*, vol. 362, pp. 739–756, 2004.

[66] J. R. Krenn, B. Lamprecht, H. Ditlbacher, G. Schider, M. Salerno, A. Leitner, and F. R. Aussenegg, "Nondiffraction-limited light transport by gold nanowires," *Europhysics Letters*, vol. 60, pp. 663–669, 2002.

[67] S. I. Bozhevolnyi, V. S. Volkov, E. Devaux, and T. W. Ebbesen, "Channel plasmon-polariton guiding by subwavelength metal grooves," *Physical Review Letters*, vol. 95, pp. 046802, 2005.

[68] D. F. P. Pilea, T. Ogawa, D. K. Gramotnev, T. Okamoto, M. Haraguchi, M. Fukui, and S. Matsuo, "Theoretical and experimental investigation of strongly localized plasmons on triangular metal wedges for subwavelength waveguiding," *Applied Physics Letters*, vol. 87, pp. 061106, 2005.

[69] S. A. Maier, M. D. Friedman, P. E. Barclay, and O. Painter, "Experimental demonstration of fiber-accessible metal nanoparticle plasmon waveguides for planar energy guiding and sensing," *Applied Physics Letters*, vol. 86, pp. 071103, 2005.

[70] S. A. Maier, P. E. Barclay, T. J. Johnson, M. D. Friedman, and O. Painter, "Low-loss fiber accessible plasmon waveguide for planar energy guiding and sensing," *Applied Physics Letters*, vol. 84, pp. 3990, 2004.

[71] W. Nomuraa, M. Ohtsub, and T. Yatsui, "Nanodot coupler with a surface plasmon polariton condenser for optical far/near-field conversion," *Applied Physics Letters*, vol. 86, pp. 181108, 2005.

[72] L. Yin, V. K. Vlasko-Vlasov, J. Pearson, J. M. Hiller, J. Hua, U. Welp, D. E. Brown, and C. W. Kimball, "Subwavelength focusing and guiding of surface plasmons," *Nano Letters*, vol. 5, pp. 1399–1402, 2005.

[73] R. M. Dickson and L. A. Lyon, "Unidirectional plasmon propagation in metallic nanowires," *Journal of Physical Chemistry B*, vol. 104, pp. 6095–6098, 2000.

[74] E. Hecht, *Optics*, Addison Wesley, San Francisco, 2002.

[75] E. D. Palik, *Handbook of Optical Constants of Solids*, Academic Press, San Diego, 1998.

[76] P. B. Johnson and R. W. Christy, "Optical constants of the noble metals," *Physical Review B*, vol. 6, pp. 4370, 1972.

[77] B. T. Draine and P. J. Flatau, "User guide to the discrete dipole approximation code ddscat.6.0," http://arxiv.org/abs/astro-ph/0309069, Sep 2003.

[78] R. Scarmozzino, A. Gopinath, R. Pregla, and S. Helfert, "Numerical techniques for modeling guided-wave phontonic devices," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 6, pp. 150, 2000.

[79] K. S. Chiang, "Review of numerical and approximate methods for the modal analysis of general optical dielectric waveguides," *Optical and Quantum Electronics*, vol. 26, pp. S113, 1994.

[80] A. Taflove and S. C. Hagness, *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, Artech House, Boston, 2005.

[81] D. Marcuse, "Solution of the vector wave equation for general dielectric waveguides by the Galerkin method," *IEEE Journal of Quantum Electronics*, vol. 28, pp. 459–465, 1992.

[82] S. J. Hewlett and F. Ladouceur, "Fourier decomposition method applied to mapped infinite domains: scalar analysis of dielectric waveguides down to modal cutoff," *IEEE Journal of Lightwave Technology*, vol. 13, pp. 375–383, 1995.

[83] S. Sujecki, T. M. Benson, P. Sewell, and P. C. Kendall, "Novel vectorial analysis of optical waveguides," *Journal of Lightwave Technology*, vol. 16, pp. 1329–1335, 1998.

[84] M. A. McCord and M. J. Rooks, "Electron beam lithography," in *Handbook*

of Microlithography, Micromachining, and Microfabrication: Volume 1, P. Rai-Choudhury, Ed., p. 139. SPIE, 1997.

[85] P. N. Minh and T. Ono, "High throughput aperture near-field scanning optical microscopy," *Review of Scientific Instruments*, vol. 71, pp. 3111, 2000.

[86] K. J. Vahala, "Optical microcavities," *Nature*, vol. 424, pp. 839–846, 2003.

[87] M. O. Scully and M. S. Zubairy, *Quantum Optics*, Cambridge, New York, 1997.

VITA

Zhijie Deng was born in Guangzhou, China. He received the B. Eng. degree in Electrical Engineering from Tsinghua University, Beijing, China in July, 2000, and the Ph.D. degree in Electrical and Computer Engineering from Texas A&M University in May, 2006. The doctoral dissertation focused on novel optical devices for information processing. The latest project investigated the surface plasmonic nano-waveguides and coupling antennas, including analysis, design, fabrication by electron beam lithography and testing. Other projects covered room temperature ultraslow light and image memories by computer generated holograms. In support of parallel work in the lab, he also helped constructing a laser scanning confocal microscope, which enabled fluorescent imaging of single atoms, detecting single photon emissions as well as electron spin resonance. He can be reached by the mailing address as follows: Dept. of Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843-3128, USA.

The typist for this thesis was Zhijie Deng.