PHOTOREFRACTIVE VOLUME HOLOGRAPHY IN ARTIFICIAL NEURAL NETWORKS

Thesis by

David Jones Brady

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California
1990
(Submitted December 19, 1989)

ACKNOWLEDGEMENTS

Professor Demetri Psaltis has been a patient teacher, a wise advisor and a loyal friend. Although I have not always been able to follow where he has led me, he has never led me astray. I would like to thank him for his support. I am sorry that I fouled him so hard and so often on the basketball court.

The work described in this thesis reflects collaboration and discussion with Dr. Ken Hsu, Dr. Xiang-Guang Gu, R. Scott Hudson, Dr. David Kagan, Daniel Raguin, Dr. Jeffery Yu, Sidney Li, and Dr. Kelvin Wagner. My stay at Caltech has been made more pleasant and my work easier as a result of the generous friendships of Mark Neifeld, Dr. Nabeel Riza, Cheol Hoon Park, Dr. Robert Snapp, Dr. Fai Ho Mok, Dr. John Hong, Alan Yamamura, Charles Stirk, Chaunyi Ji, Steven Lin, Francis Ho, Seiji Kobayashi, Subrata Rakshit, and Yong Qiao. I would like to thank Su McKinley and Helen Carrier for the kind and efficient manner in which they provide administrative support. When I began my work with this group, Dr. Gabriel Sirat told me that it was natural for the experienced members to help the new members. I hope that I have returned to the group some portion of the experience Dr. Sirat lent me.

The first three years of my studies at Caltech were supported by the Office of Naval Research and the American Society for Engineering Education through the ONR/ASEE Graduate Fellowship Program. Support in later years was provided by the Aerojet Corporation through the Caltech Program in Advanced Technologies and by the ARCS Foundation. I am deeply grateful to these organizations. I am also grateful to Ratnakar Neurgaonkar and Rockwell International, who supplied the SBN used in some of my experiments.

I cannot convey the depth of my appreciation for the unwavering support of Verl and Norma Brady, but I proudly acknowledge their help and humbly thank them for it. My wife, Rachael Alexandra, and my daughter, Katherine Anne, have also been generous. I will not say that I did this for them, for they might have been as happy if I had stayed home more. I thank them, however, for understanding.

ABSTRACT

This thesis describes the use of volume holography to implement large-scale linear transformations on distributed optical fields. Such transformations are useful in the construction of hardware for artificial neural networks. The reconstruction of multiple grating holograms in layers of thin transparencies and in continuous volume media is considered and conditions under which such holograms may be used for linear transformations are derived. The control of the nature of the transformation implemented using fractal sampling grids is reviewed and the impact of such sampling grids on the energy efficiency of the overall system is considered. Information storage in volume holograms is shown to require multiple exposures and the impact of multiple exposures on linear hologram formations in saturable media and photorefractive materials is considered. It is shown for both types of media that the overall diffraction efficiency of a recorded hologram must decrease with the square of the rank of the transformation implemented. A theory for hologram formation in photorefractive materials with multiple trapping species is developed and compared with experimental results. The impact of multiple species and fixing mechanisms on linear hologram formation is evaluated. A method for refreshing the diffraction efficiency of photorefractive holograms in adaptive systems is described and demonstrated. The construction of thick holograms for linear transformations in waveguides is considered. A novel method for controlling such holograms is described and demonstrated. Learning in holographic neural networks is considered and two experimental holographic neural systems are described. The relative strengths of optical and electronic technologies for implementations of neural interconnections are considered.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 THE ALLURE OF VOLUME HOLOGRAPHY

In the sense that it offers, within the limits of its resolution, the possibility of complete knowledge of and control over the properties of a region of space, volume holography represents the ultimate technique in the quest to understand and manipulate the electromagnetic properties of the physical world. One can imagine using this technique to construct arbitrary three-dimensional objects. One might imagine constructing such objects using a "brick by brick" approach, where volume elements are stacked layer by layer to form the object. Holographic techniques, on the other hand, form a portion of every volume element, or "voxel," in each processing step. A hologram is formed by interference between "control beams." While this thesis deals exclusively with optical holograms, the control beams might in general lie in any portion of the electromagnetic spectrum or might even consist of acoustic or particle beams. Continuing our analogy with masonry, holography is a technique for forming an object by injecting streams of clay and mortar whose superposition is the desired stack of bricks. The change in each voxel in each processing step is related via a linear transformation to the amplitudes of all the control beams.

Obviously, if the physical parameters of each voxel are independent then any method for controlling the volume must have at least one degree of freedom for each voxel. Thus, the number of degrees of freedom in a set of holographic control beams cannot be less than the number of degrees of freedom used in the brick by brick approach, *i.e.*, one degree of freedom per brick. The potential advantage of holography lies in the fact that it is not necessary to reach into the volume to form or evaluate a hologram. The volume is controlled and evaluated using only boundary conditions. While this is a crucial advantage in the control process, removing the need to individually probe each voxel is particularly significant in evaluating the state of a volume, since for typical materials no nondestructive alternative exists even in principle.

The potential of volume holography for information storage was recognized from its conception. van Heerden was first to emphasize this possibility and noted that the number of bits which can be stored in a volume is proportional to the resolution of the holographic system. For optical holography, this limit suggests the possibility of storing up to 10^{12} bits per cm^3. This possibility represents "the allure of volume holography." While many workers have been drawn to holography by this possibility, the promise of holograms for high density information storage remains largely unfulfilled. The principle reasons why volume holograms have failed to contribute are first, that the natural parallel format of the inputs and outputs to a hologram are not well matched to the sequential format of information in traditional information processing devices, second, that competing information storage technologies have developed at a breakneck pace, and third, that supporting optical technologies and devices have not been sufficiently advanced for large-scale holographic systems.

Since volume holography has existed for 30 years, most of a professional lifetime, it is easy to lose perspective and to downplay its applicability. It should be recalled, however, that information science itself is not much older than holography. Since information science is likely just reaching its paleozoic era, it seems unwise to predict the future of information storage hardware based on the experience of the last 30 years. The integrated electronic microchip, also an invention of thirty years, has just over the last decade begun to exert its influence as a component of everyday devices. Over the next 30 years, the implications of this invention in the control of ordinary devices and the construction of exotic hardware will likely continue to multiply in unforeseen ways.

The competitors of volume holography in the area of large-scale information storage technologies have changed dramatically over the last decade. Planar electronic memories have grown in scale to the point where they are able to challenge mechanically accessed magnetic memories in speed sensitive applications. The resolution and speed of magnetic memories has continued to improve as fabrication technologies, materials and magnetic heads have become more advanced. Planar optical memories, which where roughly contemporaries of holographic memories in conception, have begun to play a major role. The history of planar optical memories is particularly illuminating since it shows that even a technology which is basically a good idea from its inception is adopted only as the original idea is refined and the needs of the overall system evolve to a compatible state. While these competing technologies are fundamentally planar, the advantage of volume holography lies in the possibility of 3-D access rather than 3-D storage. Planar media can be rolled into 3-D, as is done with magnetic tape and optical film. The fact that the readout is still planar makes this approach excruciatingly slow, however. Magnetic disks are also typically stacked, but require a separate readout system for each layer. As is shown in some detail in chapter 2, a volume hologram can be viewed as a stack of optical disks read out in parallel with beams generated on a single boundary. In fact, implementation of volume storage using a stack of optical disks may be a practical approach.

Volume holographic information storage is ripe for a fresh look because of the potential it offers for large scale parallel interconnections in parallel computing systems. Novel computing architectures are of increasing interest because traditional architectures seem to be getting better at the sort of numerically intensive problems computers have historically handled well rather than adding new sorts of problems, such as abstract pattern recognition, to their repertoire. These novel architectures are typically based on parallel consideration of spatial or temporal relationships which map easily onto the parallel nature of bulk optical systems. The most intensely parallel systems are artificial distributed neural processors. Examples of such neural systems are used to explore the use of volume holographic storage in this thesis.

While the need for parallel adaptive hardware is the primary driving force in the renaissance of interest in volume holographic storage, developments in optical technology have also played a role. Advances in coherent sources and spatial light modulators have continued at a steady pace for some time. Further dramatic improvements in laser sources seem assured in light of recent developments in solid state laser technology. High power, high efficiency, narrow line solid state sources should in the near future eliminate the troublesome need for gas lasers in holographic systems. The future of spatial light modulators is less assured, but promising devices are under development and, as shown by the experimental systems described in this thesis, interesting holographic systems can be constructed using commercially available devices. While the development of holographic materials has also proceeded at a steady pace, advances in the understanding of currently available materials has probably been of more significance over the last decade. While fundamental improvements in the properties of photorefractive materials seem unlikely, it is possible that new holographic media with exciting capabilities will come out of research into organic materials.

1.2 A BRIEF HISTORY OF VOLUME HOLOGRAPHY

For the purposes of this thesis, holography can be viewed as a technology first, for changing the properties of sensitive media in proportion to the intensity of a set of control beams and second, for probing the properties of media using similar beams. The historic basis of the first task lies in conventional photography. The historic basis of the second task lies in scattering phenomena used to determine the internal structure of volumes of unknown composition. Since the relationship between holography and scattering phenomena is less commonly emphasized, this section is biased somewhat toward the history of scattering phenomena. Ballistic scattering, *i.e.*, throwing a projectile into the volume and measuring the direction at which it bounces out, is the simplest technique. When the contents of the volume are complex, information can be gathered more efficiently if complex excitations are used as probes. The properties of the volume are determined by their effect on the diffraction of the probes. Early in this century, electron and x-ray scattering were used to probe the atomic structure of materials. Of particular interest is the x-ray microscope developed by Bragg, which used a two-step process to magnify and image 2-D projections of crystal structures [2]. Unfortunately, the range of materials to which Bragg's technique could be applied was limited by the fact that he had no means of determining the phase of the scattered signals.

The problem of determining the phase of scattered signals, at least at optical frequencies, was solved by Gabor [3], who later coined the term holography to describe his discovery. Gabor's initial interest was the application to electron microscopy of the two-step imaging technique. This technique is based on recording a transparency proportional to the scattered signal. When this transparency is illuminated, the recording field is recreated at a magnification proportional to the ratio of reconstructing and recording wavelengths. Ideally, the transmittance of the transparency is proportional to the field of the diffracted x-rays. In Bragg's x-ray system, however, only the intensity of the scattered signal could be detected and used to form the transparency. For this reason, Bragg's technique could not be used to form images of crystals where this phase information was needed. Gabor was able to overcome this problem by recording the phase information in the interference between the probe and the diffracted light. Since a crystal is known to be a lattice of identical unit cells, *a priori* knowledge about the nature of crystals can in many cases be used to make the problem of determining crystal structures using information from scattered x-rays tractable without resorting to Gabor's method [4]. Scattering from an optical volume hologram can be regarded as the optical analog of x-ray crystallography, but a volume hologram need not be regular or periodic. Evaluation of a volume hologram thus must rely critically on holographic methods.

In line with the application he envisioned, the method Gabor proposed for hologram formation had the probe and the diffracted signals co-propagating. This colinear alignment made it difficult to extract the image from noise in reconstructed holograms. This problem was overcome in 1962 with the development of two-beam off-axis holography by Leith and Upatnieks [5, 6]. While we don't say much here about the relative advantages of the off-axis geometry, we should note that the systems described in this thesis use this geometry because of the analytic simplification offered by the separation of the probes and the scattered signals and because the media used in our experiments generally have modulation transfer characteristics which are maximal at relatively high spatial carrier frequencies.

A new form of holography invented by Denisyuk [7] improved the image quality of single beam holograms using a reflective geometry based on a method for making color photographs developed by Lippman [8]. In contrast to the planar slice of the interference pattern provided by Gabor's method, the grating wavevectors in the recording interference pattern in Denisyuk's holograms were normal to the surface of the holographic emulsion. In order to record a large number of fringes, the emulsion in Denisyuk's holograms was much thicker than the wavelength of the recording light. The holograms were thus volume in nature and evinced the primary attribute of volume holograms, Bragg sensitivity. The image quality was improved in these holograms because the reconstructing beam was not Bragg matched to the conjugate of the object.

Independently, van Heerden also considered holograms in volume media and estimated their storage capacity [1]. The advent of volume, as opposed to thin, holograms necessitated new wavecoupling theories. Of course, similar theories had long existed to describe X-ray diffraction and acousto-optics effects. The early extensions to volume holography were scattering theories based on the assumption of weak coupling. Phase modulated volume holograms can support very high diffraction efficiencies, however. The first theory to adequately describe such effects was the coupled wave theory developed by Kogelnik [9]. The early history of volume holography and the many theories for holographic coupling which have since arisen are reviewed by Collier *et al.* [10], by Russell [11] and by Solymar and Cooke [12].

In this thesis we are primarily concerned with holograms in photorefractive media, which are electro-optic material in which phase holograms may be formed by the displacement of photo-generated charge. Photorefractive effects, first observed in lithium niobate and lithium tantalate [13], have been observed in a wide range of electro-optic materials. The principle families of materials are ferro-electric crystals such as lithium niobate, barium titantate [14] , and strontium barium niobate [15], silinites such as the oxides of bismuth with silicon, germanium or titanium [16], and semiconductors such as gallium arsenide and indium phosphide [17, 18].

A number of theories have been used to describe the formation of photorefractive holograms. The most commonly used theory is the band transport model developed by Kuktarev [19]. This theory is reviewed in chapter 4. More detailed reviews and overviews of current topics photorefractive research are provided in [20, 21].

As discussed in section 1.1, the possibility of using holography for mass information storage was recognized from its inception. Early attempts to realize this potential may be dichotomized into techniques for conventional digital storage and associative memory techniques. Of course, optical data storage on planar media is well established, both in analog form on film and, more recently, in digital form on disks. With the exception of imaging radars, such systems have not involved holographic techniques, although recently holographic storage on optical disks has been investigated [22]. Digital storage in volume media is complicated by the fact that it is not a simple task to take an independent look at the information stored in a single voxel. Early proposals for information storage in a volume generally were "page oriented," which meant that image holograms were multiplexed in the volume at different carrier frequencies or spatial positions. Each stored page could be reconstructed by reading the hologram with the appropriate reference [23, 25, 24, 26]. Multiple image holograms must be superposed at each spatial position if the capacity of a volume system is to exceed the storage capacity of easily controlled planar media. The principle difficulties encountered in constructing page oriented memories were the loss of dynamic range when multiple holograms are recorded [27] and the lack of computing systems which could make good use of data in 2-D blocks. The first of these problems also arises in the systems considered in this thesis and is considered in some detail below. The second problem does not arise in neural processors, where the spatially distributed signals are not only tolerated, but are in fact necessary. In spite of these problems, Staebler, Burke, Phillips and Amodei were able to record as many as 500 superposed plane wave holograms in $LiNbO_3$ [28]. While the experimental work in this study was very impressive, the assumption was made in the analysis that fixing was of use in multiple hologram storage and a fixing technique was used in the experiments. The assumption that fixing increases the storage capacity is dubious for reasons explained in chapter 4. Recently, Mok, Tackitt and Stoll have recorded 500 image bearing holograms in $LiNbO_3$ with very uniform diffraction efficiency using the multiple exposure technique outlined in chapter 4 [29].

Early proposals for holographic associative memories were based on the ability of a hologram to reconstruct a complete signal from a partial reference [30, 31]. Of course, this principle is not strictly valid, since a limited region on a plane cannot have as many degrees of freedom as are available across the whole plane, but with feedback or pre-processing methods this difficulty can be overcome. In recent years several systems have been proposed which incorporate such refinements. These systems fall into the realm of optical neural processors and are reviewed in the next section. Most optical associative memories of this sort are based on the ability of a thin hologram to act as a spatial filter. This ability is compromised to some extent by the Bragg selectivity of volume holograms. The work described in much of this thesis can be viewed as an attempt to combine the storage capacity of page oriented memories with the associative capabilities of optical associative memories.

1.3 OPTICS AND ARTIFICIAL NEURAL NETWORKS

Advances in the understanding of biological neural networks and the continued ineffectiveness of traditional information processing systems in pattern recognition applications have led in recent years to a dramatic resurgence of interest in artificial neural systems [32-36]. The structure of these systems is to a large extent biologically motivated. The reason neurobiology is viewed as a source for ideas in the design of computing systems lies in the empirical distinction that exists between the numerical problems which modern digital computers handle well and the pattern extraction problems at which humans and other animals excel. Interest in optical neural computers is based on the expectation that this difference in capabilities is due to basic hardware differences. In some instances, particularly in the case of sensory systems, it is possible to construct circuits that are reasonably accurate replicas of biological systems [37]. More often, however, we only attempt to extract basic properties that may be evident in the nervous system and use this information to guide the design of computers. The hope is that if we are successful in identifying relevant properties, then the neural analogy will be a positive input to the process of designing a computer even though we do not have a detailed understanding of the circuitry of the nervous system.

The systems considered in this thesis are based on the assumption that the potential advantages of neural systems lie the following aspects of their architecture:

1. *Parallelism.* There are perhaps 10^{12} neurons in the human brain. At the simplest level of abstraction, each neuron may be viewed as an independent processor. Each processor operates asynchronously on its input signals. Typically, many processors will be simultaneously active. In conventional systems, in contrast, only a very few processors are simultaneously active.

2. *Dense Interconnections.* A neuron in the brain typically receives inputs from several hundred to several thousand other units to produce its own output which is broadcast to roughly the same number of units. It is widely believed that the synaptic strengths serve as part of the memory of the system.

3. *Learning.* Humans learn to perform most tasks by exposure to the problem and practice at solving it. Conventional computers, in contrast, generally learn by algorithmic instruction. Based on this distinction, the design of neural systems is inverse to the design of logical systems in the sense that the system designer tries to match the neural network hardware to the problem he wishes to solve but expects the neural software, *i.e.*, the synaptic strengths, to adapt itself to the specifics of the problem. The only software which the designer specifies is the "learning algorithm" which controls the adaptation of the synapses. In conventional machines the hardware is de-

signed to be general purpose and the system designer specifies software to adapt his machine to problems of interest. The traditional view that the adapting a general purpose machine is easier than building special purpose hardware is correct for traditional numerical computational problems, but is not likely to apply to pattern recognition problems which have historically been difficult to solve with simple algorithms.

The reason optical technology has been considered for applications in neural hardware is that dense interconnections can be implemented in optics with relative ease. The concept that holography could be used to mimic brain functions occurred in van Heerden's original paper on volume holography and in several of the early papers on optical associative memories [1,30, 31]. While the conceptions of the mechanisms behind brain functions which are currently popular are substantially different from that assumed by van Heerden, the basic idea that the linear interconnection capabilities of optical holography makes it a suitable technology for artificial neural interconnections remains sound. More recently a second attraction for the inclusion of optics in neural systems has arisen, which is that dynamic holography, such as occurs in photorefractive crystals, provides a very natural basis for implementation of neural learning algorithms.

In step with the revival of interest in artificial neural systems, interest in optical neural computing has grown dramatically in recent years. Implementations of Hopfield style networks were proposed by Psaltis and Farhat [38] in 1985. The first proposed implementation was based on an incoherent vector-matrix multiplier [40] and was not holographic in nature. A second architecture proposed for processing two-dimensional signals was based on more traditional holographic associative principles but again used thin storage media. Since 1985 a large number of optical implementations of artificial neural networks have been proposed [41- 71]. In these systems, the activity of the neurons is represented by the optical field or intensity in some spatial mode. The modes are interconnected by transformations stored in either holographic or image plane optical elements and a nonlinear electronic or optical device is used to threshold the activity of each mode to generate feedback and output signals. The principle differences between the various systems involve the nature of the light sources and modulators used, the choice of thin or thick interconnection media, and the choice of a thresholding element. Of course, the neural algorithm being implemented often plays a key role in determining the choices taken.

This thesis is limited mostly to a consideration of the capabilities of volume holographic elements as media for neural-style interconnections. We rely in particular on photorefractive media. These media have been used in resonator style memories [41, 52,44, 45] and in linear interconnection networks [50, 51, 56, 57, 59]. The second approach is taken here. Since we wish our results to be as general as possible, very little consideration is given to specific neural models. In fact, after reviewing the framework for volume holograms in neural systems in this chapter, we do not consider neural systems again until chapter 6.

The various components of a pair of "neurons" and their interconnection hardware is shown conceptually in Fig. 1.1(a). Each neuron receives inputs through the synapses on its dendrites, it processes these inputs in some fashion, and then it broadcasts the result on the axon where it is picked up by the dendrites of other neurons. Except at the conceptual level, we do not claim any biological authenticity for this model. The diagram in Fig. 1.1(b) shows the holographic analog of the two neurons in Fig. 1.1(a). The neurons are nonlinear

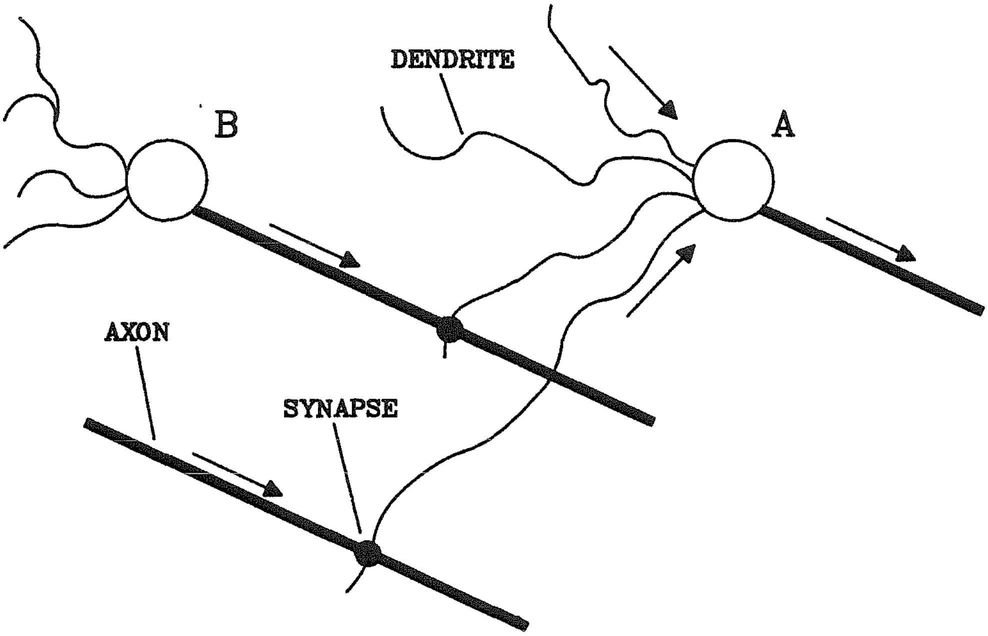


Figure 1.1(a). Conceptual structure of a neural system. Neurons are nonlinear nodes which receive inputs from other nodes via a set of "dendrites." Each neuron broadcasts its output along an "axon." Connections between axons and dendrites are made through "synapses."

devices which detect and emit light. The activity of each neuron is coded in the amplitude or intensity of the signal it emits. The input into each neuron is sensed by a detector. Interconnections between neurons are made by holographic gratings. The strength of the connection between a pair of neurons depends on the modulation depth of the grating which connects them. We can draw some direct analogies between the component structure of a neuron and its optical simulation. The output light beam serves the role of the axon, broadcasting the signal from each neuron. The holographic grating plays the role of the synapse, directing the signal from one neuron to the next and the optical pathways through which light is transferred from the hologram to the detector area of the neuron is analogous to the dendrites. As a group, the optical source and detector, and the circuits that process the detected signal are reminiscent of the soma of the neuron.

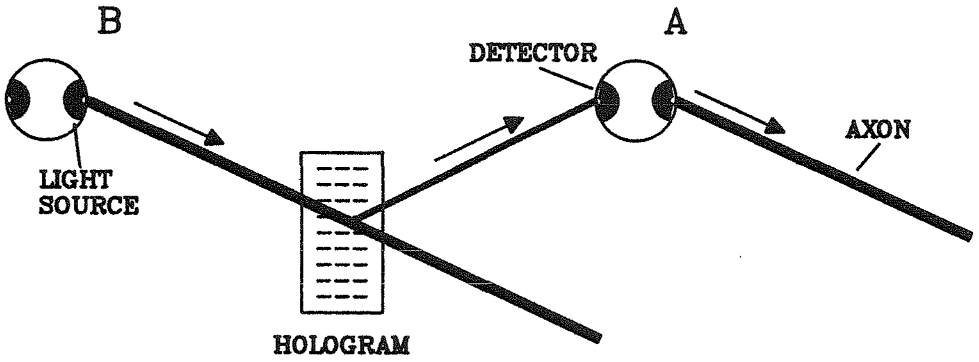


Figure 1.1(b). Optical analog of a neural system. A light source broadcasts the activity of each neuron into an optical mode which corresponds to the axon for that neuron. The synapses correspond to holograms which diffract light into optical modes incident on detectors at each neuron. The diffracted modes correspond to dendrites.

Due to the photolithographic techniques which are used to build them, arrays of opto-electronic devices are generally constrained to planes. This constraint leads to architectures for optical neural computers such as that shown in Fig. 1.2. A plane of sources at the input is connected to a plane of detectors at the output via a volume hologram. Light sources on a training plane are used in concert with the input plane to control the hologram. To form a general neural network,

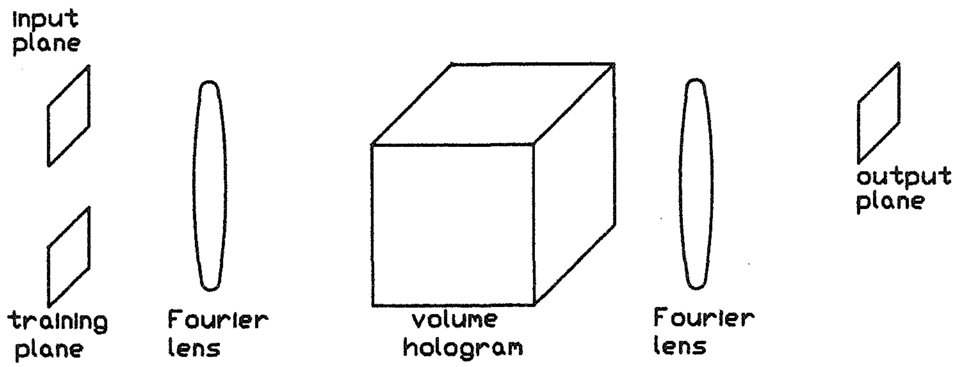


Figure 1.2. Optical neural computer architecture.

this basic module can be cascaded and feedback channels can be added. By fortunate coincidence, this layered architecture is also employed in most models for artificial neural nets. The holographic element in this system implements a linear transformation between the input and output planes. The formation and control of this element is the principle theme of this thesis.

Since we do not consider the planar elements in detail, it is useful at this point to briefly review the devices which are available or under development for this purpose. These devices can be classified according to the nonlinearities they implement and the amount of gain they provide. Transmissive devices, such as liquid crystal displays, magneto-optic light valves [72], and etalon arrays [73, 74, 75], provide no gain. Liquid crystal displays and magneto-optic devices are typically electrically controlled, though optical control is possible. Signal dependent nonlinearities are implemented in these devices by performing electrical operations on detected optical signals and using the resulting electrical signals to drive the device. Etalon arrays, on the other hand, are optically controlled. The nonlinearity implemented by these devices is determined by the architecture and detailed physics of the materials used. In order to use these devices in a multilayer network, optical gain must be provided elsewhere in the optical architecture. Blocking layer devices, such as liquid crystal light valves [76] and multichannel spatial light modulators [77], provide gains of $10^6 - 10^9$ by isolating the output of the device from the input. Control passes from the input plane to the output plane via photogenerated charge. Essentially, these devices can be regarded as a cascade of a detector array, a simple nonlinearity, and an electrically-controlled transmissive modulator. Recently, a third class of modulators has been developed in which detectors and optical sources are integrated monolithically [78]. Since they involve active sources and integrated devices, these devices can in principle provide complex nonlinear functionality and gain. While the experiments described in this thesis use liquid crystal light valves, we do not mean to argue that these devices are superior to other types. All of the above devices have potential applications in neural interconnection systems. For our purposes it is enough to note that devices adequate for the systems described here exist now and that potentially superior devices are under development.

1.4 THESIS OVERVIEW

The primary goal of this thesis is to demonstrate methods for using volume holograms to implement adaptive linear transformations. Two problems are encountered in attempting to use volume holograms in this manner. First, the number of degrees of freedom which can be used to implement a transformation in a volume is insufficient to perform an arbitrary linear transformation on arbitrary optical fields. This problem can be addressed by appropriately constraining the degrees of freedom of the readout field. Methods for designing appropriate constraints are presented in a recent thesis by Xiang-Guang Gu [79] and are briefly mentioned in chapter 2 of the present thesis. The second problem is that the number of degrees of freedom in a monochromatic optical field is insufficient to control a volume hologram. This problem is addressed by making multiple exposures to form holograms. The implications of this second problem are the the topic of this thesis.

Chapters 2 and 3 develop a theory for the formation, control and reconstruction of volume holograms for linear interconnections. As described in section 1.2, a number of authors have considered the formation and reconstruction of volume holograms. The analysis presented here differs from previous analyses in that many grating holograms are strongly emphasized. The conditions assumed in the analysis are such as to keep the holograms in a linear first-order regime. We show in these chapters that by applying discrete Fourier analysis we can illuminate a number of surprisingly simple characteristics of volume holograms. The simplicity is surprising because rigorous analysis of even single grating holograms is well known as a messy and difficult prospect. The many grating problem is easy in contrast because the modulation depths of the individual gratings are extraordinarily small. A second limitation on the many grating problem which is used to make the analysis tractable is the limitation of the spatial bandwidth of the gratings to a small convex region in the allowed grating space.

In order to develop the application of discrete Fourier analysis to volume holographic transformations in as simple a manner as possible, much of chapter 2 is concerned with the analysis of Bragg diffraction from a cascade of thin holograms. Information storage was first proposed in such a system by Pohl in 1974 [80]. Pohl suggested that the information stored in different layers be discriminated using pulses and time of flight measurements rather than the holographic methods suggested in chapter 2, however. A another approach suggested by Thaxter and Kestigian [81], consisted of layers of thin SBN holograms. The gratings stored in a layer could be masked and unmasked electronically, allowing independent holographic readout of each plane. More recently, Johnson and Tanguay have suggested information storage in a holographic cascade [84] and have analyzed this system using a beam propagation method [83, 82]. To the author's knowledge, however, a detailed consideration of this problem such as that given in chapter 2 has not been previously published. Interesting new results which this approach yields include the relationships between "distributed" and "local" holograms derived in subsection 2.2.2 and various constraints on many beam interconnection networks derived in subsection 2.2.3. The system is shown to exhibit diffraction characteristics similar to those of continuous volume holograms and is used to describe many of these characteristics. The analysis in section 2.2 is based on a lecture given to the Fourier optics class at Caltech in the spring of 1988. The third section of chapter 3 presents coupled wave theory in a discrete multiple grating context. While the perspective from the many grating point of view using discrete Fourier methods is not conventional, the results in this section are trivial extensions of previous studies, particularly Kogelnik's [9]. The final section describes what we mean by an "arbitrary linear transformation" and describes methods for constraining the input fields to achieve such transformations. The fractal sampling grids described in [50, 85, 70, 86,79] are reviewed. A new result describing the impact of sampling grids on the diffraction efficiency of interconnections systems is presented in this section.

Chapter 3 describes the use of optical fields to control both cascades of thin holograms and true volume holograms. While the determination of three-dimensional structure from holographic data has been an area of active interest for some time [88], the control of three-dimensional structures with holography has not received similar attention. In principle, control is achieved by simply inverting the scattering process. In practice, the lack of sufficient boundary conditions to do this in one step leads to constraints on the nature of the structures which can be recorded. The key points in this chapter are that multiple exposures are needed to control a hologram and that it is possible to construct an arbitrary volume hologram in a linear holographic medium if the total number of degrees of freedom in the recording exposures is equal to the total number of degrees of freedom in the recorded hologram. As an example of real holographic media, hologram formation in saturable materials is considered in section 3.3. We find that it is possible to form holograms linearly in a saturable material if an appropriate sacrifice of the total dynamic range of the material is made in each exposure. The final section of chapter 3 considers the use of polychromatic light to control volume holograms, the advantage of such control being that a polychromatic field can specify a volume hologram in one exposure. We find, however, that this advantage is likely to be negated by the difficulty of controlling a polychromatic field. Some of the results developed in this chapter were first presented in [87].

Chapter 4 applies the holographic control techniques of chapter 3 to photorefractive materials. These materials are of particular interest because semi-permanent high efficiency phase gratings can be recorded in them using reasonable optical powers. We review the formation of photorefractive holograms with a special emphasis on many grating holograms in the first two sections of the chapter. The third section considers a method for linearizing hologram formation in photorefractive media and describes experimental results based on this method. Similar exposure methods were found by Bløtekjaer [27] and by Burke and Sheng [89]. Our results, derived directly from the band transport theory, are slightly different than those in [27, 89]. Strasser *et al.* [90] present a fourth variant, but the applicability of this version is clouded by the fact that the analysis used to derive it lacks internal self-consistency. A different exposure method based on a stocastic approach was described by McRuer *et al.* [91]. This section is based on results presented in [57, 59, 92, 93]. Figure 4.1 in this section was generated by Sidney Li. We find that the cost to the dynamic range exacted by linear multiple hologram storage is identical to the cost to the dynamic range in saturable materials. A discussion, new to this thesis, of the advantages of hologram formation in the presence of a large applied field is presented. In section 4.4 we consider the impact of multiple trapping species on hologram formation and describe experimental results which show the effects of multiple species in $LiNbO_3$. We explain our experimental results using a band transport theory for multiple species materials developed earlier by Valley [94]. In order to explain our results it is necessary to change the boundary conditions Valley assumed in his analysis. The development of appropriate boundary conditions and the solution of the transport equations in the presence of an applied field form the major new contributions of this section. Figures 4.9 and 4.10 in this section were generated by Dr. David Kagan. The principal advantage of photorefractive materials for holographic interconnects is that photorefractive holograms can be recorded dynamically *in situ.* In an adaptive recording environment it is not expected that the holographic recording schedule will be optimal. In section 4.5 we suggest a method for using copying techniques to restore the diffraction efficiency of a hologram in an adaptive recording environment and describe experimental results demonstrating this method. This method was described by Brady, Hsu, and Psaltis in [92, 93]. The experiments described in this section were performed jointly by the author and by Dr. Ken Hsu. The final section of chapter 4 briefly considers the impact of fixing mechanisms on photorefractive storage.

It is interesting to note that both of the key problems in implementing holographic interconnections are avoided by thick holograms in waveguides. Since the inputs and outputs to a planar waveguided are one-dimensional, a two-dimensional waveguide supports sufficient degrees of freedom to implement an arbitrary linear transformation on its inputs. A waveguide hologram can be controlled in a single exposure by a 2-D control field incident from above the waveguide. In chapter 5 we consider holograms for linear interconnections in optical waveguides. The key result of the chapter is a method for specifying the matrix for an integrated vector matrix multiplier in one exposure using one SLM. We present experimental results in support of this method. A number of previous studies have been made of holograms in waveguides. A review of photorefractive holograms in waveguides is presented in [95]. The significant new contribution of chapter 5 is the development of a method for control of complex holograms from above the waveguide. Simple chirped gratings have previously been recorded and etched on waveguides from above the guided plane [96]. Some of the analysis in chapter 5 was first presented in [97].

Chapter 6 returns to the use of volume holograms in artificial neural networks. The first section considers neural learning algorithms and their implementation in optical networks. Sections 6.2 and 6.3 describe two experimental systems, an outer-product based associative storage system and an adaptive optical perceptron. Section 6.2 is based on work presented in [98, 70]. The perceptron system of section 6.3 was previously considered in [60, 59, 57, 99]. Perceptron learning has since be implemented by Yoshinaga *et al.* [68], by Hong *et al.* [100], and by Paek *et al.* [101]. Chapter 6 concludes with a consideration of the relative advantages of optical and electronic interconnection technologies for implementations of neural processors. Some of the results described in section 6.4 were presented in [102].

2. THE RECONSTRUCTION OF VOLUME HOLOGRAMS

2.1 INTRODUCTION

While it may seem strange to describe hologram reconstruction before we deal with hologram formation, this approach is taken because an understanding of how information is extracted from a volume is needed to guide us in later chapters when we consider how to store information in it. Rather than review the many previous studies which have considered the reconstruction of volume holograms, we focus in this chapter on perspectives on the problem which will be particularly useful to our application. The second section of this chapter considers the reconstruction of a cascade of thin holograms. This system, which is fascinating in its own right, is analyzed because it is conceptually simpler than a continuous volume hologram and yet it exhibits quite similar properties. We do not mean to argue that the systems are totally equivalent, but an understanding of the properties of a cascade of thin holograms provides good intuition into the properties of volume holograms. This analogy will be particularly useful in chapter 3, where we consider the control of volume holograms. The third section of the present chapter considers the reconstruction of multiple grating volume holograms from the point of view of coupled wave theory. Multiple grating holograms are emphasized because the information storage capacity of a hologram in signal processing applications is proportional to the number of gratings the hologram can support. The final section of this chapter uses the results of the second and third sections to describe techniques by which volume holograms may be used to implement arbitrary linear transformations.

Three basic techniques have been applied to the problem of calculating the effect of volume holograms on their read out fields. Prior to the development of coupled wave theory, integration of the scattered field under the Born approximation was the most popular technique for analyzing volume diffraction [7,1]. This approach views the diffracted field as arising from weak diffraction from the undepleted read out beam. Integrating the diffracted signals from each voxel yields an expression for the total diffracted field. This intuitively satisfying approach is tractable only for weak coupling, when the Born approximation is appropriate. At high diffraction efficiencies this approximation fails. More recently, Langbein and Lederer have developed a more sophisticated version of this approach using Green's functions [103]. The second technique involves a decomposition of the boundary conditions imposed by the incident field into eigenmodes of the holographically perturbed medium [104- 106]. This technique allows for rigorous solutions in cases where the eigenmodes can be found. Describing the eigenmodes is very difficult in all but the simplest cases. The third technique is coupled wave theory [9]. This technique involves solving a set of coupled differential equations generated by substituting spatially modulated versions of the eigenmodes of the unperturbed system into the holographically perturbed differential wave equation.

In addition to the three basic categories described above, volume holographic coupling theories may be classified according to how they treat optical polarization and anisotropy. These issues can be dealt with within the general framework of the methods described above. While polarization effects may prove useful in the detailed design of information processing systems, such effects do not add substantially to the capabilities of volume holograms at the level of abstraction employed in much of this thesis. For this reason, polarization effects are not considered here. Excellent reviews of theories dealing with the reconstruction of volume holograms are provided in [107, 11, 12].

The analysis of a cascade of thin holograms presented in the next section is based on a discrete analogy to the eigenmodal approach. As discussed in chapter 1, this analysis is similar to the beam propagation method of [83, 82]. Because coupled wave theory is the most convenient method for analyzing the information processing systems described later in this thesis, the third section of this chapter concentrates on this approach.

2.2 A CASCADE OF THIN HOLOGRAMS

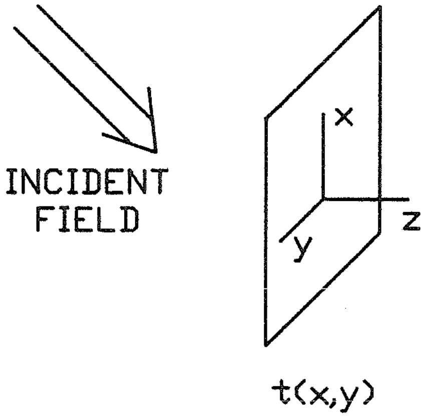


Figure 2.1. A thin hologram.

By "thin hologram" we mean a two-dimensional mask with transmittance function $t(x, y)$. A typical thin holographic system is sketched in Fig. 2.1. Under

Kichhoff boundary conditions [108], the optical field at the plane of the hologram, the plane $z = 0$ in Fig. 2.1, is the field incident on the hologram from the negative z axis multiplied by $t(x, y)$. Given the boundary conditions, the field at all points in the positive z half-space can be calculated. A volume hologram differs from a thin one in that its extent along the z axis is not negligible. Intuitively, a volume hologram can be regarded as a stack of thin holograms. Such a stack is sketched in Fig. 2.2. The stack consists of N cascaded transparencies. The transparencies lie in parallel planes spaced along the z axis by the distance l. The x and y axes lie in the plane of the first transparency. In this section we consider this stack in some detail. We find that a cascade of thin holograms exhibits the characteristic features of volume holograms, *i.e.*, Bragg diffraction and increasing diffraction efficiency with increasing overall thickness.

Absorption based volume holograms can have only limited diffraction efficiency [27] and are generally ill-suited to multiply exposed applications. For this reason, we assume throughout this thesis that we are dealing with holograms based primarily on perturbations to the real part of the index of refraction. Such "phase holograms" may be formed in thin media using, for example, gelatin or thermoplastic films. In our cascade of thin holograms we assume that the optical phase delay at (x, y) on the n^{th} hologram is given by the function $\phi_n(x, y)$. The transmittance of the n^{th} hologram is

$$t_n(x, y) = e^{j\phi_n(x,y)}. \tag{2.1}$$

We consider three functional forms for ϕ_n. In the first case, each hologram

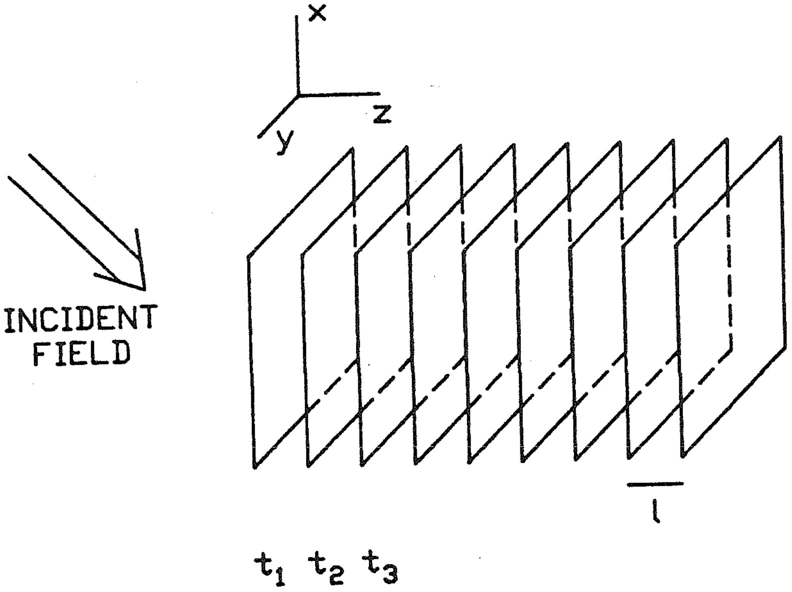


Figure 2.2. A cascade of thin holograms.

in the cascade is a sinusoidal grating, *i.e.*,

$$\phi_n = m \sin(\vec{K} \cdot \vec{\rho}) \tag{2.2}$$

$\vec{\rho}$ is a position vector in the x-y plane and $\vec{K}$ is the wavevector of the grating. The wavelength, orientation, amplitude and phase of the grating are the same in all of the holograms in the cascade. Our goal in considering this case is to determine the conditions under which the cascade implements a linear connection from a read out mode to exactly one diffracted mode. If these conditions hold, then when more complex holograms are stored in each layer of the cascade we can associate the component of the recorded holograms which is described by Eq. (2.2) with the connection between specific pairs of modes. We assume that these conditions hold in our analysis of the second two cases. The second case we consider is

$$\phi_n = m_n \sin(\vec{K} \cdot \vec{\rho} + \varsigma_n) \tag{2.3}$$

In this case, each hologram is a grating with the same grating wavevector, but the amplitudes and phases of the gratings vary. Analysis of this case illustrates the following important concepts:

1. If m_n is constant over n and $|\varsigma_n|$ grows linearly with n then the cascade implements a linear connection between an input mode propagating with wavevector $\vec{k}_r$ and an output mode propagating with wavevector $\vec{k}_s$ if and only if

$$\vec{k}_r - \vec{k}_s = \vec{K} + \frac{(\varsigma_n - \varsigma_o)}{nl}\hat{z}, \tag{2.4}$$

   where $\hat{z}$ is a unit vector along the z axis. It is important to note that $\vec{K}$ is confined to the x-y plane. If these conditions are satisfied, we say that the "distributed hologram" which connects the r^{th} input mode to the s^{th} output mode is recorded in the cascade of thin holograms. In the same vein, each thin transparency is a "local hologram."

2. We refer to

$$\vec{K}_g = \vec{k}_r - \vec{k}_s = \vec{K} + \frac{(\varsigma_n - \varsigma_o)}{nl}\hat{z} \tag{2.5}$$

   as "the grating wavevector" of the corresponding distributed hologram. Similarly, we refer to $\vec{K}$ as the grating wavevector of the local holograms. Two distributed grating wavevectors, $\vec{K}_g$ and $\vec{K}_{g'}$ are called "distinct" if there exists no pair of read out and diffracted modes which are coupled by both grating wavevectors.

3. For fixed $\vec{K}$, there exist exactly N mutually distinct grating wavevectors. This means that there exist exactly N distinct distributed holograms.

4. Returning to ϕ_n with arbitrary amplitude and phase as in Eq. (2.3), we derive a transformation which relates the amplitudes and phases of the distributed holograms stored in the cascade with the amplitudes and phases of the local holograms. Since each distributed hologram corresponds to a specific connection between pairs of modes, knowing the amplitudes and phases of the distributed holograms allows us to express the transformation implemented by the cascade.

The third case we consider is

$$\phi_n = \sum_{\vec{K}} \sum_{g=0}^{N-1} m_{ng} \sin(\vec{K}_g(\vec{K}) \cdot (\vec{\rho} + Nl\hat{z})). \tag{2.6}$$

In this case, the local holograms are a superposition of the gratings corresponding to all the local grating wavevectors in a suitable range with arbitrary amplitude and phase. The range of the local grating wavevectors is constrained by the fact that we limit ourselves to considering interconnections between a set of read out modes controlled by an input SLM and a set of diffracted modes incident on an output detector array. As suggested by Eq. (2.6), the transformation implemented by the cascade in this case can be described by decomposing the transmittances of the local holograms in terms of the distributed holograms. In summary, our goals in this section are:

1. To derive, in subsection 2.2.1, the sense in which a distributed grating may be considered a linear connection between two modes,

2. To develop, in subsection 2.2.2, the correspondence between the local transmittances and the distributed holograms, and
3. To describe, in subsection 2.2.3, how multiple distributed holograms can be superposed to form an interconnection network.

2.2.1 A phase-locked cascade of sinusoidal gratings

Suppose that the amplitude transmittance of each transparency in a cascade is

$$t(x,y) = e^{jm\sin(u_o x)}. \tag{2.7}$$

Such transparencies may be formed by perturbing the index of refraction, $n(x,y)$, of a dielectric sheet of thickness δ such that

$$n(x,y) = 1 + (n_o - 1)(1 + \alpha\sin(u_o x)), \tag{2.8}$$

where n_o is the unperturbed index of refraction, and $-1 \geq \alpha \leq 1$. The phase delay at (x,y) experienced by a monochromatic field of wavelength λ propagating through the sheet is

$$\phi(x,y) = n(x,y)k_o\delta = \phi_o(1 + \alpha(1 - \frac{1}{n_o})\sin(u_o x)), \tag{2.9}$$

where $k_o = 2\pi/\lambda$ and $\phi_o = n_o k_o \delta$. Discarding the constant term in $\phi(x,y)$, we find that this phase delay yields the transmittance of Eq. (2.7) if we define $m = \alpha(1 - \frac{1}{n_o})\phi_o$. To maintain the analogy with the usual volume holographic case, we will assume that α is very small, specifically that $\alpha \sim o(1/N)$. To insure that propagation through each transparency can be considered in the geometric regime, we assume that ϕ_o is of order 1. This implies that $m \sim o(1/N)$.

We wish to determine the field diffracted from the system of Fig. 2.2 when a plane wave of the form

$$U_o(x,z) = A_o e^{jk_o(z\cos\theta + x\sin\theta)}, \tag{2.10}$$

is incident. When $U_o(x,z)$ is incident on a thin transparency with transmittance $t(x)$ placed at $z=0$, the field at $z=0$ on the positive side of the transparency is

$$U^+ = U_o(x,0)t(x) = A_o e^{jk_o x\sin\theta} e^{jm\sin(u_o x)}. \tag{2.11}$$

By expanding $t(x)$ in the Fourier series

$$t(x) = \sum_{q=-\infty}^{\infty} J_q(m) e^{jqu_o x}, \tag{2.12}$$

where $J_q(m)$ is the q^{th} order Bessel function of the first kind [109], Eq. (2.11) can be rewritten

$$U^+ = A_o \sum_{q=-\infty}^{\infty} J_q(m) e^{j(k_o\sin\theta + qu_o)x}. \tag{2.13}$$

Eq. (2.13) is a boundary condition from which we can calculate the field in the positive half space. If we assume, as is reasonable for the present problem, that we are interested in the field for z less than the size of the aperture of the thin hologram, then it can be shown that each Fourier component of Eq. (2.13) gives rise to a plane wave. The field in this regime is

$$U^+(x,z) = A_o \sum_{q=-\infty}^{\infty} J_q(m) e^{j\vec{k}^{(q)}\cdot\vec{r}}, \tag{2.14}$$

where

$$\begin{aligned} k_z^{(q)} &= \sqrt{k_o^2 - (k_o\sin\theta + qu_o)^2} \\ k_x^{(q)} &= k_o\sin\theta + qu_o. \end{aligned} \tag{2.15}$$

The q^{th} order diffracted off of the first transparency produces the field

$$U_q^+(x,l) = A_o J_q(m) \sum_{q'=-\infty}^{\infty} J'_q(m) e^{jk_z^{(q)}l} e^{j(k_x^{(q)}+q'u_o)x}. \tag{2.16}$$

on the positive side of the second transparency. This boundary condition produces plane waves propagating at the spatial frequencies

$$\begin{aligned} k_z^{(q')} &= \sqrt{|\vec{k}^{(q)}|^2 - (k_x^{(q)} + q'u_o)^2} \\ k_x^{(q')} &= k_x^{(q)} + q'u_o. \end{aligned} \tag{2.17}$$

Substituting from Eq. (2.15), Eq. (2.17) becomes

$$\begin{aligned} k_z^{(q')} &= \sqrt{k_o^2(k_o \sin\theta + (q+q')u_o)^2} \\ k_x^{(q')} &= k_o \sin\theta + (q+q')u_o. \end{aligned} \tag{2.18}$$

Thus the second transparency maps the spectrum of plane waves described by Eq. (2.15) back onto itself. Since the x components of the wavevectors of these plane waves always consist of harmonics of u_o on the carrier $k_o \sin\theta$, the field at any z may be specified by a vector of the form

$$\vec{U}(z) = \begin{pmatrix} \vdots \\ U_q \\ \vdots \\ U_1 \\ U_o \\ U_{-1} \\ \vdots \\ U_{-q} \\ \vdots \end{pmatrix} \tag{2.19}$$

where U_q is the field amplitude of the plane wave with wavevector $\vec{k}^{(q)}$.

Each transparency implements the transformation

$$T_1 = \begin{pmatrix} & \ddots & & & & \\ \cdots & J_{-1}(m) & J_o(m) & J_1(m) & \cdots & \\ & \cdots & J_{-1}(m) & J_o(m) & J_1(m) & \cdots \\ & & \cdots & J_{-1}(m) & J_o(m) & J_1(m) \\ & & & & & \ddots \end{pmatrix} \tag{2.20}$$

on $\vec{U}(z)$. Propagation from one transparency to the next implements an additional transformation of the form

$$T_2 = \begin{pmatrix} \ddots & & & & & \\ & p_q & & & 0 & \\ & & \ddots & & & \\ & & & p_o & & \\ & & & & \ddots & \\ & 0 & & & p_{-q} & \\ & & & & & \ddots \end{pmatrix} \tag{2.21}$$

where

$$p_q = e^{jk_z^{(q)}l}. \tag{2.22}$$

After N transparencies the field is

$$\vec{U}(N) = (T_2T_1)^N\vec{U}(0) \tag{2.23}$$

where $\vec{U}(0)$ is the field incident on the first transparency.

We can expand $\vec{U}(0)$ in the eigenvectors of the transformation

$$T = T_2T_1 \tag{2.24}$$

such that

$$\vec{U}(0) = \sum_n \alpha_n\vec{v}_n, \tag{2.25}$$

where

$$T\vec{v}_n = \lambda_n \vec{v}_n \tag{2.26}$$

and α_n is a constant for each n. Then

$$\vec{U}(N) = \sum_n T^N \alpha \vec{v}_n = \sum_n \lambda_n^N \alpha_n \vec{v}_n. \tag{2.27}$$

Of course, for T_1 as given by Eq. (2.20), *i.e.*, when T_1 is of infinite dimension, this approach is unreasonable. Fortunately, it is possible to dramatically curtail the dimensionality of T_1 by considering Bragg matching constraints and the scaling of the various diffracted orders with m.

Suppose, for example, that we wish to couple light from the zeroth order to the q^{th} order. This might be accomplished either by direct diffraction via the q^{th} harmonic of the local gratings or by diffraction through several intermediate orders. Consider light diffracted into the q^{th} order via C intermediate orders. The phase of the field diffracted into the q^{th} order by the $(C+1)^{th}$ transparency is

$$\varphi_C = \sum_{i=1}^{C} k_z^{(i)} l, \tag{2.28}$$

where k^i is the wavevector of the i^{th} intermediate order. The phase of the field which is undiffracted by the first transparency and which then follows this path to the q^{th} order at the $(C+2)^{th}$ transparency is

$$\varphi_{C+1} = k_o \cos\theta l + \sum_{\{i\}} k_z^{(i)} l. \tag{2.29}$$

Assuming that the individual contribution of each transparency is small, the amplitude of the field diffracted onto the q^{th} order can grow only if each addition to the field is in phase with the existing field. Since the change in the phase of the existing field from the $(C+1)^{th}$ transparency to the $(C+2)^{th}$ transparency is $k_z^{(q)}l$, the field diffracted by the $(C+1)^{th}$ transparency is in phase with that diffracted by the $(C+2)^{th}$ transparency if $\varphi_C + k_z^{(q)}l = \varphi_{C+1}$, or, equivalently, if

$$p_q \prod_i p_i = p_o \prod_i p_i. \tag{2.30}$$

Thus, independent of the intermediate steps, the zeroth and q^{th} orders can be coupled effectively only if $p_q = p_o$.

$J_{\pm q}(m)$ is of order $|q|$ in m [110]. Thus the signal diffracted directly from the zeroth order to the q^{th} is proportional to $m^{|q|}$. A signal diffracted from the zeroth order through intermediate orders to the q^{th} order is of order $\prod_{\{i\}} m^{|i|} \leq m^{|q|}$. This means that even with proper phase matching, higher order diffraction will be very weak if m is small. Of course, if an intermediate order, r, could grow significantly before coupling to the q^{th} order then this problem could be overcome. This would require simultaneous phase matching from the zeroth order to the r^{th} order and from the r^{th} order to the q^{th} order. Since phase matching is transitive, the zeroth order would also be phase matched to the q^{th} order.

In order to analyze the phase matching requirements in more detail, consider the case in which the readout beam is undepleted. Of course, if strong coupling occurs this condition will ultimately be violated. Assuming, however, that diffraction occurs gradually, an analysis of the undepleted regime allows us to quantify the phase matching requirements. The field diffracted into the q^{th} order at the n^{th} transparency under the undepleted assumption is proportional to $p_o^n J_q(m) A_o$. After N transparencies the field in the q^{th} order is

$$\begin{aligned} U_q(N) &= \sum_{n=1}^{N} p_q^{N-n} p_o^n J_q(m) A_o \\ &= J_q(m) A_o p_q^N \left[\frac{e^{j\psi_q} + e^{j(N+1)\psi_q}}{1 - e^{j\psi_q}} \right], \\ &= J_q(m) A_o p_o^{\frac{N-1}{2}} p_q^{\frac{N+1}{2}} \left[\frac{\sin \frac{N\psi_q}{2}}{\sin \frac{\psi_q}{2}} \right] \end{aligned} \tag{2.31}$$

where $e^{j\psi_q} = p_o p_q^*$, *i.e.*, $\psi_q = (k_o \cos\theta - k_z^{(q)})l$. The function $\sin \frac{N\psi_q}{2} / \sin \frac{\psi_q}{2}$ is approximately zero unless

$$|\psi_q - 2\pi c| \leq \pi / N \tag{2.32}$$

for some integer c. If this condition is satisfied, $U_q(N)$ grows linearly with N.

Substituting $k_z^{(q)}$ from Eq. (2.15), we find

$$\psi_q = k_o l \cos\theta - \sqrt{k_o^2 l^2 \cos^2\theta - 2qu_o k_0 l^2 \sin\theta - q^2 u_o^2 l^2} \tag{2.33}$$

Note that $\psi_q = 0$ when

$$\sin\theta = -\frac{q\lambda}{2\Lambda}, \tag{2.34}$$

where $\Lambda = 2\pi / u_o$. Eq. (2.34) is the Bragg condition which appears in volume holography, x-ray diffraction, and various other fields. For volume holograms, Eq. (2.34) is the only condition under which exact phase matching can occur. For the cascade of thin holograms, additional solutions may be generated by substituting Eq. (2.33) into Eq. (2.32) . If we assume that $k_o l \gg 1$ and $\cos\theta \approx 1$,

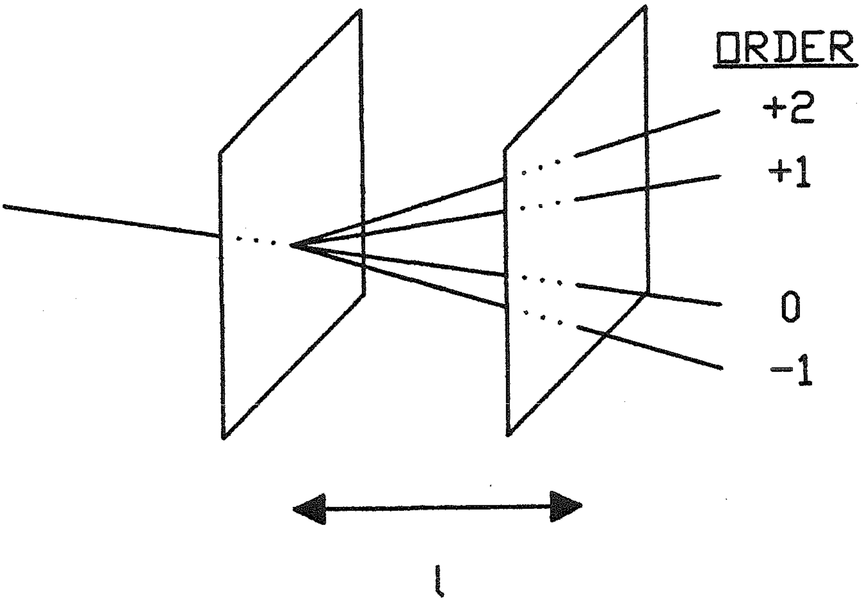


Figure 2.3. Phase matched orders for $\sin\theta = -\lambda/2\Lambda$, $\gamma = 2$, $c = 0$, $c' = 1$. The z-component of the wavevectors for the $+2$ and -1 order modes is $2\pi/l$ less than the z-component of the wavevectors for modes 0 and $+1$.

these solutions take the form

$$\sin\theta = -\frac{q\lambda}{2\Lambda} + \frac{c\Lambda}{ql}. \tag{2.35}$$

These solutions are precluded in a volume hologram because Eq. (2.32) must be satisfied for arbitrary l in a continuous medium. Note that in some cases θ can be found such that Eq. (2.35) is simultaneously satisfied for several values of q

Given θ, c and q satisfying Eq. (2.35), it can be shown that the q'^{th} order where

$$q' = \frac{1}{2}\left[(q - \frac{c}{q}\gamma) \pm \sqrt{(q - \frac{c}{q}\gamma)^2 + 4c'\gamma}\right] \tag{2.36}$$

and $\gamma = 2\Lambda^2/\lambda l$, is also phase matched at θ. c and c' are integer constants. This condition is illustrated for $\sin\theta = -\lambda/2\Lambda$ and $\gamma = 2$ in Fig. 2.3. As shown in the figure, Eq. (2.35) is satisfied when the optical path along the z axis between transparencies for the q^{th} order differs by an an integral multiple of 2π from the optical path for the 0^{th} order. For the parameters selected, this condition is satisfied for $q = \pm 1, 2$.

It is possible to simultaneously couple from the zeroth order mode to more than one higher order mode in two ways. The first method, true phase matching based on identical phase shifts between transparencies, has no analog in continuous volume holography. This is the method described in the previous paragraph. A second means by which coupling can occur to more than one diffracted mode arises when the cascade is not "thick" enough to prevent coupling between adjacent modes. The simplest case arises when $\psi_{-1} - \psi_1 < 2\pi/N$. If $\psi_1 = 0$ then,

$$\psi_{-1} = k_o l \cos\theta(1 - \sqrt{1 - 2(\frac{u_o}{k_o \cos\theta})^2}) \approx \frac{u_o^2 l}{k_o \cos\theta}, \tag{2.37}$$

where we assume $\Lambda \gg \lambda$. The +1 and −1 orders are both phase matched if $\psi_{-1} < 2\pi/N$, which is the case if

$$Nl < \frac{2\Lambda^2}{\lambda}\cos\theta. \tag{2.38}$$

This case corresponds to the Raman-Nath regime for volume holography.

Diffraction from the zeroth order to the first order is of particular interest. We assume N is large enough that Eq. (2.38) is violated and that $\sin\theta = -\lambda/2\Lambda$. In this case only the first diffracted order and associated solutions of Eq. (2.36) are phase matched. If γ is greater than 2 then all such solutions are at least two orders away from the zeroth and first orders and can thus be ignored in light of the small size of m. The appropriateness of discarding higher orders can be verified by considering the fact that the total power in all these higher orders cannot be greater than $[N\sum_{q=2}^{\infty} m^q]^2 \approx N^2m^4$. The transformation T may in this case be collapsed to the form

$$T = \begin{pmatrix} J_o(m)p & J_1(m)p \\ J_{-1}(m)p & J_o(m)p \end{pmatrix} \tag{2.39}$$

where $p_1 = p_o = p$. The eigenvalues of this matrix are

$$\lambda_\pm = p(J_o(m) \pm j J_1(m)) \tag{2.40}$$

or

$$\lambda_\pm = e^{jk_o l\cos\theta} e^{\pm j \arctan\frac{J_1(m)}{J_o(m)}} \tag{2.41}$$

where we make use of the fact that to first order in m

$$J_o^2(m) + J_1^2(m) = 1 \tag{2.42}$$

The eigenvectors corresponding to $\lambda_\pm$ are

$$\vec{v}_\pm = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ \pm j \end{pmatrix}. \tag{2.43}$$

Assuming that

$$\vec{U}(0) = \begin{pmatrix} 0 \\ A_o \end{pmatrix} \tag{2.44}$$

we find

$$\vec{U}(0) = -j\frac{A_o}{\sqrt{2}}[\vec{v}_+ - \vec{v}_-] \tag{2.45}$$

and

$$\vec{U}(N) = -j\frac{A_o}{\sqrt{2}}[\lambda_+^N \vec{v}_+ - \lambda_-^N \vec{v}_-]. \tag{2.46}$$

Thus

$$\begin{aligned}\vec{U}(N) &= A_o p^N \begin{pmatrix} \frac{e^{jN\arctan\frac{J_1(m)}{J_o(m)}} - e^{-jN\arctan\frac{J_1(m)}{J_o(m)}}}{2j} \\ \frac{e^{jN\arctan\frac{J_1(m)}{J_o(m)}} + e^{-jN\arctan\frac{J_1(m)}{J_o(m)}}}{2} \end{pmatrix} \\ &= A_o e^{jk_olN\cos\theta} \begin{pmatrix} \sin(N\arctan\frac{J_1(m)}{J_o(m)}) \\ \cos(N\arctan\frac{J_1(m)}{J_o(m)}) \end{pmatrix}.\end{aligned} \tag{2.47}$$

To lowest order in m,

$$\vec{U}(N) = A_o e^{jk_olN\cos\theta} \begin{pmatrix} \sin(\frac{Nm}{2}) \\ \cos(\frac{Nm}{2}) \end{pmatrix}. \tag{2.48}$$

The intensity of the light in the zeroth order mode after the cascade is

$$|A_o|^2\cos^2(\frac{Nm}{2}).$$

The intensity of the light in the first order mode is $|A_o|^2\sin^2(\frac{Nm}{2})$. If $Nm = \pi$ the diffraction efficiency of the system reaches 100 percent to first order in m. Since we would like the diffraction to be linear, we will assume that m is much less than π/N, in which case the diffracted field is linear in m and the diffraction

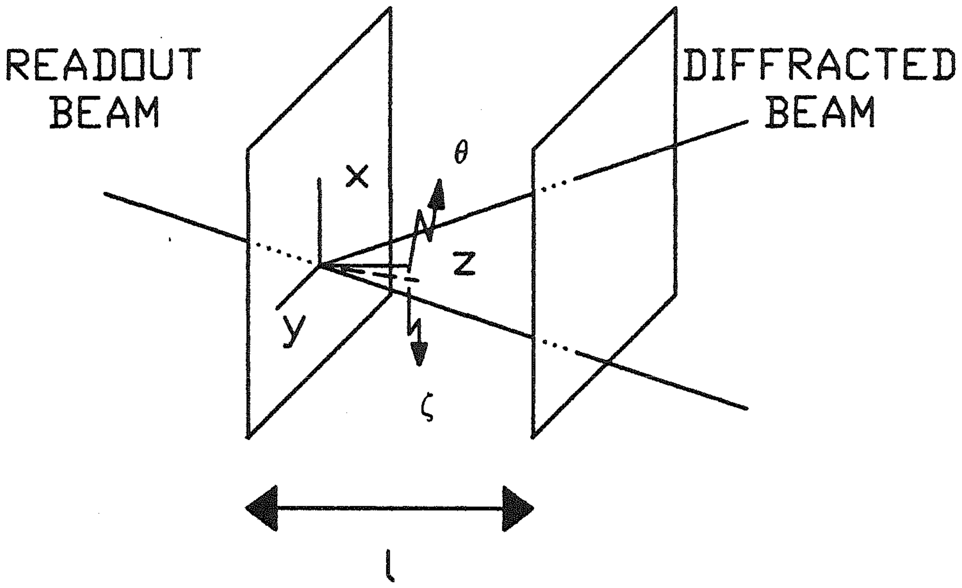


Figure 2.4. Read out geometry in the general case.

efficiency for the intensity is approximately $N^2m^2/4$. These results are, of course, quite similar to results describing two beam coupling in volume holography.

The observant reader will have noticed that while we promised to consider an arbitrary local grating vector in the x-y plane in this subsection, we have thus far considered only a grating wavevector oriented along the x-axis. Since we can always define an x axis parallel to the local grating wavevector, the problem we have thus far considered does, in fact, correspond to the general case, except that the propagation vector of the read out field need not lie in the same plane as the grating wavevector and $\hat{z}$. To analyze the most general case we must add a $\hat{y}$ component to $\vec{k}^{(0)}$. This situation may be analyzed most simply if we define ζ, the angle between the projection of $\vec{k}^{(0)}$ in the z-y plane and the z axis. The geometry of the read out beam is sketched in Fig. 2.4. The incident field can be expressed as

$$U_o(x,y,z) = A_o e^{jk_o(z\cos\theta\cos\zeta + x\sin\theta + y\cos\theta\sin\zeta)}. \tag{2.49}$$

$\frac{\pi}{2} - \theta$ is the angle between $\vec{k}^{(0)}$ and the x axis. If we define new coordinate axes, x', y', and z', to correspond to the original axes rotated about the x axis by $-\zeta$, U_o may be expressed as

$$U_o(x, y, z) = A_o e^{jk_o(z'\cos\theta + x'\sin\theta)}. \tag{2.50}$$

Since x' is identical to x, the description of the transmittance of the local holograms given by Eq. (2.12) is unchanged in the new coordinate system. This means that, just as was shown above, the holograms in the cascade diffract the incident field into a spectrum of plane waves which can be described by a vector $\vec{U}(z)$ with components U_q, where U_q is the amplitude of the plane wave propagating with wavevector

$$\begin{aligned} k_{z'}^{(q)} &= \sqrt{k_o^2 - (k_o \sin\theta + qu_o)^2} \\ k_{x'}^{(q)} &= k_o \sin\theta + qu_o. \end{aligned} \tag{2.51}$$

In the x, y, z coordinate system the wavevector of the q^{th} plane wave is

$$\begin{aligned} k_z^{(q)} &= \cos\zeta\sqrt{k_o^2 - (k_o \sin\theta + qu_o)^2} \\ k_x^{(q)} &= k_o \sin\theta \cos\zeta + qu_o \\ k_y^{(q)} &= \sin\zeta\sqrt{k_o^2 - (k_o \sin\theta + qu_o)^2}. \end{aligned} \tag{2.52}$$

Just as above, coupling from the zeroth order mode to the q^{th} order can be strong only if the change in phase between transparencies for the two modes differs by an integral product of 2π. Since oblique incidence reduces the optical path between transparencies along the z axes the phase matching condition given by Eq. (2.35), must be modified. The new condition is

$$\sin\theta = -\frac{q\lambda}{2\Lambda} + \frac{c\Lambda}{ql\cos\zeta}. \tag{2.53}$$

Note that the phase matching condition for $c = 0$ is unchanged. If $\sin\theta =$

$-q\lambda/2\Lambda$, phase matching is maintained as the holographic system is rotated about the grating wavevector. This fact results in "shift invariance" normal to the grating wavevector. The effects of this property on information storage systems is discussed below.

In this subsection we have considered diffraction from a cascade of sinusoidal gratings with locked phases and amplitudes. While we have assumed that the local grating wavevector lies along the x axis, we have considered an arbitrary incidence angle for the read out field. The case of arbitrary grating orientation and arbitrary incidence can be solved by simple rotation of the coordinate axes. We have found that if Nl is large enough that Eq. (2.38) is violated, $\gamma > 2$, and $Nm \ll \pi$ then the cascade couples a Bragg matched incident beam to only one diffracted order. In this case, the cascade of thin local holograms implements a "distributed hologram" which connects a specific class of inputs to specific outputs. Due to shift invariance normal to the grating wavevector, the number of inputs which can be diffracted by the system is large. In section 2.4 we consider methods for constraining the inputs and outputs so that the connection formed by a distributed hologram can be associated with a specific pair of input and output modes.

2.2.2 A cascade of sinusoidal gratings with variable amplitude and phase

In this subsection we consider the case

$$t_n(x,y) = e^{jm_n \sin(\vec{K}\cdot\vec{\rho}+\varsigma_n)}, \tag{2.54}$$

We can determine the diffraction properties of this system by writing the perturbation at each hologram in terms of the perturbations associated with holograms with known diffraction properties. Consider, for example, the hologram which couples

$$R(\vec{r}) = R_o e^{j(\vec{k}_r \cdot \vec{r})} \tag{2.55}$$

and

$$S(\vec{r}) = S_o e^{j(\vec{k}_s \cdot \vec{r})}, \tag{2.56}$$

where $|\vec{k}_r| = |\vec{k}_s| = k_o$. This hologram is formed by interfering the two signals in a linear medium, where by "linear medium" we mean a material in which the index of refraction changes in proportion to the recording intensity. The recording intensity in this case is

$$I(\vec{r}) = |R_o|^2 + |S_o|^2 + 2|R_o||S_o| \sin(\vec{K}_g \cdot \vec{r} + \varphi_{RS}), \tag{2.57}$$

where $\vec{K}_g = \vec{k}_r - \vec{k}_s$ and $\varphi_{RS} - \pi/2$ is the difference between the phases of R_o and S_o. We may assume without loss of generality that $\varphi_{RS} = 0$. If we use this intensity to record the local holograms in a cascade, the index of refraction at the n^{th} transparency becomes

$$n(x, y) = 1 + (n_o - 1)(1 + \alpha \sin(\vec{K} \cdot \vec{\rho} + nK_{gz}l)), \tag{2.58}$$

where $\vec{K}$ is the component of $\vec{K}_g$ in the x-y plane and K_{gz} is the z component of $\vec{K}_g$. The transmittance of the n^{th} hologram is

$$\begin{aligned} t_n(x, y) &= e^{jm \sin(\vec{K} \cdot \vec{\rho} + nK_{gz}l)} \\ &= \sum_{q=-\infty}^{\infty} J_q(m) e^{jq(\vec{K} \cdot \vec{\rho} + nK_{gz}l)}. \end{aligned} \tag{2.59}$$

Eq. (2.59) is equivalent to Eq. (2.7) except that the phase of the grating on the transparencies increases linearly with n.

As described in subsection 2.2.1, coupling from one mode to another is possible only if the field diffracted at each transparency is in phase with the field which is already in the diffracted order. With the additional phase shift at each hologram the phase matching condition given by Eq. (2.32) becomes

$$|\psi_q - qK_{gz}l - 2\pi c| \leq \pi/N. \tag{2.60}$$

For first order coupling between $R(\vec{r})$ and $S(\vec{r})$, $\psi_1 = K_{gz}l$ and Eq. (2.60) is satisfied. Eq. (2.60) may also be satisfied for higher order beams, but such coupling is very weak if the conditions for two beam coupling described in the previous subsection are satisfied. For the rest of this section we assume that the amplitudes of the the holograms are small enough that we need only consider first order couplings.

In a single transparency, a grating wavevector couples an incident beam with wavevector $\vec{k}^{(i)}$ to a diffracted beam with wavevector $\vec{k}^{(i)} + \vec{K}$. According to Eq. (2.60), incident light with wavevector $\vec{k}^{(i)}$ is diffracted by a cascade of thin holograms to light propagating with the wavevector

$$\vec{k}^{(j)} = \vec{k}^{(i)} + \vec{K} + (K_z + 2\pi c/l)\hat{z} \tag{2.61}$$

This diffraction can occur only if $\vec{k}^{(j)}$ corresponds to a propagating mode for the wavelength of the read out light. The endpoints of the wavevectors of the propagating modes for a given wavelength lie on a spheroid (or two connected spheroids in the anisotropic case) known as the wave normal surface. For a homogeneous isotropic system the normal surface is a sphere of radius k. According to Eq. (2.61), a pair of modes is coupled if their wavevectors are separated on the normal surface by $\vec{K}_g + (2\pi c/l)\hat{z}$.

As defined in the introduction to this section, two different grating wavevectors are distinct if they do not couple the same pairs of optical beams. Eq. (2.60) allows us to quantize this definition. If a pair of modes satisfies Eq. (2.60) for the grating wavevector $\vec{K}_g$, then the equation cannot be satisfied for the grating wavevector $\vec{K}_{g'}$ if

$$|[\vec{K}_g - \vec{K}_{g'}]_z l - 2\pi c| > \frac{2\pi}{N} \tag{2.62}$$

for all integral values of c. The maximum number of vectors $\vec{K}_g$ for which this condition can be satisfied is N. A set of N wavevectors which are mutually distinct in the sense of Eq. (2.62) is given by

$$\vec{K}_n = \vec{K} + \frac{2n\pi}{Nl}\hat{z}, \tag{2.63}$$

where n is an integer between 0 and $N-1$. This set of grating wavevectors is sketched on the normal surface in Fig. 2.5. By construction, no two of these grating wavevectors can couple the same pair of beams.

At this point it is useful to clarify the distinction between "local" and "distributed" holograms. By the n^{th} local hologram we mean the transmittance of the n^{th} transparency. By the n'^{th} distributed hologram we mean the component of the transmittances of the local holograms which is described by

$$t_n(x,y) = e^{jm\sin(\vec{K}\cdot\vec{\rho}+n\vec{K}_{n'z}l)} \tag{2.64}$$

The N local holograms can be described in terms the N distributed holograms, as we now show. Suppose that the N distinct distributed holograms have been linearly superposed in the cascaded local hologram system. Let $a_{n'}$ and $\zeta_{n'}$ represent

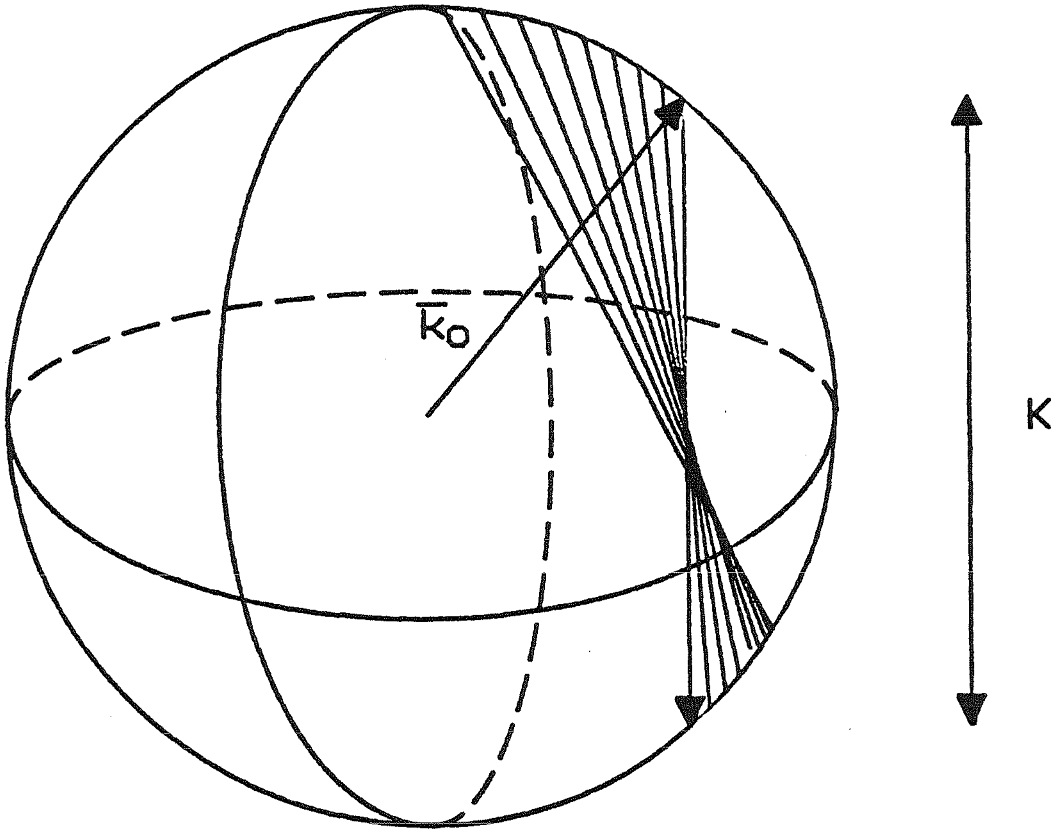


Figure 2.5. Wave normal surface and grating wavevectors. l/λ is assumed to be 2.

the amplitude and phase with which the n'^{th} distributed hologram is recorded. Ignoring higher order terms under the approximations derived in subsection 2.2.1, the local hologram stored in the n^{th} transparency is

$$t_n = \sum_{n'=1}^{N'} a_{n'}(1 + jm\sin(\vec{K}\cdot\vec{\rho} + \frac{2\pi nn'}{N} + \zeta_{n'})), \tag{2.65}$$

where m is a constant of order $1/N$. We relate the amplitudes of the local and distributed holograms by comparing Eq. (2.65) with Eq. (2.54). Letting $b_n = -j(m_n/m)e^{j\varsigma_n}$, where m_n and ς_n are the amplitude and phase of the n^{th}

local hologram, we find

$$\vec{B} = \overline{\overline{P}}\vec{A}, \tag{2.66}$$

where

$$\vec{B} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix}, \tag{2.67}$$

$$\vec{A} = \begin{pmatrix} a_0 e^{j\zeta_0} \\ a_1 e^{j\zeta_1} \\ a_2 e^{j\zeta_2} \\ \vdots \\ a_N e^{j\zeta_N} \end{pmatrix}, \tag{2.68}$$

and

$$\overline{\overline{P}} = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & e^{j\frac{2\pi}{N}} & e^{j\frac{4\pi}{N}} & \dots & e^{j\frac{2\pi(N-1)}{N}} \\ 1 & e^{j\frac{4\pi}{N}} & e^{j\frac{8\pi}{N}} & \dots & e^{j\frac{4\pi(N-1)}{N}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{j\frac{2\pi(N-1)}{N}} & e^{j\frac{4\pi(N-1)}{N}} & \dots & e^{j\frac{2\pi(N-1)^2}{N}} \end{pmatrix} \tag{2.69}$$

$\overline{\overline{P}}$ is a Vandermonde matrix. Matrices of this form are known to be nonsingular. By inverting $\overline{\overline{P}}$, $\vec{A}$ corresponding to arbitrary $\vec{B}$ can be found.

To determine the diffraction properties of an arbitrary series of gratings with spatial frequency $\vec{K}$, we first use $t(x, y)$ as expressed in Eq. (2.54) to determine $\vec{B}$. We then find $\vec{A}$ using $\vec{A} = \overline{\overline{P}}^{-1}\vec{B}$. A readout signal can then be decomposed into plane wave components phase matched to each of the N distributed holograms described by $\vec{A}$. As described in subsection 2.2.1, each distributed hologram corresponds to a linear connection from an input mode to an output mode. The strength of the connection between the fields is $a_n m$. The diffraction efficiency of the connection is $|a_n|^2(Nm)^2/4$.

2.2.3 A cascaded hologram interconnection network

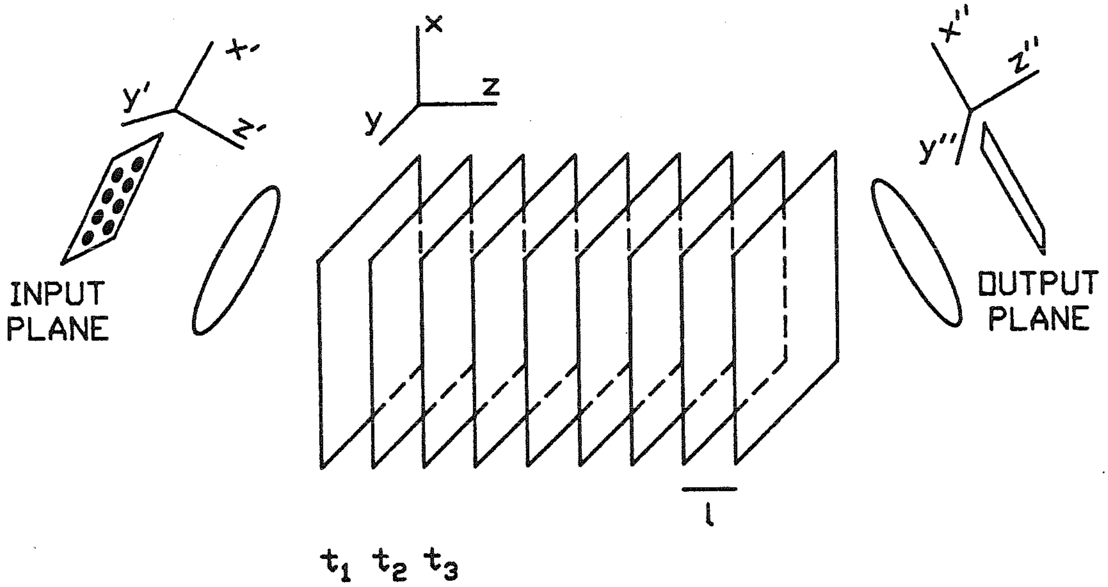


Figure 2.6. Thin hologram cascade optical interconnection system.

As described in the introduction to this section, we consider the case

$$\phi_n = \sum_{\vec{K}} \sum_{g=0}^{N-1} m_{ng} \sin(\vec{K}_g(\vec{K}) \cdot (\vec{\rho} + Nl\hat{z})). \tag{2.70}$$

in this subsection. Since we are primarily interested in systems in which a very large number of distributed holograms are recorded for use as linear interconnections, the range of distributed grating wavevectors $\vec{K}_g$ which we wish to consider is limited by the architecture of the interconnection system. A typical architecture for such a system is sketched in Fig. 2.6. A cascade of thin holograms is addressed by light generated on an input 2-D spatial light modulator (SLM). The thin holograms are spaced by l along the z axis. The origin of the z axis is in the center of the first hologram, which defines the x-y plane. The input plane lies in the x' y' plane. The z' direction is normal to the input plane. The primed coordinate system corresponds to the unprimed coordinate system rotated by an angle θ about the y axis. Readout signals are generated by a 2-D pixel array on the input plane. The pixels are arranged on a regular 2-D grid with a separation between pixels of Δ. Each pixel occupies an area δ^2. The signals generated by the pixel array are collimated by a lens of focal length F. The lens lies in a plane parallel to the x'-y' plane. The holograms diffract light from the input signals onto a set of output signals. The output signals which exit the final thin hologram are focused onto a detector array by a second lens, also of focal length F. The output plane lies in the x''-y'' plane. The z'' direction is normal to this plane. The double primed coordinate system corresponds to the unprimed coordinate system rotated by an angle $-\theta$ about the y axis. The detectors are also arranged on a 2-D grid and spaced by Δ. The apertures of the holograms, the SLM, and the detector array are of length L on a side.

If λ/δ is smaller than L/F, then the field generated in the volume by a single pixel is approximately a plane wave. The amount by which $\lambda F/\delta L$ must exceed one depends on the effective aperture of the holograms, which depends on θ. If $L = Nl$, for example, it can be shown that the plane wave approximation holds if

$$\sin\theta < \frac{((\frac{\lambda F}{\delta L})^2 - 1)}{((\frac{\lambda F}{\delta L})^2 + 1)}. \tag{2.71}$$

The plane wave generated by the pixel at $x' = i'\Delta$, $y' = j'\Delta$ propagates with the wavevector

$$\begin{aligned}\vec{k}_{i'j'} &= \frac{i'k_o\Delta}{F}\hat{x}' + \frac{j'k_o\Delta}{F}\hat{y}' + k_o\big[1 - (\frac{i'\Delta}{F})^2 - (\frac{j'\Delta}{F})^2\big]^{\frac{1}{2}}\hat{z}' \\ &= (\frac{i'k_o\Delta}{F}\cos\theta - k_o\big[1 - (\frac{i'\Delta}{F})^2 - (\frac{j'\Delta}{F})^2\big]^{\frac{1}{2}}\sin\theta)\hat{x} + \frac{j'k_o\Delta}{F}\hat{y} \\ &\quad + (\frac{i'k_o\Delta}{F}\sin\theta + k_o\big[1 - (\frac{i'\Delta}{F})^2 - (\frac{j'\Delta}{F})^2\big]^{\frac{1}{2}}\cos\theta)\hat{z}\end{aligned} \tag{2.72}$$

Simlarly, the signal which is detected at the pixel at $x'' = i''\Delta$, $y'' = j''\Delta$ propagates with the wavevector

$$\begin{aligned}\vec{k}_{i''j''} &= \frac{i''k_o\Delta}{F}\hat{x}'' + \frac{j''k_o\Delta}{F}\hat{y}'' + k_o\big[1 - (\frac{i''\Delta}{F})^2 - (\frac{j''\Delta}{F})^2\big]^{\frac{1}{2}}\hat{z}'' \\ &= (\frac{i''k_o\Delta}{F}\cos\theta + k_o\big[1 - (\frac{i''\Delta}{F})^2 - (\frac{j''\Delta}{F})^2\big]^{\frac{1}{2}}\sin\theta)\hat{x} + \frac{j''k_o\Delta}{F}\hat{y} \\ &\quad + (-\frac{i''k_o\Delta}{F}\sin\theta + k_o\big[1 - (\frac{i''\Delta}{F})^2 - (\frac{j''\Delta}{F})^2\big]^{\frac{1}{2}}\cos\theta)\hat{z}\end{aligned} \tag{2.73}$$

A distributed sinusoidal grating with wavevector

$$\vec{K}_{i'j'i''j''} = \vec{k}_{i'j'} - \vec{k}_{i''j''} \tag{2.74}$$

diffracts light from the $i'j'^{th}$ input pixel to the $i''j''^{th}$ output pixel.

The $(i', j')^{th}$ pixel is not the only input pixel which is phase matched to the grating wavevector $\vec{K}_{i'j'i''j''}$. Recalling Eq. (2.60), we find that any incident plane wave whose wavevector, $\vec{k}_r$ satisfies

$$|[\vec{k}_r]_z - \sqrt{{k_o}^2 - |[\vec{k}_r - \vec{K}_{i'j'i''j''}]_t|^2} - [\vec{K}_{i'j'i''j''}]_z| < \frac{\pi}{Nl} \tag{2.75}$$

will be phase matched. $[\vec{k}_r - \vec{K}_{i'j'i''j''}]_t$ is the component of $\vec{k}_r - \vec{K}_{i'j'i''j''}$ in the

x-y plane. Eq. (2.75) can be transformed to

$$\frac{Nl|(\vec{k}_r - \vec{k}_{i'j'}) \cdot \vec{K}_{i'j'i''j''}|}{\pi|[\vec{k}_r]_z - [\vec{K}_{i'j'i''j''}]_z|} < 1. \tag{2.76}$$

Thus, any input corresponding to a wavevector which is displaced from $\vec{k}_{i'j'}$ in a direction orthogonal to $\vec{K}_{i'j'i''j''}$ will be phase matched to read out the $(i'j'i''j'')^{th}$ grating.

Assuming that the curvature of the normal surface over the bandwidth accessed by the input and output planes is small we can visualize the gratings between the input and output planes with the help of Fig. 2.7. The figure shows the end points on the wave normal surface of the wavevectors of the input and output signals. We assume that the angular bandwidth of the signal is small enough that each set of signals may be represented by a $N' \times N'$ grid of points, where $N' = L/\Delta$. The separation between the two grids, $2k_o \sin\theta$, is assume to be much larger than the widths of the grids and is not shown to scale in the figure. $(2N'-1)^2$ different planar wavevectors can be drawn between points on the two grids, as shown in the figure. The out-of-plane component of a wavevector between two points depends on the position of the two points.

Two grating wavevectors with the same transverse components are distinct if no $\vec{k}_r$ exists which satisfies Eq. (2.76) for both gratings. By substituting the corresponding input wavevectors into Eq. (2.76), it can be shown that the two grating wavevectors sketched in Fig. 2.8 are distinct if

$$\sin\theta > \frac{\pi F}{4\Delta N k_o l}, \tag{2.77}$$

where terms greater than first order in Δ/F are discarded. Since the grating wavevectors sketched in Fig. 2.8 have the greatest inclination to the horizontal

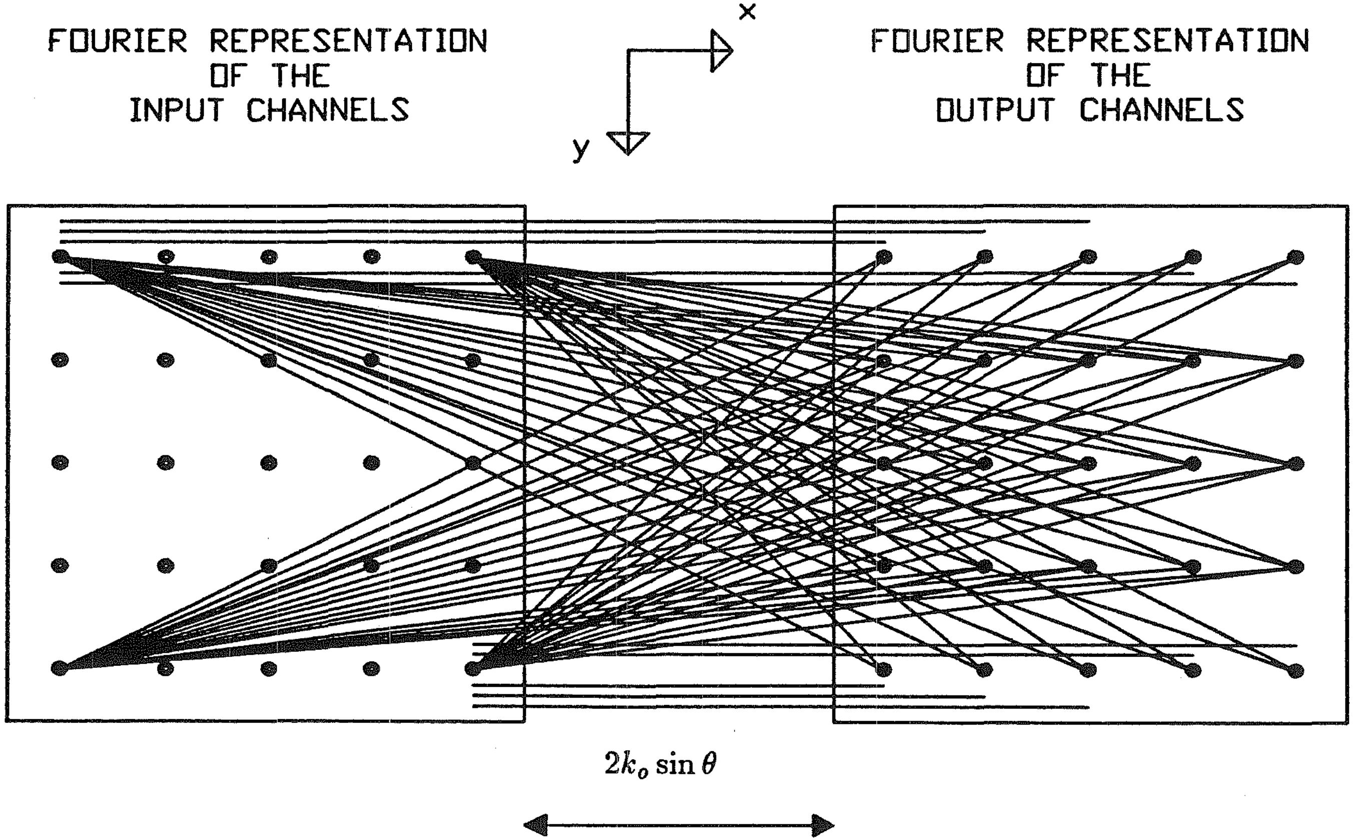


Figure 2.7. Planar representation of grating wavevectors.

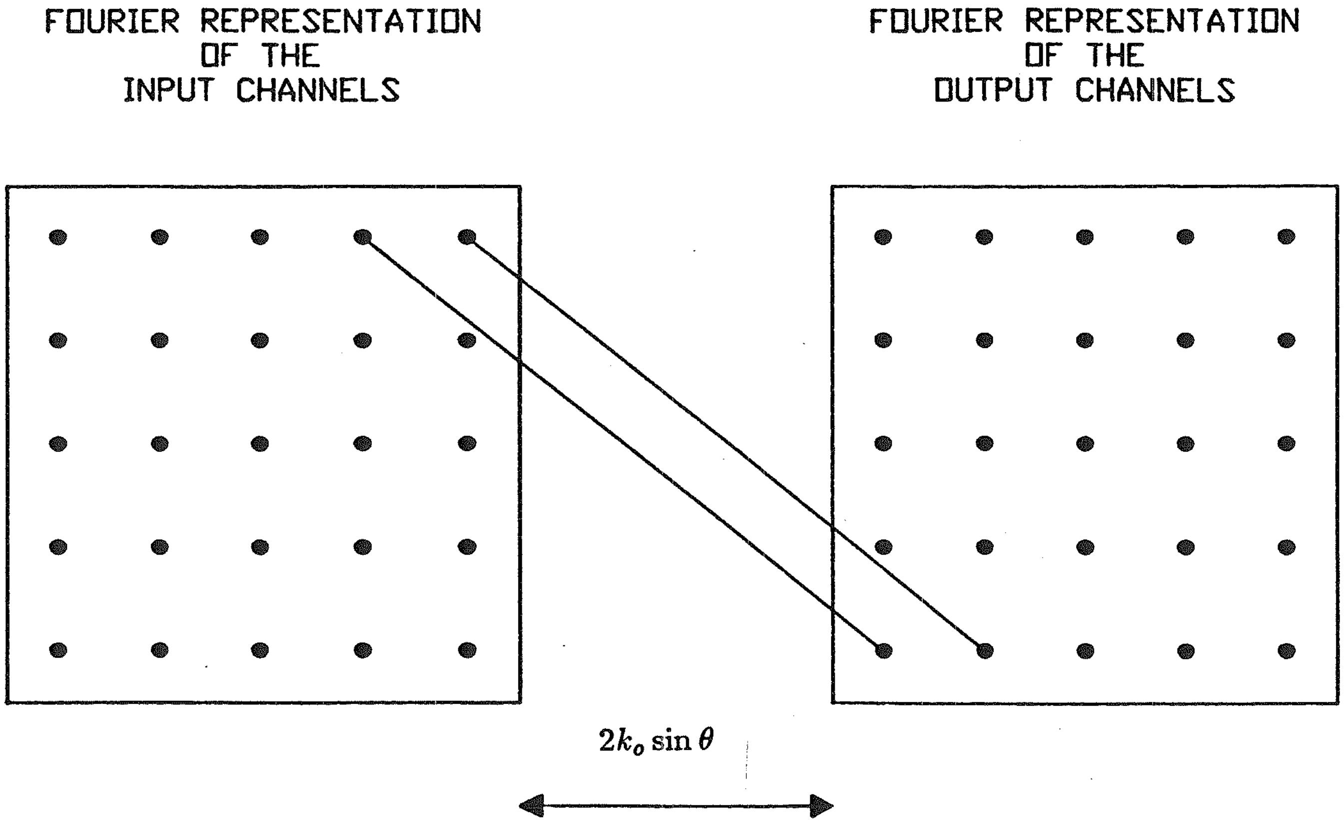


Figure 2.8. Grating wavevectors with maximal inclination.

of any pair of wavevectors between the two grids, it can be shown a distributed grating cannot be phase matched to more than one pixel in each row of the input grid if Eq. (2.77) is satisfied. (We assume that $N' \leq N$.) Combining this result with Eq. (2.71), we require

$$\frac{\lambda F}{8N\Delta l} \leq \sin\theta \leq \frac{((\frac{\lambda F}{\delta L})^2 - 1)}{((\frac{\lambda F}{\delta L})^2 + 1)}. \tag{2.78}$$

A lower bound on the number of distinct distributed gratings the system can support is provided by Fig. 2.9. Each of the grating wavevectors shown in this figure corresponds to a distinct grating. The number of such grating wavevectors is $2N'^3 - N'^2$.

Suppose that the transmittance of the n^{th} transparency is a linear superposition of the transmittances corresponding to the distinct distributed gratings between light from the input and output planes. For the gratings shown in Fig. 2.9, we can write such a superposition as

$$\begin{aligned} t_n(\vec{\rho}) &= \\ &\sum_{i'=-\frac{N'-1}{2}}^{\frac{N'-1}{2}} \sum_{i''=-\frac{N'-1}{2}}^{\frac{N'-1}{2}} \sum_{j''=-\frac{N'-3}{2}}^{\frac{N'-1}{2}} a_{i'-\frac{N'-1}{2}i''j''}(1 + jm\sin(\vec{K}_{i'-\frac{N'-1}{2}i''j''} \cdot (\vec{\rho} + nl\hat{z}))) \\ &+ \sum_{i'=-\frac{N'-1}{2}}^{\frac{N'-1}{2}} \sum_{i''=-\frac{N'-1}{2}}^{\frac{N'-1}{2}} \sum_{j''=-\frac{N'-1}{2}}^{\frac{N'-1}{2}} a_{i'\frac{N'-1}{2}i''j''}(1 + jm\sin(\vec{K}_{i'\frac{N'-1}{2}i''j''} \cdot (\vec{\rho} + nl\hat{z}))), \end{aligned} \tag{2.79}$$

where we have again discarded higher order terms. Note $t_n(\vec{\rho})$ could also be expressed in terms of the $(2N' - 1)^2$ distinct local grating wavevectors in the plane of each hologram. The components of the $N \times (2N' - 1)^2$ local gratings can be related to the components of the distributed gratings using the results of subsection 2.2.2.

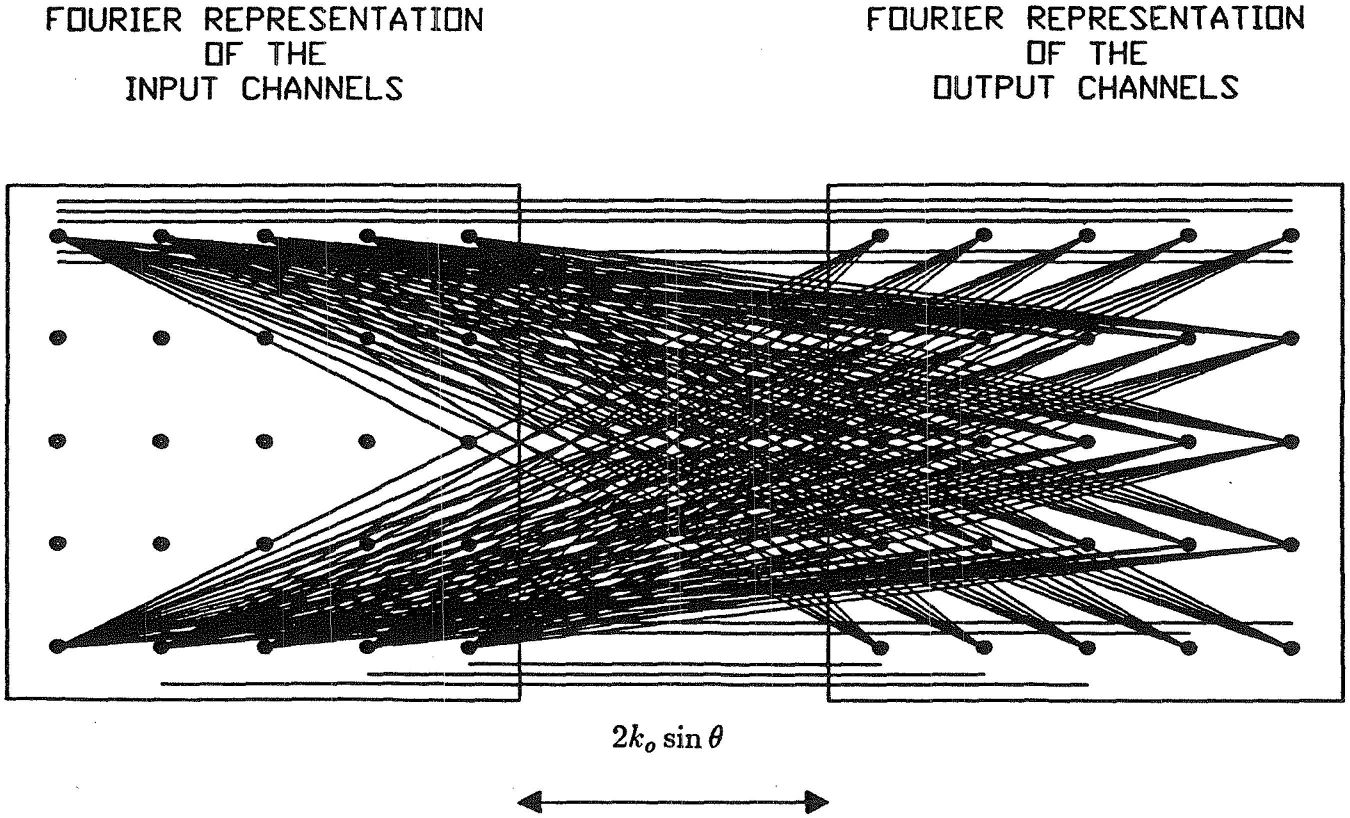


Figure 2.9. Distinct gratings in the thick case.

The read out field may be written

$$U(\vec{r}) = \sum_{i'} \sum_{j'} R_{i'j'} e^{\vec{k}_{i'j'} \cdot r}, \tag{2.80}$$

where $R_{i'j'}$ is the amplitude of the signal generated by the $(i'j')^{th}$ input pixel. Since we are interested in distributed holograms which are weak enough to remain in the linear regime, we can compute the diffracted field by forming a linear superposition of the fields diffracted from each incident plane wave. We assume that the incident waves are undepleted. In this case, the diffracted field is

$$U'(\vec{r}) = \sum_{i'} \sum_{j'} \sum_{i''} \sum_{j''} m a_{i'j'i''j''} R_{i'j'} e^{\vec{k}_{i''j''} \cdot r}. \tag{2.81}$$

The amplitude of the field incident on the $(i'', j'')^{th}$ output pixel is

$$S_{i''j''} = \sum_{i'} \sum_{j'} m a_{i'j'i''j''} R_{i'j'}. \tag{2.82}$$

Each of the N'^4 parameters $a_{i'j'i''j''}$ corresponds to one of the $2N'^3 - N'^2$ grating amplitudes. Obviously, since $\sim N'^3$ degrees of freedom are used to implement N'^4 connections, the transformation between the input and output fields described by Eq. (2.82) is a highly constrained form of linear transformation. The implementation of less constrained transformations with this system is explored in section 2.4.

The nature of the transformation implemented by Eq. (2.82) is clearer in the special case when the separation between the input and output grids in the Fourier space is large enough that all the gratings are effectively parallel to the x axis. In this case, all the inputs in each column of the input grid will be phase matched to the same gratings and the transformation implemented by the system will be shift invariant along the y axis. The condition under which this case holds can be visualized using Fig. 2.10. The two gratings sketched in this figure, $\vec{K}_{-N'-1,N'-1,N'-1,0}$ and $\vec{K}_{-N'-1,0,N'-1,-N'-1}$, exhibit the greatest inclination with respect to the x axis of any pair of gratings from points in the same column with identical planar components. If these two gratings are not distinct then no pair of gratings with identical planar components originating from the same input column can be distinct. The line normal to the gratings passing through the origin of the upper grating is shown as a dashed line in Fig. 2.10. If the origin of the lower grating is closer to this line than its nearest neighbor in the adjacent column of inputs, then the two gratings are not distinct. Simple geometry shows this to be the case if

$$\sin 2\theta > \frac{N'L}{4F}. \tag{2.83}$$

If we want $N' = 100$, this constraint forces F/L to be at least 25. If Eq. (2.83) is satisfied, then all the signals from a given input column are diffracted onto the output grid by the same gratings and signals in different columns are not diffracted by the same gratings. Eq. (2.82) becomes

$$S_{i''j''} = \sum_{i'} \sum_{j'} m a_{i'j'i''(j'-j'')} R_{i'j'}. \tag{2.84}$$

in this case. The field distribution on the i''^{th} column of the output array is a sum over i' of the correlations integrated over y of the field distributions on the i'^{th} column of the input array with $a_{i'i''j_d}$. In spite of the simplicity of this result, this situation is not generally desirable because satisfying Eq. (2.83) severely restricts the scale of N'.

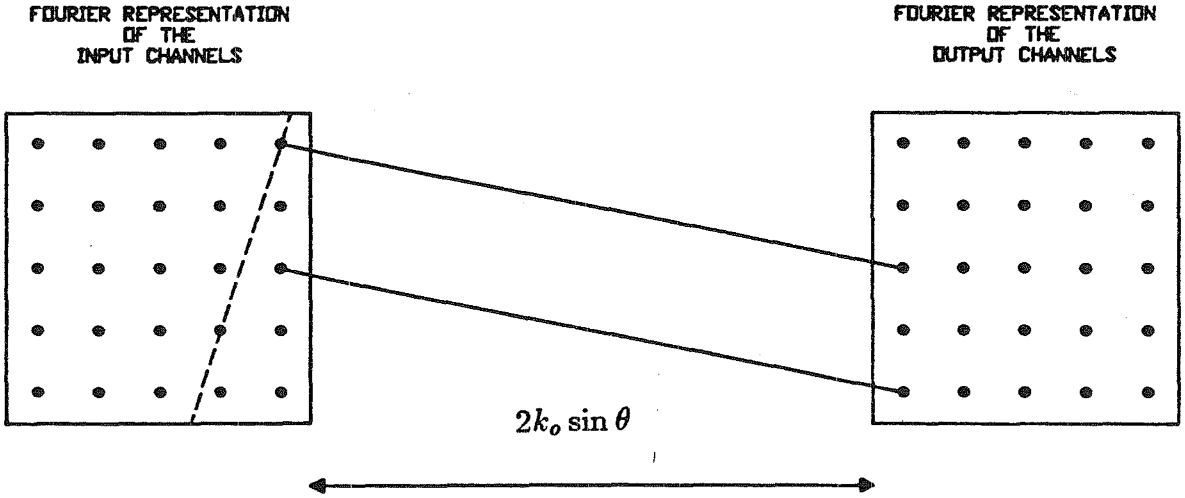


Figure 2.10. Condition for vertical degeneracies.

Before leaving this section, we review some of the approximations we have made and consider the system which satisfies them. It is reasonable to assume that $\delta = 3\lambda$, $\Delta = 10\lambda$, $Nl = L$, and $F = 5L$. If $\lambda = 1$ μm and $L = 1$ cm then $N' = N = L/(10\lambda) = 10^3$. Substituting in Eq. (2.78) yields $28° > \theta > 4°$. θ is further constrained by the assumption made after Eq. (2.36) that $\gamma > 2$. In the present case this assumption implies $\theta < 9°$. This last assumption was made to insure that only first order diffraction occurs with measurable efficiency. If higher order diffracted fields do not overlap the output fields this assumption may not be necessary. If these conditions are met, the analysis leading up to Eq. (2.82) is valid. The number of distinct gratings which can be stored in the holograms is $\sim 10^9$.

2.3 COUPLED WAVE THEORY FOR MANY BEAM VOLUME HOLOGRAMS

We now turn our attention to diffraction in a continuous volume hologram in which a perturbation, $\Delta\epsilon$, to the optical permittivity, ϵ, has been recorded. In general, ϵ and $\Delta\epsilon$ are two-dimensional tensors. Our goal in this section is to use coupled wave theory to find the optical fields diffracted by a volume hologram given boundary conditions and $\Delta\epsilon$. We first consider the case

$$\Delta\epsilon = m \sin(\vec{K} \cdot \vec{r}). \tag{2.85}$$

Just as in the previous section, we derive conditions under which a sinusoidal grating implements a linear connection between two modes and we consider the formation of interconnection networks based on superposition of grating interconnections. We are able to show that the same principles of Bragg matching and rotational degeneracy as were derived in the previous section apply to the volume case. In the final section of this chapter we consider methods for blocking degenerate interconnections and thereby allow us to implement arbitrary linear transformations on distributed fields.

The optical fields in a volume satisfy the Maxwell equations

$$\begin{aligned} \nabla \times \vec{E} &= -\frac{\partial \vec{B}}{\partial t} \\ \nabla \times \vec{H} &= \frac{\partial \vec{D}}{\partial t} \\ \vec{\nabla} \cdot \vec{D} &= 0 \\ \vec{\nabla} \cdot \vec{B} &= 0 \end{aligned} \tag{2.86}$$

and the material equations

$$\begin{aligned} \vec{B} &= \mu \vec{H} \\ \vec{D} &= (\epsilon + \triangle\epsilon)\vec{E}. \end{aligned} \tag{2.87}$$

Using the identity

$$\nabla \times \nabla \times \vec{E} = \vec{\nabla}(\vec{\nabla} \cdot \vec{E}) - \nabla^2 \vec{E} \tag{2.88}$$

and assuming a monochromatic field at frequency $\omega = k_o/\sqrt{\mu\epsilon}$, Eqs. (2.86) yield

$$\nabla^2 \vec{E} - \vec{\nabla}(\vec{\nabla} \cdot \vec{E}) + \mu(\epsilon + \triangle\epsilon)\omega^2 \vec{E} = 0 \tag{2.89}$$

Using Poisson's equation and assuming an isotropic medium, we find

$$\vec{\nabla} \cdot (\epsilon + \triangle\epsilon)\vec{E} = (\epsilon + \triangle\epsilon)\vec{\nabla} \cdot \vec{E} + \vec{E} \cdot \vec{\nabla} \triangle \epsilon = 0. \tag{2.90}$$

To first order in the perturbation,

$$\vec{\nabla}(\vec{\nabla} \cdot \vec{E}) = -\epsilon^{-1} \vec{\nabla}(\vec{E} \cdot \vec{\nabla} \triangle \epsilon). \tag{2.91}$$

This term is often dropped from the coupled wave equation under the assumption that the polarization of the field is perpendicular to the grating wavevector. In most situations this geometry is desired to maximize the modulation depth of recorded holograms. In photorefractive holograms, however, the polarization is often parallel to the grating wavevector due to the need to adjust the geometry to the available electro-optic coefficients. In this case this term cannot necessarily be dropped.

Suppose that $\triangle\epsilon = m\sin(\vec{K}\cdot\vec{r})$. Just as in the previous section, a read out beam with wavevector $\vec{k}_r$ is coupled to a diffracted beam with wavevector

$$\vec{k}_s = \vec{k}_r - \vec{K}. \tag{2.92}$$

For this coupling to occur with high efficiency, $\vec{k}_s$ must lie on the normal surface. When $\vec{k}_s$ is on the normal surface we say that $\vec{k}_r$ is Bragg matched to read out the grating. We define the mismatch parameter

$$\vartheta = ({k_o}^2 - |\vec{k}_s|^2) \tag{2.93}$$

in order to include in our analysis the possibility that the read out beam is misaligned. The coupled wave approach consists of assuming that

$$\vec{E} = R(z)\hat{e}_r e^{j\vec{k}_r\cdot\vec{r}} + S(z)\hat{e}_s e^{j\vec{k}_s\cdot\vec{r}}, \tag{2.94}$$

where $\hat{e}_r$ and $\hat{e}_s$ are unit vectors along the polarizations of the corresponding plane waves. We can now solve for $\vec{E}(\vec{r})$ by substituting Eq. (2.94) in Eq. (2.89).

Substituting Eq. (2.94) into Eq. (2.91) we find

$$\begin{aligned}\vec{\nabla}(\vec{\nabla}\cdot\vec{E}) = -\epsilon^{-1}\Big[& j\frac{m}{2}R(z)\hat{e}_r\cdot\vec{K}[(\vec{k}_r+\vec{K})e^{j(\vec{k}_r+\vec{K})\cdot\vec{r}} + (\vec{k}_r-\vec{K})e^{j(\vec{k}_r-\vec{K})\cdot\vec{r}}] \\ & j\frac{m}{2}S(z)\hat{e}_s\cdot\vec{K}[(\vec{k}_s+\vec{K})e^{j(\vec{k}_s+\vec{K})\cdot\vec{r}} + (\vec{k}_s-\vec{K})e^{j(\vec{k}_s-\vec{K})\cdot\vec{r}}] \\ & + c.c.\Big]\end{aligned} \tag{2.95}$$

The components of Eq. (2.95) which depend on $(\vec{k}_r+\vec{K})$ and $(\vec{k}_s-\vec{K})$ correspond to higher order disturbances. Since these wavevectors are unlikely to lie on the normal surface, These terms can be ignored in the typical case. The term which depends on $(\vec{k}_r - \vec{K}) \approx \vec{k}_s$ oscillates with the same spatial dependence as the diffracted mode but is polarized orthogonally to the propagating mode and thus cannot contribute to the diffracted field. Similarly the term which depends on $(\vec{k}_s + \vec{K})$ oscillates with the spatial dependence of the read out mode along a polarization orthogonal to the polarization of the mode and thus cannot affect the field in that mode. We know that these two terms are orthogonal to the polarizations of the modes because they are polarized along the propagation directions of the corresponding beams. In isotropic media the polarization of the fields is orthogonal to the propagation direction and the terms in Eq. (2.95) can be dropped from the wave equation.

Separating Fourier components of the remaining terms in Eq. (2.89) we find

$$\frac{\partial^2 R}{\partial z^2}\hat{e}_r + 2jk_{rz}\frac{\partial R}{\partial z}\hat{e}_r - j\frac{m}{2}k_o^2 S\hat{e}_s = 0 \tag{2.96}$$

and

$$\frac{\partial^2 S}{\partial z^2}\hat{e}_s + 2jk_{sz}\frac{\partial S}{\partial z}\hat{e}_s + \vartheta S + j\frac{m}{2}k_o^2 R\hat{e}_r = 0. \tag{2.97}$$

If the amplitude of the perturbation is small compared to the dielectric constant we may assume that the length scale for coupling between the two beams is much greater than the optical wavelength. In this case the second order derivatives of R and S are much smaller than $k_{[r,s]}z$ times the first derivatives. The assumption that this is the case is known as the "slowly varying envelope approximation." If

$S(0) = 0$, the solution to these equations under this approximation is

$$\begin{bmatrix} R(z) \\ S(z) \end{bmatrix} = \begin{bmatrix} R(0)e^{j\xi}[\cos(\sqrt{\nu^2 + \xi^2}) - \frac{j\sin(\sqrt{\nu^2+\xi^2})}{\sqrt{1+\frac{\nu^2}{\xi^2}}}] \\ -\sqrt{\frac{k_{rz}}{k_{sz}}}e^{j\xi}\frac{R(0)\sin(\sqrt{\nu^2+\xi^2})}{\sqrt{1+\frac{\xi^2}{\nu^2}}} \end{bmatrix} \tag{2.98}$$

where

$$\nu = \frac{{k_o}^2 z}{4\sqrt{k_{rz}k_{sz}}}|m\hat{e}_r \cdot \hat{e}_s|. \tag{2.99}$$

and

$$\xi = \frac{\vartheta z}{2k_{sz}}. \tag{2.100}$$

This result appears as Eq. (42) in [9].

As in the previous section, we assume that ν is small enough that diffraction remains in the linear regime. Linearity is assured if

$$\frac{\sin(\sqrt{\nu^2 + \xi^2})}{\sqrt{\nu^2 + \xi^2}} \approx 1. \tag{2.101}$$

For small ν, this condition implies

$$|\xi| < \frac{\pi}{2} \tag{2.102}$$

or

$$|\vartheta| < \frac{\pi k_{sz}}{L_z}, \tag{2.103}$$

where L_z is the thickness of the hologram. Eq. (2.103) specifies the range of read out angles over which phase matching is maintained. The analogous result for cascaded thin holograms was given by Eq. (2.60).

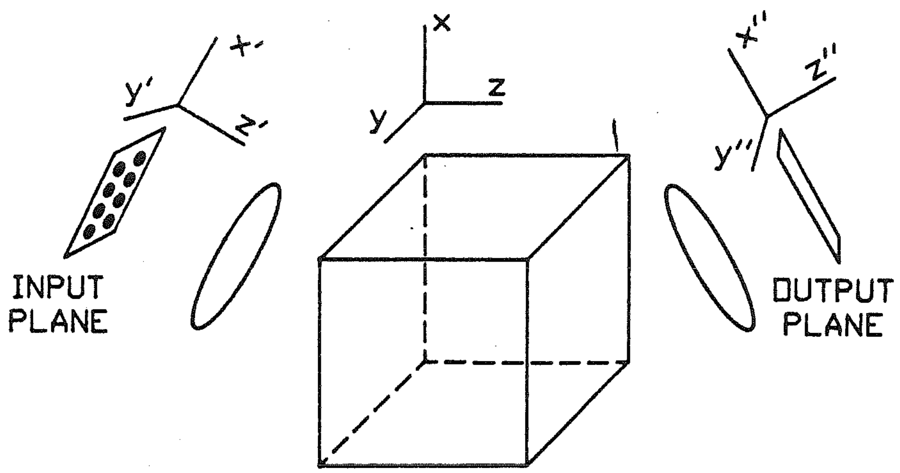


Figure 2.11. Volume holographic interconnection system.

Consider the system of Fig. 2.6 with a volume hologram substituted for the cascade of thin holograms. Such a system is sketched in Fig. 2.11. The system is again addressed by light from a spatial light modulator in the x'-y' plane and the diffracted signal is detected by a detector array in the x''-y'' plane. The parameters Δ, δ, L, θ, and N' are all as defined in subsection 2.2.3. The optical wavevectors in the readout and diffracted fields are exactly as described above, as are the corresponding grating wavevectors.

In Eq. (2.77) we presented a condition for a thin transparency cascade interconnection network which ensured that pixels in different columns of the same row could not be diffracted by the same grating wavevector. A similar condition can be derived for a volume hologram interconnection network by requiring that θ be large enough that light from a given input pixel cannot be diffracted by a grating which diffracts light from an adjacent pixel in the same row. A read out beam generated at a pixel in column $(i' \pm 1)$ is not diffracted by $\vec{K}_{i'j'i''j''}$ for all j', i'' and j'' if $|\vartheta_{(i'+1)}| > \pi k_{sz}/L_z$, where

$$\begin{aligned}\vartheta_{(i'+1)} &= {k_o}^2 - |\vec{k}_{(i'+1)j'} - \vec{K}_{i'j'i''j''}|^2 \\ &= -2{k_o}^2 \frac{\Delta}{F} \sin 2\theta\end{aligned} \quad . \tag{2.104}$$

We have discarded second order terms in Δ/F. Thus, gratings corresponding to different i' are distinct if

$$\sin\theta > \frac{\pi F}{4\Delta k_o L_z}. \tag{2.105}$$

Eq. (2.105) is the volume holographic analog of Eq. (2.77).

The distinction between local and distributed holograms which we developed for the thin hologram cascade is not needed in considering continuous volume holograms. A connection in a continuous volume hologram is made by a Fourier component of the perturbation which forms the hologram. The Fourier expansion of this perturbation can be written

$$\begin{aligned}\Delta\epsilon = &\sum_{i'=-\frac{N'-1}{2}}^{\frac{N'-1}{2}} \sum_{i''=-\frac{N'-1}{2}}^{\frac{N'-1}{2}} \sum_{j''=-\frac{N'-3}{2}}^{\frac{N'-1}{2}} a_{i'-\frac{N'-1}{2}i''j''} \exp(\vec{K}_{i'-\frac{N'-1}{2}i''j''} \cdot \vec{r}) \\ + &\sum_{i'=-\frac{N'-1}{2}}^{\frac{N'-1}{2}} \sum_{i''=-\frac{N'-1}{2}}^{\frac{N'-1}{2}} \sum_{j''=-\frac{N'-1}{2}}^{\frac{N'-1}{2}} a_{i'\frac{N'-1}{2}i''j''} \exp(\vec{K}_{i'\frac{N'-1}{2}i''j''} \cdot \vec{r}).\end{aligned} \quad . \tag{2.106}$$

This expression is the continuous analog of Eq. (2.79). With the substitution of $\Delta\epsilon$ for $t_n(\vec{\rho})$, the analysis of subsection 2.2.3 applies to a volume holographic interconnection network exactly as to a cascaded thin hologram network. In particular, Eq. (2.82) also describes the output signals for the volume system. The condition under which a volume hologram implements the transformation given by Eq. (2.84) is identical to Eq. (2.83). In section 2.4 we explore in more detail the the nature of the transformations which can be implemented with both the volume system and the thin hologram cascade.

We saw in subsection 2.2.2 that a local planar grating in a cascade of N thin transparencies corresponds to N distinct distributed gratings. In a volume hologram, the same rule of thumb holds true, except that the effective N can be much larger because the resolution along z in a volume system approaches λ, as opposed to l in the cascade. As an example of an interconnection network based on continuous media, assume that $\delta = 3\lambda$, $\Delta = 3\lambda$, $L_z = L$, $F = 5L$, $\lambda = 1\mu\text{m}$ and $L = 1$ cm. The improvement in Δ over the cascade is possible because of the higher resolution in z. Note that the parameter γ, which limited θ in the thin case, does not apply in the volume case. The range allowed for θ using these parameters is $28° > \theta > 12°$. Under these conditions approximately 10^{11} distinct gratings may be stored in the volume.

2.4 ARBITRARY LINEAR TRANSFORMATIONS IN VOLUME INTERCONNECTION NETWORKS

By an arbitrary linear transformation we mean a transformation which maps an input containing N_1 degrees of freedom into an output containing N_2 degrees of freedom using $N_1 N_2$ independent weights. Unfortunately, volume holography cannot be used to implement arbitrary linear transformations between arbitrary incident and diffracted fields. This is because N'^4 degrees of freedom, corresponding in this case to N'^4 distinct gratings, would be needed to implement an arbitrary linear transformation between the N'^2 components of the input field and the N'^2 components of the output field. As we have seen, only $o(N^3)$ distinct gratings wavevectors connect the two field distributions. However, by limiting the degrees of freedom in the input and output fields, it is possible to design systems which implement arbitrary linear transformations between the input and output field distributions. In this section we discuss two approaches to designing such systems, one method based on the "fractal sampling grids" and a second method based on holograms between orthogonal spatial modes.

A linear transformation may be represented as a vector-matrix multiplication. The input and output vectors in the present case correspond to the input and output fields. The amplitude of each Fourier component of the field is represented by a corresponding vector component. The matrix components represent the grating amplitudes. As we have seen, if the input and output fields are unconstrained, the components of the grating matrix which transforms them are highly correlated. Assuming that the amplitude of each of the gratings is an independent variable, a volume holographic transformation may be arbitrary if each grating is used to form exactly one connection between an input and an output component.

A sampling grid is an overlay which selects active pixels from the input or output planes. A transformation which the holographic system implements between the active pixels is arbitrary if none of the grating components of the hologram implements a connection between two different pairs of pixels. In designing a sampling grid, the goal is to select a maximal number of pixels for a given ratio of input nodes to output nodes such that an arbitrary linear transformation is still possible. The number of components in the interconnection matrix is equal to the product of the number of input and output points used. Based on the constraint on the scale of the matrix mentioned above, we will be satisfied with a sampling grid if this product is $o(N'^3)$. A simple example of a sampling grid which meets this criterion is afforded by Fig. 2.9. If only the top and bottom rows of the input plane are used for reading out the hologram and if the bottom row of the output plane is not used, each grating is used to form exactly one connection. The input vector is of dimension $2N'$. The output vector is of dimension $N'(N'-1)$. The interconnection matrix consists of $2N'^2(N'-1)$ independent components. This number is N'^2 less than the maximum because the N'^2 gratings which connect the top row of the input to the bottom row of the output must be discarded due to the fact that the gratings which connect the top row of the input to the top row of the output also connect the bottom row of the input to the bottom row of the output.

In many applications it is desirable for the dimension of the input and output vectors to be nearly the same. This is especially true if transformations are to be cascaded, as in multilayer neural networks. A simple grid derived from the asymmetric grid of the previous paragraph which balances the number of pixels on the input and output planes is shown in Fig. 2.12. The input plane consists of $\sqrt{N'}$ rows of N' active points. The rows are spaced by $\sqrt{N'}$ inactive rows. The output plane also consists of $\sqrt{N'}$ rows of N' active points. The rows are grouped together in the center of the plane. The number of active points on the input plane and on the output plane is $N'^{\frac{3}{2}}$. The interconnection matrix consists of N'^3 independent components.

The "fractal" nature of sampling grids arises from the fact that a grid appropriate for a value of N' may be used to generate a grid for a larger value of N'. This is done by substituting a grid which meets the design criteria onto its own

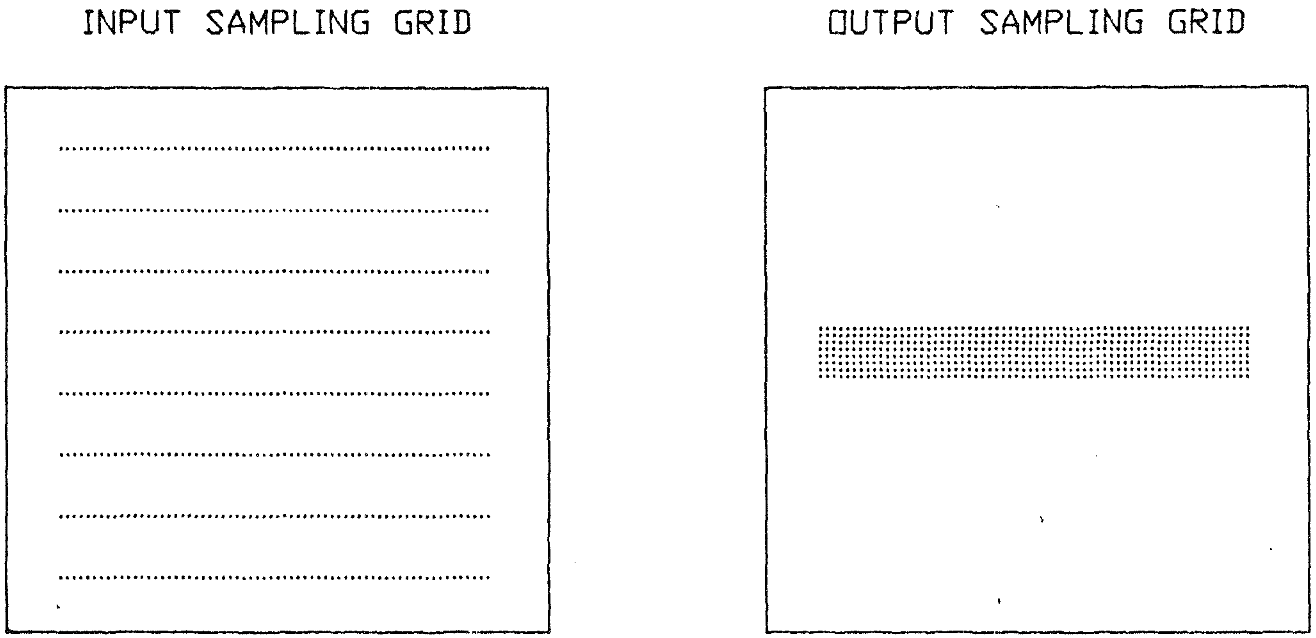


Figure 2.12. Symmetric sampling grid. $N' = 64$.

active nodes. It can be shown that this new grid, of dimension ${N'}^2$ also meets the design criteria. This means that given one grid which allows $o({N'}^3)$ connections, a suitable grid may be found for larger scale systems. The grids which are generated display the "self similar" property of fractal patterns. Many variations of suitable fractal sampling grids exist. In particular, grids which uniformly sample the coupled planes and grids with variable ratios of input and output points have been derived. We define the dimension of the input plane to be

$$d_1 = \frac{\log N_1}{\log N'} \tag{2.107}$$

and the dimension of the output plane to be

$$d_2 = \frac{\log N_2}{\log N'}, \tag{2.108}$$

where N_1 and N_2 are the number of active points on the input and output sampling grids, respectively. It can be shown that sampling grids can be constructed with any desired dimensions subject to the following constraints [79]:

$$\begin{aligned} 0 \leq d_1 \leq 2 \\ 0 \leq d_2 \leq 2 \\ d_1 + d_2 \leq 3 \end{aligned} \tag{2.109}$$

Thorough reviews of the use of fractal sampling grids are presented in [85, 79, 70].

When a sampling grid is used to mask the output detector array, some of the energy diffracted from the active inputs will be diffracted onto unused portions of the output array and will be wasted. We can explore the impact of this fact

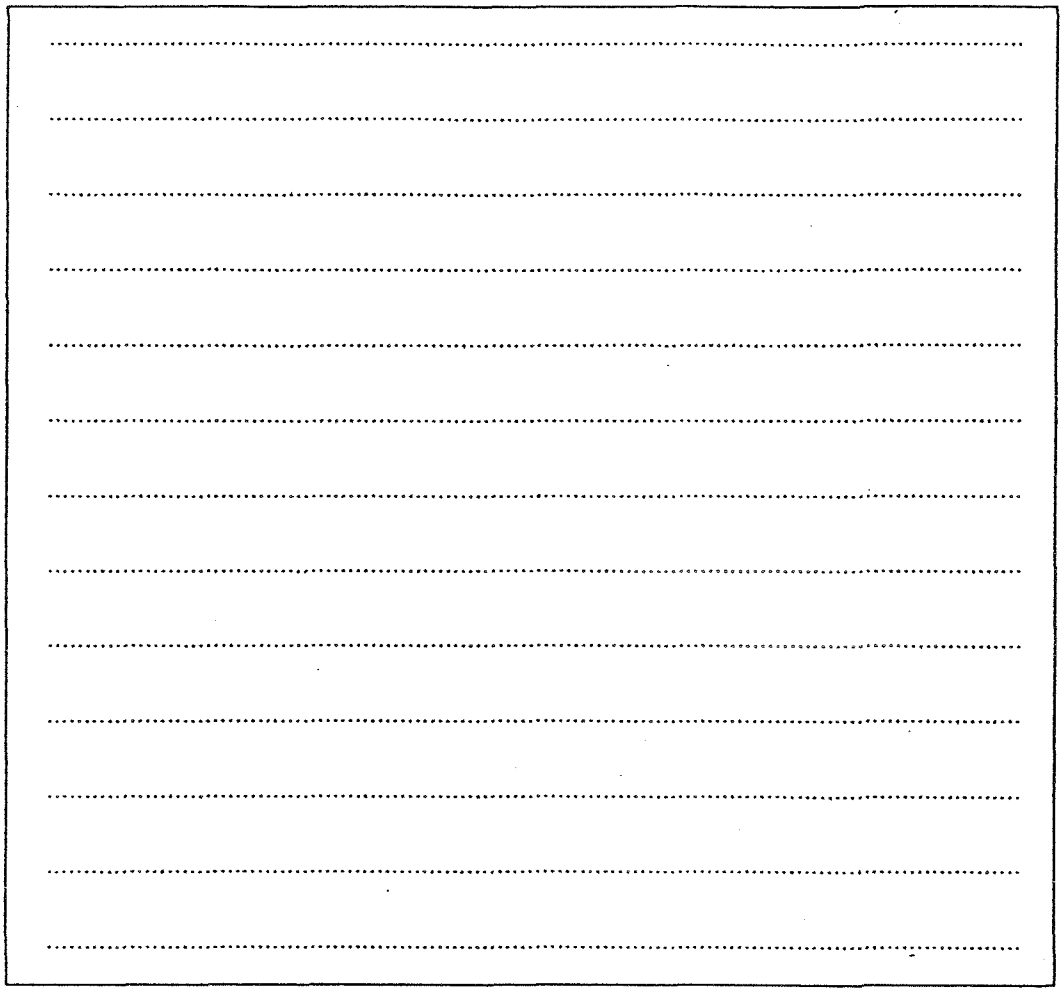


Figure 2.13(a). Example of input plane of a fractal sampling grid.

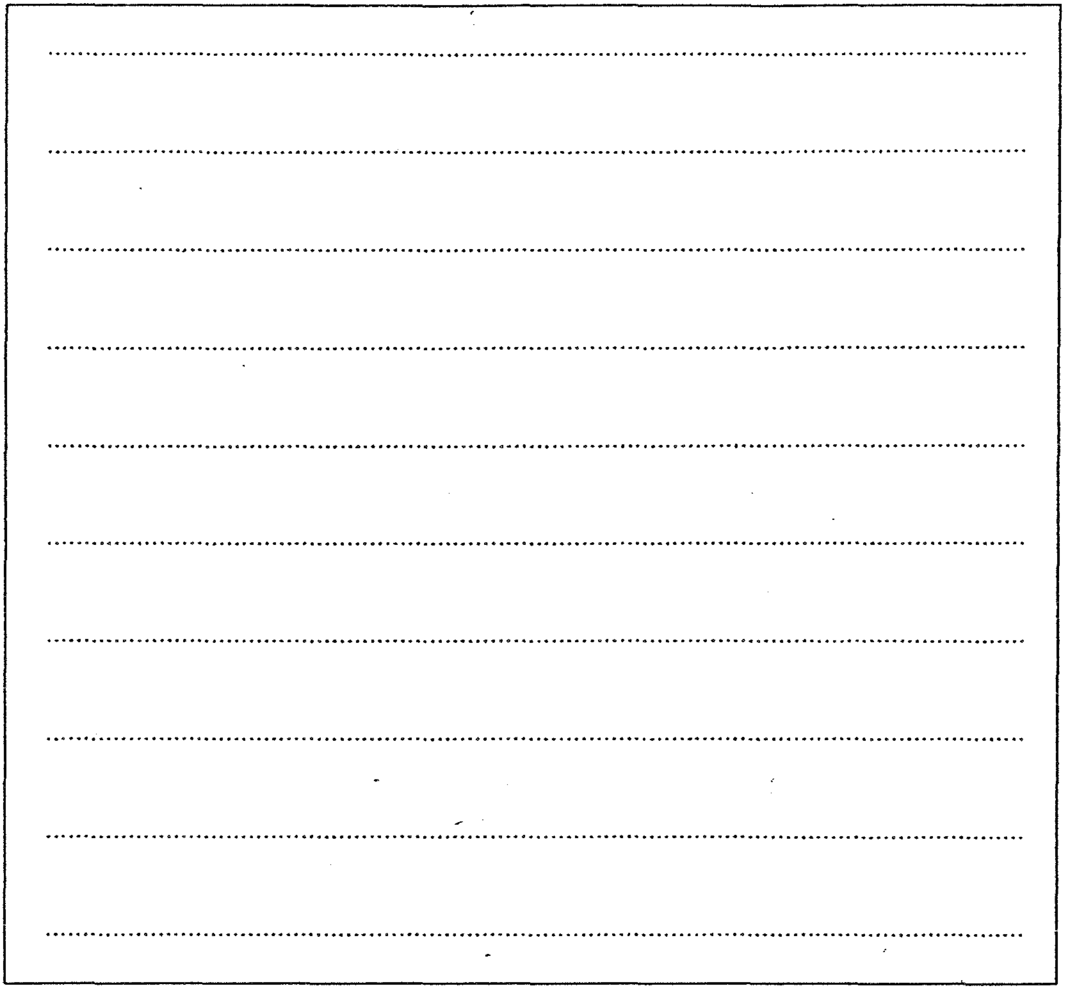


Figure 2.13(b). Example of output plane of a fractal sampling grid.

on the overall energy efficiency of the system as a function of the dimensions of the input and output grids using sampling grids of the sort sketched in Fig. 2.13. These grids consist of rows of N' input pixels spaced in x by $s_1 - 1$ unused rows and rows of N' output pixels separated by $s_2 - 1$ unused rows. The number of active pixels on the input grid is $N_1 = {N'}^2/s_1$. The input dimension is

$$d_1 = 2 - \frac{\log s_1}{\log N'}. \tag{2.110}$$

The number of active pixels on the output grid is $N_2 = {N'}^2/s_2$. The output dimension is

$$d_2 = 2 - \frac{\log s_2}{\log N'}. \tag{2.111}$$

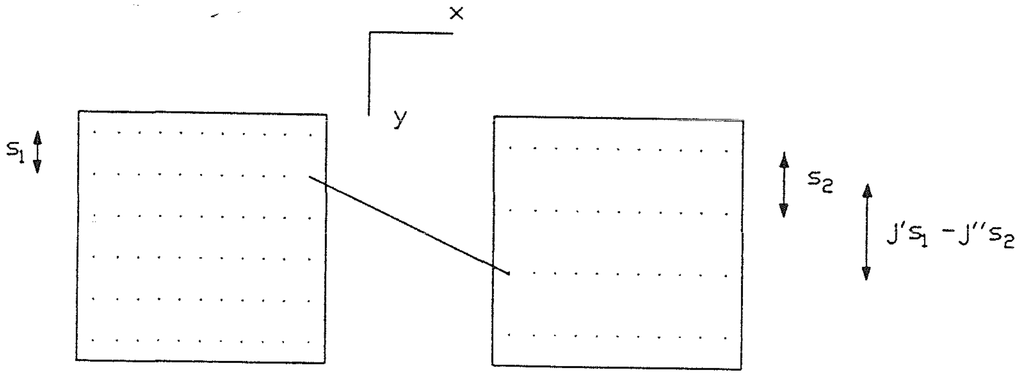


Figure 2.14. Geometry of sampling grid.

As sketched in Fig. 2.14 , the y component of the grating wavevector between the i'^{th} pixel in the j'^{th} active input row and the i''^{th} pixel in the j''^{th} active output row is

$$[\vec{K}_{i'j'i''j''}]_y = (j's_1 - j''s_2)\frac{k_o\Delta}{F} \tag{2.112}$$

Letting $Q = F[\vec{K}_{i'j'i''j''}]_y/k_o\Delta$, Eq. (2.112) can be cast in the form of a Diophantine equation,

$$Q = j's_1 - j''s_2. \tag{2.113}$$

Given integers Q, s_1 and s_2, a simple theorem in number theory [111] tells us that there exist integers j' and j'' which satisfy Eq. (2.113) if and only if d divides Q with no remainder, where d is the greatest common divisor of s_1 and s_2. Assuming that this condition is satisfied, the theorem further states that all solutions for j' and j'' are described by

$$\begin{aligned} j' &= j'_o + (\frac{s_2}{d})p \\ j'' &= j''_o + (\frac{s_1}{d})p, \end{aligned} \tag{2.114}$$

where (j'_o, j''_o) is a solution and p is an arbitrary integer. If

$$\frac{s_1 s_2}{d} > N' \tag{2.115}$$

then at most one of these solutions is such that j' and j'' are both between $-(N'-1)/2$ and $(N'-1)/2$. This means that if Eq. (2.115) is satisfied the separation between the j'^{th} input row and the j''^{th} output row is unique, *i.e.*, no two different pairs of active input and output rows are separated by the same distance in y. Since we have assumed $\sin\theta$ large enough that a grating cannot be phase matched to more than one input in a row, a grating cannot couple more than one active input-output pair if Eq. (2.115) is satisfied.

The number of independent connections which can be made between the active inputs and outputs on these grids is

$$S = N_1 N_2 = \frac{N'^4}{s_1 s_2} \leq \frac{N'^3}{d}, \tag{2.116}$$

where we have substituted from Eq. (2.115). For the grids shown in Fig. 2.13 $s_1 = 10$, $d_1 = 1.53$, $s_2 = 13$, $d_1 = 1.47$, $d = 1$, $N' = 130$, and $S \approx 2.2 \times 10^6$. As a second example, grids with $s_1 = 26$, $d_1 = 1.33$, $s_2 = 5$, $d_2 = 1.67$, and $d = 1$ are shown in Fig. 2.15. N' and S are the same as in Fig. 2.13.

We now turn our attention to calculating the ratio of the total power diffracted to active pixels on the output grid to the total power diffracted to inactive pixels. Suppose that the recorded hologram is described by the matrix $\overline{\overline{W}}$. Since we are using fractal grids to break all the degeneracies, every component of $\overline{\overline{W}}$ may be assumed to be independent. Diffraction from the input plane to the output plane may be regarded as the multiplication of an input vector $|R\rangle$ by $\overline{\overline{W}}$. The amplitudes of the diffracted modes can be represented by the vector $|S\rangle$, where

$$|S\rangle = \overline{\overline{W}}\,|R\rangle\,. \tag{2.117}$$

Each component of the vectors $|R\rangle$ and $|S\rangle$ corresponds to the field at a specific pixel. Using singular value decomposition $\overline{\overline{W}}$ can be expressed

$$\overline{\overline{W}} = \sum_{q=1}^{r} \beta_q \, |S^q\rangle \, \langle R^q|\,, \tag{2.118}$$

where the vectors $|S^q\rangle$ are N_2 dimensional and orthonormal, the vectors $|R^q\rangle$ are N_1 dimensional and orthonormal, β_q is a constant and r is the rank of $\overline{\overline{W}}$. An

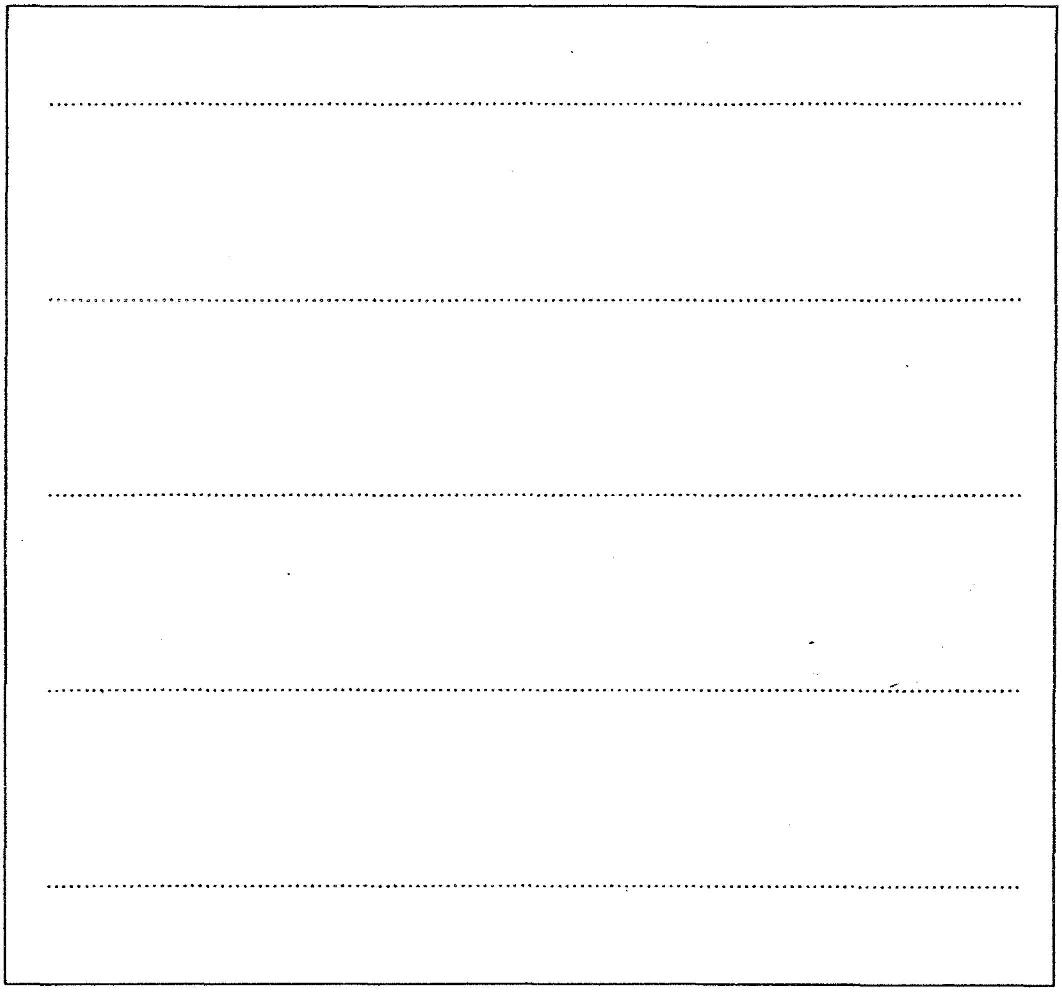


Figure 2.15(a). Example of input plane of a fractal sampling grid.

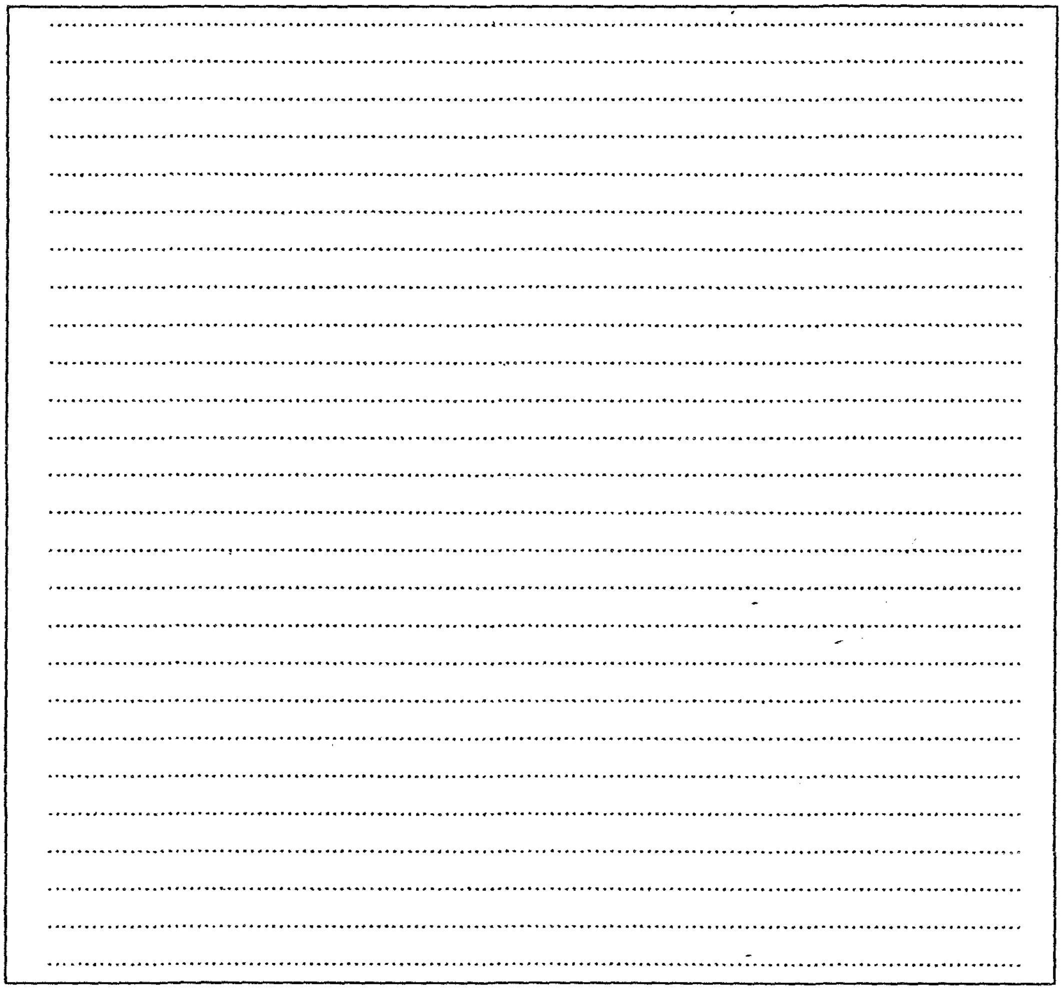


Figure 2.15(b). Example of output plane of a fractal sampling grid.

arbitrary input vector $|R\rangle$ can be decomposed in basis vectors which include the vectors used in the singular value decomposition as a subset.

$$|R\rangle = \sum_{p=1}^{N_1} \alpha_p \left|R^p\right\rangle. \tag{2.119}$$

The orthonormal vectors $\left|R^{r+1}\right\rangle \ldots \left|R^{\frac{N'^2}{s_1}}\right\rangle$ span the null space of $\overline{\overline{W}}$. The field on the active output pixels can similarly be expressed

$$|S\rangle = \sum_{q=1}^{r} \alpha_q \beta_q \left|S^q\right\rangle. \tag{2.120}$$

The power diffracted onto the active pixels of the output array is

$$P_o = \langle S|S\rangle = \sum_{q=1}^{r} |\alpha_q|^2 |\beta_q|^2. \tag{2.121}$$

Each distinct grating stored in the hologram corresponds to a component of the matrix $\overline{\overline{W}}$. In addition to diffracting light from a pixel in the j'^{th} active input row to a pixel in the j''^{th} active output row, a grating between these rows also diffracts light from one active pixel in the other $N'/s_1 - 1$ active rows to an unused output pixel. The fields diffracted onto the unused pixels add incoherently. The total power diffracted onto the unused output pixels is

$$P_u = \left(\frac{N'}{s_1} - 1\right) |\bar{R_{i'j'}}|^2 \sum_{i',j',i'',j''} |w_{(i'j')(i''j'')}|^2, \tag{2.122}$$

where $|\bar{R_{i'j'}}|^2$ is the mean over i' and j' of the activity at the $i'j'^{th}$ input pixel and $w_{(i'j')(i''j'')}$ is the strength of the connection between the $(i'j')^{th}$ input pixel and the $(i''j'')^{th}$ output pixel. Assuming that $\langle R|R\rangle = 1$, $|\bar{R_{i'j'}}|^2 = s_1/N'^2$.

Substituting from Eq. (2.118) and using the orthonormality of $|S^q\rangle$ and $|R^q\rangle$, we find that

$$\sum_{i',j',i'',j''} |w_{(i'j')(i''j'')}|^2 = \sum_q |\beta_q|^2. \tag{2.123}$$

Thus,

$$P_u = \frac{(N' - s_1)}{N'^2} \sum_q |\beta_q|^2. \tag{2.124}$$

The ratio of the power diffracted onto the active pixels of the output to power diffracted onto unused pixels is

$$\frac{P_o}{P_u} \approx \frac{N'^2 \sum_q |\alpha_q|^2 |\beta_q|^2}{(N' - s_1) \sum_q |\beta_q|^2} \approx \frac{N'^2 \sum_{q=1}^{r} |\alpha_q|^2}{(N' - s_1) r}, \tag{2.125}$$

where we assume that $|\beta_q|$ is approximately constant over q. If the input field is completely random then $\langle |\alpha_q|^2 \rangle = s_1 / N'^2$ and

$$\frac{P_o}{P_u} \approx \frac{s_1}{(N' - s_1)} = \frac{1}{(s_2 - 1)}, \tag{2.126}$$

where we have assumed that $s_1 s_2 = N'$, *i.e.*, $d_1 + d_2 = 3$. In terms of the dimension of the output grid, d_2,

$$\frac{P_o}{P_u} = \frac{1}{(N'^{2-d_2} - 1)}. \tag{2.127}$$

For random inputs, P_o/P_u is less than or equal to 1 for all s_2 except $s_2 = 1$, which corresponds to the entire output plane being active. A plot of P_o/P_u verses d_2 is shown in Fig. 2.16.

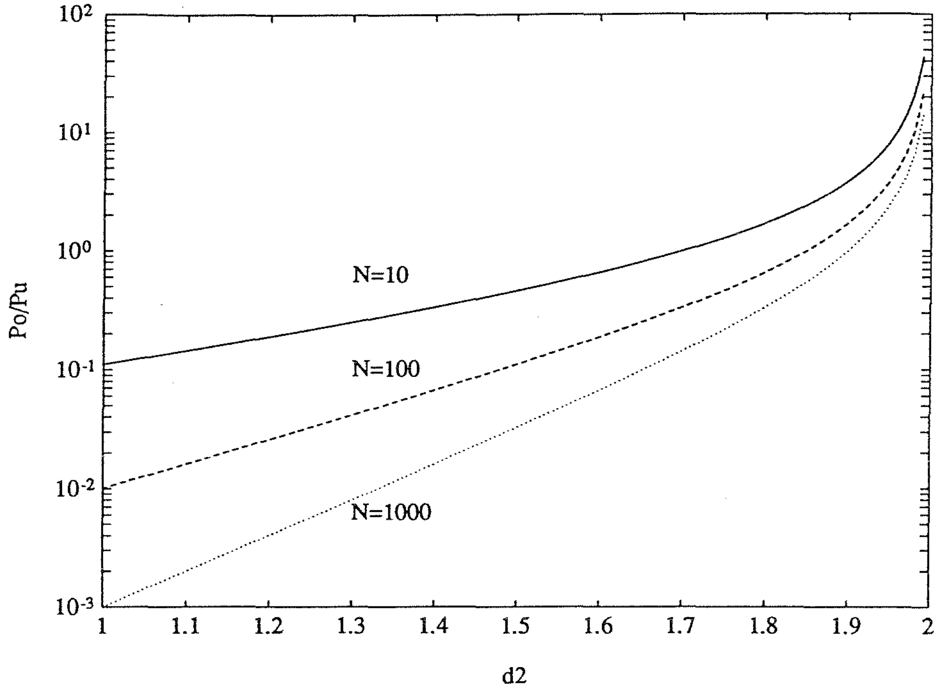


Figure 2.16. $\frac{P_o}{P_u}$ verses d_2 for a random input signal. Plots are shown for various values of N'.

In the applications described later in this thesis, $|R\rangle$ is typically close to a specific $|R^q\rangle$, in which case $|\alpha_q|$ is near one and

$$\begin{aligned} \frac{P_o}{P_u} &= \frac{N'^2}{(N'-s_1)r} \\ &= \frac{N'}{(1-N'^{d_2-2})r} \end{aligned} \tag{2.128}$$

r must be less than or equal to the lesser of N_1 and N_2. Generally, it is desirable to make r as large as possible, since r corresponds to the number of different input vectors which yield a strong output. It is also desirable to keep P_o/P_u as large as possible, since the diffraction efficiency from the active inputs to the active outputs cannot exceed this ratio. If we assume that r is as large as possible, *i.e.*, $r = min(N_1, N_2)$, then P_0/P_u decreases monotonically as d_1 increases. Assuming that $d_1 < d_2$, that $d_1 + d_2 = 3$ and $r = N_1$,

$$\frac{P_o}{P_u} = \frac{1}{\frac{N_1}{N_\prime} - 1} = \frac{1}{N'^{2-d_2} - 1} \tag{2.129}$$

If we want to increase the number of degrees of freedom in the input we must increase d_1. Particular significance may be associated with the input dimension for which the total power diffracted onto the active points for an eigenvector input is equal to the total power diffracted to unused output pixels. The value of d_1 for which this is the case is

$$d_{1p} = 1 + \frac{\log 2}{\log N'} \tag{2.130}$$

A plot of d_{1p} verses N' is shown in Fig. 2.17.

Sampling grids are not a unique solution to the problem of removing degeneracies from volume holographic interconnection systems. Other sets of spatial

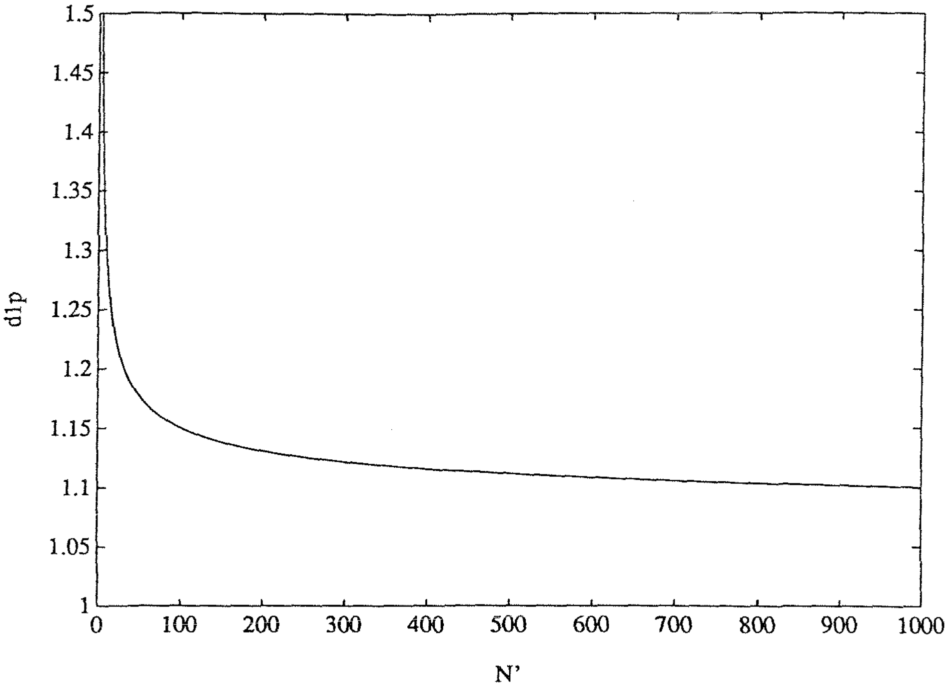


Figure 2.17. d_{1p} verses $\log N'$.

modes, such as the spherical waves suggested in [56], could also yield independent connections. The number of independent connections cannot, however, exceed the bounds derived in this chapter. This is because any connection between modes may be described by a linear superposition of distinct gratings. While the question of whether or not other sets of modes can have greater energy efficiency than fractal sampling grids has not been completely resolved, whenever the input is multimode degeneracies will cause some fraction of the read out energy to be lost to inactive areas of the output plane if such areas exist.

3. FORMATION AND CONTROL OF VOLUME HOLOGRAMS

3.1 INTRODUCTION

In this chapter we consider techniques for controlling the physical parameters of 3-D systems using light. The principle difficulty encountered in the control of such systems is that there is no simple means of reaching into a volume to specify the parameters of a single volume element, or voxel, without affecting the parameters of other voxels. Methods which can be used to overcome this problem include layered growth using lithographic and epitaxial techniques and critical control techniques in which control signals pass through the entire volume but effect only a single voxel. Critical control techniques are especially appropriate to specify parameters using multiphoton processes [112, 113], in which case the control signals may consist of beams at different wavelengths which cross only in one voxel or focusing beams which exceed critical intensity only in one voxel [114]. In this chapter we discuss a third approach based on holographic techniques. The principle difference between holographic control and other techniques is that in holographic systems there is no one-to-one relationship between the state of a specific control signal and the state of a specific voxel. The control signals in an optical holographic system correspond to fields on the boundary of the system. The state of a given voxel is a function of the fields generated in it over time. The relationship between the final states of the voxels and the control signals is specified by a global transformation in space and time. To drive the voxels to some desired state, this transformation must be inverted to find appropriate control signals. The advantages of this approach are that it is much easier to implement than layered growth and that it is much less material dependent than critical control.

We define the number of degrees of freedom in a system to be the number of independent variables which are needed to specify its state. If we set one parameter at each point in the volume, the number of degrees of freedom stored in a volume hologram is equal to the number of voxels. Obviously, a set of control signals cannot independently specify more degrees of freedom in a system then are present in the control signals themselves. Volume holographic control is based on specifying fields on the boundaries of the volume. These fields contain one degree of freedom for each resolution element, or pixel, on the boundary. In a monochromatic optical system, the resolution of the signals used to control and evaluate the state of a volume is lower bounded by λ. The number of pixels with independently specifiable states on the aperture of such a system is $o(A/\lambda^2)$, where A is the aperture of the system and λ is the wavelength. The number of voxels which can be independently determined in such a system is $o(V/\lambda^3)$ [1, 115, 85]. However, these voxels cannot be independently specified unless the number of degrees of freedom in the control signals is raised to V/λ^3. This can be done by changing the boundary conditions in time or by using multiple wavelengths in the control signals. In discrete steps, at least $V/\lambda A$ independent exposures or wavelengths are needed to control the volume. The case $\lambda A > V$ corresponds to thin holograms, which can be specified in one exposure.

In the following three sections of this chapter we consider the control of volume holograms using multiple exposures. The second section considers the formation of arbitrary linear holograms in the cascade of thin transparencies, where by "arbitrary", we mean that all the degrees of freedom of the holographic system are independently specified, and by "linear" we mean that the transmittance of each transparency is proportional to the energy distribution which passed through it during the recording process. We assume that the system is controlled by light incident through the first transparency. This system is analyzed because it allows us to illustrate from an almost intuitive level both why multiple exposures are needed and how such exposures can be used to created an arbitrary hologram. The results of this section are extended to continuous volume holograms in linear media in the third section. Our goals in the third section are to describe more precisely what we mean by "arbitrary" and "linear" in the context of continuous media and to show how arbitrary holograms can be recorded in such media. In the fourth section we consider saturable holographic media. Since such media respond linearly only over a limited range of exposures, the dynamic range of the holograms which can be stored in them is limited. The principle result of this section is that the diffraction efficiency of a hologram stored in a saturable medium is inversely proportional to R^2, where R is the rank of the hologram, *i.e.*, the minimum number of exposures needed to record it. The final section of this chapter considers the control of holograms using polychromatic signals. Polychromatic control allows a hologram to be made in one exposure. The diffraction efficiency of a polychromatically controlled hologram could, in principle, be as high as R^{-1}. However, the difficultly of generating appropriate polychromatic control signals makes this potential difficult to realize.

3.2 CONTROL OF A CASCADE OF THIN HOLOGRAMS

In thin holography, a hologram is simply a faithful recording of a spatially modulated intensity pattern. In a cascaded system, once the recording image on the first transparency is known, the image on each successive transparency can be determined using a propagation operator. This means that if we know the information being stored in one hologram, we know the information being stored in all the holograms. To store independent information in each hologram, it is necessary to vary the image over several exposures.

As an example of this process, suppose that we wish to record a grating with wavevector $\vec{K}$ in each of N cascaded holograms. The amplitude and phase of the grating are to be independently set at each hologram. We record the grating in a series of N' exposures using the interference pattern between the plane waves

$$R^{n'} = R_o e^{j\vec{k}_{rn'}\cdot\vec{r}}. \tag{3.1}$$

and

$$S^{n'} = s_{n'} e^{j\vec{k}_{sn'}\cdot\vec{r}+\phi_{n'}}. \tag{3.2}$$

to record a grating in each hologram in the n'^{th} exposure. Since we are limiting ourselves to recording gratings with wavevector $\vec{K}$, we require that the component of $\vec{k}_{rn'} - \vec{k}_{sn'}$ in the x-y plane be equal to $\vec{K}$ for all n'. In propagating from one hologram to the next the phases of $R_{n'}$ and $S_{n'}$ change by $[\vec{k}_{rn'}]_z l$ and $[\vec{k}_{sn'}]_z l$, respectively. Thus, the spatially modulated terms in the intensity pattern on the n^{th} transparency during the n'^{th} exposure are

$$t_{nn'} = s_{n'} R_o e^{(\vec{K}\cdot\vec{\rho}+(n-1)l\kappa_{n'}-\phi_{n'})}, \tag{3.3}$$

and $t^*_{nn'}$, where $\kappa_{n'} = [\vec{k}_{rn'} - \vec{k}_{sn'}]_z$. As defined above, by "linear" medium we mean a holographic material in which the transmittance of each transparency is proportional to the energy distribution used to record it. If this is the case, the spatially modulated part of the transmittance of the n^{th} transparency after N' exposures is the real part of

$$t_n(\rho) = \sum_{n'=1}^{N'} t_{nn'}. \tag{3.4}$$

Defining the N-dimensional vector $\vec{H}$ with $t_n(\rho)$ as its n^{th} component, Eq. (3.4) can be expressed

$$\vec{H} = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ e^{j2\kappa_1 l} & e^{j2\kappa_2 l} & e^{j2\kappa_3 l} & \dots & e^{j2\kappa_{N'} l} \\ e^{j3\kappa_1 l} & e^{j3\kappa_2 l} & e^{j3\kappa_3 l} & \dots & e^{j3\kappa_{N'} l} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e^{j(N-1)\kappa_1 l} & e^{j(N-1)\kappa_2 l} & e^{j(N-1)\kappa_3 l} & \dots & e^{j(N-1)\kappa_{N'} l} \end{pmatrix} \begin{pmatrix} s_1 e^{-j\phi_1} \\ s_2 e^{-j\phi_2} \\ s_3 e^{-j\phi_3} \\ \vdots \\ s_{N'} e^{-j\phi_{N'}} \end{pmatrix} \tag{3.5}$$

Suppose that we are given $\vec{H}$. We can invert Eq.(3.5) to determine the constants $s_{n'}$ and $\phi_{n'}$ if $N' = N$ and if the constants $\kappa_{n'}$ are such that the matrix is nonsingular. We saw in section 2.2.2 that N distinct values of $\kappa_{n'}$ exist. If we assume, as in that section, that $\kappa_{n'} l = 2(n'-1)\pi/N$ then Eq. (3.5) becomes

$$\vec{H} = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & e^{j\frac{2\pi}{N}} & e^{j\frac{4\pi}{N}} & \dots & e^{j\frac{2(N-1)\pi}{N}} \\ 1 & e^{j\frac{3\pi}{N}} & e^{j\frac{6\pi}{N}} & \dots & e^{j\frac{3(N-1)\pi}{N}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{j\frac{(N-1)\pi}{N}} & e^{j\frac{2(N-1)\pi}{N}} & \dots & e^{j\frac{(N-1)^2\pi}{N}} \end{pmatrix} \begin{pmatrix} s_1 e^{-j\phi_1} \\ s_2 e^{-j\phi_2} \\ s_3 e^{-j\phi_3} \\ \vdots \\ s_{N'} e^{-j\phi_{N'}} \end{pmatrix} \tag{3.6}$$

This matrix in this equation is of the Vandermonde form and is known to be nonsingular [116]. This equation can thus be used to derive an exposure scheme to store an arbitrary $\vec{H}$. Note that we have assumed that the number of input-output beam pairs separated by $\vec{K}$ at the input aperture of the cascade is greater than or equal to the number of layer, N. If N exceeds the range of N' then it would be impossible to make enough distinct exposures through the aperture to control all N plates. In the remainder of this chapter we assume that $N = N'$.

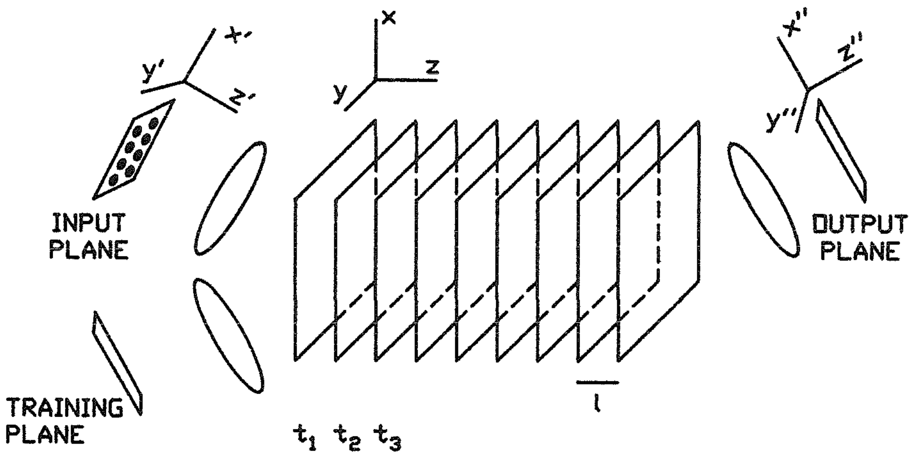


Figure 3.1. Control system for a thin hologram cascade.

Inverting the matrix in Eq. (3.6) provides us with a technique for devising a set of N exposures which will form an arbitrary set of gratings with the wavevector $\vec{K}$ in the cascaded system. There are of order N^2 possible vectors $\vec{K}$ which we can form in each thin hologram. The gratings corresponding to each of these wavevectors can be recorded in parallel in the system of Fig. 2.6, which is repeated as Fig. 3.1. Within the spatial bandwidth of the distributed holograms, this approach allows us to arbitrarily specify the transmittances of the transparencies in the cascade in $o(N)$ steps. This is equivalent to arbitrarily specifying the amplitudes and phases of the $2N^3 - N^2$ distributed holographic gratings which can be stored in this system. Fig. 3.1 is the same as Fig. 2.6 except that a training plane has been added to generate the signals incident on the output plane. The interference pattern between signals generated on the input plane and signals generated on the training plane creates a hologram on each of the cascaded transparencies.

Suppose that only the top and bottom rows of pixels on the input plane are active, as in Fig. 2.9. Since each distributed hologram couples only one input in this case, it is trivially seen that the amplitudes and phases of the distributed holograms can be specified in $2N$ exposures. This is done by recording in sequence the connections from each input to all the outputs to which it is connected. During an exposure, a single input pixel is active and the training pixels corresponding to the outputs to which it is connected are active with amplitudes and phases proportional to the amplitudes and phases desired in the associated connections. The distributed holograms cannot be recorded in fewer than $2N$ exposures because the rank of the transformation being recorded is $2N$. Since the rank of the transformation cannot be lowered in this case without sacrificing some of the distributed holograms, this is the minimal number of exposures needed to specify all the distributed holograms between the input and the output planes.

3.3 FORMATION OF VOLUME HOLOGRAMS IN LINEAR MEDIA

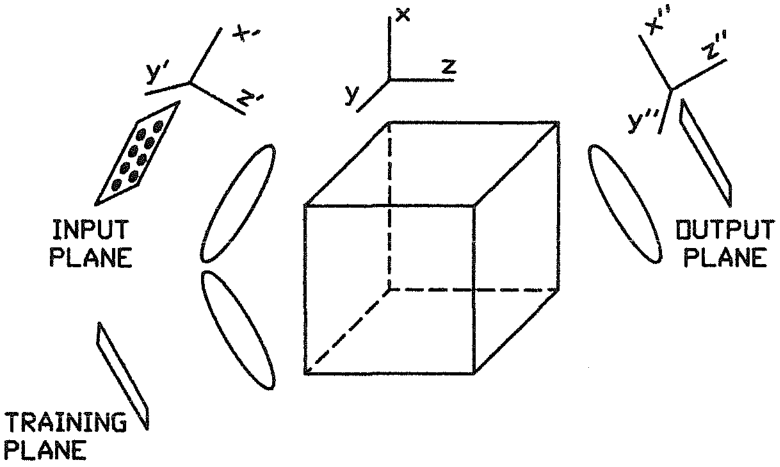


Figure 3.2. Volume holographic interconnection system

A hologram in a continuous medium is a perturbation, $A(\vec{r})$ to the optical properties of the volume. A medium is "linear" if a hologram can be formed in it in which is proportional to the exposure. In terms of the recording intensity, $I(\vec{r},t)$, $A(\vec{r})$ in a linear medium is

$$A(\vec{r}) = A_o \int_0^T \frac{I(\vec{r},t)}{E_o} dt, \tag{3.7}$$

where A_o and E_o are constants. In this section we consider once again the system of Fig. 2.10, which is repeated as Fig. 3.2, and we argue that a perturbation $A(\vec{r})$

which contains a maximal number independently specified connections can be linearly recorded in $o(N)$ exposures in this system.

As explained in chapter 2, the field in the volume of the hologram is

$$U(\vec{r},t)=\sum_{i'j'}\psi_{i'j'}(t)e^{j\vec{k}_{i'j'}\cdot\vec{r}}+\sum_{i''j''}\phi_{i''j''}(t)e^{j\vec{k}_{i''j''}\cdot\vec{r}}. \tag{3.8}$$

This field creates an intensity pattern in the volume which is proportional to $|U(\vec{r},t)|^2$. Assuming a linear medium, a hologram forms which is proportional to the temporal integral of this intensity pattern. We are primarily interested in the components of the hologram which couple beams from SLM$_1$ onto beams from SLM$_2$ and *vice versa.* Thus we will ignore the low spatial frequency components of $|U(\vec{r},t)|^2$ and consider the recorded perturbation to be

$$A(\vec{r})=\frac{A_o}{E_o}\int[\sum_{i'j'}\sum_{i''j''}\phi_{i'j'}(t)\psi^*_{i''j''}(t)e^{j(\vec{k}_{i'j'}-\vec{k}_{i''j''})\cdot\vec{r}}+c.c.]dt. \tag{3.9}$$

The maximum number of independent Fourier components in $A(\vec{r})$ corresponds to the number of distinct grating wavevectors appearing in Eq. (3.9). The wavevector of the grating formed between light from the $i'j'^{th}$ input pixel and the $i''j''^{th}$ output pixel is

$$\begin{aligned}\vec{K}_{i'j'i''j''}=&[(u_{i'}-p_{i''})\cos\theta-(\beta_{i'j'}+\gamma_{i''j''})\sin\theta]\hat{x}+[v_{j'}-q_{j''}]\hat{y}\\&+[(\beta_{i'j'}-\gamma_{i''j''})\cos\theta+(p_{i''}+u_{i'})\sin\theta]\hat{z}.\end{aligned} \tag{3.10}$$

where $u_{i'}=2i'\pi/L$, $v_{j'}$, $p_{i''}$ and $q_{j''}$ are similarly defined, $\beta_{i'j'}=k-u_{i'}^2/2k-v_{j'}^2/2k$ and $\gamma_{i''j''}=k-p_{i''}^2/2k-q_{j''}^2/2k$. In a well designed optical system, *i.e.*, a system where the input aperture is somewhat less than the focal length, the quadratic terms do not contribute significantly to the magnitude of $\vec{K}_{i'j'i''j''}$. In this case,

$$\vec{K}_{i'j'i''j''} \approx [(u_{i'} - p_{i''})\cos\theta - 2k\sin\theta]\hat{x} + [v_{j'} - q_{j''}]\hat{y} + [(p_{i''} + u_{i'})\sin\theta]\hat{z}. \quad (3.11)$$

Let G correspond to the number of distinct values $\vec{K}_{i'j'i''j''}$ can assume. By the number of "distinct values" of $\vec{K}_{i'j'i''j''}$ we mean the number of points separated by at least π/L in the range of $\vec{K}_{i'j'i''j''}$. $u_{i'}$, $v_{j'}$, $p_{i''}$, and $q_{j''}$ assume $N = L/\Delta = L^2/(\lambda F)$ values between $-\pi(N-1)/L$ and $\pi(N-1)/L$, where we assume that Δ is equal to the resolution of the optical system, $\lambda F/L$. For given $u_{i'}$ and $p_{i''}$, $v_{j'} - q_{j''}$ can assume $2N-1$ values. Thus, G is equal to $2N-1$ times S, the number of distinct values that $(u_{i'} - p_{i''})cos\theta\hat{x} + (u_{i'} + p_{i''})\sin\theta\hat{z}$ can assume. Letting $\xi = (u_{i'} - p_{i''})cos\theta$ and $\zeta = (u_{i'} + p_{i''})\sin\theta$, we can estimate S as

$$S \approx \int\int_{\substack{1<i'<N \\ 1<i''<N}} \frac{d\xi d\zeta}{(\frac{\pi}{L})^2} = 4(N-1)^2\sin 2\theta. \quad (3.12)$$

Substituting Δ into Eq. (2.90), we find $\sin 2\theta > 1/4$ and $G > (2N-1)(N-1)^2 \approx 2N^3$.

In order to arbitrarily specify the connections between the inputs and the outputs in the system of Fig. 3.2, the discrepancy between the number of distinct gratings, $2N^3$, and the number of degrees of freedom in a single exposure, $2N^2$, must be recovered by varying the recording fields in time. The simplest method for doing this is to use discrete exposures, in which case,

$$A(\vec{r}) = A_o \sum_{n=1}^{N_e}[\sum_{i'j'}\sum_{i''j''} \frac{t_n}{E_o}\phi_{i'j'}(n)\psi^*_{i''j''}(n)e^{j(\vec{k}_{i'j'} - \vec{k}_{i''j''})\cdot\vec{r}} + c.c.]dt, \quad (3.13)$$

where t_n is the exposure time for the n^{th} exposure and N_e is the total number of exposures. The $2N^3$ distinct gratings in the system of Fig. 3.2 may be controlled in $2N$ exposures using a training plane and the top and bottom rows of the input plane, as described in the previous section. However, if the active pixels of the input and output planes are constrained to lie on sampling grids, it is difficult to produce holograms with coherent fan-in if the hologram is controlled by pixels which will be inactive during reconstruction. We saw in chapter 2 that a hologram which interconnects appropriate sampling grids may be represented by a matrix, $\overset{=}{W}$. $\overset{=}{W}$ can be represented as the sum of R outer-products

$$\overset{=}{W} = \sum_{n=1}^{R} \beta_n \left|\phi^n\right\rangle \left\langle\psi^n\right|, \tag{3.14}$$

R is the rank of the matrix. We can record the hologram corresponding to $\overset{=}{W}$ in R exposures by setting $\phi_{i'j'}(n)$ and $\psi_{i''j''}(n)$ equal to the corresponding components of $\left|\phi^n\right\rangle$ and $\left|\psi^n\right\rangle$.

In this section we have shown that holograms can be recorded in linear media which utilize almost all the degrees of freedom of the volume. The number of exposures needed to form such holograms is equal to the rank of the transformation implemented. In the following section and in chapter 4 we consider the costs that the need to make multiple exposures imposes on the dynamic range of holograms recorded in real media.

3.4 VOLUME HOLOGRAMS IN SATURABLE MEDIA

Real holographic media perform linearly only over a limited range of exposures. The disadvantage of using multiple exposures to record a hologram in such materials is that the dynamic range of the information bearing component of the holographic perturbation is limited by the number of exposures. In this section we consider the simplest form of real holographic media, *i.e.*, media which respond linearly up to a saturation exposure. Photographic film belongs to this class of materials, at least in the sense that it can often be assumed that film responds linearly up to a certain exposure E_{max}. Information is recorded in saturable media by converting some specially prepared light sensitive resource to an insensitive form. In film the resource is silver halide grains, which are converted into metallic silver. Once the exposure exceeds that for which almost all the light sensitive material is converted, further exposure does not affect the medium.

Consider a hologram recorded in a saturable medium using N_e exposures. The recording architecture is assumed to be that of Fig. 3.2. The field during the n^{th} exposure is

$$U^{(n)}(\vec{r}) = \sum_{i'j'} \psi^{(n)}_{i'j'} e^{j\vec{k}_{i'j'}\cdot\vec{r}} + \sum_{i''j''} \phi^{(n)}_{i''j''} e^{j\vec{k}_{i''j''}\cdot\vec{r}}. \tag{3.15}$$

The corresponding intensity is

$$\begin{aligned} I^{(n)}(\vec{r}) = & \sum_{i'j'} |\psi^{(n)}_{i'j'}|^2 + \sum_{i'j'} \sum_{\substack{p'\neq i' \\ q'\neq j'}} \psi^{(n)}_{i'j'} {\psi^{(n)}_{p'q'}}^* e^{j(\vec{k}_{i'j'}-\vec{k}_{p'q'})\cdot\vec{r}} \\ & + \sum_{i''j''} |\phi^{(n)}_{i''j''}|^2 + \sum_{i''j''} \sum_{\substack{p''\neq i'' \\ q''\neq j''}} \phi^{(n)}_{i''j''} {\phi^{(n)}_{p''q''}}^* e^{j(\vec{k}_{i''j''}-\vec{k}_{p''q''})\cdot\vec{r}} \\ & + \sum_{i'j'} \sum_{i''j''} \psi^{(n)}_{i'j'} {\phi^{(n)}_{i''j''}}^* e^{j(\vec{k}_{i'j'}-\vec{k}_{i''j''})\cdot\vec{r}} + \sum_{i'j'} \sum_{i''j''} {\psi^{(n)}_{i'j'}}^* \phi^{(n)}_{i''j''} e^{-j(\vec{k}_{i'j'}-\vec{k}_{i''j''})\cdot\vec{r}} \end{aligned} \tag{3.16}$$

We wish to record a hologram which is a linear superposition of the intensity patterns in all the exposures. This is the case in a saturable material is the exposure time is constant over all exposures and the total exposure is less than the saturation value. In this case,

$$\begin{aligned} A(\vec{r}) =& \frac{A_o t_o}{E_o} \sum_{n=1}^{N_e} \Bigg[\sum_{i'j'} \sum_{i''j''} \psi_{i'j'}^{(n)} \phi_{p''q''}^{(n)}{}^{*} e^{j(\vec{k}_{i'j'}-\vec{k}_{i''j''})\cdot\vec{r}} \\ &+ \sum_{i'j'} \sum_{i''j''} \psi_{i'j'}^{(n)*} \phi_{p''q''}^{(n)} e^{-j(\vec{k}_{i'j'}-\vec{k}_{i''j''})\cdot\vec{r}} \Bigg], \end{aligned} \tag{3.17}$$

where t_o is the exposure time.

To ensure that the hologram is recorded linearly, we require that

$$t_o \sum_{n=1}^{N_e} I^{(n)}(\vec{r}) < E_{max}. \tag{3.18}$$

at almost all $\vec{r}$. If the Fourier components of the recording field are not correlated, then both the mean and the standard deviation of the intensity are given by

$$I_o^{(n)} = \sum_{i'j'} |\psi_{i'j'}^{(n)}|^2 + \sum_{i''j''} |\phi_{i''j''}^{(n)}|^2. \tag{3.19}$$

Statistics of this sort are well known in speckle patterns [117]. Since $I^{(n)}(\vec{r})$ is very unlikely to exceed a few times its standard deviation, we can ensure that hologram formation occurs linearly by choosing t_o such that

$$t_o = \alpha \frac{E_{max}}{\sum_n I_o^{(n)}}, \tag{3.20}$$

where α is a number between zero and one. Substituting into Eq. (3.17) yields

$$\begin{aligned} A(\vec{r}) =& \frac{\alpha A_{max}}{\sum_{n'} I_o^{(n')}} \sum_{n=1}^{N_e} \Bigg[\sum_{i'j'} \sum_{i''j''} \psi_{i'j'}^{(n)} \phi_{p''q''}^{(n)}{}^{*} e^{j(\vec{k}_{i'j'}-\vec{k}_{i''j''})\cdot\vec{r}} \\ &+ \sum_{i'j'} \sum_{i''j''} \psi_{i'j'}^{(n)*} \phi_{p''q''}^{(n)} e^{-j(\vec{k}_{i'j'}-\vec{k}_{i''j''})\cdot\vec{r}} \Bigg] \end{aligned} \tag{3.21}$$

where $A_{max} = (A_o E_{max}/E_o)$. A_{max} is the maximum amplitude of the perturbation which can be recorded linearly.

Our goal in this section is to discover how the effective amplitude of $A(\vec{r})$ scales with the rank of the hologram being recorded. Since we have seen that the sampling grids of Fig. 2.9 can be used to record a hologram in a minimal number of exposures, we can assume without loss of generality that sampling grids are used to record the hologram of Eq. (3.21). In this case, the transformation implemented by the hologram can be expressed as a sum of outer-products. Writing Eq. (3.21) in this form yields the transformation matrix

$$\overline{\overline{W}} = w_o \sum_{n=1}^{N_e} \frac{|\phi^n\rangle\langle\psi^n|}{\sum_{n'}[\langle\psi^{n'}|\psi^{n'}\rangle + \langle\phi^{n'}|\phi^{n'}\rangle]}. \tag{3.22}$$

The choice of the vectors $|\phi^n\rangle$ and $|\psi^n\rangle$ used to record $\overline{\overline{W}}$ is not unique. However, as reflected by the sum over the mean recording intensities in the denominator of Eq. (3.22), the choice of the recording vectors does affect the amplitude of the recorded hologram. In general, we would like to use the recording vectors which yield the maximal amplitude in the recorded hologram.

Orthonormal bases $\{|u_n\rangle\}$ and $\{|v_n\rangle\}$ can be found in the input and output spaces, respectively, by singular value decomposition such that

$$\overline{\overline{W}} = \Gamma \sum_{n=1}^{R'} \beta_n |u_n\rangle\langle v_n|, \tag{3.23}$$

where R is the rank of $\overline{\overline{W}}$, β_n is a positive real number, where Γ is a positive real scaling factor [118]. The inclusion of Γ in this expression allows us to impose the normalization condition

$$\sum_{n=1}^{R} \beta_n = 1. \tag{3.24}$$

An arbitrary set of recording vectors $|\phi^n\rangle$ and $|\psi^n\rangle$ can be expressed in terms of sum over the bases vectors $|u_n\rangle$ and $|v_n\rangle$, respectively, as

$$\begin{aligned} |\phi^n\rangle &= \sum_{n'}^{R} s_{nn'} |u_{n'}\rangle \\ |\psi^n\rangle &= \sum_{n'}^{R} r_{nn'} |v_{n'}\rangle \end{aligned} \tag{3.25}$$

Substituting in Eq. (3.22) yields

$$\overline{\overline{W}} = w_o \sum_{n=1}^{N_e} \sum_{n'}^{R} \sum_{n''}^{R} \frac{s_{nn'} r^*_{nn''} |u_{n'}\rangle \langle v_{n''}|}{\sum_m^{N_e} \sum_{m'}^{R} [(|r_{mm'}|^2 + |s_{mm'}|^2)]}. \tag{3.26}$$

For specific vectors in the orthonormal recording bases $|u_p\rangle$ and $|v_p\rangle$, Eq. (3.23) yields

$$\langle u_p| \overline{\overline{W}} |v_p\rangle = \Gamma \beta_p \tag{3.27}$$

According to Eq. (3.26)

$$\langle u_p| \overline{\overline{W}} |v_p\rangle = w_o \sum_{n=1}^{N_e} \frac{s_{np} r^*_{np}}{\sum_m^{N_e} \sum_{m'}^{R} [(|r_{mm'}|^2 + |s_{mm'}|^2)]}. \tag{3.28}$$

Equating these results we find

$$\beta_p \Gamma = w_o \sum_{n=1}^{N_e} \frac{s_{np} r^*_{np}}{\sum_m^{N_e} \sum_{m'}^{R} [(|r_{mm'}|^2 + |s_{mm'}|^2)]}. \tag{3.29}$$

Summing over p, Eq. (3.29) becomes

$$\Gamma = w_o \frac{\sum_{m=1}^{N_e} \sum_{m'=1}^{R} [s_{mm'} r^*_{mm'}]}{\sum_m^{N_e} \sum_{m'}^{R} [(|r_{mm'}|^2 + |s_{mm'}|^2)]}. \tag{3.30}$$

To maximize the amplitude of the recorded hologram, we pick the coefficients $r_{mm'}$ and $s_{mm'}$ which maximize the scaling factor, Γ. This maximization is achieved by comparing terms in the numerator of the right-hand side of Eq. (3.30) with terms in the same parameters in the denominator. If we maximize the ratios of terms in the numerator with their corresponding terms in the denominator, then the value of the overall fraction is maximized. Thus we wish to choose values of $r_{mm'}$ and $s_{mm'}$ which maximize the fraction

$$\frac{|s_{mm'}r^*_{mm'}|}{|r_{mm'}|^2 + |s_{mm'}|^2}$$

From the triangle inequality we know that the maximum of this ratio occurs when $r_{mm'} = s_{mm'}$. We also know that a solution for suitable recording vectors with coefficients $r_{mm'}$ and $s_{mm'}$ which satisfy this condition exists, namely $r_{mm'} = s_{mm'} = \delta_{mm'}$. Assuming $r_{mm'} = s_{mm'}$,

$$\Gamma = \frac{w_o}{2}. \tag{3.31}$$

Thus, when the hologram is recorded with the maximal strength in its components,

$$\bar{\bar{W}} = \frac{w_o}{2} \sum_{n=1}^{R'} \beta_n \left|u_n\right\rangle \left\langle v_n\right| . \tag{3.32}$$

This transformation can be recorded between the vectors $|u_n\rangle$ and $|v_n\rangle$ if the recording time of the n^{th} exposure is proportional to β_n. When the outer-products between the basis vectors are all stored with equal weight, the normalization condition yields $\beta_n = 1/R$ and

$$\bar{\bar{W}} = \frac{w_o}{2R} \sum_{n=1}^{R'} \left|u_n\right\rangle \left\langle v_n\right| . \tag{3.33}$$

The diffraction efficiency, *i.e.*, the ratio of the diffracted intensity to the undiffracted intensity, when a pattern $|\psi\rangle$ on the input plane reconstructs the hologram is

$$\eta_\psi = \frac{\langle\psi| \overline{\overline{W}}^T \overline{\overline{W}} |\psi\rangle}{\langle\psi|\psi\rangle}. \tag{3.34}$$

Assuming that $|\psi\rangle$ has no component in the nul space of $\overline{\overline{W}}$, $|\psi\rangle$ can be decomposed as $|\psi\rangle = \sum_n^R r_n |v_n\rangle$ so that

$$\eta_\psi = \eta_{max} \frac{\sum^R \beta_n^2 |r_n|^2}{\sum^R |r_n|^2}. \tag{3.35}$$

η is maximal when $r_m = 1$, where β_m is the largest singular value of $\overline{\overline{W}}$. Our purpose in using a volume rather than a thin medium is to store as many different associations as possible, we are not typically interested in holograms where one singular value is much larger than the others. If all the singular values are approximately equal, then the maximum diffraction efficiency is achieved when any basis vector is used to reconstruct the hologram. When $|v_n\rangle$ is active on the input plane the diffraction efficiency is

$$\eta_n = \eta_{max}\beta_n^2 \approx \frac{\eta_{max}}{R^2}, \tag{3.36}$$

where we have again applied the normalization condition on β_n.

In summary, arbitrary volume holograms can be recorded linearly in saturable materials by dividing the dynamic range of the material equally between the exposures needed to record the hologram. The hologram is recorded with maximal strength if the orthogonal basis vectors generated by singular value decomposition are used to record it. Assuming that each exposure is weighted equally, the maximum diffraction efficiency of the hologram falls with the square of the rank of the transformation recorded, R. If our goal is to store associations between orthogonal patterns, this means that the diffraction efficiency when R associations are recorded is R^2 times less than the diffraction efficiency when 1 association is recorded. Because various techniques, such as placing the hologram in an cavity, can be used to improve the diffraction efficiency of a hologram, the fact that the amplitude of the perturbation falls with R is more significant in the sense that it limits the dynamic range, and thus the storage capacity, of the hologram. In chapter 4 we show that a similar loss of dynamic range under multiple exposure occurs in photorefractive materials, which are not saturable in the sense described above. The rest of this chapter is devoted to considering holograms formed using multiple independent wavelengths instead of multiple exposures.

3.5 CONTROL OF VOLUME HOLOGRAMS WITH POLYCHROMATIC SIGNALS

As an alternative to multiple exposure, a hologram may be controlled in a single exposure using polychromatic light. Since the number of degrees of freedom in a single wavelength component of a polychromatic field cannot exceed C, as derived above, at least $R = G/C$ distinct wavelengths are needed to control an arbitrary hologram. In this section we show that a hologram recorded in N_e exposures can be recorded in one exposure using signals carried independently on N_e wavelengths. Polychromatic control is potentially superior to monochromatic multiply-exposed control because coherence between different wavelength components could yield greater modulation depth and thus better dynamic range in the perturbation. Unfortunately, the difficulty of simultaneously and independently specifying signals in many wavelengths may outweigh the potential advantages of polychromatic control in practical systems.

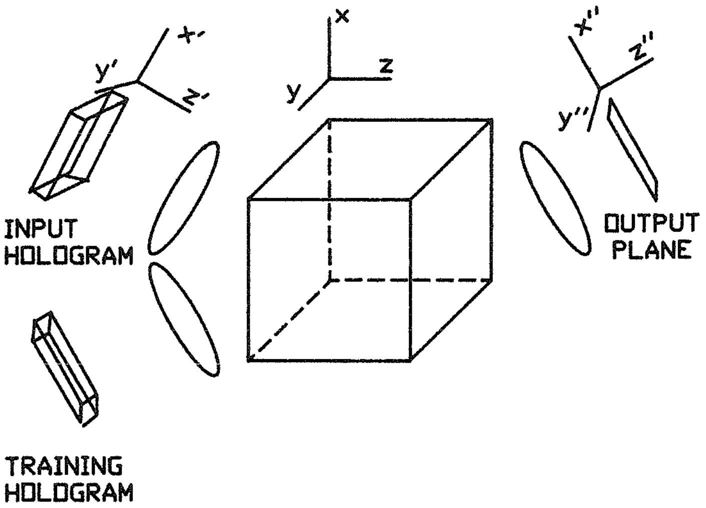


Figure 3.3. Polychromatically controlled volume interconnect system.

Consider polychromatic control of the system of Fig. 3.2. The SLMs of Fig. 3.2 must be replaced by devices capable of separately controlling each wavelength component of the control beams. One means of constructing such devices is to use N_e distinct SLMs, each illuminated by a different wavelength, to generate the input and training signals. The input and training planes would consist of superimposed images of the signals on all the corresponding SLMs. A second approach would be to replace the input and training SLMs with volume holograms which store an appropriate image at each wavelength. Each hologram is simultaneously reconstructed by N_e beams with different wavelengths. The reconstructed fields of different wavelengths must, of course, be independent. A sketch of a system incorporating such holograms is shown in Fig. 3.3. Supposing for the moment that one of these methods can be used to generate appropriate images, we assume that a field of the form given in Eq. (3.15) is generated in the volume at each wavelength. The total field in the volume is

$$U(\vec{r},t) = \sum_{n=1}^{N_e} \left[\sum_{i'j'} \psi_{i'j'}{}^{(n)} e^{j(\vec{k_{i'j'}}^{(n)} \cdot \vec{r} - \omega^{(n)} t)} + \sum_{i''j''} \phi_{i''j''}{}^{(n)} e^{j(\vec{k_{i''j''}}^{(n)} \cdot \vec{r} - \omega^{(n)} t)} \right], \tag{3.37}$$

where $\vec{k_{i'j'}}^{(n)}$ and $\omega^{(n)}$ are the wavevector of the $i'j'^{th}$ component and the frequency for the n^{th} wavelength. The perturbation that forms in the volume is

$$A(\vec{r}) = \frac{A_o}{E_o} \int_0^T |U(\vec{r},t)|^2 dt. \tag{3.38}$$

If $T \gg \left(\omega^{(n)} - \omega^{(n')}\right)^{-1}$ for all $n' \neq n$, then interference patterns due to each wavelength add incoherently. In this case

$$A(\vec{r}) = \frac{A_o T}{E_o} \sum_{n=1}^{N_e} I^{(n)}(\vec{r}), \tag{3.39}$$

where $I^{(n)}(\vec{r})$ is exactly as written in Eq. (3.16) with the proviso that n now corresponds to the n^{th} wavelength rather than the n^{th} exposure. Assuming that the statistics for the field amplitudes are the same as in the previous section, the requirement that the exposure remain less than E_{max} forces T to be less than or equal to the value given in Eq. (3.18). This means that hologram formation using N_e independently controlled mutually incoherent wavelengths yields exactly the perturbation described by Eq. (3.21) and that, in particular, the dependence of the modulation terms in $A(\vec{r})$ on N_e is exactly the same as for control using multiple exposures.

The modulation depth of the signal terms in $A(\vec{r})$ might be increased if the integration time T were short enough that modulation terms could form between different wavelengths. Suppose that $\omega^{(m)} = \omega_o + m\omega$, where $\omega_o \gg \omega$. In this case,

$$\begin{aligned} A(\vec{r}) =& \frac{A_o T}{E_o} \sum_{m=1}^{N_e} \sum_{m'=1}^{N_e} e^{-j\frac{\omega T(m-m')}{2}} \frac{2\sin\left(\frac{\omega T(m-m')}{2}\right)}{\omega T(m-m')} \\ & \times \left[\sum_{i'j'} \sum_{p'q'} \psi_{i'j'}^{(m)} \psi_{p'q'}^{(m')^*} e^{j(\vec{k}_{i'j'}{}^{(m)} - \vec{k}_{p'q'}{}^{(m')})\cdot\vec{r}} \right. \\ & + \sum_{i''j''} \sum_{p''q''} \phi_{i''j''}^{(m)} \phi_{p''q''}^{(m')^*} e^{j(\vec{k}_{i''j''}{}^{(m)} - \vec{k}_{p''q''}{}^{(m')})\cdot\vec{r}} \\ & + \sum_{i'j'} \sum_{i''j''} \psi_{i'j'}^{(m)} \phi_{i''j''}^{(m')^*} e^{j(\vec{k}_{i'j'}{}^{(m)} - \vec{k}_{i''j''}{}^{(m')})\cdot\vec{r}} \\ & \left. + \sum_{i'j'} \sum_{i''j''} \psi_{i'j'}^{(m')^*} \phi_{i''j''}^{(m)} e^{-j(\vec{k}_{i'j'}{}^{(m')} - \vec{k}_{i''j''}{}^{(m)})\cdot\vec{r}} \right], \end{aligned} \tag{3.40}$$

where we have discarded integrals over $e^{\pm 2j\omega_o T}$ under the assumption that $\omega_o T \gg 1$. If $N_e \omega T < 1$ then

$$e^{-j\frac{\omega T(m-m')}{2}} \frac{2\sin\left(\frac{\omega T(m-m')}{2}\right)}{\omega T(m-m')} \approx 1 \tag{3.41}$$

for all m', m. To ensure that each wavelength represents an independent recording channel we must require that $|\vec{k}_{i'j'}^{(m)} - \vec{k}_{i'j'}^{(m')}| \geq 2\pi/L$. Assuming that the propagation directions are nearly colinear, this is equivalent to assuming $||\vec{k}_{i'j'}^{(m)}| - |\vec{k}_{i'j'}^{(m')}|| \geq 2\pi/L$. Since $|\vec{k}_{i'j'}^{(m)}| = (\omega_o + m\omega)/c$, this constraint implies that $\omega \geq 2\pi c/L$.

To find a set of mode amplitudes which can be used to record a given hologram, we set the terms in each grating wavevector of Eq. (3.40) equal to the corresponding terms in the Fourier representation of the desired value of $A(\vec{r})$

and solve for the mode amplitudes, *i.e.*, we generate a set of equations constraining the mode amplitudes by setting

$$\sum_{k_{i'j'}^{(m)}} \psi_{i'j'}^{(m)} \phi^*(\vec{k}_{i'j'}^{(m)} - \vec{K}_g)$$

equal to the coefficient of Fourier component of $A(\vec{r})$ with wavevector $\vec{K}_g$. By $\phi^*(\vec{k}_{i'j'}^{(m)} - \vec{K}_g)$ we mean the coefficient of the output mode propagating with wavevector $\vec{k}_{i'j'}^{(m)} - \vec{K}_g$. The difference between monochromatic and polychromatic recording with cross-wavelength coherence is that in the coherent polychromatic case,

$$\psi_{i'j'}^{(m)} \phi^*(\vec{k}_{i'j'}^{(m)} - \vec{K}_g)$$

can be nonzero even when $\vec{k}_{i'j'}^{(m)} - \vec{K}_g$ does not lie on the same normal surface as $\vec{k}_{i'j'}^{(m)}$. In this sense, an effective thickness is added to the normal surface and the control space becomes three-dimensional.

The three-dimensional nature of the control space with coherent polychromatic control allows us to record an arbitrary holographic transformation using a single reference beam. A grating with wavevector $\vec{K}_g$ is controlled by the interference pattern formed by the reference and signal field separated from the reference by $\vec{K}_g$ in the three-dimensional control space. Since the amplitude and phase of the reference are fixed, the relative phase and amplitude of the grating correspond to the phase and amplitude of the signal field. The modulation depth for gratings recorded by this approach is maximum when the total intensity in all the signal fields is equal to the reference field. In this case, the mean intensity of a signal, $\bar{I}_s$ is N_e times less than the reference intensity, I_r. The amplitude of a signal field is $\sqrt{N_e}$ times less than the amplitude of the reference field. The mean modulation depth, $\bar{m}$, for a grating in the interference pattern is

$$\bar{m} = \frac{2\sqrt{I_r \bar{I}_s}}{I_r + N_e \bar{I}_s} = \frac{1}{\sqrt{N_e}}. \tag{3.42}$$

In both the saturable media considered in this chapter and in the photorefractive media considered in chapter 4 the maximum amplitude of a perturbation formed in a single exposure is linearly proportional to the modulation depth of the recording interference pattern. With this fact in mind, Eq. (3.42) suggests that coherent polychromatic control would allow us to record an arbitrary volume hologram with a fall off in amplitude proportional to the square root of the rank of the hologram and a fall off in diffraction efficiency proportional to the rank of the hologram. The diffraction efficiency would thus be R times better than for the same hologram stored using multiple exposures.

The principle difficulty involved in recording coherently with polychromatic light is the problem of limiting T. If we assume that $N_e = 1000$, $L = 1$ cm and $\lambda = 1$ μm, then the conditions which follow Eq. (3.40) imply that

$$10^{-15}\text{sec} \ll T < 10^{-13}\text{sec}. \tag{3.43}$$

One method of limiting T to this range would be to use a holographic medium which is excited in a two-step process. If the hologram were sensitized by a short pulse during recording with the signals of Eq. (3.37) then the integration time could be suitably limited. An alternative approach which is not material dependent is to assume that the recording signals themselves are riding on such a pulse. Unfortunately, it would be extremely difficult to generate $N_e \sim 1000$

pulses shifted coherently in frequency with respect to one another and spatially modulated in 2-D. The prime candidate for controlling such a set of pulses, a volume hologram, cannot perform the task in a straight forward way because the temporal bandwidth of the pulses must be wider than the Bragg sensitivity of the hologram, resulting in extreme dispersion problems. Thus, while the combination of temporal and spatial signals to control volume holograms is of considerable interest, pulsed control seems unlikely to directly challenge sequential control in the near term.

4. FORMATION OF PHOTOREFRACTIVE VOLUME HOLOGRAMS

4.1 INTRODUCTION

Photorefractive holograms consist of photogenerated space charge distributions in electro-optic materials. The charge patterns modulate the index of refraction of the material via the electro-optic effect. The formation of photorefractive holograms may be described by dynamical equations coupling the free carrier densities, the local trap densities, the current density, the space charge field, and the optical fields. Physical models which yield suitable dynamical equations have been developed by Kukhtarev *et al.* [19], Moharam *et al.* [119], and Feinberg *et al.* [120].

In this thesis we consider information storage in photorefractive crystals. In order to effectively control the information recording process, we would like hologram formation to occur as linearly as possible. In this chapter we consider Kukhtarev's band transport model with particular emphasis on situations in which the holographic component of the charge pattern grows linearly with exposure. In chapter 3 we argued that information storage in volume holograms requires multiple exposures and we developed an exposure procedure for recording holograms which implement arbitrary transformations between sampled planes. Here, we develop an exposure schedule which allows us to apply the results of chapter 3 to hologram formation in photorefractive crystals.

The second section of this chapter is a review of the single active trapping species, single carrier version of the band transport model. Solutions to the band transport equations have been analyzed in great detail by many authors [20, 21].

While the analysis in this chapter mainly follows previous solutions, the emphasis on many grating holograms is not common in the photorefractive literature. We do not consider dynamic beam coupling effects in our discussion. Such effects can be suppressed by writing polarization switching gratings or by writing in the drift dominated regime. In the third section of this chapter we consider cases in which the formation of photorefractive holograms may be linearized. The key to being able to record linearly is an exposure schedule which uniformly weights each exposure. The fourth section considers hologram formation in the presence of multiple photoactive trapping species. Multiple species lead to multiple response times in the recording process, which complicates the design of an exposure schedule. In the fifth section we consider a copying technique for improving the diffraction efficiency of linearly recorded photorefractive holograms in adaptive systems. The final section considers the impact of fixing mechanisms on hologram formation.

4.2 DYNAMICS OF THE PHOTOREFRACTIVE EFFECT

In the simplest case, a photorefractive hologram forms as a result of charge migration via a single carrier from neutral local trapping sites to ionized sites of the same species. The dynamics of hologram formation in this case are described by the charge continuity equation

$$\frac{\partial n}{\partial t} = \frac{\partial N_D^+}{\partial t} + \frac{1}{e}\vec{\nabla} \cdot \vec{J}, \tag{4.1}$$

the charge generation rate equation

$$\frac{\partial N_D^+}{\partial t} = (\beta + sI)(N_D - N_D^+) - \gamma n N_D^+, \tag{4.2}$$

the current equation

$$\vec{J} = e\mu n\vec{E} + \mu k_B T\vec{\nabla}n + \vec{J}_{ph}, \tag{4.3}$$

and the Poisson equation

$$\vec{\nabla} \cdot \epsilon\vec{E} = -e(n + N_A^- - N_D^+), \tag{4.4}$$

where

n is the excited carrier number density.

N_D is the donor species number density.

N_D^+ is the ionized donor number density.

N_A^- is the number density of ionized acceptors.

$\vec{J}$ is the current density.

$\vec{E}$ is the quasi-static electric field.

e is the electron charge.

β is the donor thermal ionization rate.

sI is the photoionization rate.

I is the illuminating intensity.

γ is the carrier-trap recombination rate.

μ is the carrier mobility.

k_B is Boltzman's constant.

T is the temperature.

$\vec{J}_{ph}$ is the photovoltaic current.

ϵ is the static permittivity.

While we assume for the present that the acceptors are not photoionizable, N_A^- must be nonzero in order for the amplitude of the spatial variation in N_D^+ to be nonzero.

We are again interested in forming holograms in the system sketched in Fig. 3.3. As described in chapters 2 and 3, the optical field is generated on a pair of SLMs. For reasons described below we will assume that cross-gratings recorded between signals generated on the same SLM may be ignored. In order to simplify our analysis, we will assume that the ratio of the total intensity generated on the input SLM to the total intensity generated on the training SLM is $m_o \ll 1$. The recording intensity is

$$I(\vec{r}) = I_o \left(1 + \sum_{ij} m_{ij} \cos(\vec{K}_{ij} \cdot \vec{r} + \varphi_{ij}) \right). \tag{4.5}$$

m_{ij} and φ_{ij} are the modulation depth and phase of the fringe pattern between the i^{th} input beam and the j^{th} training beam. I_o is the spatial mean of the recording intensity. The recording intensity can, of course, be affected by self-diffraction effects, in which case the wave equation for the optical field is coupled with the band transport equation. We will assume, however, that such effects are suppressed. In practice, suppression is often possible by judicious choice of the recording geometry, the beam polarizations, and the applied field.

We can expand the spatial and temporal dependence of n, N_D^+, $\vec{E}$, and $\vec{J}$, in

Fourier series in harmonics of the wavevectors $\vec{K}_{ij}$:

$$n(\vec{r},t) = n_o + \sum_{\substack{p_{00}=-\infty \\ \vec{p}\neq 0}}^{\infty} \sum_{p11=-\infty}^{\infty} \cdots \sum_{p_{N_1N_2}=-\infty}^{\infty} n_{\vec{p}} e^{j\sum_{ij} p_{ij}\vec{K}_{ij}\cdot\vec{r}} \tag{4.6}$$

$$N_D^+(\vec{r},t) = N_{D\,o}^+ + \sum_{\substack{p_{00}=-\infty \\ \vec{p}\neq 0}}^{\infty} \sum_{p11=-\infty}^{\infty} \cdots \sum_{p_{N_1N_2}=-\infty}^{\infty} N_{D\,\vec{p}}^+ e^{j\sum_{ij} p_{ij}\vec{K}_{ij}\cdot\vec{r}} \tag{4.7}$$

$$\vec{E}(\vec{r},t) = \vec{E}_o + \sum_{\substack{p_{00}=-\infty \\ \vec{p}\neq 0}}^{\infty} \sum_{p11=-\infty}^{\infty} \cdots \sum_{p_{N_1N_2}=-\infty}^{\infty} \vec{E}_{\vec{p}} e^{j\sum_{ij} p_{ij}\vec{K}_{ij}\cdot\vec{r}} \tag{4.8}$$

$$\vec{J}(\vec{r},t) = \vec{J}_o + \sum_{\substack{p_{00}=-\infty \\ \vec{p}\neq 0}}^{\infty} \sum_{p11=-\infty}^{\infty} \cdots \sum_{p_{N_1N_2}=-\infty}^{\infty} \vec{J}_{\vec{p}} e^{j\sum_{ij} p_{ij}\vec{K}_{ij}\cdot\vec{r}}, \tag{4.9}$$

where the ij^{th} component of $\vec{p}$ is p_{ij}. Substituting the Fourier expansions into the band transport equations and matching zeroth order Fourier components we find

$$\frac{dn_o}{dt} = \frac{dN_{D\,o}^+}{dt} \tag{4.10}$$

$$\begin{aligned} \frac{dN_{D\,o}^+}{dt} =& (\beta + sI_o)(N_D - N_{D\,o}^+) - \gamma n_o N_{D\,o}^+ \\ & - \sum_{ij} \left[\frac{1}{2} s I_o \frac{m_{ij}}{2} N^+_{D\,(p_{ij}=-1)} e^{j\varphi_{ij}} + \frac{1}{2} s I_o \frac{m_{ij}}{2} N^+_{D\,(p_{ij}=1)} e^{-j\varphi_{ij}} \right. \\ & \left. + \gamma n_{(p_{ij}=1)} N^+_{D\,(p_{ij}=-1)} + \gamma n_{(p_{ij}=-1)} N^+_{D\,(p_{ij}=1)} \right], \end{aligned} \tag{4.11}$$

$$\vec{J}_0 = e\mu n_o \vec{E}_o + \vec{\kappa}\alpha_o I_o + \sum_{ij} e\mu n_{(p_{ij}=-1)} \vec{E}_{(p_{ij}=1)} + \sum_{ij} e\mu n_{(p_{ij}=1)} \vec{E}_{(p_{ij}=-1)} \tag{4.12}$$

$$n_o + N_A^- = N_{D\,o}^+, \tag{4.13}$$

where we have noted that the magnitude of q^{th} order Fourier components will be of q^{th} order in the modulation depths and we have kept terms to second order.

$\vec{E}_o$ is an applied field. We have assumed that the photovoltaic current is linear in the illuminating intensity and the absorption, α. $\vec{\kappa}$ is the Glass constant [121].

In deriving the above equations we have made use of the fact that the first order terms in the Fourier expansion are of first order in the modulation depth of their driving terms. While the second order terms in the above equations are small compared to the zeroth order terms, the sum in Eq. (4.11) may scale with the zeroth order terms. If the phases of the Fourier components in the intensity pattern are uncorrelated and all the Fourier components obey the same statistics, then this sum is unlikely to exceed N_g times the variance of the squared modulation terms, m_o/N_g. N_g is the number of distinct Fourier components in the recording intensity. This sum can be neglected only if we make the assumption that $m_o \ll 1$.

The "quasi-steady approximation" [122] consists of the assumption that the charge excitation and recombination dynamics are much faster than the charge migration effects which cause the buildup of the space charge field. Under this approximation the carrier density n may be assumed to respond instantaneously to the current state of the illuminating field and the trap density. Stated in this way the approximation differs slightly from the traditional quasi-steady approximation, which has been taken to imply only that the spatial mean of the carrier density responds instantaneously to the illuminating intensity and is constant during illumination. The distinction between the two cases is developed in section 4.4. For the present problem, the approximation consists of discarding the higher order terms in the driving equations for the zeroth order Fourier coefficients and assuming that the time derivatives of these coefficients are zero.

Under this approximation,

$$(\beta + sI_o)(N_D - N^+_{D\,o}) = \gamma n_o N^+_{D\,o} \tag{4.14}$$

$$\vec{J}_o = e\mu n_o \vec{E}_o + \vec{\kappa}\alpha_o I_o \tag{4.15}$$

$$n_o + N^-_A = N^+_{D\,o}, \tag{4.16}$$

Typically the recombination rate is much greater than the charge excitation rate, in which case $N^+_{D\,o} \approx N^-_A$ and

$$n_o \approx \frac{sI_o N_D}{\gamma N^-_A}. \tag{4.17}$$

The first order Fourier components are described by differential equations in the odd order Fourier coefficients. The third and higher order terms can be dropped relative to the first order terms using exactly the same approximation as was used to drop the second order terms in the driving equations for the zeroth order components. Let $x_{ij,1}$ represent the first order Fourier coefficient of the parameter x for the grating wavevector $\vec{K}_{ij}$. To first order in modulation depth, the band transport equations for the first order Fourier coefficients corresponding to the ij^{th} component of the illuminating intensity are

$$\frac{dn_{ij,1}}{dt} = \frac{dN^+_{D\,ij,1}}{dt} + \frac{j\vec{K}_{ij} \cdot \vec{J}_{ij,1}}{e} \tag{4.18}$$

$$\frac{dN^+_{D\,ij,1}}{dt} = sI_o \frac{m_{ij}}{2} e^{j\varphi_{ij}} (N_D - N^+_{D\,o}) - sI_o N^+_{D\,ij,1} - \gamma n_o N^+_{D\,ij,1} - \gamma n_{ij,1} N^+_{D\,o} \tag{4.19}$$

$$\vec{J}_{ij,1} = e\mu n_{ij,1}(\vec{E}_o + j\vec{E}_{d,ij}) + e\mu n_o \vec{E}_{ij,1} + \vec{\kappa}\alpha_o I_o \frac{m_{ij}}{2} e^{j\varphi_{ij}} + \vec{\kappa}\alpha_{ij,1} I_o \tag{4.20}$$

$$j\epsilon\vec{K}_{ij}\cdot\vec{E}_{ij,1} = -e(n_{ij,1} - N^{+}_{D\,ij,1}), \tag{4.21}$$

where we assume that the thermal excitation rate is much less than the photoexictation rate and we define the diffusion field $\vec{E}_{d,ij} \equiv k_B T\vec{K}_{ij}/e$. The Fourier coefficients for the absorption can be related to the trap densities using the relationship

$$\alpha = s\hbar\omega V(N_D - N^{+}_{D}), \tag{4.22}$$

where V is the volume of the crystal. Thus,

$$\alpha_{ij,1} = s\hbar\omega V N^{+}_{D\,ij,1} = \alpha_o \frac{N^{+}_{D\,ij,1}}{(N_D - N^{+}_{D\,o})}. \tag{4.23}$$

The quasi-steady approximation, as stated above, amounts to an assumption that n is a "slave" variable. This approximation allows us to assume that $dn_{ij,1}/dt = 0$, in which case Eqs. (4.18)-(4.21) can be reduced to a first order differential equation in the first order Fourier coefficient for the field. Since there are no driving terms for the field normal to the grating wavevector, we assume that each first order Fourier coefficient is parallel to the corresponding grating wavevector. In this case, we can substitute the scalar $E_{ij,1}$ for $\vec{E}_{ij,1}$. Solving for the field in the above equations we find

$$\tau_t \frac{dE_{ij,1}}{dt} = -E_{ij,1} + m_{ij}e^{j\varphi_{ij}}E_s \tag{4.24}$$

where

$$\tau_t = \frac{N^{-}_{A}}{sI_o N_D}\frac{E_{o\|} + j(E_{d,ij} + E_m)}{E_{o\|} - \frac{N^{-}_{A}}{N_D}E_{ph} + j(E_{d,ij} + E_q)} \tag{4.25}$$

and

$$E_s = -\frac{1}{2}\frac{[E_{o\|} + E_{ph} + jE_{d,ij}]}{1 - j(E_{o\|} + \frac{N_A^-}{N_D}E_{ph} + jE_{d,ij})/E_q} \tag{4.26}$$

where

$$E_{o\|} = \frac{\vec{E}_o \cdot \vec{K}_{ij}}{|\vec{K}_{ij}|}$$

$$E_{ph} = \frac{\vec{\kappa} \cdot \vec{K}_{ij} \alpha_o I_o}{|\vec{K}_{ij}| e \mu n_o}$$

$$E_m = \frac{\gamma N_{D_o}^+}{\mu |\vec{K}_{ij}|}$$

$$E_q = \frac{e N_{D_o}^+}{\epsilon |\vec{K}_{ij}|}$$

E_q is the saturation limit for the space charge field based on the maximum amplitude of the ionized trapping species.

Assuming constant mean intensity, the solution to Eq. (4.24) is

$$E_{ij,1}(t) = e^{\frac{-t}{\tau_t}} \frac{E_s}{\tau_t} \int_0^t m_{ij}(s) e^{j\varphi_{ij}(s)} e^{\frac{s}{\tau_t}} ds + e^{\frac{-t}{\tau}} E_{ij,1}(0) \tag{4.27}$$

If the intensity is not constant, the solution is complicated by the intensity dependence of the effective photovoltaic field. In materials with no photovoltaic field or in which $N_A^- \ll N_D$, Eq. (4.24) can be recast as a differential equation in the mean exposure by defining $p = I_o t$. In this case,

$$E_{ij,1}(p) = e^{\frac{-p}{\gamma_t}} \frac{E_s}{\gamma_t} \int_0^p m_{ij}(s) e^{j\varphi_{ij}(s)} e^{\frac{s}{\gamma_t}} ds + e^{\frac{-p}{\gamma_t}} E_{ij,1}(0) \tag{4.28}$$

where $\gamma_t = I_o \tau_t$. Eqs. (4.27) and (4.28) form the basis of our discussion of the linear formation of many grating photorefractive holograms. Note that the time constant for diffusion dominated holograms is function of the grating frequency.

In the interconnect architectures considered here this fact is unlikely to create problems because the bandwidth of the interconnections is typically small compared to the mean grating frequency.

4.3 LINEAR FORMATION OF PHOTOREFRACTIVE HOLOGRAMS

Our goal in this section is to determine the effect of storing multiple grating volume holograms in photorefractive media. The analysis is very similar the analysis used in section 3.4 to consider holograms in saturable media. Our first task in this section is to establish the link between the space charge field of the previous section and the holographic perturbation analyzed in section 2.3. Once we have shown the correspondence, we revert back to the abstract representation of the perturbation which was used in chapter 3.

The space charge field modulates the index of refraction of a photorefractive crystal via the linear electro-optic effect. This effect is described by electro-optic coefficients, r_{ijk}, which are traditionally defined in terms of the perturbation to the impermeability tensor, η,

$$\triangle\eta_{ij} = r_{ijk}E_k, \tag{4.29}$$

where E_k is the component of the space charge field along the k^{th} axis. The corresponding perturbation to the permittivity tensor is

$$\triangle\epsilon = -\epsilon \mathbf{r}\vec{E}\epsilon. \tag{4.30}$$

We saw in chapter 2 that for low diffraction efficiencies the field diffracted by a hologram is linearly proportional to $\hat{e}_2\cdot\triangle\epsilon\hat{e}_1$ where $\hat{e}_1$ and $\hat{e}_2$ are the polarizations of the coupled beams. The field diffracted by the grating with wavevector $\vec{K}_{ij}$ is proportional to $-\epsilon_{eff}^2 r_{eff} E_{ij,1}$ where by definition

$$\epsilon_{eff}^2 r_{eff} = \hat{e}_2 \cdot \epsilon \mathbf{r} \frac{\vec{K}_{ij}}{|\vec{K}_{ij}|} \epsilon \hat{e}_1. \tag{4.31}$$

A plot of $(n^3 r_{eff}/2)^2$ as a function of the angle between $\vec{K}_{ij}$ and the c axis for $BaTiO_3$ is shown in Fig. 4.1. Both beam polarizations are assumed to lie in the plane of incidence.

In a holographic interconnection system such as that sketched in Fig. 3.3, the spatial bandwidth of the input and training signals may be assumed to be small enough that the effective coupling constant does not vary over the bandwidth used in grating space. Small variations in the effective coupling constant might be compensated by adaptive feedback in neural network applications. Typically, the angle between the input and training planes is selected to position the volume used in grating space near a maxima of the effective coupling constant.

Assuming that the effective coupling constant is uniform over the region of interest in grating space, the connection between an input and an output pixel is completely described by the amplitude of the first Fourier component of the space charge field of the corresponding grating. Thus, a one-to-one correspondence exists between the amplitudes of the first order Fourier components of the fields and the Fourier coefficients of the perturbation $A(\vec{r})$ discussed in chapter 3. Using the formalism developed in chapter 3, we see from Eq. (4.27) that for constant mean recording intensity the analog of Eq. (3.13) for photorefractive holography

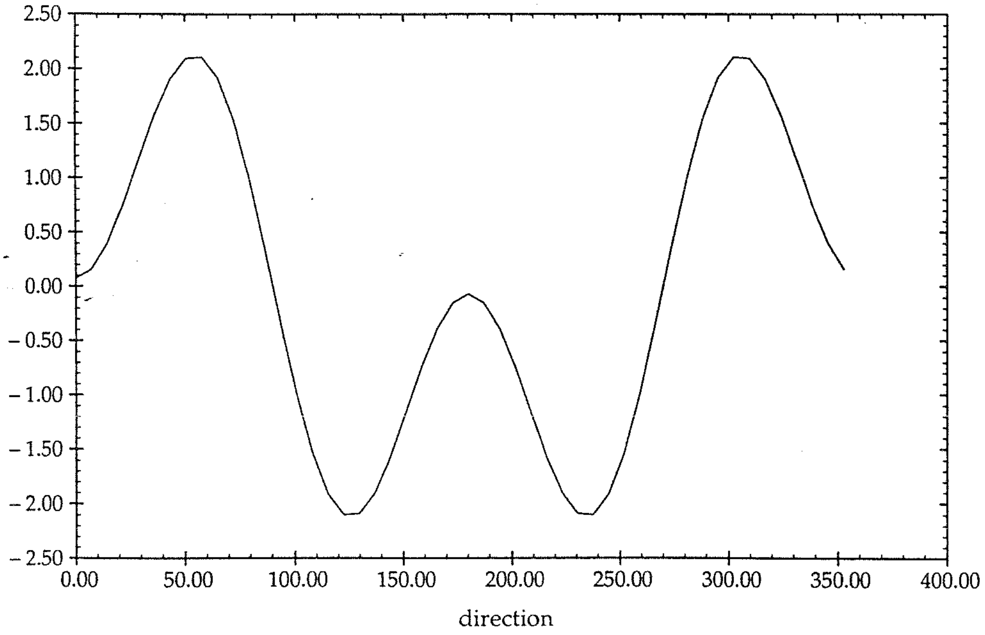


Figure 4.1. Effective coupling constant versus the angle between the grating wavevector and the c axis in $BaTiO_3$.

is

$$A(\vec{r},t) = e^{\frac{-t}{\tau_t}} \frac{A_o}{E_o} \int_{t_o}^{t} e^{\frac{s}{\tau_t}} [\sum_{i'j'} \sum_{i''j''} \phi_{i'j'}(s) \psi^*_{i''j''}(s) e^{j(\vec{k}_{i'j'} - \vec{k}_{i''j''})\cdot\vec{r}}] ds + c.c., \tag{4.32}$$

where $E_o = I_o \tau_t$. The exponential decay of previously recorded signals as more information is recorded creates a "window" in exposure such that after a certain characteristic exposure previously stored information is lost.

Consider a hologram recorded in discrete exposures. During each individual exposure the recording signals are constant. In this case Eq (4.32) becomes

$$\begin{aligned} A(\vec{r}) &= e^{\frac{-t}{\tau_t}} \frac{A_o}{E_o} \sum_{n=1}^{N_e} \int_{T_n}^{T_n+t_n} e^{\frac{s}{\tau_t}} [\sum_{i'j'} \sum_{i''j''} \phi_{i'j'}(n) \psi^*_{i''j''}(n) e^{j(\vec{k}_{i'j'} - \vec{k}_{i''j''})\cdot\vec{r}}] ds + c.c. \\ &= \frac{A_o}{I_o} \sum_{n=1}^{N_e} (1 - e^{\frac{-t_n}{\tau_t}}) e^{-\sum_{n'=n+1}^{N_e} t_{n'} \tau_t} [\sum_{i'j'} \sum_{i''j''} \phi_{i'j'}(n) \psi^*_{i''j''}(n) e^{j(\vec{k}_{i'j'} - \vec{k}_{i''j''})\cdot\vec{r}}] \\ &\quad + c.c. \end{aligned} \tag{4.33}$$

where t_n is the time of the n^{th} exposure, $T_n = \sum_{n'=1}^{n-1} t_{n'}$, and N_e is the total number of exposures recorded.

In order to form a hologram which is a linear sum of the recording signals over all the exposures, t_n must be selected such that the factor

$$f_n = (1 - \exp(-t_n/\tau_t)) \exp(-\sum_{n'>n} t_{n'}/\tau_t)$$

is constant for all n. This may be achieved in a relatively simple manner if τ_t is real. From Eq. (4.25) we see that this is the case when $E_o = E_{ph} = 0$. Choosing

$$t_n = \tau_t \log\left(\frac{1 + (n-1)\chi}{1 + (n-2)\chi}\right), \tag{4.34}$$

where $\chi = 1 - \exp(-t_1/\tau_t)$, yields

$$(1 - e^{\frac{-t_n}{\tau_t}})e^{-\sum_{n'=n+1}^{N_e} t_{n'}\tau_t} = \frac{\chi}{1 + (N_e - 1)\chi} \tag{4.35}$$

and

$$A(\vec{r}) = \frac{A_o \chi}{I_o(1 + (N_e - 1)\chi)} \sum_{n=1}^{N_e} [\sum_{i'j'} \sum_{i''j''} \phi_{i'j'}(n) \psi^*_{i''j''}(n) e^{j(\vec{k}_{i'j'} - \vec{k}_{i''j''})\cdot\vec{r}}] + c.c. \tag{4.36}$$

A particularly simple result occurs when $t_1 \gg \tau_t$, in which case $\chi \approx 1$. By following this exposure schedule, the formation of photorefractive holograms becomes linear in the sense discussed in chapter 3. Using the formalism developed there, an arbitrary volume hologram can be recorded in a crystal. Unfortunately, this linearization of the recording process exacts a cost to the dynamic range of the recorded hologram, as evidenced by the dependence of the amplitude of the perturbation on N_e^{-1}. Recall that the same inverse dependence on the number of exposures recorded was found in chapter 3 for saturable media. In chapter 3 this factor was found to cause a R^{-2} fall off in the maximum diffraction efficiency which can be achieved for a given interconnect pattern. Since the perturbation of Eq. (4.36) is linear in modulation depth, as was the perturbation in section 3.4, the analysis of that section applies to the present case. In particular, the maximum diffraction efficiency of a rank R hologram recorded in this way is at least R^{-2} times less than the maximum diffraction efficiency for a rank 1 hologram.

If uncorrelated information is stored in each exposure, the inverse dependence of the amplitude of the perturbation on the number of exposures results in a fall off in diffraction efficiency of photorefractive hologram proportional to the square of the number of exposures. This behavior is demonstrated in Fig. 4.2, where the power diffracted from plane wave holograms recorded in a cerium doped strontium barium niobate (SBN) crystal is shown. The holograms were written with grating wavevectors approximately along the c axis of the crystal. The writing beams and the reconstruction beams were extraordinarily polarized plane waves with a wavelength of 514 nm. Each hologram was recorded between two plane waves. The ratio of the intensities of the two beams was 100 to 1. The weaker beam propagated toward the c axis. The thickness of the crystal along the optical path was 2 mm. The power diffracted by each of two holograms written sequentially to equal diffraction efficiency is one quarter the saturation diffracted power from a single plane wave hologram. The power diffracted by each of three holograms written to equal diffraction efficiency, shown on a vertical scale stretched by a ratio of 5/2, is one ninth the saturation diffraction efficiency for a single hologram. The power diffracted per hologram relative to the saturation diffracted power for a single hologram when 111 holograms were recorded using this approach is shown in Fig. 4.3. The each spot in this figure is approximately 10^4 times weaker in intensity than the spot which forms for the single exposure hologram of Fig. 4.2. The characteristic exposure for this crystal, γ_t, was 119 mJ/cm^2 gratings written with 1400 lines/mm. Inverse square dependence of the diffraction efficiency of recorded holograms in more complex holographic recording systems is demonstrated in chapter 6.

If τ_t is complex, the additional degree of freedom provided by shifting the phase between signals from the input and training planes must be employed to achieve linear recording. Let ξ_n be a global phase shift of the fringe pattern between input and output plane in the n^{th} exposure. t_n and ξ_n are selected to

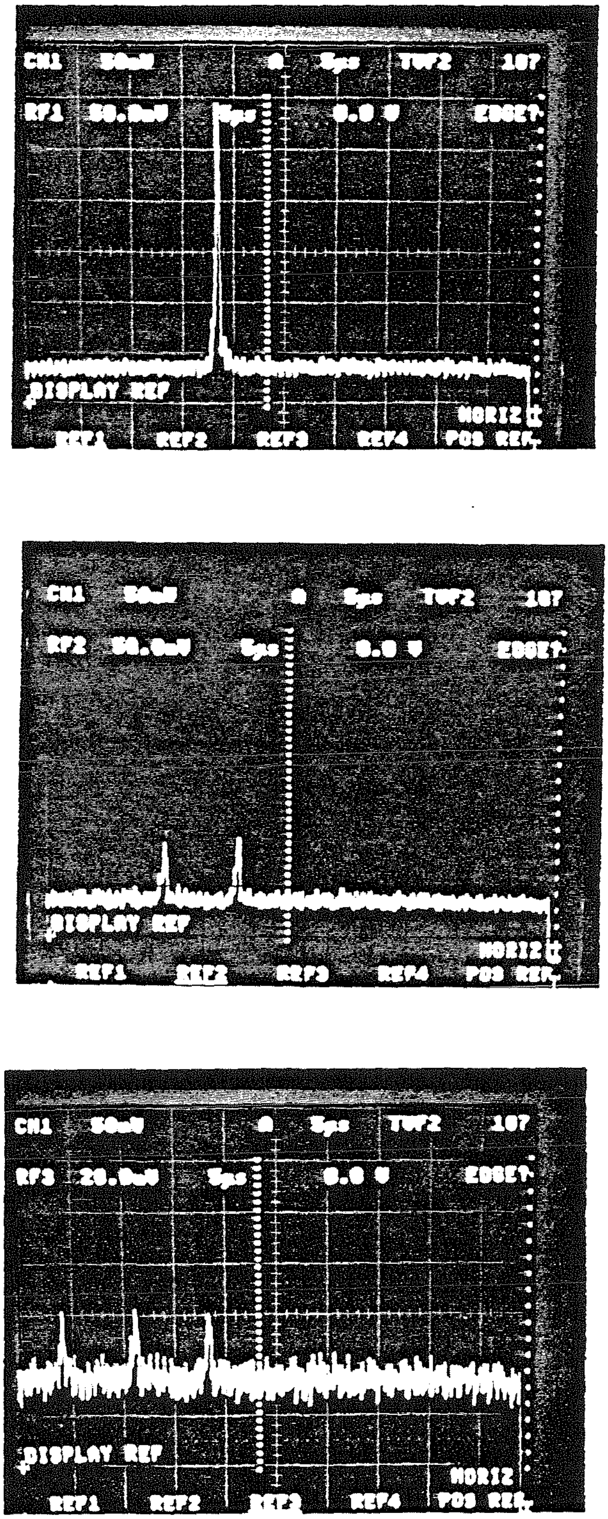


Figure 4.2. Diffracted power per hologram for one, two and three holograms.

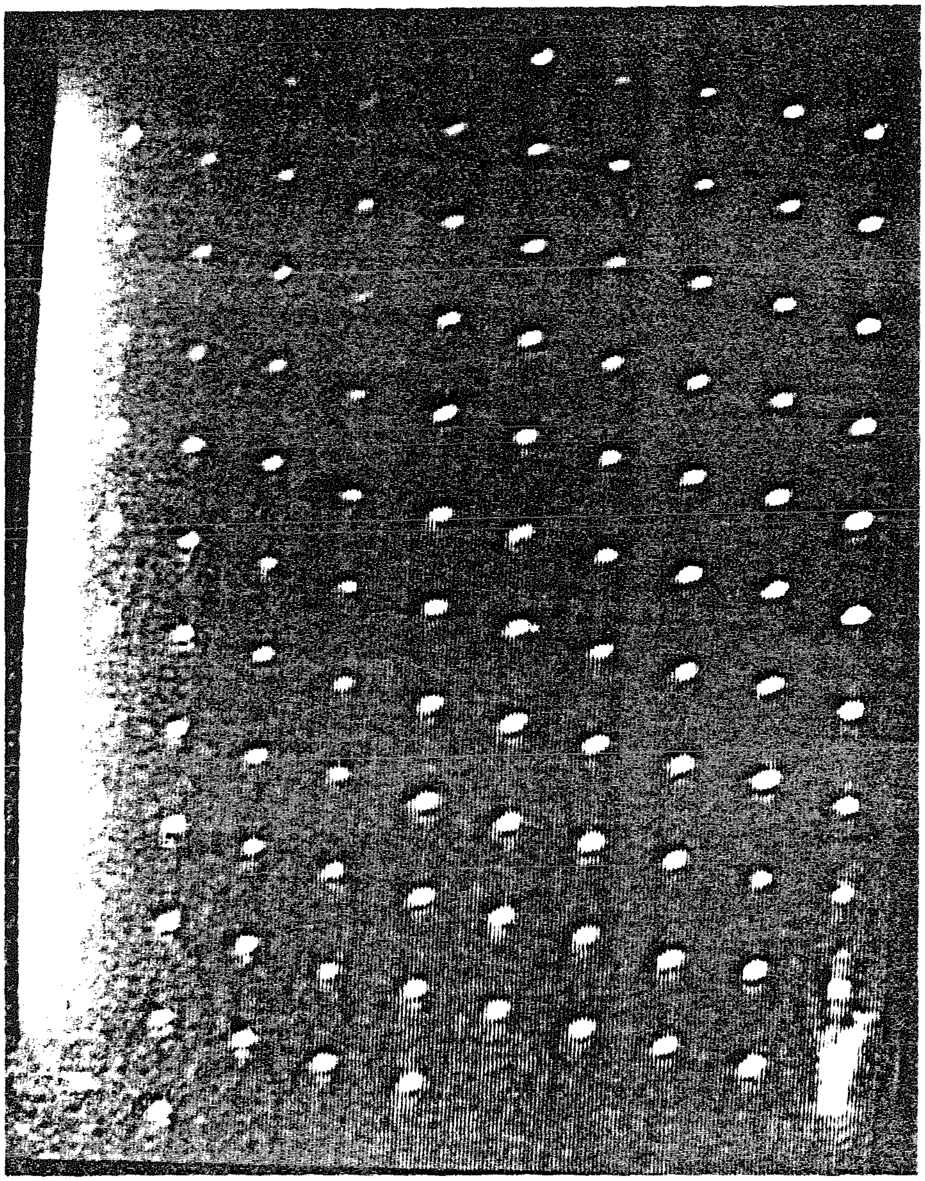


Figure 4.3. Focused spots diffracted from 111 plane wave holograms written sequentially to equal diffraction efficiencies.

make

$$f_n = e^{j\xi_n}(1 - e^{-t_n \tau_e} \exp(-j\omega_r t_n)) \exp(- \sum_{n'>n} (t_{n'}/\tau_e + j\omega_r t_{n'})) \tag{4.37}$$

constant for all n. $1/\tau_e$ and ω_r are the real and imaginary parts of $1/\tau_t$, respectively. In order to maximize the diffraction efficiency of the recorded hologram, t_1 should be selected such that $|1 - \exp(-t_1/\tau_e)\exp(-j\omega_e t_1)|$ is maximal. If $\omega_r \tau_e \gg 1$ the maximum occurs at $t_1 \approx \pi/\omega_r$. For all values of ω_r the maximum occurs for $t_1 < \pi/\omega_r$. ξ_1 can be selected arbitrarily. Given t_1 and ξ_1, t_n and ξ_n may be discovered using a recursive process based on the requirement that $f_{n+1} = f_n$. This approach yields

$$e^{j\xi_{n+1}}(e^{\frac{t_{n+1}}{\tau_e}} e^{j\omega_r t_{n+1}} - 1) = e^{j\xi_n}(1 - e^{\frac{-t_n}{\tau_e}} e^{j\omega_r t_n}). \tag{4.38}$$

Separating the real and complex parts of Eq. (4.38) yields two transcendental equations in the two unknowns t_{n+1} and ξ_{n+1}. These equations can be solved numerically without difficulty.

To compare hologram recording with a complex time constant with the result obtained in Eq. (4.36), we would like to estimate t_n as n grows large. Since the effect of a nonzero imaginary component of τ_t early in the writing process is to increase the rate at which the perturbation grows, we find that the exposure time for the n^{th} hologram is reduced by the complex term. For real values of τ_t, t_n asymptotically approaches τ_t/n. If we postulate that $t_n = \tau_e/n - \epsilon_n$ for complex τ_t, then by substituting in the constraint $|f_n|^2 = |f_{n+1}|^2$ we find that ϵ_n decreases monotonically as n increases. Asymptotically we find that t_n approaches τ_e/n.

This behavior is shown in Fig. 4.4, which is a plot of t_n versus n for various values of $\omega_r \tau_e$. As expected the curves merge asymptotically.

The value of $|f_n|^2$ when $N_e \gg 1$ holograms are recorded in a system with a complex time constant is

$$\begin{aligned} |f_n|^2 &= 1 + e^{\frac{-2}{N_e}} - 2e^{-\frac{1}{N_e}} \cos(\frac{\omega_r \tau_e}{N_e}) \\ &\approx \frac{1}{N_e^2}(1 + \omega_r^2 \tau_e^2) \end{aligned} \tag{4.39}$$

Substituting in Eq. (4.33) we find

$$A(\vec{r}) = \frac{A_o\sqrt{1+\omega_r^2\tau_e^2}}{I_o N_e} \sum_{n=1}^{N_e} [\sum_{i'j'} \sum_{i''j''} \phi_{i'j'}(n) \psi^*_{i''j''}(n) e^{j(\vec{k}_{i'j'} - \vec{k}_{i''j''})\cdot\vec{r}}] + c.c. \tag{4.40}$$

Clearly, critical importance is attached to the size of $\omega_r \tau_e$. Eq. (4.25) yields

$$\omega_r \tau_e = \frac{E_{o\|}(E_q - E_m)}{E_{o\|}^2 + (E_d + E_q)(E_d + E_m)}, \tag{4.41}$$

We have neglected photovoltaic terms for simplicity. Since $\omega_r \tau_e$ scales with the applied field, we may assume for large $\omega_r \tau_e$ that the applied field is much greater than the diffusion field. Maximizing $|\omega_r \tau_e|$ with respect to $E_{o\|}/E_q$ yields

$$|\omega_r \tau_e|_{max} = \frac{1}{2}|\sqrt{\frac{\gamma\epsilon}{e\mu}} - \sqrt{\frac{e\mu}{\gamma\epsilon}}|. \tag{4.42}$$

when

$$\frac{E_{o\|}}{E_q} = \sqrt{\frac{\gamma\epsilon}{e\mu}}. \tag{4.43}$$

Note that E_o must exceed E_q to make $\omega_r \tau_e$ large. When this is the case, the amplitude of the first Fourier component saturates at E_q. According to Eq. (4.42),

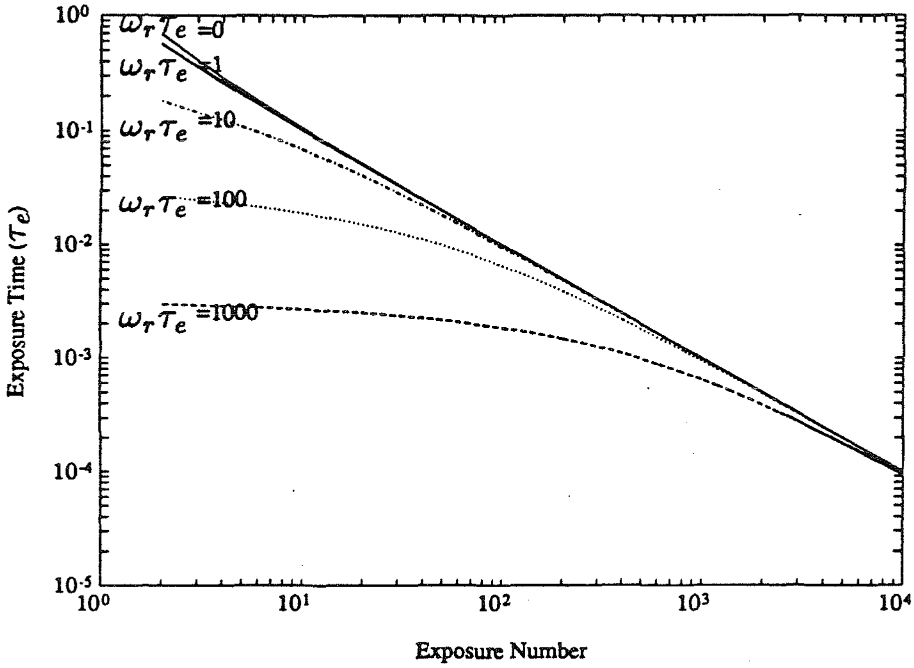


Figure 4.4. Exposure time versus exposure number for various values of $\omega\tau_r$. Exposure time is found by solving Eq. (4.38).

the imaginary component of the response time is likely to be large in materials where ϵ is large. Precise estimation of the maximum value of $\omega_r \tau_e$ which can be achieved is hampered by the lack of good estimates for γ. Maximum values of 50 to 100 high permittivity materials such as SBN and $BaTiO_3$ are not unreasonable, however. In this case, applying a driving field in excess of the maximum field that the charge density can support may still increase the storage capacity of a crystal for multiple hologram storage.

Consider once more our assumption that the field diffracted by a grating is linear in the grating amplitude. We saw in section 2.3 that the field diffracted by a volume hologram is linear in the sine of the product of the perturbation amplitude and the thickness of the hologram. Our analysis in this section and in section 3.3 has shown that the perturbation amplitude in photorefractives and in saturable materials must fall at least linearly with the rank of the hologram being recorded. The sine dependence of the diffracted field on the perturbation amplitude means that the diffraction efficiency does does not begin to fall with the square of the rank of the hologram until the rank is large enough that the perturbation amplitude is much less than the amplitude needed approach 100% diffraction efficiency, A_{100}. In photorefractive materials, the maximum perturbation amplitude can be well in excess of A_{100} [125].

Fundamental limitations on the dynamic range of photorefractive media due to the finite trap density force the amplitude of the perturbation to fall linearly with $\sqrt{R}$ [59]. If

$$\frac{A_o \omega_r \tau_e}{A_{100}} > \sqrt{R} \tag{4.44}$$

then these limitations, rather than the control problem, limit the amplitude of the perturbation. For R on the range of 1000 to 10,000, Eq. (4.44) can be satisfied using currently available materials. This means that the techniques described in this section should allow us to record holograms of rank up to 10,000 with diffraction efficiency near 10^{-4}, the same diffraction efficiency as occurs to each point in Fig. 4.3. Depending on the numerical aperture of the input and output planes, storing holograms of such high rank allows us to store from 10^9-10^{11} independently-controlled distinct gratings per cm^3 in photorefractive media. In neural applications, these holograms would store 10^3-10^5 associations between images.

4.4 MULTIPLE ACTIVE SPECIES

Typically, a photorefractive crystal is heavily doped with a single impurity during growth. As we saw in section 4.2, however, even when only one trapping species is photoactive, a second species is needed to allow the active species to be partially ionized before a hologram is recorded. Partial ionization is necessary for vacant trapping sites to be available for displaced carriers. In the most general case, more than one trapping species is photoactive. For each active species, a charge generation equation describing the ionization and recombination dynamics for that species must be added to the band transport equations. When the equations are linearized, the system is found to have the same number of characteristic time constants as independent charge generation equations. In this section we describe experimental results which show evidence of multiple time constants for hologram formation in a $LiNbO_3$ crystal, we review a simple model for hologram formation in a material with two active species and we discuss the effect of multiple time constants on linear hologram formation.

A system appropriate for recording the dynamics of hologram formation and erasure in photorefractive materials is sketched in Fig. 4.5. A hologram is written by a pair of writing beams for a specified time. The diffraction efficiency of the hologram is monitored continuously using a longer wavelength probe beam. At the end of the recording time the writing beams are switched off and an erase beam with the same total intensity as the writing beams incident at a non-Bragg matched angle is switched on. The probe continues to monitor the diffraction efficiency as the grating decays.

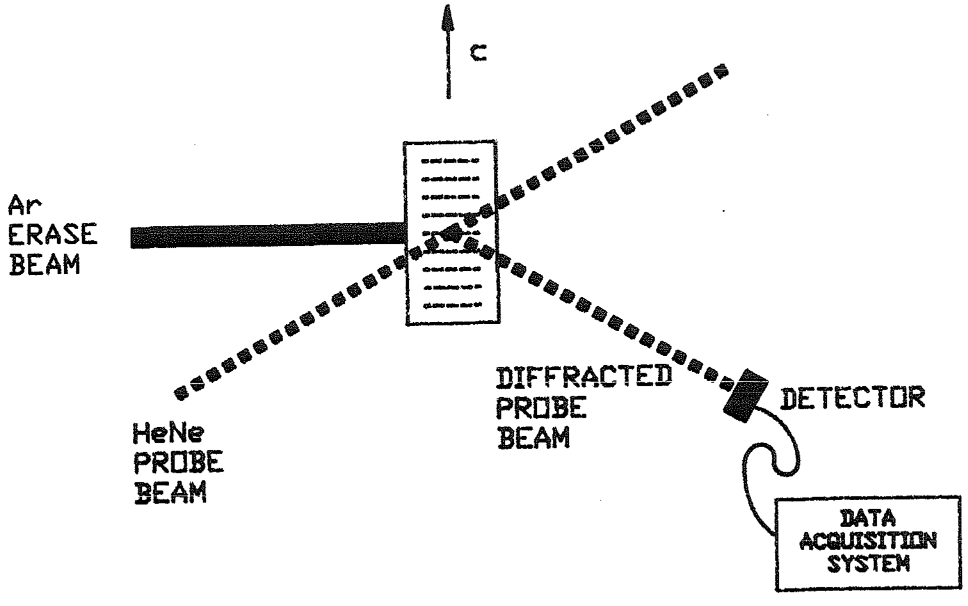


Figure 4.5. Experimental system for recording the dynamics of hologram formation and erasure.

The single trapping species theory outlined in the previous section predicts a growth in the diffraction efficiency of the probe proportional to

$$1 + \exp(-2t/\tau_e) - 2\exp(-t/\tau_e)\cos(\omega_r t)$$

during the writing phase and decay proportional to $\exp(-2t/\tau_e)$ during the erasure phase. In experiments with iron doped $LiNbO_3$ crystals we have observed dramatically different results. A plot of an experimental curve is shown in Fig. 4.6. The transition from writing to erasing occurs after the hologram has been exposed for 2000 seconds and is marked by a vertical line. The key thing to notice is the rise in diffracted power which occurs at the beginning of the erasure process. The crystal used in this experiment was doped with 0.015% Fe in the melt and was reduced after poling. The writing intensity for the plot shown was 50 mW/cm^2. The grating spacing was 2315 lines/mm. The grating vector was along the z axis. The writing and erasing beams were ordinarily polarized to reduce beam coupling effects. The probe was extra-ordinarily polarized. The ratio between the intensities of the writing beams was 10 to 1.

Previously, increases in diffraction efficiency during the erasure process have been associated with self-reinforcement effects during Bragg matched erasure [124] and back-coupling effects in very high diffraction efficiency holograms [125]. The first of these effects is suppressed in the present experiment by the use of a non-Bragg matched erasure beam. The second effect does not apply because the diffraction efficiency did not exceed 0.005. The behavior observed in this experiment can be explained by the presense in the crystal of two separate grating components which develop with different time constants and whose relative phase is nonzero. At saturation the two components are out of phase. When erasure begins, the more quickly decaying component is erased and the slowly decaying

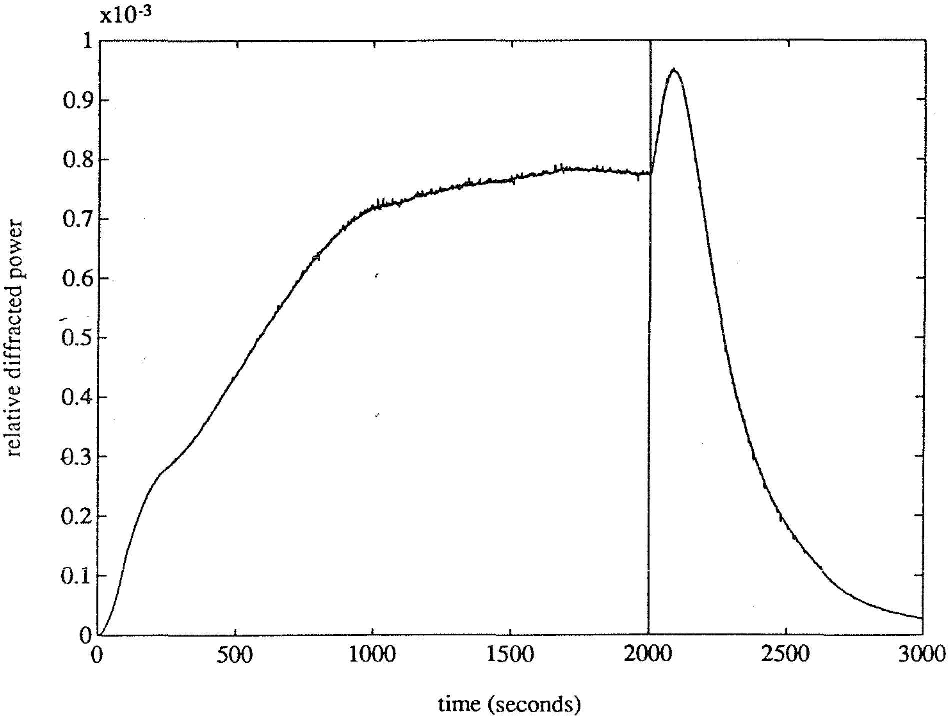


Figure 4.6. Diffracted probe power versus exposure time.

component is “unmasked.” This two time constant theory can be supported by comparing erasure data for holograms written for a short time with erasure data for holograms written to saturation. Fig. 4.7 is a log plot of experimental results under the experimental conditions described above. The upper curve shows the decay of a grating which was recorded for 2000 seconds and then erased for 1735 seconds. The data shown represents the diffracted power from 1735 seconds of erasure to 2035 seconds. The lower curve shows the decay of a grating which was written for 10 seconds and then erased for 300 seconds. The key thing to note is that the slope of the lower curve, corresponding to the time constant with which the diffraction efficiency decays, is much more negative than the slope of the upper curve. This behavior is expected if two grating components evolve with distinct time constants. For short exposure and erasure times, relatively more of the quicker grating component is recorded. When the hologram is erased, the decay curve reflects the presence of this component. In a saturated hologram, after erasure has been underway for some time the faster grating component is much reduced compared to the slow component and the decay is dominated by the slower decay rate.

Band transport equations for the simplest case of hologram formation in the presence of multiple trapping species were developed by Valley in 1983 [94]. Valley's theory was used to explain indications of multiple time constant decay found in holograms recorded in BSO by Mullen [126]. Similar results have since been found in $LiNbO_3$ [127] and GaAs [128]. Valley's theory consists essentially of the assumption that N_A^- is also a photoactive parameter. Note that in addition to adding more trapping species to the dynamical theory, a second carrier species could also be added. Since the carrier distributions are typically slaved to the

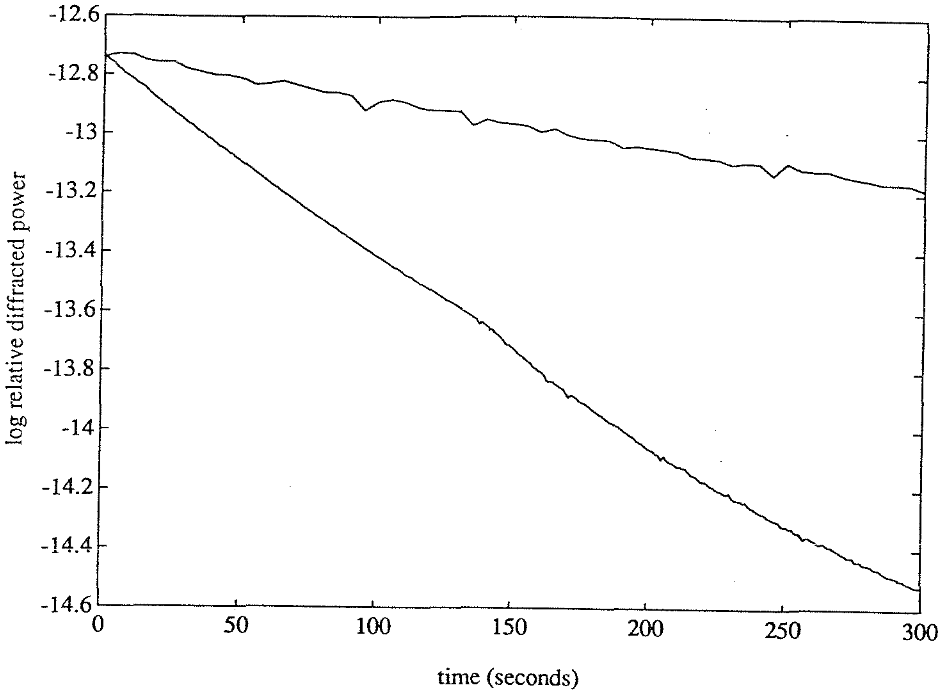


Figure 4.7. Log relative diffracted power versus erase time for a hologram written and erased for a long time (upper curve) and immediately upon erasure for a hologram written for a short time (lower curve).

more slowly evolving variables, however, the presense of multiple carriers alone does not add characteristic time constants to the dynamical behavior. For this reason we limit ourselves to a single carrier here. Multiple carrier theories are described in [129]. The band transport equations in this case are

$$\frac{\partial n}{\partial t}+\frac{\partial N_A^-}{\partial t}-\frac{\partial N_D^+}{\partial t}=\frac{\vec{\nabla}\cdot\vec{j}}{e} \tag{4.45}$$

$$\frac{\partial N_D^+}{\partial t}=(\beta_D+s_D I)(N_D-N_D^+)-\gamma_D n N_D^+ \tag{4.46}$$

$$\frac{\partial N_A^-}{\partial t}=-(\beta_A+s_A I)N_A^-+\gamma_A n(N_A-N_A^-) \tag{4.47}$$

$$\vec{j}=e\mu n\vec{E}+\mu k_B T\vec{\nabla}n+\vec{\kappa}\alpha I \tag{4.48}$$

$$\vec{\nabla}\cdot\epsilon\vec{E}=-e(n+N_A^- -N_D^+) \tag{4.49}$$

These equations reduce to the single species equations in the limit as $\beta_A, s_A \to 0$.

Suppose that

$$I(\vec{r})=I_o(1+m\cos(\vec{K}\cdot\vec{r})). \tag{4.50}$$

Linearizing using the same techniques as in the previous section and again making the quasi-steady approximation yields conditions for the zeroth order charge distributions

$$n_o N_{D_o}^+=n_D(N_D-N_{D_o}^+) \tag{4.51}$$

$$n_o(N_A-N_{A_o}^-)=n_A N_{A_o}^- \tag{4.52}$$

$$n_o+N_{A_o}^-=N_{D_o}^+, \tag{4.53}$$

where $n_D = s_D I/\gamma_D$ and $n_A = s_A I/\gamma_A$. Given n_D and n_A, these equations can be solved for n_o, $N_{D\,o}^{+}$ and $N_{A\,o}^{-}$. The dynamics for the first order Fourier components are described by

$$\frac{dn_1}{dt} = \frac{dN_{D\,1}^{+}}{dt} - \frac{dN_{A\,1}^{-}}{dt} + \frac{jKJ_1}{e} \tag{4.54}$$

$$\frac{dN_{D\,1}^{+}}{dt} = s_D I_o \frac{m}{2}(N_D - N_{D\,o}^{+}) - s_D I_o N_{D\,1}^{+} - \gamma_D n_o N_{D\,1}^{+} - \gamma_D n_1 N_{D\,o}^{+} \tag{4.55}$$

$$\frac{dN_{A\,1}^{-}}{dt} = -s_A I_o \frac{m}{2} N_{A\,o}^{-} - s_D I_o N_{A\,1}^{-} - \gamma_D n_o N_{A\,1}^{-} + \gamma_D n_1 (N_A - N_{A\,o}^{-}) \tag{4.56}$$

$$J_1 = e\mu n_1(\vec{E}_o + j\vec{E}_d) + e\mu n_o \vec{E}_1 + \kappa\alpha_o I_o \frac{m}{2} + \kappa\alpha_1 I_o \tag{4.57}$$

$$j\epsilon K E_1 = -e(n_1 + N_{A\,1}^{-} - N_{D\,1}^{+}), \tag{4.58}$$

where we have assumed that the applied field and $\vec{\kappa}$ are parallel to $\vec{K}$ and we have defined $K = |\vec{K}|$.

In the previous section we used the quasi-steady approximation to reduce the second order dynamical equations to a first order equation. The solution to this equation involved only one boundary condition, the initial state of the space charge field. Eqs.(4.54)-(4.58), in contrast, represent a third order dynamical system. In Valley's analysis, the time constants of a third order equation for the development of the space charge field are determined. The system is then reduced to second order based on the discovery that one of the time constants is much smaller than the other two. In order to specify the dynamics of the second order system it is necessary to specify two boundary conditions, typically the initial state of the field and the initial time derivative of the field amplitude.

Unfortunately, Valley's approach leads to an *ad hoc* assumption of

$$\frac{dE_1}{dt}\Big|_{t=0} = -\frac{E_1}{\tau_{di}}, \tag{4.59}$$

where $\tau_{di} = \epsilon / e\mu n_o$, for the initial condition of the derivative of the field when erasure begins. This condition is inconsistent with our empirical results, in which the derivative of the amplitude of the field is initially positive.

This difficulty points out an inconsistency in the the quasi-steady approximation as it has commonly been applied in the literature. The approximation has been used to imply that the zeroth order of the carrier density responds instantaneously to optical illumination, but the same assumption has not been applied to higher order Fourier components of the carrier density. When the approximation is applied in this manner, a fast time constant corresponding to the carrier response time is found when the first order equations are solved. If the approximation is applied consistently, however, all time derivatives of the carrier density should be set to zero. In this case, the carrier density is said to be "slaved" by the slowly varying parameters [130]. In this case, Eqs. (4.54)-(4.58) reduce immediately to a second order system. More importantly, boundary conditions for the field and its derivative can be obtained directly by assuming continuity in the slowly varying parameters and instantaneous response in the carrier density to ambient conditions.

If we set the time derivative of the carrier density to zero, Eqs. (4.54)-(4.58) can be reduced to three independent equations in the dynamic variables N_{D1}^{+} and N_{A1}^{-} and in the parameter n_1. These equations are

$$\frac{dN_{D1}^{+}}{dt} = \frac{m s_D I_o}{2}\frac{\epsilon K}{e}E_D - \frac{N_{D1}^{+}}{\tau_{ID}} - \frac{n_1}{\tau_D} \tag{4.60}$$

$$\frac{dN_{A1}^{-}}{dt} = -\frac{ms_A I_o}{2}\frac{\epsilon K}{e}E_A - \frac{N_{A1}^{-}}{\tau_{IA}} + \frac{n_1}{\tau_A} \tag{4.61}$$

and

$$n_1 = \frac{j}{(\frac{1}{\tau_{dr}} + j\frac{1}{\tau_{diff}})}\left[\frac{dN_{D1}^{+}}{dt} - \frac{dN_{A1}^{-}}{dt} + \frac{N_{D1}^{+} - N_{A1}^{-}}{\tau_{di}} + j\frac{m}{2\tau_{di}}\frac{\epsilon K}{e}E_{ph}\right], \tag{4.62}$$

where the various parameters are defined below and where we have eliminated the current density by differentiating Eq. (4.58) and comparing the result with Eq. (4.54). This yields

$$\epsilon\frac{dE_1}{dt} = -J_1 \tag{4.63}$$

We can solve Eqs. (4.60)-(4.62) for $N_{D1}^{+}(t)$ and $N_{A1}^{-}(t)$ if we are able to specify the initial condition of these variables. For writing in a material where no grating exists, $N_{D1}^{+}(t=0) = N_{A1}^{-}(t=0) = 0$. When we erase a pre-recorded grating, we may assume that $N_{D1}^{+}(t)$ and $N_{A1}^{-}(t)$ remain continuous even when the driving terms are discontinuous.

Since it is the dynamical behavior of the space charge field which is of interest for holographic applications, we solve Eqs. (4.60)-(4.62) and Eq. (4.58) for the field. This yields

$$\frac{d^2E_1}{dt^2} + \frac{1}{T_1}\frac{dE_1}{dt} + \frac{1}{T_2^2}(E_1 - mE_s) = 0, \tag{4.64}$$

where

$$\frac{1}{T_1} = T_3\left(\left(\frac{1}{\tau_{IA}} + \frac{1}{\tau_{ID}}\right)\left(\frac{1}{\tau_{diff}} - j\frac{1}{\tau_{dr}}\right) + \frac{1}{\tau_D\tau_{di}} + \frac{1}{\tau_D\tau_{IA}} + \frac{1}{\tau_A\tau_{di}} + \frac{1}{\tau_A\tau_{ID}}\right) \tag{4.65}$$

$$\frac{1}{T_2^2} = T_3\left(\frac{1}{\tau_{IA}\tau_{ID}}\left(\frac{1}{\tau_{diff}} - j\frac{1}{\tau_{dr}}\right) + \frac{1}{\tau_{IA}\tau_{di}\tau_D} + \frac{1}{\tau_{ID}\tau_{di}\tau_A}\right) \tag{4.66}$$

$$\frac{1}{T_3} = \frac{1}{\tau_{diff}} - j\frac{1}{\tau_{dr}} + \frac{1}{\tau_D} + \frac{1}{\tau_A} \tag{4.67}$$

$$\begin{aligned} E_s = \frac{1}{4}T_3T_2^2\Bigg[&(E_o + jE_d)\left(\frac{s_D I_o \mu K E_D}{\tau_{IA}} + \frac{s_A I_o \mu K E_A}{\tau_{ID}}\right) \\ &+ \frac{1}{\tau_{di}}\left(\frac{1}{\tau_{IA}\tau_D} + \frac{1}{\tau_{ID}\tau_A}\right)E_{ph}\Bigg] \end{aligned} \tag{4.68}$$

$\tau_D = (\gamma_D N_{D\,o}^+)^{-1}$ is the recombination time for species D.

$\tau_A = (\gamma_A(N_A - N_{A\,o}^-))^{-1}$ is the recombination time for species A.

$\tau_{ID} = (s_D I_o + \gamma_D n_o)^{-1}$ is the inverse of the sum of the photoproduction rate and the ion recombination rate for species D.

$\tau_{IA} = (s_A I_o + \gamma_A n_o)^{-1}$ is the inverse of the sum of the photoproduction rate and the ion recombination rate for species A.

$\tau_{di} = \frac{\epsilon}{e\mu n_o}$ is the dielectric relaxation time.

$\tau_{diff} = \frac{e}{K^2\mu k_B T}$ is the diffusion time.

$\tau_{dr} = (\mu K E_o)^{-1}$ is the drift time.

$E_D = \frac{e(N_D - N_{D\,o}^+)}{\epsilon K}$.

$E_A = \frac{eN_{A\,o}^-}{\epsilon K}$.

Assuming that the field and its first derivative are continuous, the solution to Eq. (4.64) is

$$
\begin{aligned}
(\tau_+ - \tau_-)E_1(t) = e^{-\frac{t}{\tau_+}} &\left[E_s \int_0^t m e^{\frac{s}{\tau_+}} ds + \tau_+ E_1(0) + \tau_+\tau_- \frac{dE_1}{dt}|_0\right] \\
- e^{-\frac{t}{\tau_-}} &\left[E_s \int_0^t m e^{\frac{s}{\tau_-}} ds + \tau_- E_1(0) + \tau_+\tau_- \frac{dE_1}{dt}|_0\right]
\end{aligned}
\tag{4.69}
$$

where

$$
\frac{1}{\tau_\pm} = \frac{1}{2T_1}\left[1 \pm \sqrt{1 - \frac{4T_1^2}{T_2^2}}\right]. \tag{4.70}
$$

If there is no grating at $t = 0$, $E_1(0) = 0$. Solving for the initial condition for the first derivative of the field using Eqs. (4.58) and (4.60)-(4.62) under the assumption that $N_{D\,1}^{+}(t=0) = N_{A\,1}^{-}(t=0) = 0$ yields

$$
\frac{dE_1}{dt}|_0 = \frac{T_3 m}{2}\left[\left(E_o + jE_d\right)(s_D I_o \mu K E_D + s_A I_o \mu K E_A) + \frac{E_{ph}}{\tau_{di}}\left(\frac{1}{\tau_D} + \frac{1}{\tau_A}\right)\right] \tag{4.71}
$$

In the experiments described above, a hologram is recorded at constant modulation depth for a fixed time, t_o. The modulation depth is then switched to zero and the hologram is erased. Switching the modulation depth to zero causes a discontinuity in the time derivative of the field amplitude. This discontinuity must be explicitly accounted for in our solution to the second order system. The field amplitude during recording is

$$
\begin{aligned}
E_1(t) = &\frac{m\tau_-}{\tau_- - \tau_+} E_s(1 - e^{-\frac{t_o}{\tau_-}}) + \frac{\tau_+\tau_-}{\tau_- - \tau_+}\frac{dE_1}{dt}|_0 e^{-\frac{t_o}{\tau_-}} \\
&- \frac{m\tau_+}{\tau_- - \tau_+} E_s(1 - e^{-\frac{t_o}{\tau_+}}) - \frac{\tau_+\tau_-}{\tau_- - \tau_+}\frac{dE_1}{dt}|_0 e^{-\frac{t_o}{\tau_+}},
\end{aligned}
\tag{4.72}
$$

where $dE_1/dt|_0$ is given by Eq. (4.71). During erasure the field amplitude is

$$E_1(t) = \frac{\tau_-}{\tau_- - \tau_+}\left(E_1(0) - \tau_+\frac{dE_1}{dt}|_0\right)e^{-\frac{t}{\tau_-}} - \frac{\tau_+}{\tau_- - \tau_+}\left(E_1(0) - \tau_-\frac{dE_1}{dt}|_0\right)e^{-\frac{t}{\tau_+}}, \tag{4.73}$$

where the initial condition for the field on erasure is found by assuming that its value is continuous when we switch from writing to erasing. The initial condition of the derivative of the field is found by solving Eqs. (4.58) and (4.60)-(4.62) under the assumption that the ionized trap densities are continuous. In the case of a hologram written to saturation this approach yields

$$\frac{dE_1}{dt}|_{se} = -\frac{T_3 m}{2}\left[\left(E_o + jE_d\right)(s_D I_o \mu K E_D + s_A I_o \mu K E_A) + \frac{E_{ph}}{\tau_{di}}\left(\frac{1}{\tau_D} + \frac{1}{\tau_A}\right)\right] \tag{4.74}$$

for the derivative at the start of the erasure cycle. The derivative at the beginning of the erasure cycle is the negative of the derivative at the beginning of the write cycle because at saturation the derivative due to the homogeneous equation is just sufficient to cancel the driving terms. When the driving terms are turned off the derivative is equal to the negative of their contribution to the derivative. The initial derivative on writing contains only the driving terms.

The diffraction efficiency of a recorded hologram increases when erasure begins if the initial value of $d|E_1|^2/dt$ is positive. If we write to saturation then this initial value can be evaluated using Eq. (4.74):

$$
\begin{aligned}
\frac{d|E_1|^2}{dt} &= mE_s\frac{dE^*}{dt}|_{se} + mE_s^*\frac{dE}{dt}|_{se} \\
&= -\frac{|T_3|^2m^2}{4}\Bigg[\Re\{T_2^2\}\left[(E_o^2+E_d^2)(\frac{s_DI_o\mu KE_D}{\tau_{IA}}+\frac{s_AI_o\mu KE_A}{\tau_{ID}})\right. \\
&\quad \times (s_DI_o\mu KE_D + s_AI_o\mu KE_A) \\
&\quad \left. + \frac{|E_{ph}|^2}{\tau_{di}}(\frac{1}{\tau_{IA}\tau_D}+\frac{1}{\tau_{ID}\tau_A})(\frac{1}{\tau_D}+\frac{1}{\tau_A})\right] \\
&\quad + \Re\{T_2^2E_{ph}(E_o - jE_d)\}\left[(\frac{s_DI_o\mu KE_D}{\tau_{IA}}+\frac{s_AI_o\mu KE_A}{\tau_{ID}})(\frac{1}{\tau_D}+\frac{1}{\tau_A})\right. \\
&\quad \left. + (s_DI_o\mu KE_D + S_AI_o\mu KE_A)(\frac{1}{\tau_{IA}\tau_D}+\frac{1}{\tau_{ID}\tau_A})\right]\Bigg]
\end{aligned}
\tag{4.75}
$$

Note that if E_{Ph} is zero then the sign of the derivative in Eq. (4.75) must be negative, since the term inside the large bracket which does not contain E_{ph} contains only squared values and positive constants. If E_{ph} is not zero then whether or not the diffraction efficiency increases on erasure depends on the effective phase of the photovoltaic field, on the relative doping densities of the trapping species and on the relative magnitudes of the applied field and the effective diffusion and photovoltaic fields. Sensitivity to the magnitude of the photovoltaic field has been observed in our experiments. Fig. 4.8 is an experimental writing and erasure curve. The writing intensity in this case is 100 mW/cm^2, twice the intensity of Fig. 4.6. The exposure time is 1000 sec., half as long as in Fig. 4.6. Note that the bump on erasure is much less pronounced in this case.

The simulated diffraction efficiency during the recording and erasure process when the photovoltaic field is 180° out of phase with the illumination pattern is shown in Fig. 4.9. We relied as much as possible on published parameter values for LiNbO$_3$ in generating this curve. The effective photovoltaic field was assumed to be 1.2 kilovolts/cm at a recording intensity of 56mW. The key parameter values

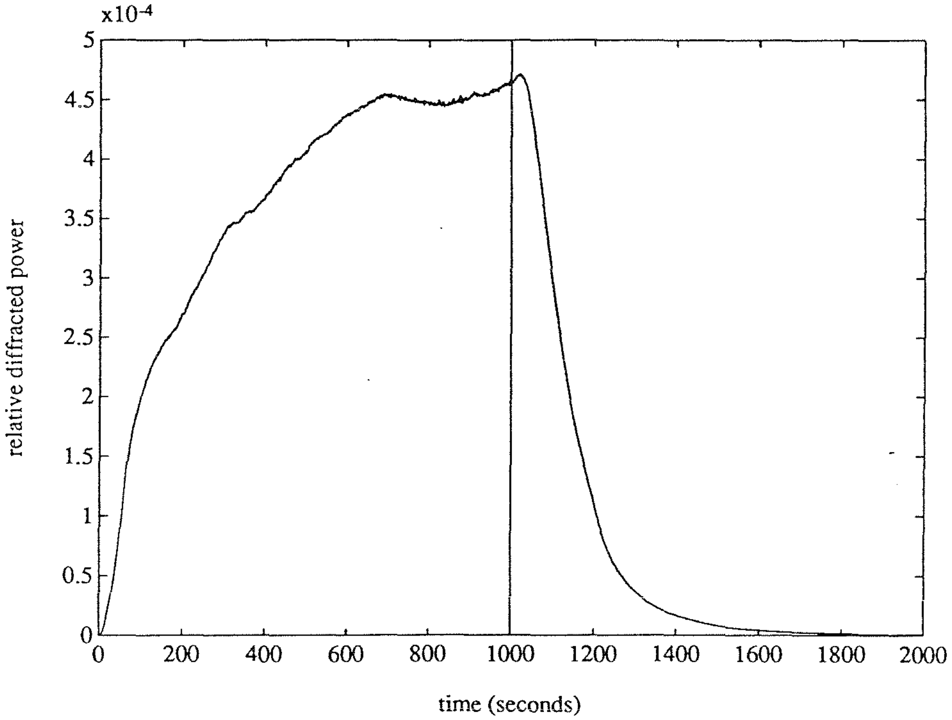


Figure 4.8. Experimental writing and erasure curve in Iron doped $LiNbO_3$ under the same conditions as in Fig. 4.6, except that the recording and erasing intensity is doubled and the exposure time is halved.

used are listed in table 4.1. Note that the simulation shows a hump in the initial stages of writing, which we do not observe in our experiments. This hump occurs because the initial derivative of the field is out of phase with the final state of the field. This derivative eventually switches sign and the field amplitude, but not the ionized trap densities, passes through zero before reaching saturation. In the simulation shown in Fig. 4.10 the photovoltaic term is assumed to be 112° out of phase with the recording interference pattern. Since the initial derivative of the field is not exactly 180° out of phase in this case, the magnitude of the field does not pass through zero during the writting cycle. A characteristic hump is observed in the erasure cycle. A qualitatively similar experimental curve plotted as a dashed line is superposed on the simulation in Fig. 4.10. The writing intensity for the experimental curve was 50 mW/cm^2 at 488 nm. The spatial frequency of the recorded grating was 2315 lines/mm.

Table 4.1: Parameters for Figs. 4.9 and 4.10.

ϵ	$\mu(V^{-1}sec.^{-1}cm.^2)$	$\frac{s_d}{\gamma_d}$	$\frac{s_a}{\gamma_a}$	N_d	N_a
29	0.8	10^{-15}	10^{-14}	10^{19}	10^{16}

While our theoretical results lead us to conclude that the qualitative behavior observed in our experiments can be expected of materials in which multiple species are active, it is not reasonable to try and fit theoretical parameters on the basis of our experiments. This is because the problem of fitting curves with multiple exponential terms is extremely ill-posed and because we have no evidence excluding the possibility that more than two species of traps and more than one carrier are active in our crystal. A detailed analysis of hologram formation in our crystals would have to be based principally on non-holographic data.

We now turn to the question of linear hologram formation in materials with

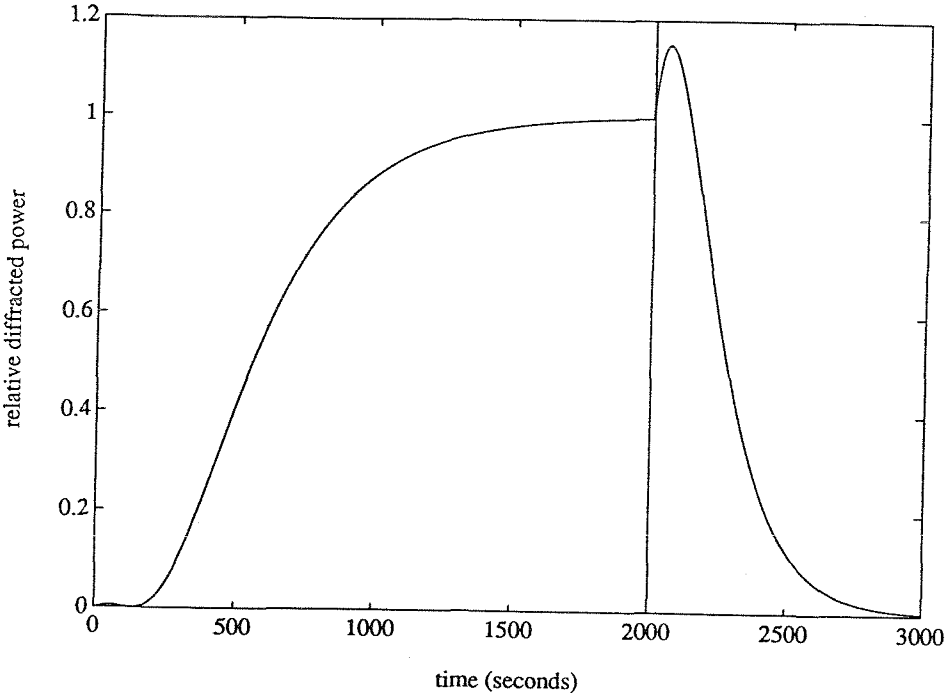


Figure 4.9. Simulation of recording and erasure in a multiple species environment. The photovoltaic field is assumed to be 180° out of phase with the fringe pattern of the recording intensity.

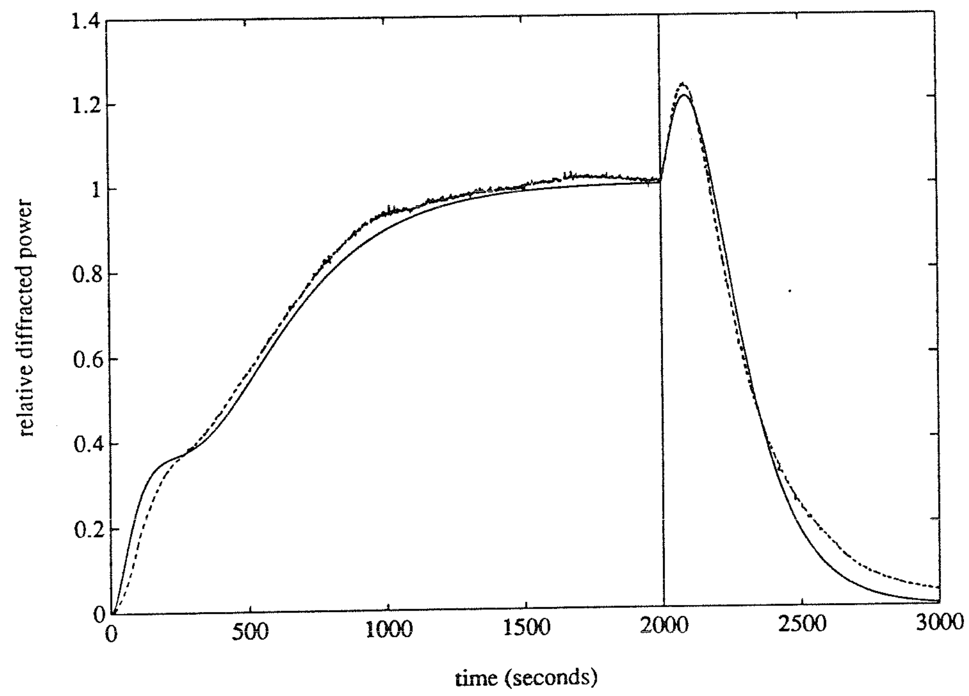


Figure 4.10. Simulation of recording and erasure process with a phase shift of 112° in the photovoltaic field relative to the phase of the illumination pattern. The solid line is a simulation and the dashed line is the experimental curve of Fig. 4.6.

multiple trapping species. Since the state of the field depends on multiple variables in these materials, it is no longer the case that fields which are of the same amplitude at one time will remain equal during erasure. In fact, a pair of field amplitudes cannot be made to follow the same path in time in this system unless their exposure history is exactly the same. Of course, if one time constant is much longer than the other, the exposure cycle for a single time constant can be followed and the shorter time constant can be ignored since it does not effect the magnitude after substantial erasure. If the two time constants are more nearly equal, a method suggested by McRuer *et al.* for the single species case could be applied to record holograms linearly [91]. Under this model each exposure lasts for a very short time compared to the response times. Associated pairs of images are recorded in a repeating sequence until the system reaches saturation. The dynamics of hologram formation in a material containing T active trapping species is described by a set of T coupled equations similar to Eqs. (4.60) and (4.61). These equations can be expressed in vector form as

$$\frac{d\vec{N}}{dt} = \bar{\bar{Q}}\vec{N} + \vec{D}, \tag{4.76}$$

where the i^{th} component of $\vec{N}$ is N_i, the first Fourier component of the number density of the ionized sites for the i^{th} species. $\bar{\bar{Q}}$ is a matrix describing the coupling and $\vec{D}$ contains the driving terms for the growth of the grating. The component of the recorded hologram which corresponds to information recorded in the q^{th} exposure cycle is recorded for a time δ, erased for a time $(N_e - 1)\delta$ and then again recorded. This sequence repeats until the system reaches reaches an equilibrium state. At equilibrium, the amount by which the charge pattern corresponding to the q^{th} exposure decays in making $(N_e - 1)$ uncorrelated exposures is exactly canceled by the amount by which this charge pattern is increased by the q^{th} exposure in each cycle. If δ is very small, the decay of the information stored by the q^{th} exposure due to the uncorrelated exposures can be expressed as $\triangle_e \vec{N} = (N_e - 1)\delta \overline{\overline{Q}} \vec{N}$. The growth of the grating due to the q^{th} exposure is $\triangle_r \vec{N} = \delta(\overline{\overline{Q}} \vec{N} + \vec{D})$. At saturation, $\triangle_e + \triangle_r = 0$, which implies

$$\vec{N} = \frac{\overline{\overline{Q}}^{-1} \vec{D}}{N_e} \tag{4.77}$$

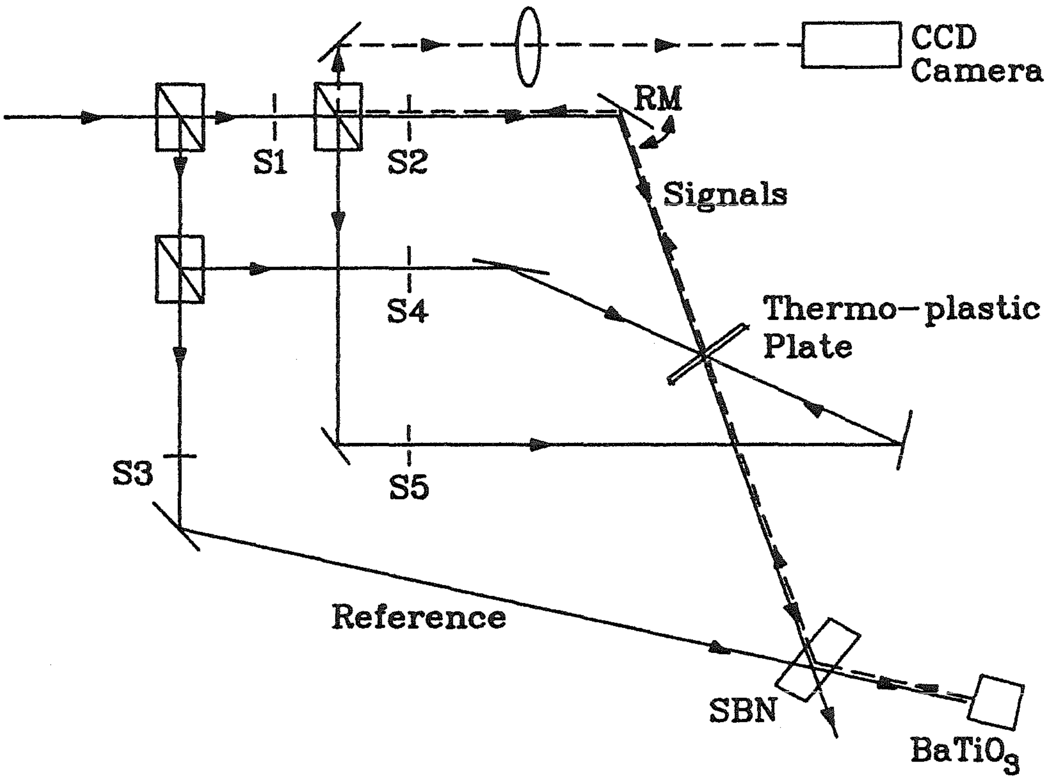
Since we know from the Poisson equation that the amplitude of the space charge field is linear in the trapping densities, we find once again that the amplitude of the perturbation for N_e independent associated pairs of exposure patterns is N_e times less than amplitude for one exposure.

The dynamics of hologram control in the presence of multiple species can be substantially more complicated than the single trap case. Various other recording schemes using resonances between the time constants and phase shifts in the recording beams are interesting topics for future investigation.

4.5 PERIODICALLY REFRESHED PHOTOREFRACTIVE HOLOGRAMS

In this section we describe a system in which part of the decrease in the diffraction efficiency of a multiply exposed hologram is recovered by periodic copying between two holographic media. Palais and Wise [131] have previously demonstrated that copying a weak hologram can dramatically increase the diffraction efficiency of the hologram. Johnson et. al. [132] have applied this technique to improve the diffraction efficiency of multiply exposed silver halide films. In this section we demonstrate that periodic copying between two dynamic media improves the diffraction efficiency by a factor of N_e. A related result is that periodic copying between two holograms results in a stable diffraction efficiency when an indefinitely long sequence of exposures is performed. Long sequences of exposures are used in adaptive holographic systems, such as those described in chapter 6.

The architecture of the system is shown in Fig. 4.11. A series of holograms between a reference plane wave, $R1$, and a set of signal beams are recorded in a SBN:Ce crystal. Shutters $S4$ and $S5$ are closed during this operation. We use plane waves as signal beams in our experiments. Different beams are generated by rotation of the mirror RM. Let I_r and I_s represent the intensities of the reference and signal beams, respectively. The diffraction efficiency of the recorded holograms is monitored by illuminating the crystal with the phase conjugated reference, $R1^*$. The path of the diffracted read out beam to the output CCD is shown as a dashed line in the figure. A self-pumped $BaTiO_3$ phase conjugate mirror is used to generate the conjugate wave. In addition to providing automatic alignment of the conjugate beam, the PCM compensates for phase distortions due to imperfections in the SBN. When the diffraction efficiency of the holograms becomes unacceptably low, the recorded holograms are copied from the SBN to a second holographic medium, which in our experiments is a thermoplastic plate. The thermoplastic hologram is formed using the light diffracted by the SBN hologram and reference wave, $R2$. Shutters $S2$ and $S4$ are closed. The hologram written on the thermoplastic plate is copied back to the SBN using reference beams $R1$ and $R2^*$ (see Fig. 4.11). The intensity of $R2^*$ is selected to make the intensity of the signal beam diffracted from the thermoplastic equal to the original signal intensity, I_s. Shutters $S1$ and $S5$ are closed during this



PERIODICALLY REFRESHED RECORDING SYSTEM

Figure 4.11. Experimental layout for a periodic refreshing system.

step. The result is a rejuvenated hologram of each of the signal beams in the SBN. At this point, we begin adding a new series of holograms to those already stored in the SBN. When the diffraction efficiency of the holograms again falls unacceptably low, we repeat the copying process.

We begin the experiment by recording a series of N_1 holograms in the SBN following the schedule of Eq. (4.34) for $\chi = 1$. After the exposure sequence, the amplitude of the space charge for each hologram is $A_n = A_o/M_1$. If N_1 is large enough, then the diffraction efficiency per hologram is small and the diffraction efficiency of the n^{th} hologram, η_n, is proportional to $|A_n|^2$. This implies that $\eta_n = \eta_o/{N_1}^2$, where η_o is proportional to $|A_o|^2$. After N_1 exposures, we copy the holograms summed in the SBN onto the thermoplastic. We then copy back to the SBN as described above. The amplitude of the saturation space charge in a photorefractive hologram is proportional to the modulation depth with which it is recorded. Since the intensity of the signal beam reconstructed from the thermoplastic hologram is equal to I_s, the intensity of each of the N_1 equally recorded components is I_s/N_1. The modulation depth for each component is the ratio of $\sqrt{I_r I_s/N_1}$ to the total recording intensity, $(I_r + I_s)$. Since this modulation depth is $\sqrt{N}_1$ times less than the modulation depth for one hologram, the amplitude of the space charge grating for each component is $A_n = A_o/\sqrt{N}_1$. η_n is thus η_o/N_1. The total diffraction efficiency, $\eta_T = N_1\eta_n$, is restored to the saturation value of η_o.

We now begin recording a new series of holograms in the SBN. The first hologram of each new series is written until its amplitude is equal to the partially erased amplitude of the copied holograms. Let $A_n^{(j)}$ and t_{jn} be the amplitude and exposure time of the n^{th} hologram recorded in the SBN in the j^{th} recording cycle.

We would like t_{jn} to be such that $A_n^{(j)}$ is constant for all j and m. Assuming that this is the case, we use an argument similar to that described in the previous paragraph to conclude that $A_m^{(j)} = A_o/\sqrt{\sum_{l=1}^{j-1} N_l}$ at the start of the j^{th} recording cycle, *i.e.*, just after the hologram is copied for the $(j-1)^{th}$ time. N_j is the number of exposures we make in the j^{th} cycle. The perturbation in the crystal at the start of the j^{th} recording cycle is

$$\begin{aligned}\Phi(\vec{r}) &= \sum_{l=1}^{j-1}\sum_{n=1}^{N_l} A_n^{(l)} f_n^{(l)}(\vec{r}) \\ &= \frac{A_o}{\sqrt{\sum_{l=1}^{j-1} N_l}} \sum_{l=1}^{j-1}\sum_{n=1}^{N_l} f_n^{(l)}(\vec{r}),\end{aligned} \tag{4.78}$$

where $f_n^{(l)}(\vec{r})$ is a normalized distribution which describes the spatial structure of the n^{th} hologram in the j^{th} recording cycle. The amplitude of the first hologram in the j^{th} recording cycle is equal to the partially erased amplitude of the copied holograms if

$$A_o(1 - \exp(-t_{j1}\alpha I)) = \frac{A_o}{\sqrt{\sum_{l=1}^{j-1} N_l}} \exp(-t_{j1}\alpha I). \tag{4.79}$$

Letting $\chi^{(j)} = 1 - \exp(-t_{j1}\alpha I)$, Eq. (4.79) yields $\chi^{(j)} = 1 - (\sqrt{\sum_{l=1}^{j-1} N_l}/(1 + \sqrt{\sum_{l=1}^{j-1} N_l}))$. Letting $\chi = \chi^{(j)}$ in the exposure schedule of Eq. (4.34) we make N_j exposures in the j^{th} cycle to record $\sum_{l=1}^{j} N_l$ holograms with equal amplitudes. In order to maintain a constant diffraction efficiency from the thermoplastic, N_j is selected such that the total diffraction efficiency of the summed hologram on the SBN falls back to its value after the first N_1 exposures, *i.e.*, η_o/N_1. Each time $N_e = \sum N_i$ holograms are copied back and forth, the diffraction efficiency for each hologram is restored to $1/N_e$.

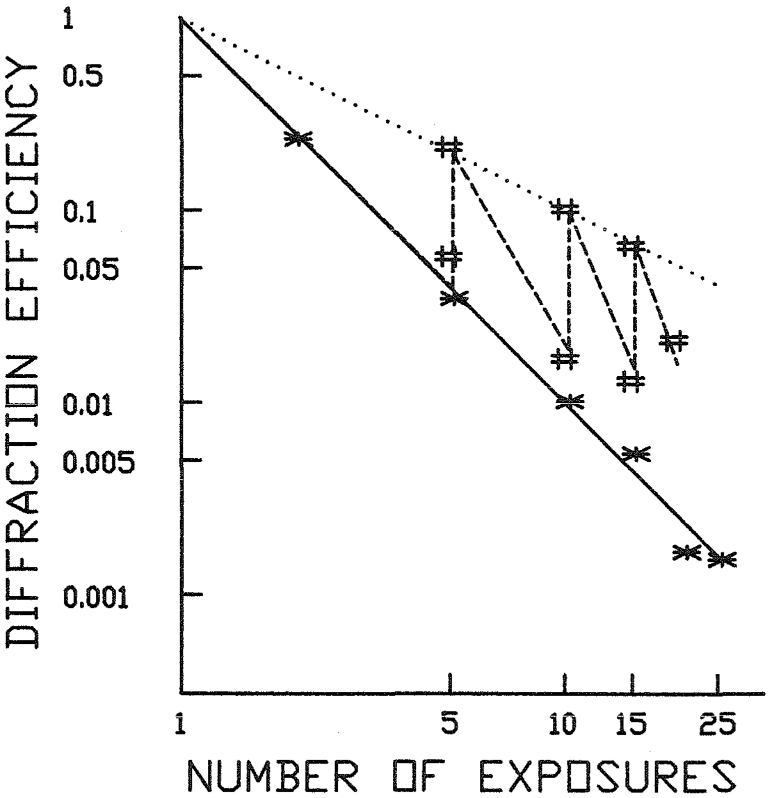


Figure 4.12. Log relative diffraction efficiency per hologram versus log number of exposures.

Fig. 4.12 is a log-log plot of the mean diffraction diffraction efficiency per hologram relative to the diffraction efficiency for a single exposure versus the number of exposures. The figure shows experimental results for recording up to 25 holograms in a SBN:Ce crystal. The diffracted phase conjugate reference for each hologram was monitored by the CCD shown in Fig. 4.11 . The solid line in Fig. 4.12 corresponds to the theoretical N_e^{-2} decay in the diffraction efficiency per hologram. The $*$s are experimental data points for the mean diffracted power of holograms stored according to this schedule. The dashed line shows the theoretical path followed by the diffraction efficiency per hologram with periodic copying when $N_1 = 5$. The # s show experimental data points found when this approach was followed. The dotted line corresponds to a decrease in diffraction efficiency proportional to N^{-1}. Thermoplastic holograms were made after 5, 10 and 15 exposures.

One might like to implement periodically refreshed storage using two photorefractive media. Unfortunately, copying between two photorefractive crystals is complicated by the decay of the hologram being copied while the second hologram is recorded. Suppose that we wish to copy a hologram between two identical photorefractive crystals. As we read out the first hologram, the light intensity diffracted from it decays exponentially at a rate proportional to the read out intensity. The rate at which the second hologram builds up is determined by the overall intensity incident on it. If the reference beam for the second hologram is too bright then the modulation depth is too weak; if it is too low then the second hologram does not build up sufficiently before the first hologram decays. If we set the intensity of the reference beam such that the modulation depth is constant at the second hologram, then the overall diffraction efficiency of the copied hologram can exceed the efficiency of the original only if the strength of the perturbation produced in the crystal by a single exposure, exceeds the strength required to achieve (in principle) 100% diffraction efficiency. This can be accomplished by selecting a crystal with the appropriate combination of electro-optic coefficient, dielectric constant, index, and thickness. For example, the critical thickness for holograms using the r_{51} electro-optic coefficient in barium titanate is approximately 2 mm. Since crystals of such thickness are readily available, it should be possible to extend the copying technique to an all photorefractive system.

We have described periodic copying between two holograms using two key simplifications: the holographic media are planar, which implies that they can be copied in one exposure, and there is a fixed number of exposures M that need to be done. The implementation of learning algorithms in optical neural networks is an important class of problems that requires an arbitrary number of exposures. From Eq. (4.78) we see that the perturbation that accumulates in the crystal is simply the average of all the exposures, independent of their number. If the individual exposures are statistically independent, then the sum in Eq. (4.78) will grow in proportion to $\sqrt{M}$. This will precisely counterbalance the $N^{-1/2}$ factor in Eq. (4.78) yielding a steady state diffraction efficiency that is independent of M. If the individual exposures are correlated (a case we have not analyzed in the section), then the sum can grow faster than $\sqrt{M}$. In this case the reduction in diffraction efficiency due to an additional exposure would have been less that what we calculated. The copying process recovers this smaller loss in diffraction efficiency and the steady state overall efficiency remains unchanged and independent of M.

The extension of periodically refreshed recording to holograms utilizing volume degrees of freedom is complicated by the fact that a volume hologram cannot in general be evaluated or copied in a single exposure. If the complete hologram is not known *a priori*, *i.e.*, if we wish to record holograms adaptively, it is not possible to ensure that the minimal number of exposures will be used in recording. The copying technique described here allows us to recover the loss in diffraction efficiency due to exposing a hologram more than R times, where R is the rank of the hologram. In our experiments R was limited to 1 by our use of a single reference beam. For holograms of higher rank, R references would be required. Using appropriate reference signals to copy a stored hologram results in an increase in the diffraction efficiency of each component of the hologram by the factor N_e/R^2, compared to the efficiency obtained without copying, where N_e is the number of exposures used to form the original hologram. In the simplest case the steady state efficiency is again independent of N_e.

4.6 FIXING MECHANISMS AND LINEAR HOLOGRAM FORMATION

Before leaving the topic of linear volume hologram formation in photorefractive materials, we briefly consider the impact of fixing techniques. Thermal fixing has been observed in $LiNbO_3$ [133, 28], where the fixing mechanism is associated with hydrogen migration [134]. Electrical fixing has been observed in $BaTiO_3$ and SBN [135, 136, 81], where the mechanism was associated with domain reversal. Fixing has been achieved in $LiNbO_3$ by heating the crystal. Within a certain temperature range it was found that the ion conductivity is greater than the carrier conductivity and ions migrate to cancel the photorefractive grating. When the crystal is exposed to read out light, the photorefractive grating is partially erased and the ion grating is unmasked. Since the ion grating is not light sensitive, the resulting hologram remains until the hologram is reheated. In SBN fixing has been achieved by heating and by applying a strong field. The suspected mechanism is domain reversal to cancel the photorefractive grating. As in $LiNbO_3$, the photorefractive grating is erased under read out illumination. Since the domains are not light sensitive, the resulting hologram remains until the crystal is repoled.

While such fixing mechanisms may play an important role in stablizing holographic interconnection systems, they are not generally of use in increasing the storage capacity of photorefractive holograms. The reason is that, if the fixing is not permanent, previously fixed holograms will decay as newly recorded holograms are fixed. This decay will result in the same N_e^{-1} fall off in field amplitude as was found above without fixing. If the fixing mechanism is permanent, it will also be saturable and the N_e^{-1} fall off for saturable effects will come into play. Thus, our conclusion with regard to fixing mechanisms is that, while they are of substantial interest in hologram stablization, they need not play a crucial role in our consideration of hologram formation.

5. HOLOGRAPHIC INTERCONNECTIONS IN WAVEGUIDES

5.1 INTRODUCTION

Many of the difficulties which we have encountered in considering the control of bulk volume holograms are avoided by thick holograms in waveguides. While the scale of the interconnections systems which can be implemented in waveguides is relatively small, holograms in waveguides are of interest because they are relatively easy to control and because they may be interfaced with integrated opto-electronic devices. Previous uses of thick holograms in waveguides have included grating couplers and distributed feedback lasers [137]. Holograms for dynamic applications have also been considered, especially in photorefractive crystals. A review of work on photorefractive holograms in waveguides is presented by Wood *et al.* [95].

The potential for information storage in integrated volume holograms was considered by Jannson [138], who showed that the number of degrees of freedom which can be stored in a planar waveguide hologram scales with the area of the hologram divided by the square of the guided wavelength. In this chapter, we consider integrated holograms for large-scale linear transformations. We rederive Jannson's result in this context and describe a novel method for recording holograms in a waveguide using unguided light. The use of unguided light dramatically simplifies the problem of forming appropriate holograms with high dynamic range. We present experimental results for holograms in photorefractive waveguides formed by titanium indiffusion in $LiNbO_3$.

A thick hologram in a waveguide may be regarded as a vector-matrix multiplier mapping an input vector corresponding to the incident field to an output vector corresponding to the diffracted field via a matrix represented by the interconnecting hologram. In the second section of this chapter we develop this analogy in detail and derive basic relationships for this system. The third section describes methods for controlling a hologram in the waveguide with unguided light. The fourth section describes an experimental demonstration of multiple grating holograms written with unguided light and reconstructed with guided light. The concluding section briefly considers applications for integrated volume holograms.

5.2 AN INTEGRATED VECTOR-MATRIX MULTIPLIER

An architecture for an integrated optical vector-matrix multiplier is shown in Fig. 5.1 [97]. Light from each of N_1 input channel waveguides is coupled into a slab waveguide and collimated by an integrated lens. The collimated beams are diffracted by a volume hologram distributed over an area A of the slab waveguide. The diffracted signals are focused by a second integrated lens into a set of N_2 output channel waveguides. In this section we derive conditions such that a unique holographic grating couples light from each input channel to each output channel. If the gratings are weak enough that the hologram does not deplete the input beams, the fraction of the l^{th} input which is diffracted to the m^{th} output is linear in the corresponding grating amplitude and the diffraction of light from the inputs to the outputs may be regarded as a vector-matrix multiplication.

As shown Fig. 5.1, we define the z axis to be the principle axis of propagation in the waveguide. The x axis is normal to the plane of the waveguide and

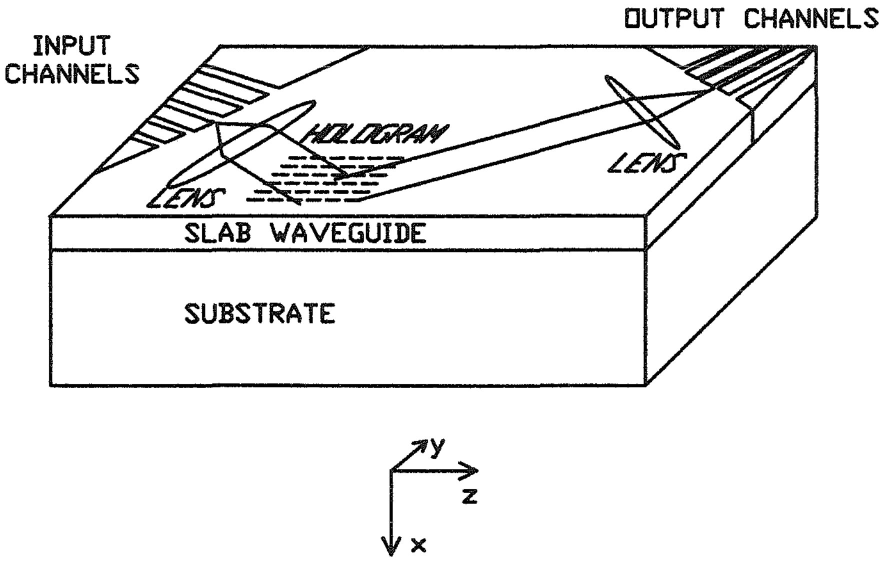


Figure 5.1. Integrated optical vector-matrix multiplier architecture.

the y axis is transverse to the z axis in the plane of the waveguide. The origin of the coordinate system is at the entrance face of the hologram at the center of the waveguide. We assume that the waveguide has a boundary with air at $x = -d/2$ and a substrate boundary at $x = d/2$. In the absence of a holographic perturbation, the regions $x < -d/2$, $-d/2 < x < d/2$ and $d/2 < x$ are homogeneous and isotropic. The indices of refraction of these regions are n_1, n_2 and n_3, respectively. We assume that $n_1 < n_3 < n_2$. The guided modes of this system have been analyzed by a number of workers [139, 137]. In considering the system of Fig. 5.1, the most important results of this analysis are that the dependence of the guided modes on x is separable from the dependence on z and y and that the wave normal curve for the longitudinal components of the guided modes is a circle. We will assume that d and (n_1, n_2, n_3) are such that the slab waveguide in Fig. 5.1 supports only one mode in x.

This system is most easily visualized in the Fourier domain. The signal from the l^{th} input channel is transformed into a plane wave propagating in a unique direction, with wavevector $\vec{k}^{(l)}$. Similarly, the signal incident on the m^{th} output channel is associated with the wavevector $\vec{k}^{(m)}$. The hologram is a perturbation, $\Delta n(\vec{r})$, to the index of the refraction of the waveguiding material. We can expand $\Delta n(\vec{r})$ in a Fourier series:

$$\Delta n(\vec{r}) = \sum_g \kappa_g e^{j\vec{K}_g \cdot \vec{r}}, \tag{5.1}$$

where κ_g and $\vec{K}_g$ are the amplitude and the wavevector of the g^{th} Fourier component. Although only one index is shown, the sum in Eq. (5.1) is over three dimensions in the space containing $\vec{K}_g$. Since the purpose of the hologram is to couple guided modes, we will not be interested in Fourier components corresponding to $\vec{K}_g$ with large components in the out-of plane ($\hat{x}$) direction. Note that $\Delta n(\vec{r}) = 0$ in the region $x < -d/2$.

At a given wavelength, the longitudinal components of the wavevectors of the guided modes of a planar waveguide are constrained to lie on a set of wave normal curves in the guiding plane. In a single mode homogeneous isotropic waveguide, these curves take the form of a circle of radius $k = 2\pi n_{eff}/\lambda$, where λ is the free space wavelength of the guided light and n_{eff} is the effective index of the waveguide. A set of input and output wavevectors on this wave normal circle is sketched in Fig. 5.2. Each of these wavevectors corresponds to an eigenmode of the waveguide. In the presence of a holographic perturbation, the plane waves

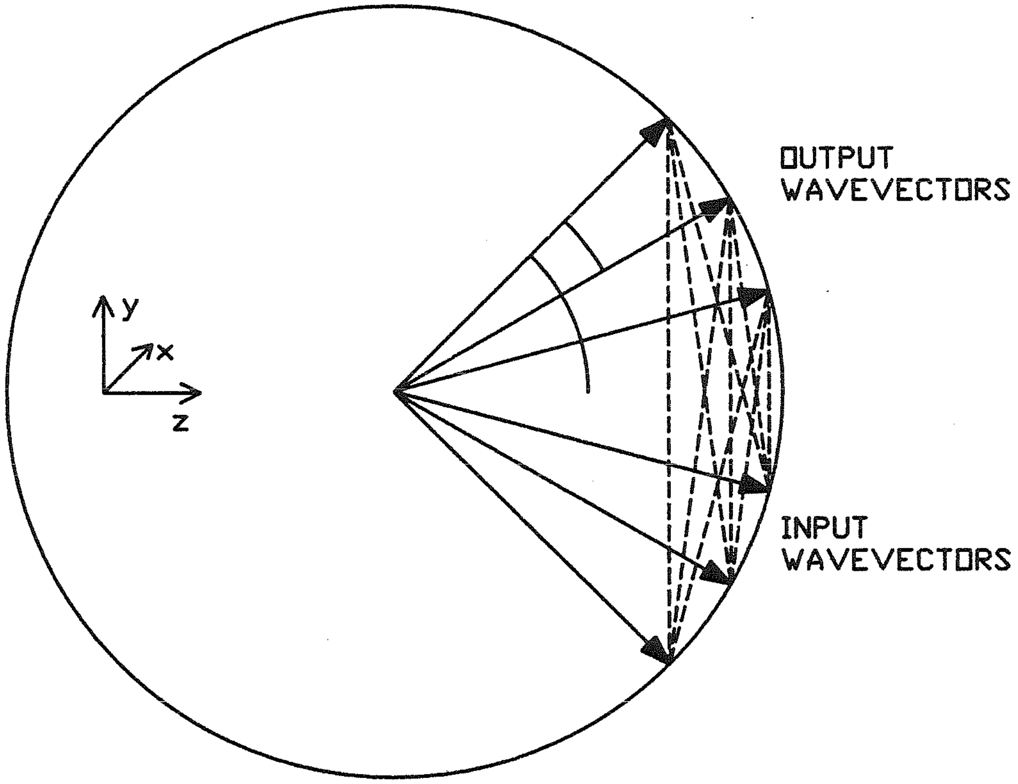


Figure 5.2. Optical wavevectors of the coupled guided beams on the wave normal curve. The connecting grating wavevectors are shown as dashed lines.

corresponding to these wavevectors cease to be eigenmodes. If the perturbation is weak, fields in the waveguide can be described in terms of the plane wave eigenmodes weighted by slowly varying amplitude functions. These amplitude functions are determined using coupled wave theory [9]. The coupling strength between the l^{th} input and the m^{th} output is proportional to the amplitude of the Fourier component of Δn with wavevector

$$\vec{K}_{lm} = \vec{k}^{(m)} - \vec{k}^{(l)}. \tag{5.2}$$

In Fig. 5.2, this condition means that the wavevector of the grating which couples an input to an output must join the endpoints of the coupled optical wavevectors. A set of grating wavevectors coupling the fields represented in the figure is shown by the dashed lines.

Let ϕ be the angle between the z axis and the central plane wave components of the incident and diffracted fields. α is the angle between adjacent plane wave components. The angle between the wavevector corresponding to the l^{th} input channel and the z axis is $-(\phi + l\alpha)$. In this framework, the wavevector of the l^{th} input is

$$\begin{aligned} \vec{k}^{(l)} &= k\cos(\phi + l\alpha)\hat{z} - k\sin(\phi + l\alpha)\hat{y} \\ &\approx k[\cos(\phi) - l\alpha\sin(\phi)]\hat{z} - k[\sin(\phi) + l\alpha\cos(\phi)]\hat{y}, \end{aligned} \tag{5.3}$$

where $\hat{x}$, $\hat{y}$ and, $\hat{z}$ are unit vectors along the corresponding axes and we have made the paraxial approximation; $l\alpha \ll 1$. The angle between the wavevector corresponding to the m^{th} output channel and the z axis is $\phi + m\alpha$. The corresponding wavevector is

$$\begin{aligned} \vec{k}^{(m)} &= k\cos(\phi + m\alpha)\hat{z} + k\sin(\phi + m\alpha)\hat{y} \\ &\approx k[\cos(\phi) - m\alpha\sin(\phi)]\hat{z} + k[\sin(\phi) + m\alpha\cos(\phi)]\hat{y}. \end{aligned} \tag{5.4}$$

The field, $\vec{E}(\vec{r})$, in the waveguide may be represented by a weighted sum of the incident and diffracted fields. We represent the x dependence of the guided modes by the normalized transverse field distribution $\mathcal{E}(x)$. Coupling from the input modes to the output modes results in a z dependence in the amplitudes of the fields in each mode. Letting $\Psi_l(z)$ and $\Phi_m(z)$ represent the amplitudes of the fields corresponding to the l^{th} and m^{th} input and output channels, respectively, the field in the waveguide is

$$\vec{E}(\vec{r}) = \mathcal{E}(x) \sum_l \Psi_l(z) e^{j\vec{k}^{(l)}\cdot\vec{\rho}} \hat{e}_l + \mathcal{E}(x) \sum_m \Phi_m(z) e^{j\vec{k}^{(m)}\cdot\vec{\rho}} \hat{e}_m, \tag{5.5}$$

where $\vec{\rho}$ is the position vector in the y-z plane $\hat{e}_l$ and $\hat{e}_m$ represent the polarization vectors of the corresponding modes. In the absence of a hologram Ψ_l and Φ_m are independent of z. The effect of the holographic coupling is determined by substituting $\vec{E}(\vec{r})$ into the wave equation

$$\nabla \times \nabla \times \vec{E} - n^2(\vec{r}) k^2 \vec{E} = 0, \tag{5.6}$$

where $k = 2\pi/\lambda$ and $n(\vec{r})$ is the index of refraction. Applying the slowly varying envelope approximation and keeping only the first order in Δn, this approach yields

$$\begin{aligned} &\sum_l j k_z^{(l)} e^{j\vec{k}^{(l)}\cdot\vec{\rho}} \frac{\partial \Psi_l(z)}{\partial z} + \sum_m j k_z^{(m)} e^{j\vec{k}^{(m)}\cdot\vec{\rho}} \frac{\partial \Phi_m(z)}{\partial z} \\ &\quad = \sum_g n_o(x) k^2 \kappa_g e^{j\vec{K}_g\cdot\vec{r}} \left[\sum_{l'} \Psi_{l'}(z) e^{j\vec{k}^{(l')}\cdot\vec{\rho}} + \sum_{m'} \Phi_{m'}(z) e^{j\vec{k}^{(m')}\cdot\vec{\rho}} \right], \end{aligned} \tag{5.7}$$

where n_o is the index distribution in the absence of the perturbation. The next step in coupled mode analysis is to match terms at identical spatial frequencies, which allows us to convert Eq. (5.6) into a set of coupled linear equations, with each equation describing how one of the modes is coupled to the rest. The difficulty in this case arises from the fact that the wavevectors of the guided modes are confined to the $y-z$ plane, whereas the grating wavevectors can point in any direction (notice the distinction between $\vec{\rho}$ and $\vec{r}$ in Eq. (5.7)). When the recorded grating has a strong component in the x direction, guided modes can only be coupled to radiation or substrate modes. Since we are interested in coupling guided modes to one another, we must keep the grating wavevector approximately confined in the $y-z$ plane. If the guided modes are tightly confined, Eq. (5.7) can be limited to the vicinity of the guiding region. Integrating Eq. (5.7) across the waveguide in the x direction, we obtain:

$$\begin{aligned}\sum_{l} j k_z^{(l)} e^{j\vec{k}^{(l)}\cdot\vec{\rho}} \frac{\partial \Psi_l(z)}{\partial z} + \sum_{m} j k_z^{(m)} e^{j\vec{k}^{(m)}\cdot\vec{\rho}} \frac{\partial \Phi_m(z)}{\partial z} \\ = \sum_{\vec{K}_g} \frac{\sin\left(K_{gx} d/2\right)}{K_{gx} d/2} n_2 k^2 \kappa_g e^{j\vec{K}_g\cdot\vec{\rho}} \left[\sum_{l'} \Psi_{l'}(z) e^{j\vec{k}^{(l')}\cdot\vec{\rho}} + \sum_{m'} \Phi_{m'}(z) e^{j\vec{k}^{(m')}\cdot\vec{\rho}}\right]\end{aligned} \tag{5.8}$$

The coupling terms due to a given $\vec{K}_g$ are significant only if $|K_{gx}|$ is small compared to $4\pi/d$. We will assume that this condition is satisfied for all $\vec{K}$ corresponding to nonzero κ_g. In this case, $\sin\left(K_{gx}d/2\right)/K_{gx}d/2 \approx 1$. Eq. (5.8) must hold independently for each distinct Fourier component. If we multiply through the equation by $e^{-j\vec{k}^{(s)}\cdot\vec{\rho}}$ and integrate over the area of the hologram, the harmonic dependence, $e^{-j\vec{k}^{(r)}\cdot\vec{\rho}}$, of each term in Eq. (5.8) is replaced by

$$\left|\int\int e^{j(\vec{k}^{(r)}-\vec{k}^{(s)})\cdot\vec{\rho}} d\vec{\rho}\right| = \frac{\sin\left(\Delta k_{rs(y)} L_y/2\right)}{\Delta k_{rs(y)}/2} \frac{\sin\left(\Delta k_{rs(z)} L_z/2\right)}{\Delta k_{rs(z)}/2}, \tag{5.9}$$

where $\Delta\vec{k}_{rs} = \vec{k}^{(r)} - \vec{k}^{(s)}$ and the integral is over the extent of the hologram in y and z, L_y and L_z. The separation in Fourier space between the wavevectors of the input and output modes is chosen such that the integral of Eq. (5.9) effectively vanishes if $r \neq s$. This implies that

$$\alpha > \frac{\lambda}{2\pi n_{eff} L_y \cos\phi}, \frac{\lambda}{2\pi n_{eff} L_z \sin\phi}. \tag{5.10}$$

In this case, multiplying both sides of Eq. (5.8) by $e^{-j\vec{k}^{(m)}\cdot\vec{\rho}}$ and integrating yields

$$\begin{aligned} & j k_z^{(m)} L_y L_z \frac{\partial \Phi_m(z)}{\partial z} = \\ & \quad \sum_g \sum_l \frac{\sin(\Delta K_{glm(y)} L_y/2)}{\Delta K_{glm(y)}/2} \frac{\sin(\Delta K_{glm(z)} L_z/2)}{\Delta K_{glm(z)}/2} n_2 k^2 \kappa_g \Psi_l(z), \end{aligned} \tag{5.11}$$

where $\Delta\vec{K}_{glm} = \vec{K}_g - (\vec{k}^{(l)} - \vec{k}^{(m)})$. Note that we assume that κ_g is zero for $\vec{K}_g$ which would cross-couple the outputs. The coupling term between the l^{th} output mode and m^{th} input mode is strong only if $\sin(\Delta K_{glm(y)} L_y/2)/(\Delta K_{glm(y)}/2)$ and $\sin(\Delta K_{glm(z)} L_z/2)/(\Delta K_{glm(z)}/2)$ are nonvanishing. This is the case if

$$|\Delta\vec{K}_{glm(y)}| < \frac{\pi}{L_y} \tag{5.12}$$

and

$$|\Delta\vec{K}_{glm(z)}| < \frac{\pi}{L_z} \tag{5.13}$$

For the spacing between modes given by Eq. (5.10), these conditions are satisfied only if $\vec{K}_g = \vec{K}_{lm} \equiv \vec{k}^{(l)} - \vec{k}^{(m)}$. Discarding terms for which this condition is not satisfied, Eq. (5.11) becomes

$$j k_z^{(m)} \frac{\partial \Phi_m(z)}{\partial z} = \sum_l n_2 k^2 \kappa_{lm} \Psi_l(z), \tag{5.14}$$

where κ_{lm} is the amplitude of the grating at spatial frequency $\vec{K}_{lm}$. A similar equation may be derived to describe coupling from the output modes to the input modes.

Since we are interested in using the system of Fig. 5.1 to implement linear transformations, we assume that the diffraction efficiency of the hologram is very weak. In this case, the input signals may be assumed to be undepleted by the hologram. Let $H_{lm} = -jL_z n_2 k^2 \kappa_{lm}/k_z^{(m)}$. The solution to Eq. (5.14) under the assumption of undepleted inputs is

$$\vec{\Phi} = \bar{\bar{H}}\vec{\Psi}, \tag{5.15}$$

where $[\vec{\Phi}]_m = \Phi_m(L_z)$, $[\vec{\Psi}]_l = \Psi_l(0)$ and we assume that $\Phi_m(0) = 0$. Since each component of $\bar{\bar{H}}$ corresponds to the an independent grating, the hologram may be used to implement an arbitrary linear transformation.

Assuming a diffraction limited system, the spatial bandwidth of the input and output fields is $B = 2\pi L_y \cos\phi/\lambda f$, where f is the focal length of the integrated lenses. The number of modes which may be used in the input and output fields, R, is equal B divided by the separation in Fourier space between modes, $2\pi n_{eff}\alpha/\lambda$. Substituting for α from Eq. (5.10) and assuming that $L_z \sin\phi = L_y \cos\phi$, $R = A\sin\phi\cos\phi/\lambda f$, where $A = L_z L_y$. The number of connections a hologram in the plane can make between channels is

$$R^2 = \frac{A^2 \sin 2\phi}{2\lambda^2 f^2}. \tag{5.16}$$

If f^2 is of the same order as A, R^2 scales as A/λ^2 [138]. Assuming $L_y \cos\theta/f = 1$, $A = 1\text{cm}^2$ $\lambda = .5\mu\text{m}$, and $\phi = \pi/6$, $R^2 = 10^8$.

5.3 FORMATION OF INTEGRATED VOLUME HOLOGRAMS

In this section we consider methods for recording a hologram in a slab waveguide using unguided light incident from above the waveguide (along the x axis of Fig. 5.1). Recording with unguided light yields two advantages over recording with guided light. First, unguided recording avoids unintended perturbations to the waveguide. If the waveguide is sensitive to the the guided beams it is difficult to avoid crosstalk between input beams and between output beams and damage to the waveguide outside of the holographic region. With unguided control, the unintended gratings which give rise to crosstalk are not recorded. The problem of damage to the waveguide can be overcome if the waveguide is not sensitive to the guided beams. In photorefractive materials, for example, there is a threshold wavelength beyond which holograms may no longer be written. This allows us to use choose a guided wavelength which does not damage the waveguide while using a shorter unguided wavelength to control the hologram. The second advantage of unguided control is that it allows us to control the hologram in the guided plane with a single exposure. In contrast, multiple exposures are needed to form an arbitrary hologram with guided recording beams. As we saw in previous chapters, recording with multiple exposures is in general undesirable because it greatly complicates dynamic control of the hologram and because it is difficult to maintain a large dynamic range in multiply exposed holograms.

We consider two algorithms for recording a hologram in the guided plane. In the first case each of the N_1N_2 gratings in the plane is formed using a distinct pair of control beams. In the second case the gratings are formed using a single reference beam and N_1N_2 beams generated by a spatial light modulator (SLM). The advantage of the first approach is that all the gratings may be recorded in the plane with no out-of-plane components. The second approach sacrifices precision in the spatial orientation of the recorded gratings in order to simplify the generation of the recording beams and improve the modulation depth with which each grating is recorded.

The distinction between the two recording techniques can be clarified using a simple graphical technique. We have seen that for a grating with wavevector $\vec{K}_g$, the Bragg condition, Eq. (5.2), specifies the coupled optical wavevectors, $\vec{k}^{(l)}$ and $\vec{k}^{(m)}$, uniquely in the guided plane. Eq. (5.2) may, however, be satisfied for a given $\vec{K}_g$ by many pairs of unguided optical wavevectors. Consider the three-dimensional wave normal surface sketched in Fig. 5.3. We assume for simplicity that this surface is a sphere. Any pair of points on the surface which are joined by $\vec{K}_g$ are Bragg matched to the grating. As sketched in Fig. 5.3, the locus of such pairs is a cone whose axis is parallel to the grating wavevector. The cone intersects the normal surface in two parallel circles separated from each other by the grating wavevector.

Consider a 2-D version of Fig. 5.3 formed by projecting the wave normal surface onto the guided plane. Such a figure is sketched in Fig. 5.4 . The circle in this figure is the wave normal curve in the guided plane. The two parallel lines are the projections of the circles on the cone of Fig. 5.3 . We will refer to these lines as the "degeneracy lines" of $\vec{K}_g$. The degeneracy lines are separated by $\vec{K}_g$. The optical wavevectors in the plane which are coupled by the grating lie on the wave normal curve at the points where it intersects the degeneracy lines. The backward propagating conjugates of the coupled pair, shown in the

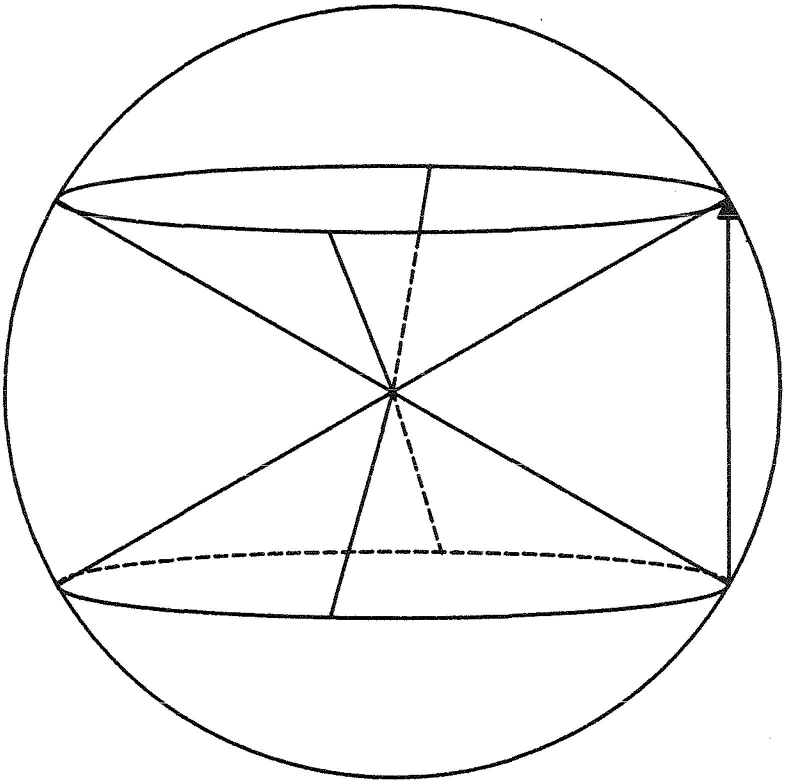


Figure 5.3. Wave normal surface showing a grating wavevector and its associated degeneracy cone.

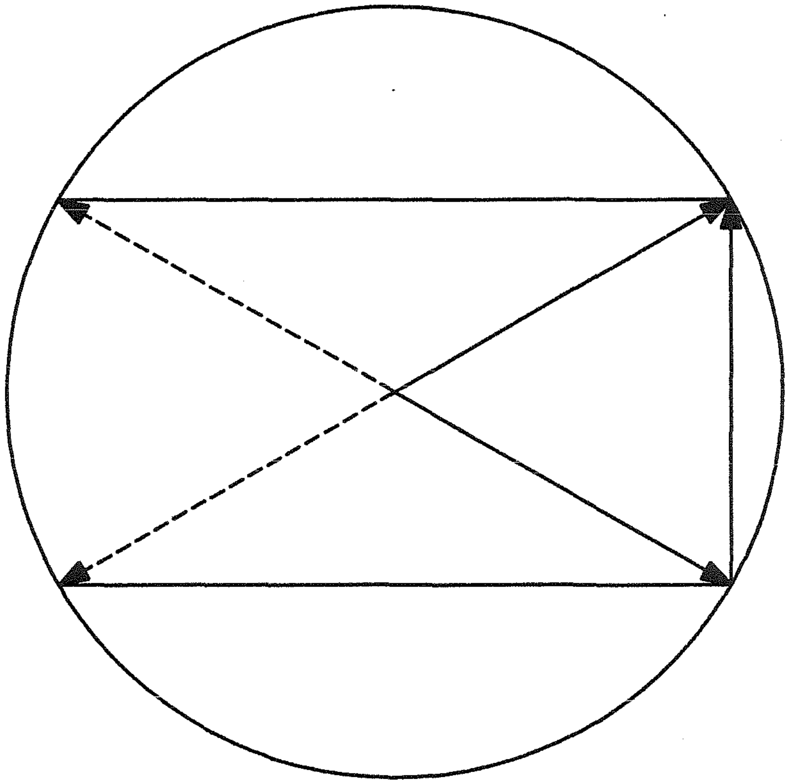


Figure 5.4. Wave normal surface projected onto a plane.

figure as dashed lines, are also coupled. Fig. 5.4 is equivalent to Fig. 5.3 in the sense that there is a one-to-one correspondence between points on the interior of the wave normal curve in Fig. 5.4 and points on the hemisphere of the normal surface above the guided plane. Thus, each point on the interior of the wave normal curve corresponds to an unguided beam incident from above the plane. Each point on the wave normal curve corresponds to a guided beam.

We are now prepared to consider in more detail the two recording methods mentioned above. In the first case, we record $\vec{K}_g$ using a pair of wavevectors separated by $\vec{K}_g$ on opposite degeneracy lines. As an example, Fig. 5.5(a) shows the projection in the guided plane of a pair of unguided optical wavevectors which could be used to record the grating of Fig. 5.4. Of course, our goal is to simultaneously record many gratings in the plane. Fig. 5.5(b) shows a set of grating wavevectors (as dotted lines) and the degeneracy lines associated with them. For simplicity, the optical wavevectors coupled by these gratings are not shown. To record these gratings from above the plane, we must select an appropriate pair of optical signals for each pair of degeneracy lines. Cross-talk between the recording signals need not concern us because gratings between optical wavevectors which are not on opposite degeneracy lines will not in general be Bragg matched to couple the signals in the plane. In Fig. 5.5(c) we have translated the grating wavevectors of Fig. 5.5(b) along the associated degeneracy lines. Each endpoint of a grating wavevector corresponds to an unguided optical wavevector which is to be used to control that grating.

This recording method is unattractive for two reasons. First, since each grating is written by a separate pair of control beams, $2N_1N_2$ distinct signals must be controlled independently. Second, the modulation depth of the fringe pattern

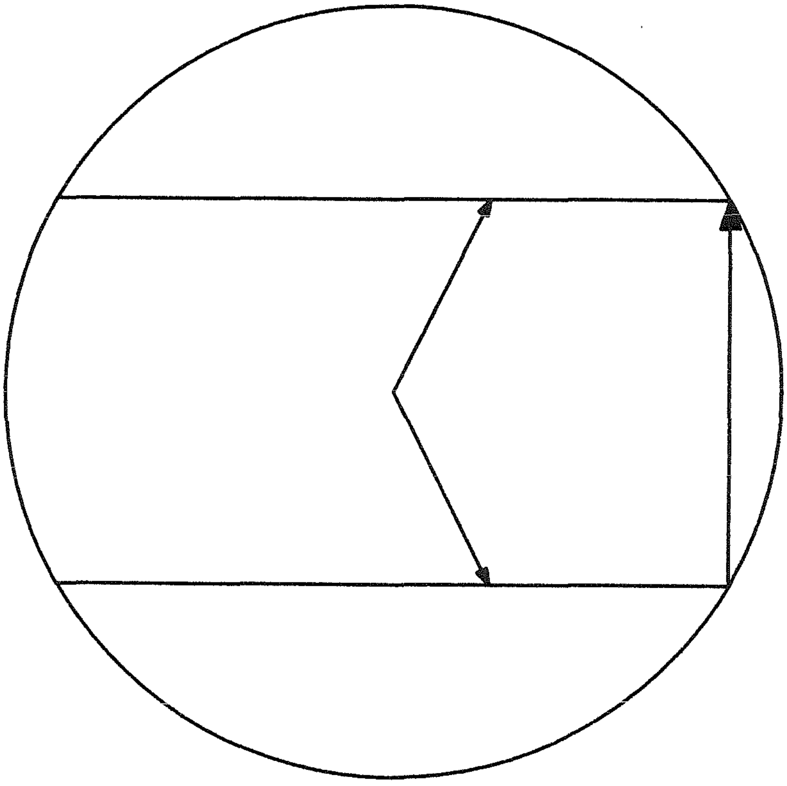


Figure 5.5(a). Projection in the guided plane of the normal surface and a pair of unguided propagating wavevectors. The grating wavevector and the degeneracy lines in the plane are also shown.

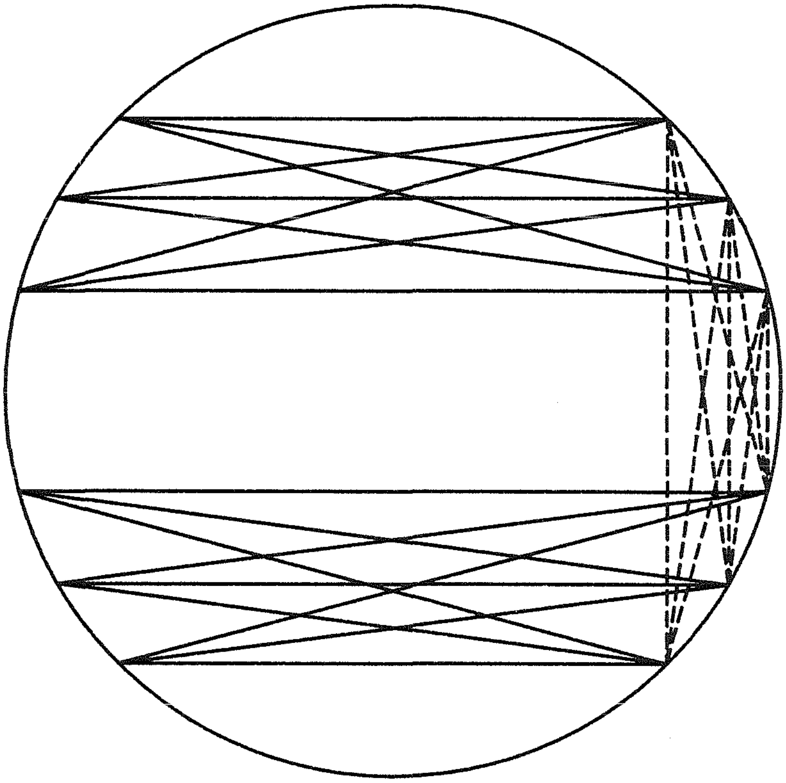


Figure 5.5(b). Set of grating wavevectors and their associated degeneracy lines. The grating wavevectors interconnect three input guided channels with three output channels.

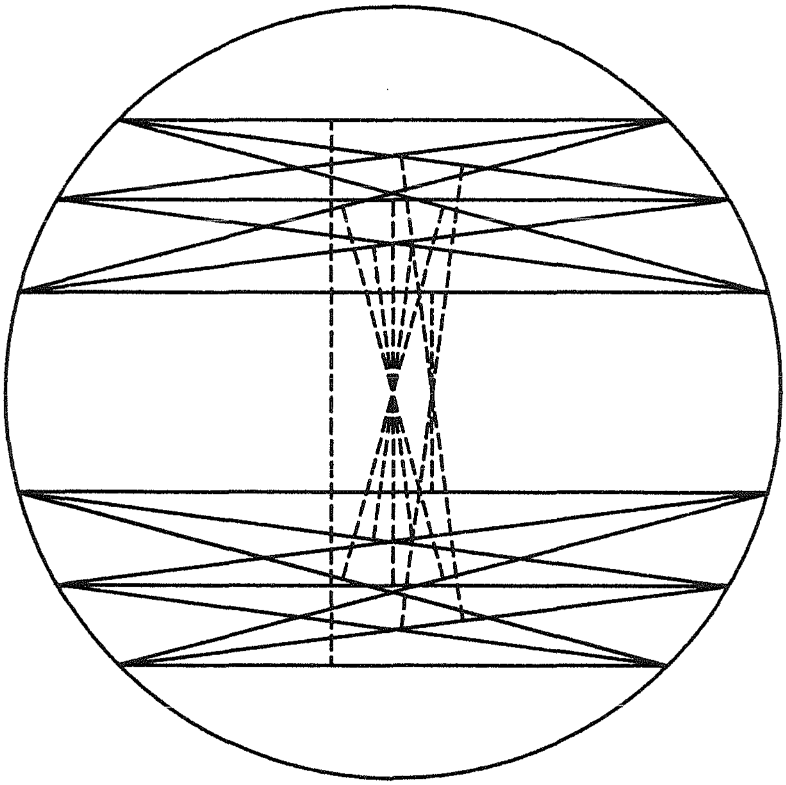


Figure 5.5(c). Grating wavevectors of Fig. 5.5(b) displaced along the degeneracy lines.

corresponding to each grating is weak when all N_1N_2 gratings are recorded with separate pairs of control beams. A much stronger modulation depth is achieved if all N_1N_2 gratings are written using a single reference beam [97]. However, a grating can be written from above the plane only when both of the writing beams lie on its degeneracy lines. Thus, a set of grating wavevectors may be recorded with a single reference beam only if there exists a point which lies on a degeneracy line of each and every the grating. Since knowledge of two wavevectors on a degeneracy line, *i.e.*, a reference and an input, uniquely specifies a grating wavevector, such an intersection point is impossible if there is fan-in or fan-out in the interconnection matrix. The second recording method we consider overcomes this dilemma by accepting a tilt out of the guided plane in the recording gratings. Such a tilt is allowable because the readout beams are confined to a thin guiding region. As we saw in Eq. (5.8), the component of the grating wavevector out of the guiding plane may be as large as π/d before the coupling strength of an otherwise Bragg matched grating is seriously affected. Since the waveguide may be only a few wavelengths thick, guided beams may reconstruct gratings with fairly sizable out-of-plane components.

If we are not concerned about Bragg mismatch out of the plane of the waveguide, we can record the interconnection pattern of Fig. 5.5(b) with any group of optical signals which record gratings with planar components matched to the desired gratings. The projection in the guided plane of the endpoints of the wavevectors of such a set of signals is similar to similar to Fig. 5.5(c) in the sense that there exists a line between a pair of endpoints which corresponds to each of the desired grating wavevectors. If we do not worry about Bragg mismatch out of the plane, however, signals which record a grating need not lie on its

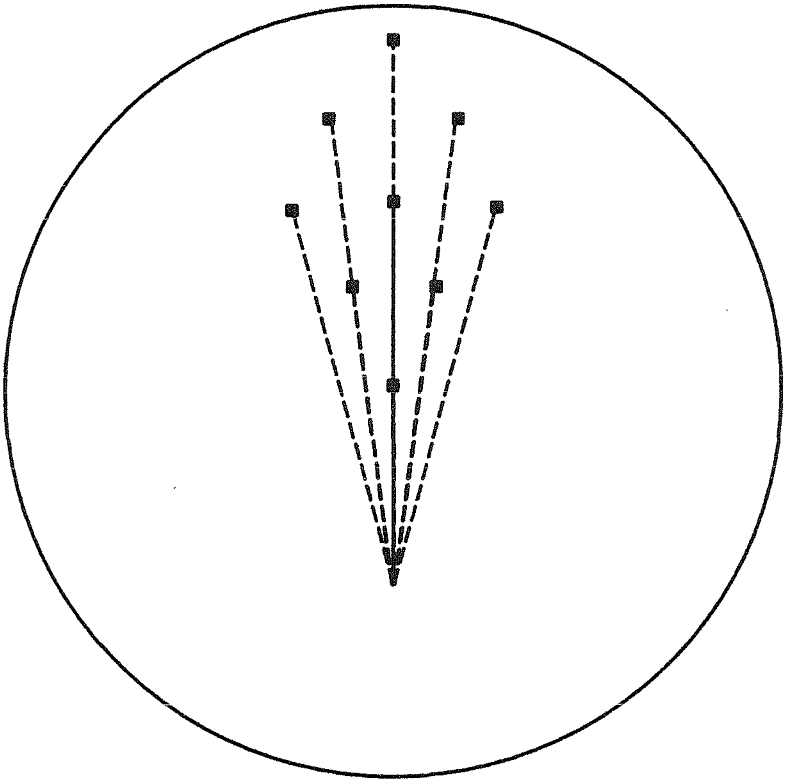


Figure 5.6. Projection in the guided plane of the endpoints of a set of unguided recording beams for recording the interconnection matrix of Fig. 5.2 using a single reference.

degeneracy lines. Since we would like to record all the gratings using a single reference, we consider the set of recording beams with wavevectors corresponding to the endpoints shown in Fig. 5.6. The grating wavevectors shown in this figure correspond to the interconnection pattern of Fig. 5.5(b) with an endpoint of each the grating wavevectors shifted onto a single reference point. Since the grating wavevectors no longer join points on their degeneracy lines, the actual gratings recorded between the reference beam and the signal beams have out-of-plane components. However, if the out-of-plane component of the actual grating wavevector recorded for each desired grating wavevector is less than π/d, then the gratings recorded using the beams suggested by Fig. 5.6 will implement the interconnection pattern shown in Fig. 5.5(b). The remainder of this section is dedicated to showing that this constraint on the out-of-plane components can be satisfied for gratings recorded in a specific recording architecture.

An architecture for recording a hologram in the waveguide using a single reference is shown in Fig. 5.7. The hologram is formed between the Fourier transform of the signal recorded on a spatial light modulator and a plane wave reference. The SLM consists of a 2-D array of S independently controllable pixels. Each pixel controls a single grating in the waveguiding plane. We refer to the pixel that controls the $(lm)^{th}$ grating as the $(lm)^{th}$ pixel. The SLM lies in the y'-z' plane. The optical axis of the SLM-Fourier lens system is along x'. The optical axis for light propagating in the waveguide is along z. x is normal to the waveguide and y is transverse to the propagation direction in the plane of the waveguide. The $x'y'z'$ coordinate system corresponds to the xyz system rotated by an angle $-\theta$ about z. The reference beam is assumed to propagate in a direction normal to z and at an angle θ with respect to x. The geometry in

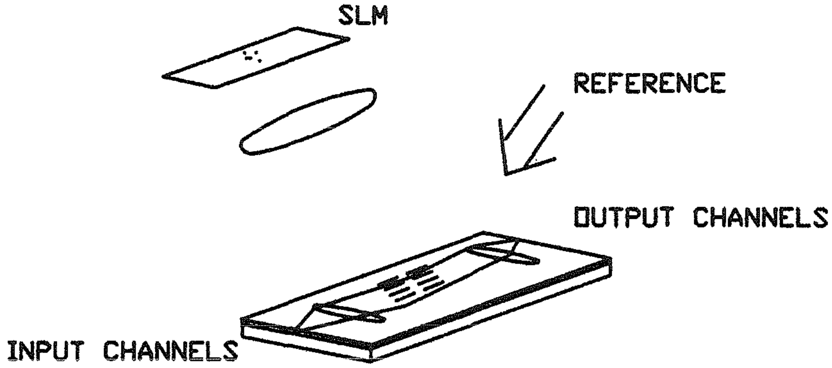


Figure 5.7. Architecture for control of guided volume holograms with a single reference and SLM

the plane is the same as in Fig. 5.1.

The cross sectional area at the plane of the waveguide of a beam of light collimated from one pixel of the SLM is $\lambda_r^2 F^2/\delta$, where δ is the area of a single pixel, λ_r is the wavelength of the recording light, and F is the focal length of the Fourier lens. If

$$\lambda_r^2 F^2/\delta > A, \tag{5.17}$$

where A is the area of the holographic interaction region in the waveguide, then the light generated by a pixel centered at (y', z') is to a good approximation a plane wave propagating with the wavevector

$$
\begin{aligned}
\vec{k} &= k_r(1 - \frac{(u^2+v^2)}{2k_r^2})\hat{x'} + u\hat{y'} + v\hat{z'} \\
&= [k_r(1 - \frac{(u^2+v^2)}{2k_r^2})\cos\theta + u\sin\theta]\hat{x} \\
&\quad + [u\cos\theta - k_r(1 - \frac{(u^2+v^2)}{2k_r^2})\sin\theta]\hat{y} + v\hat{z},
\end{aligned}
\tag{5.18}
$$

where $u = \frac{k_r y'}{F}$, $v = \frac{k_r z'}{F}$, amd $k_r = \frac{2\pi}{\lambda_r}$. Let $\vec{k}_{lm}$ be the wavevector for light collimated from the $(lm)^{th}$ pixel. u_{lm} and v_{lm} correspond to u and v for the $(lm)^{th}$ pixel. Let H_{lm} represent the optical field at the lm^{th} pixel . The field due to the SLM near the plane of the waveguide may be written

$$
\begin{aligned}
U(x,y,z) =& e^{jk_r(x\cos\theta - y\sin\theta)} \\
&\times \sum_{lm} H_{lm} e^{-jk_r(x\cos\theta - y\sin\theta)\frac{(u_{lm}^2+v_{lm}^2)}{2k_r^2}} e^{j(u_{lm}(y\cos\theta + x\sin\theta) + v_{lm}z)}.
\end{aligned}
\tag{5.19}
$$

The reference field is $R = e^{jk_r(x\cos\theta + y\sin\theta)}$. The interference pattern between the signal and reference fields is

$$
\begin{aligned}
I(x,y,z) =& e^{-j2k_r y\sin\theta} \\
&\times \sum_{lm} H_{lm} e^{-jk_r(x\cos\theta - y\sin\theta)\frac{(u_{lm}^2+v_{lm}^2)}{2k_r^2}} e^{j(u_{lm}(y\cos\theta + x\sin\theta) + v_{lm}z)} \\
&\quad + c.c.
\end{aligned}
\tag{5.20}
$$

Assuming that the holographic perturbation is linear in $I(x,y,z)$, the amplitude of the Fourier component of the perturbation at spatial frequency

$$
\begin{aligned}
\vec{K}_{lm} =& (u_{lm}\sin\theta - \cos\theta\frac{(u_{lm}^2+v_{lm}^2)}{2k_r})\hat{x} \\
&- (2k_r\sin\theta - u_{lm}\cos\theta - \sin\theta\frac{(u_{lm}^2+v_{lm}^2)}{2k_r})\hat{y} + v_{lm}\hat{z}
\end{aligned}
\tag{5.21}
$$

is proportional to H_{lm}.

Suppose that we wish to control the interconnection pattern of Fig. 5.2 using the architecture of Fig. 5.7. To control the lm^{th} grating with the lm^{th} pixel we must select u_{lm}, v_{lm}, and θ such that the components of $\vec{K}_{lm}$ in the guided plane given by Eq. (5.21) are equal to $\vec{K}_{lm}$. By assuming that $(u_{oo}, v_{oo}) = (0,0)$ and discarding second order terms we find $k_r \sin\theta = k \sin\phi$ and

$$\begin{aligned} u_{lm} &= (l+m)\alpha k \frac{\cos\phi}{\cos\theta} \\ v_{lm} &= (l-m)\alpha k \sin\phi. \end{aligned} \tag{5.22}$$

$\vec{K}_{lm}$ and ϕ are the same as in the previous section. Comparing Eq. (5.22) with Eq. (5.18) we see that the lm^{th} grating controlled by the field on a pixel centered at $(y'_{lm}, z'_{lm}) = ((l+m)\alpha F(\lambda_r \cos\phi/\lambda\cos\theta), (l-m)\alpha F(\lambda_r/\lambda)\sin\phi)$. Gratings which couple the l^{th} input to the output channels are controlled by the pixels on the line

$$y' = -\frac{z'}{\tan\phi\cos\theta} + 2l\alpha F \frac{\lambda_r \cos\phi}{\lambda\cos\theta}. \tag{5.23}$$

Gratings which couple the from input channels to the m^{th} output are controlled by pixels on the line

$$y' = \frac{z'}{\tan\phi\cos\theta} + 2m\alpha F \frac{\lambda_r \cos\phi}{\lambda\cos\theta}. \tag{5.24}$$

To ensure that the single reference exposure method does not violate the Bragg condition out of the plane, we require that the x component of Eq. (5.21) be less than π/d:

$$|(u_{lm}\sin\theta - \cos\theta \frac{(u_{lm}^2 + v_{lm}^2)}{2k_r})| < \frac{\pi}{d}. \tag{5.25}$$

Let r be the radius of the active area of the SLM. Eq. (5.25) is satisfied for all l

and m if

$$|(\frac{k_r r}{F}\sin\theta + \cos\theta\frac{k_r r^2}{2F^2})| < \frac{\pi}{d}. \tag{5.26}$$

Assuming that $\tan^2\theta \ll (k_r d)$, Eq. (5.26) simplifies to

$$r < \sqrt{2\pi}\frac{F}{\sqrt{k_r d}}. \tag{5.27}$$

The number of gratings in the guided plane that we can control using the SLM is $S = \pi r^2/\delta$. Substituting from Eq. (5.27) we find

$$S < \frac{\lambda_r F^2}{d\delta}. \tag{5.28}$$

Substituting from Eq. (5.17), we see that Eq. (5.28) is satisfied if

$$S = \frac{A}{\lambda_r d}. \tag{5.29}$$

S differs from the number of gratings we can distinguish in the plane, given as R^2 in Eq. (5.16) , only by the factor $\lambda^2/\lambda_r d$.

5.4 EXPERIMENTAL RESULTS

In our experiments the SLM and lens of Fig. 5.7 are simulated by a square-wave grating. The experimental setup is sketched in Fig. 5.8. The grating lies in the y'-z' plane. The transmittance of the grating is a one-dimensional function, $t(\xi)$, which may be described by a Fourier series over harmonics of a fundamental frequency $K_f\hat{\xi}$, where $\hat{\xi}$ is a unit vector in the y'-z' plane. Assuming that a

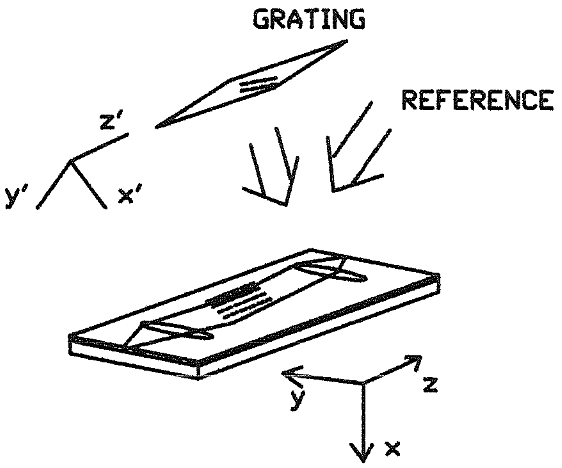


Figure 5.8. Experimental out-of-plane recording system.

plane wave propagating along the positive x' axis is incident on the grating, the wavevector of the n^{th} harmonic of the optical field diffracted from the grating is

$$\vec{p^{(n)}} = k_r\sqrt{1-\frac{n^2K_f^2}{k_r^2}}\hat{x}' + nK_f\hat{\xi}\cdot\hat{y}'\hat{y}' + nK_f\hat{\xi}\cdot\hat{z}'\hat{z}'. \tag{5.30}$$

Comparing Eq. (5.30) with Eq. (5.18), we define $u_n = nK_f\hat{\xi}\cdot\hat{y}'$ and $v_n = nK_f\hat{\xi}\cdot\hat{z}'$. The field transmitted by the grating is

$$U(x,y,z) = \sum_n H_n e^{j\vec{p^{(n)}}\cdot\vec{r}}. \tag{5.31}$$

A hologram is formed in the waveguiding plane between this field and the reference beam $R = e^{jk_r(x\cos\theta + y\sin\theta)}$. The amplitude of the Fourier component of this hologram at spatial frequency $\vec{K}_n = \vec{p}^{(n)} - k_r(\cos\theta\hat{x} + \sin\theta\hat{y})$ is proportional to H_n.

The component of the hologram at spatial frequency $\vec{K}_n$ couples the l^{th} input to the m^{th} output only if $\vec{K}_n = \vec{K}_{lm}$. Substituting u_n and v_n from Eq. (5.27) into Eq. (5.19) and solving for l and m we find $\vec{K}^{(n)} = \vec{K}_{lm}$ when

$$\begin{aligned} l &= \frac{nK_f}{\alpha k}\left(\cos\gamma\frac{\cos\theta}{\cos\phi} + \frac{\sin\gamma}{\sin\phi}\right) \\ m &= \frac{nK_f}{\alpha k}\left(\cos\gamma\frac{\cos\theta}{\cos\phi} - \frac{\sin\gamma}{\sin\phi}\right) \end{aligned} \tag{5.32}$$

where $\gamma = \cos^{-1}(\hat{\xi}\cdot\hat{y}')$. If $\tan\gamma = -\cos\theta\tan\phi$ then $l = 0$ and the n^{th} harmonic controls the coupling from the 0^{th} input channel to the m^{th} output channel, where m is given by Eq. (5.32). $\hat{\xi}$ is parallel to the line of Eq. (5.23) in this case. Similarly, if $\tan\gamma = \cos\theta\tan\phi$, $m = 0$ and the n^{th} harmonic controls the coupling from the output channel corresponding to $m = 0$ to the l^{th} input channel.

Fig. 5.9 is a sketch of the grating wavevectors formed in the guided plane by the five lowest order Fourier components of a recording grating at various rotations, γ, of the grating in the y'-z' plane. A section of the wave normal curve and the positions of the five grating wavevectors recorded on the normal surface are shown for each value of γ. In the case shown $\phi = 15°$, $\frac{k_r}{k} = 1.5$, and $\frac{K_f}{k} = 0.02$. At $\gamma = -\tan^{-1}(\cos\theta\tan\phi) = -14.8°$, a single input is coupled to 5 different outputs.

We have recorded photorefractive holograms in single mode titanium indiffused waveguides on nominally pure y-cut $LiNbO_3$ substrates. Holograms were recorded from above the waveguide using the 514 nm light from an Ar^+ laser. Guided 633 nm HeNe light was used to reconstruct the holograms. The ratio of

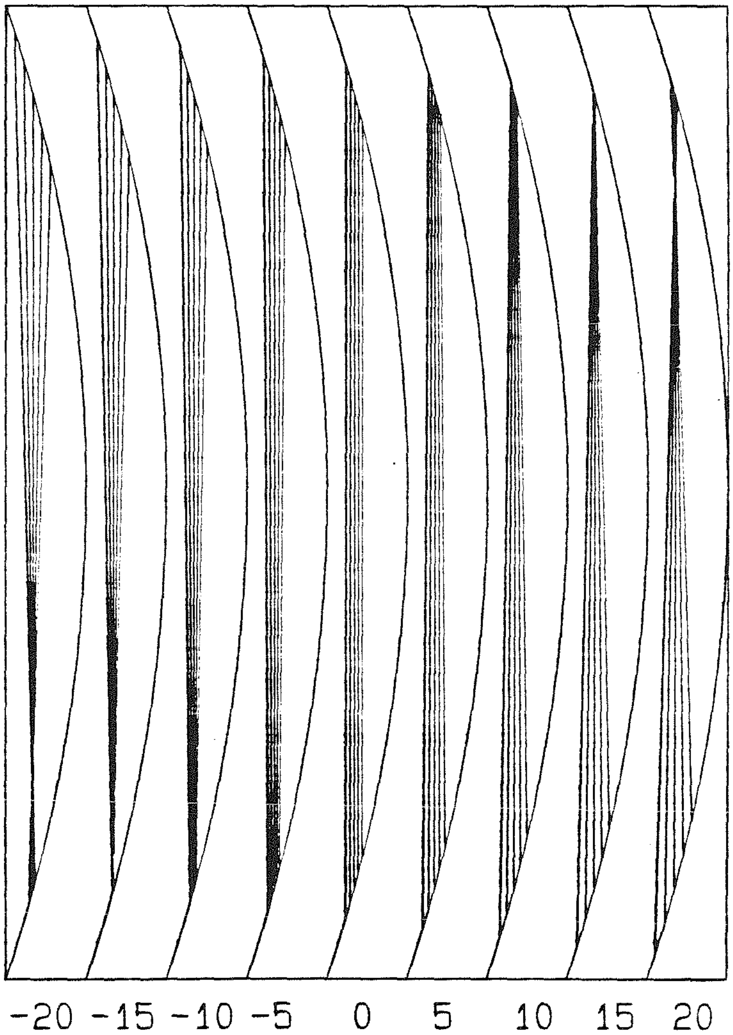


Figure 5.9. Grating wavevectors in the plane verses rotation of the recording grating.

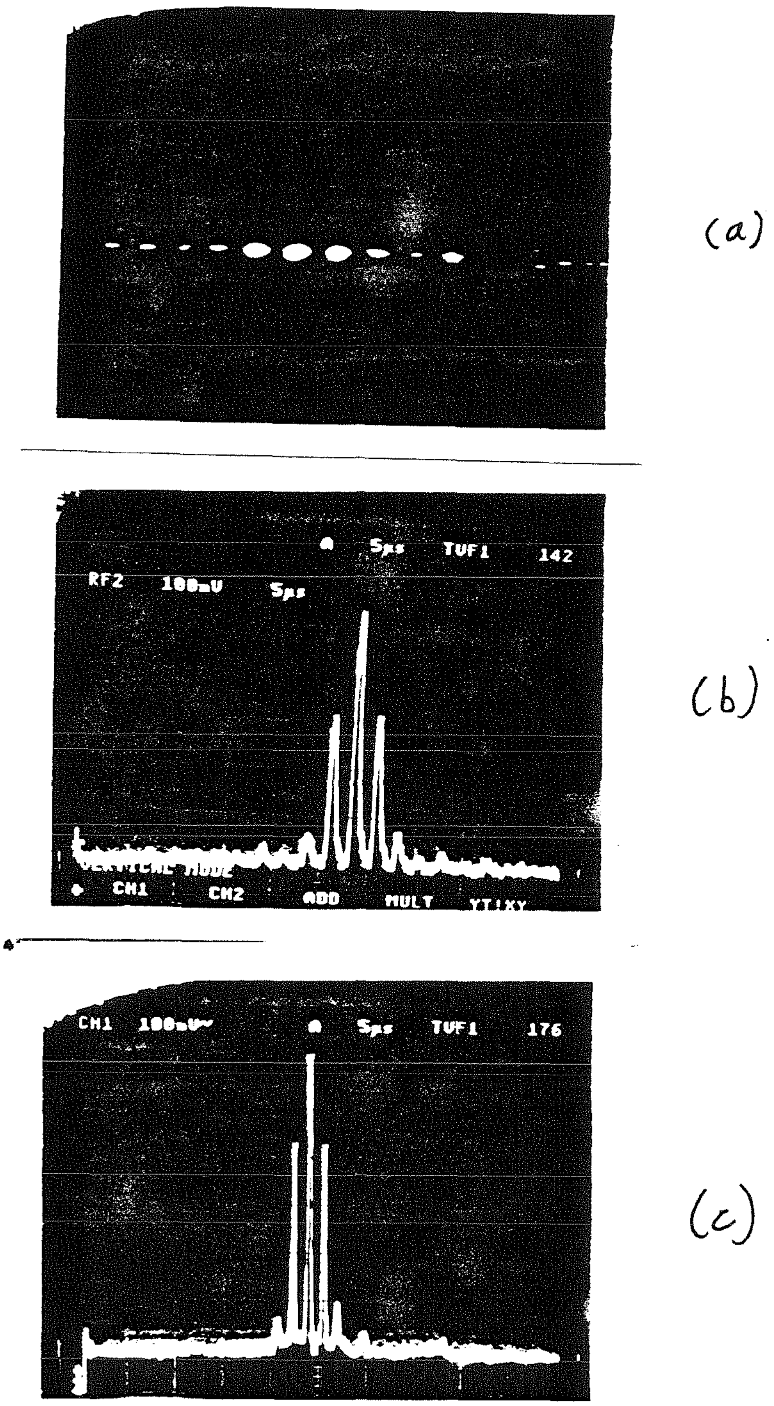


Figure 5.10. Diffracted HeNe signals at the output of a waveguide. The diffracting hologram was formed using unguided Ar^+ control beams.

the guided HeNe wavelength to the Ar^+ wavelength in the crystal was ≈ 1.5. The area of slab waveguide used to write the holograms was $\approx 1cm^2$. As in Fig. 5.8, holograms were formed between a plane wave reference and light diffracted from a grating. For the experiments described here, we used a Ronchi grating with 100 lines/inch. By rotating the grating in the y'-z' plane, we were able to observe variations in the pattern of grating wavevectors stored in the plane similar to those shown in Fig. 5.9. For ϕ was 1.5°, $\theta = 1°$, and $\gamma = -1.5°$ only one input was diffracted. The diffraction pattern end-coupled out of the waveguide for this input is shown in Fig. 5.10(a). The 14 lowest diffracted orders are visible. (The angles inside the waveguide are given. The angle between the Ar^+ beams incident on the substrate was 5°. Small angles were used because the sample we used is not guiding along y.)

We have assumed that the amplitude of the hologram we write in the waveguide is proportional to the recording intensity. This is not quite the case for photorefractive holograms, which are linear in the modulation depth of the recording fringe pattern. If the intensity is fairly uniform, however, the distinction need not concern us here. Under the assumption of linear recording, the strength of a connection in the plane is proportional to H_{lm}. This is confirmed in Fig. 5.10(b) and (c). Fig. 5.10(b) shows the squared spectrum of the Ronchi grating used to make the hologram of Fig. 5.10(a). Fig. 5.10(c) shows, on a slightly different horizontal scale, the power in the central spots of Fig. 5.10(a). The noise in the base lines of Fig. 5.10 is the dark noise of a CCD camera.

5.5 APPLICATIONS

Volume holograms in waveguides are applicable to systems which require fixed or adaptive linear transformations. While the out-of-plane control method described here provides a simple method for producing integrated holograms which fixed transformations, the use of this method for dynamic control of connections in a plane is also promising. Such a dynamically controlled vector-matrix multiplier could find application in switching networks or in adaptive artificial neural networks.

Semiconducting photorefractive materials are especially attractive for dynamic implementations of integrated volume holography. This is because photorefractive response times in these materials are relatively fast and, at least in the case of GaAs, integrated technologies are well developed. Using GaAs, we may expect to be able to monolithically construct architectures such as that shown in Fig. 5.11. In this system an array of laser diodes is dynamically connected to an array of photodetectors using a photorefractive hologram with out-of-plane control. Since photorefractive effects arise on fairly short time scale in GaAs [18], one could expect to reconfigure the interconnection matrix in this device in a few microseconds. Depending on the number of input channels which could be integrated, from 10^4 to 10^8 weighted interconnections could be stored in this device. One problem with integrating a large number of input channels might be the integration of more than 100 lasers on a single chip. This problem could be overcome by fanning out the laser outputs to feed several input channels, each channel being controlled by a simple modulator. The principle difficulty arising in the fabrication of this device is the well known problem of monolithically inte-

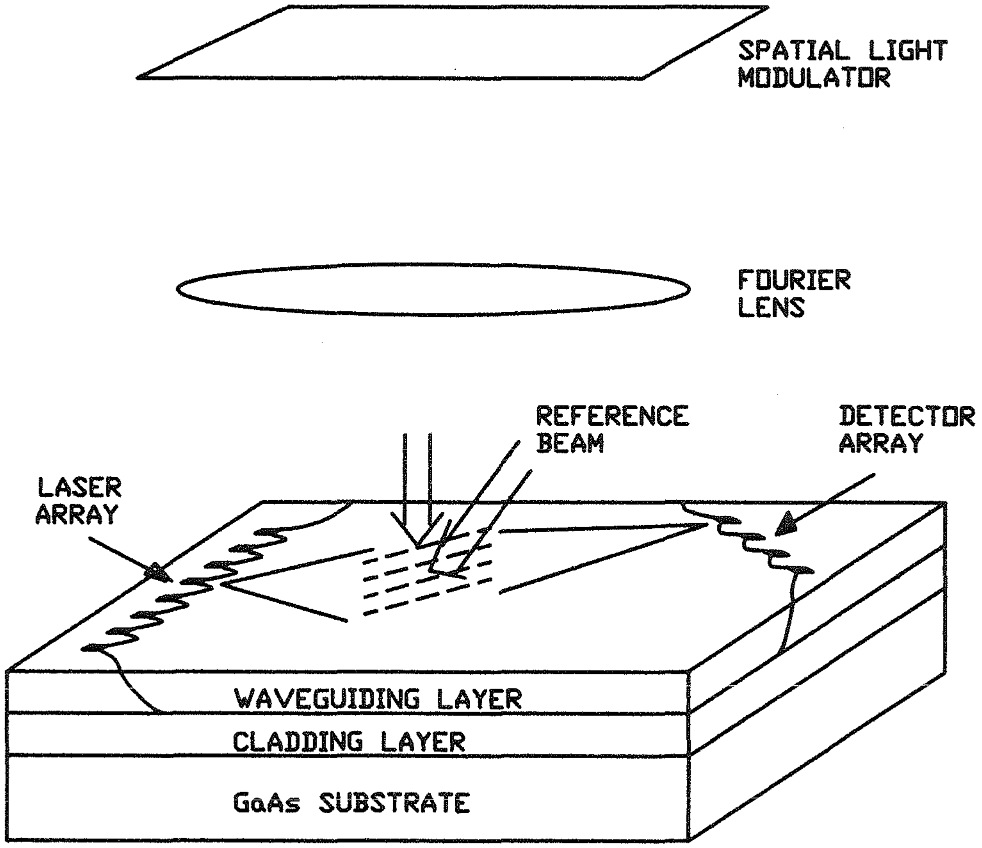


Figure 5.11. Monolithic integration of active devices and photorefractive waveguides in GaAs.

grating active and linear waveguiding regions. While considerable further work is needed to overcome this difficulty, many grating integrated holograms clearly offer interesting new possibilities to holographic information processing.

6. LEARNING IN OPTICAL NEURAL NETWORKS

6.1 INTRODUCTION

The potential advantages of neural computation over alternative computational strategies lie in the ability of neural machines to "learn" to solve problems. In contrast to conventional computers, where the system designer must have detailed knowledge of the algorithm used to solve the problem, a neural computer adapts itself to solving a given task without specific instructions. The detailed programming of a neural machine consists of specifying a learning algorithm to control the adaptation of the network. The network learns to produce the appropriate response to a class of inputs by being presented with a sufficient number of examples. The presentation of these examples causes the strength of the connections between neurons that comprise the network to be modified according to the specifics of the learning algorithm. A successful learning procedure will result in a trained network that responds correctly to the examples it has seen previously and also to other inputs that are in some sense similar to the known patterns. When we consider a physical realization of a neural network model, we have two options in incorporating learning capability. The first is to build a network with fixed but initially programmable connections. An auxiliary conventional computer is used to "learn" the correct values of the connection strengths. Once learning is complete the network is programmed by the computer. While this approach may be reasonable for some applications, the second alternative, a system with continuously modifiable connections is potentially much more powerful.

Both of these options can be implemented in volume holographic systems. In chapter 3, we saw how a known linear transformation can be recorded in a hologram. This procedure could be used to load a set of precalculated connection strengths into a network. In the second section of this chapter, we describe an experimental system for loading a prescribed set of weights into a hologram. Since the weights are never changed after programming, the only measure of the network's performance is the size of the network which can be constructed at reasonable cost to operate with reasonable speed. Due to the constraints on the diffraction efficiency and dynamic range of multiply exposed holograms described in section 2.4, section 3.4, and section 4.3, the potential of holographic systems for implementation of large scale interconnection networks is somewhat less overwhelming than it might seem at first glance. A comparison of optical and electronic neural interconnections is presented in the last section of this chapter. For dynamic learning applications, however, the ease with which learning can be implemented in holographic systems substantially improves the attractiveness of optical systems.

Most models of dynamic neural learning are based on Hebb's conjecture that if a pair of neurons are consistently simultaneously active then strength of the connection between them should increase [140]. In dynamic holographic systems, the connection between a pair of modes grows in proportion to the activity of the two modes. This fundamental correspondence between Hebbian learning and hologram formation is perhaps the most compelling motivation for the development of holographic neural systems. In the third section of this chapter, an experimental dynamic photorefractive learning system is described. Ultimately, more complex learning systems could be implemented by separating the short term behavior of the holographic system, controlled by the learning rule, from the long term behavior, controlled by the periodic refreshing technique described in section 4.5. This approach would allow direct linear implementation of learning processes with arbitrarily many exposures while avoiding unduly large losses in the dynamic range of the stored hologram.

It is convenient to separate the wide range of learning algorithms that have been discussed in the literature into three categories: prescribed learning, error driven learning and self organization. We draw the distinction between these algorithms with the aid of Fig. 6.1, where a general network is drawn with the vector $\underline{x}(k)$ as its input and $\underline{y}(k)$ the output at the k^{th} iteration (or time interval). The vector $\underline{z}(k)$ is used to represent the activity of the internal units and $w_{ij}(k)$ is the connection strength between the i^{th} and the j^{th} unit. Let $\underline{x}^{(m)}$, $m = 1 \ldots M$, be a set of specified input vectors and let $\underline{y}^{(m)}$ be the responses which the network must produce for each of these input vectors.

A prescribed learning algorithm calculates the strength of each weight simply as a function of the vectors $\underline{x}^{(m)}$ and $\underline{y}^{(m)}$:

$$w_{ij} = f_{ij}(\underline{x}^{(m)}, \underline{y}^{(m)}) \qquad m = 1 \ldots M \tag{6.1}$$

This type of procedure is relatively simple ("easy learning"). It is perhaps the most sensible approach in a single layer network. The widely used outer-product algorithm [34, 35] is an example of this type of learning algorithm, as are some schemes which utilize the pseudoinverse [34,141, 142]. Despite its simplicity, prescribed learning is limited in several important respects. First, while prescribed learning is well understood for single layer systems, the existing algorithms for two layers are largely localized representations; each input $\underline{x}^{(m)}$ activates a single internal neuron [143, 144, 145]. Moreover, the entire learning procedure usually has to be completed *a priori*. This last limitation is not encountered in the

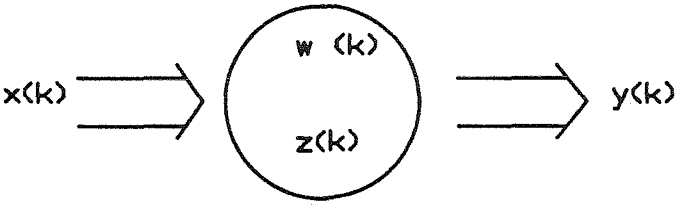


Figure 6.1. General neural network architecture.

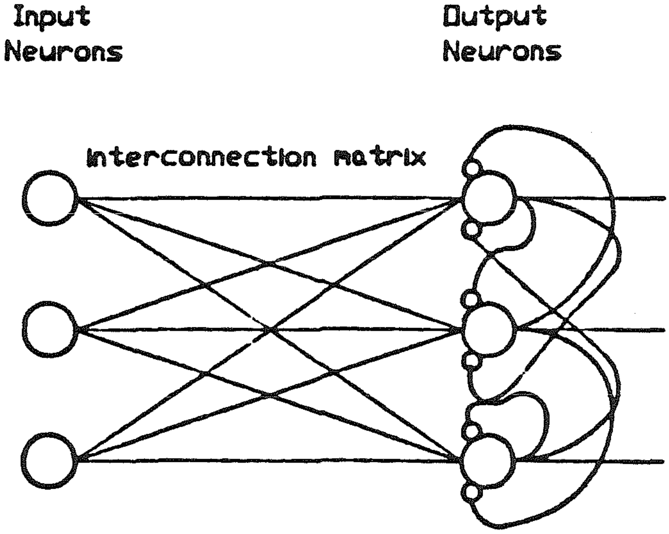


Figure 6.2. Two layer network with lateral inhibition. Connections ending with an open circle are inhibitory.

simplest form of prescribed learning, the outer-product rule:

$$w_{ij} = \sum_{m=1}^{M} x_i^{(m)} y_j^{(m)} \tag{6.2}$$

In this case new memories may be programmed by simply adding the outer-products of new samples to the weight matrix. Note that once the interconnection matrix has been determined by a prescribed learning algorithm, it may be expressed in the form of a sum of at most N outer-products, where N is the total number of neurons in each layer. As we saw in chapters 2 and 3, volume holograms record interconnections matrices represented by sums of outer-products in a very natural way. Thus, matrices which can be expressed in this form are particularly simple to implement in optics [38, 39, 42, 43, 49].

Error driven learning is distinguished by the fact that the output of the system, $\underline{y}(k)$, is monitored and compared to the desired response $\underline{y}^{(m)}$. In each learning step, an incremental change, Δw_{ij} is made to the weight between the i^{th} and j^{th} neuron to reduce the error.

$$\Delta w_{ij}(k) = f_{ij}[\underline{x}^{(m)}, w_{rs}(k), \underline{y}^{(m)}] \tag{6.3}$$

The Δw_{ij} is calculated from the vectors $\underline{x}^{(m)}$ and $\underline{y}^{(m)}$ and the current setting of the weight matrix $w_{rs}(k)$ (from which the state of the entire network can be calculated). The perceptron [146] and adaline [147] algorithms are examples of error driven learning for single layer networks. Interest in such learning algorithms has been renewed recently by the development of procedures suitable for multilayered networks [148, 33, 149]. Error driven algorithms ("hard learning") are more difficult to implement than prescribed learning since they require a large number of iterations before errors can be reduced to sufficiently low levels. In multilayered systems, however, this type of learning can provide an effective mechanism for matching the available resources (connections and neurons) to the requirements of the problem. In optical realizations error driven algorithms are more difficult to implement than prescribed approaches due to the need for dynamically modifiable interconnections and the incorporation of an optical system that monitors the performance and causes the necessary changes in the weights. While this problem could be avoided by performing learning off line in computer simulations and recording the optimized interconnection matrix as in prescribed learning, this approach has the disadvantage that once again the matrix is fixed *a priori*, thus preventing the network from being adaptive.

In the case of self organizing learning algorithms we require not that the specified inputs produce a particular response but rather that they satisfy a general restriction, often imposed by the structure of the network itself. Since there is no *a priori* expected response, the learning rule for self organizing systems is simply

$$\Delta w_{ij}(k) = f_{ij}[\underline{x}^{(m)}, w_{rs}(k)] \tag{6.4}$$

This type of learning procedure can be useful, for instance, at intermediate levels of a network where the purpose is not to elicit an external response but rather to generate appropriate internal representations of the information that is presented as input to the network. There is a broad range of self organizing algorithms, the simplest of which is probably lateral inhibition to enforce grandmother cell representations [34, 150]. The objective of the learning procedure is to have each distinct pattern in an input set of neurons activate a single neuron in a second set. In the architecture shown in Fig. 6.2 this is accomplished via inhibitory connections between the neurons in the second set. Once a particular neuron in the second layer is partially turned on for a specific pattern it prevents the connections to the other neurons in the second set from assuming values that will result in activity at more than one neuron. The details of the dynamics of such procedures can be quite complex (*e.g.*, [151]), as can corresponding optical implementations. An advantageous feature of optics in connection with self organization is that global training signals, such as fixed lateral inhibition between all the neurons in a given layer, can easily be broadcast with optical beams.

While the linearization techniques described earlier in this thesis make it possible to implement any learning scheme in volume holography, little is gained if the system used to control the hologram must perform more complex computations than the hologram itself. For this reason, it is generally desirable for mapping between the abstract learning algorithm and its physical realization be as direct as possible. In the following two sections, we describe experimental holographic neural systems which implement learning in a very direct manner. The first system implements prescribed outer-product storage. The second implements perceptron style learning.

6.2 A PHOTOREFRACTIVE OUTER-PRODUCT MEMORY

In chapters 3 and 4 we saw how connections between a large number of processing nodes can be formed in a volume hologram. The simplest demonstration of neural processing using photorefractive connections is an outer-product style memory based on the exposure schedule described in chapter 4. We have used the sampling grids shown in Fig. 6.3 to construct an outer-product style asso-

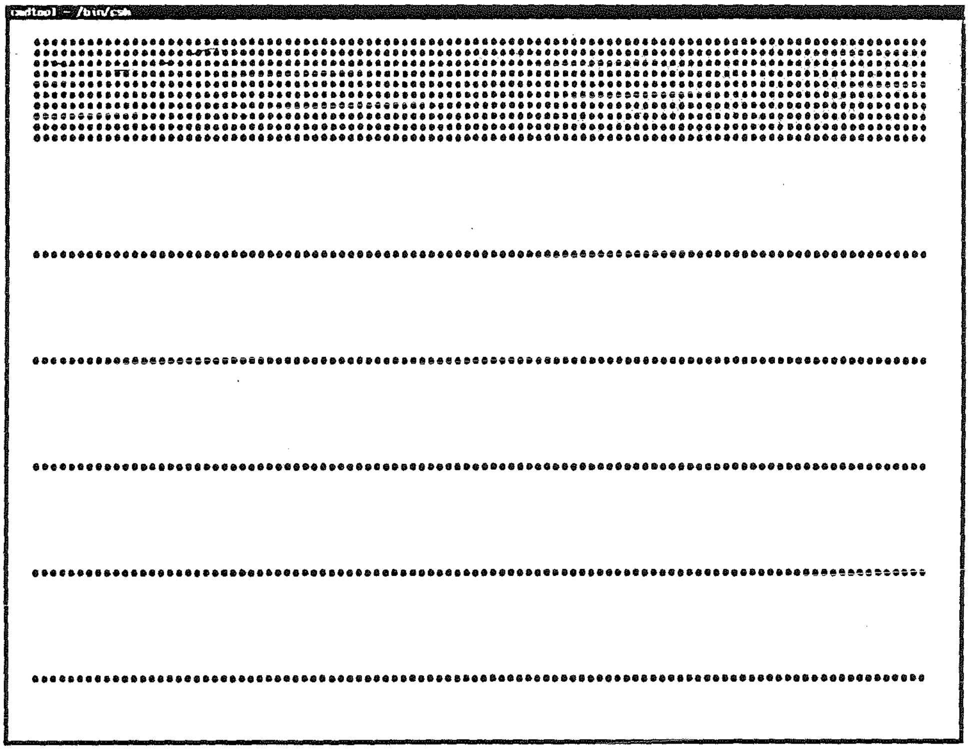


Figure 6.3. Sampling grids for outer-product memory. The input consists of 10 tightly packed rows of 100 pixels, as shown at top. The output consists of 5 rows of 100 pixels spaced by 10 unused rows, as shown below.

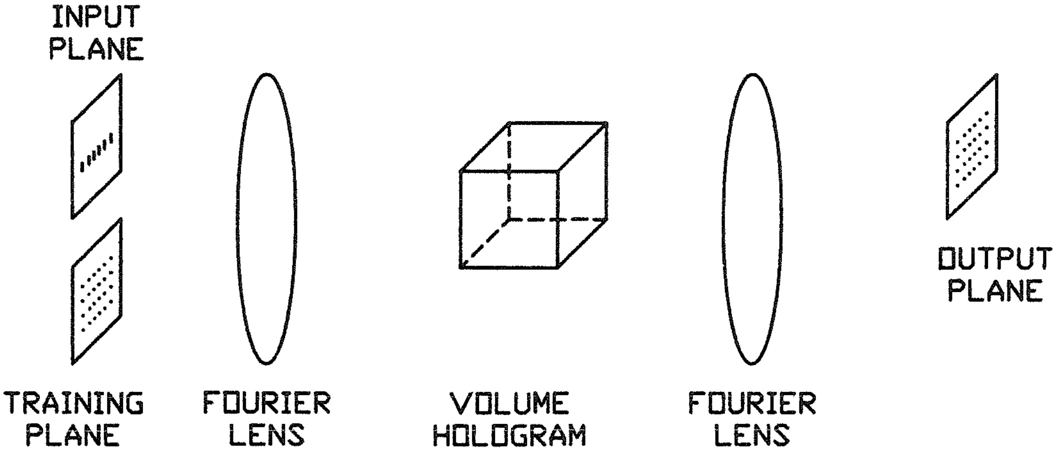


Figure 6.4. Photorefractive outer-product memory architecture.

ciative interconnect system. The experimental system is sketched in 6.4. The images were introduced into the system using a computer-controlled liquid crystal television monitor read out by an argon ion laser operating at a wavelength of 488 nm. A liquid crystal light valve was placed at the input plane to improve the contrast of the images before they were introduced to the holographic system. The holographic medium was a cesium doped strontium (0.6) barium (0.4) niobate crystal.

Volume holograms between associated patterns on the input and output grids were formed in sequence. Because the patterns were constrained to lie on sampling grids, the transformation stored in each hologram was a true outer-product of the associated patterns. The m^{th} associated pair of patterns was recorded with an exposure E_m given by

$$E_m = E_o \log(\frac{m}{m-1}), \tag{6.5}$$

where E_o is the saturation exposure. The first associated pair was recorded to saturation. In our crystal, $E_o \approx 200$ mJ/cm^2. As described in chapter 4, this exposure schedule was necessary to insure that the hologram stored in the volume was a linear sum of the outer-products of each of the associated input-output pairs.

The patterns stored in the system consisted of random code-name pairs such as those shown in Fig. 6.5. The names were arranged on 10 × 100 pixel arrays. The random codes were arranged on five 100 pixel rows spaced by 11 unused rows. In an effort to increase the orthogonality of the recorded images, a piezo-electric mirror was used to add a random phase to the active pixels for each name. We stored up to 20 image pairs using this system. Between recording cycles, the fidelity with which previously recorded images could be reconstructed was evaluated by introducing stored names at the input and detecting reconstructed codes on a CCD camera at the output plane. For a small number of exposures, the fidelity of the reconstruction (recall) was good, confirming the effectiveness of the sampling grids at ensuring independent interconnections. As more names were stored it became increasing difficult to reconstruct previously recorded patterns. This was partly due to correlations between the stored patterns, but was chiefly a result of a decrease in the diffraction efficiency per stored pattern inherent in the multiple exposure process.

Figure 6.5. Examples of pairs of associated images.

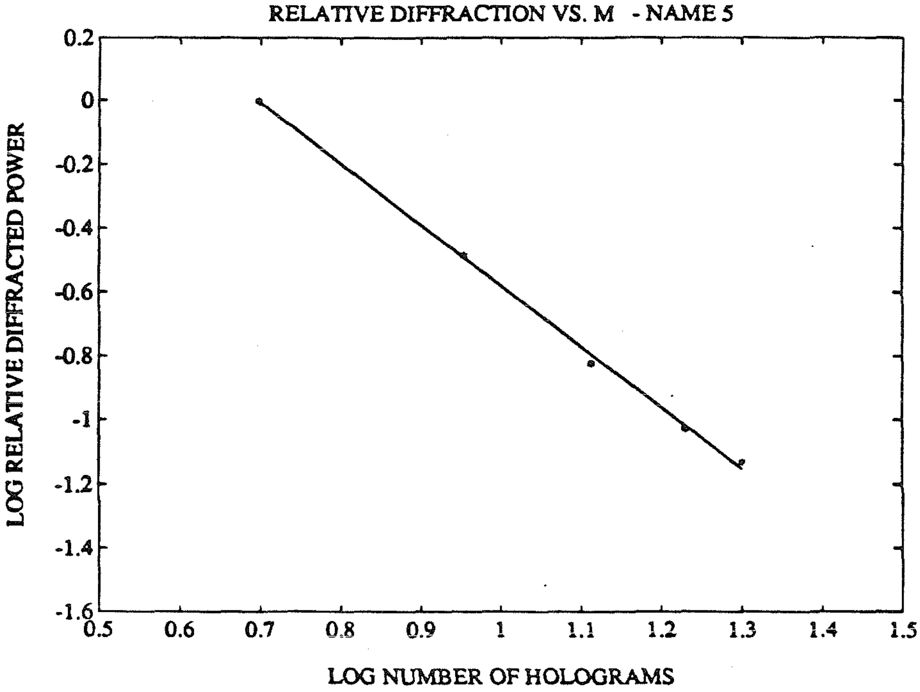


Figure 6.6. Log diffraction efficiency for the 5^{th} association *vs.* log number of associations. The slope of the least squares fit is -1.95

As predicted in chapter 4, we observed a fall-off in the diffraction efficiencies of the stored holograms as more holograms were loaded into the system. Due to the fact that the stored patterns were not orthogonal, the diffraction efficiencies of the different stored images did not degrade uniformly. We found decay rates versus the log of the number of stored patterns which varied from -1.2 to just over -2. This fall off in diffraction efficiency is illustrated in Fig. 6.6, which shows the efficiency with which one of the patterns is reconstructed vs the number of patterns stored on a log-log scale. The slope of the experimental best-fit line is -1.95. The theory derived in chapter 4 predicts a slope of -2. This decrease in diffraction efficiency under multiple exposure will limit the number of images which can be stored in currently available photorefractive materials to the range of several hundred to one thousand [125].

6.3 A PHOTOREFRACTIVE PERCEPTRON

The perceptron [146, 36] is an example of a supervised adaptive neural model which can be implemented using photorefractive crystals. A perceptron consists of a set of input neurons with activities described by a vector $\vec{x}$ which drive a single output neuron via a weight vector $\vec{w}$. The activity of the output vector is high if and only if $\vec{w} \cdot \vec{x} > \omega_o$ where ω_o is a fixed threshold level. A perceptron can be trained to separate a set of input vectors into two classes by various methods, the simplest of which involves updating the weight vector according to

$$\begin{aligned} \vec{w}(n+1) &= \vec{w}(n) + \alpha\vec{x} \\ \omega_o(n+1) &= \omega_o(n) + \alpha \end{aligned} \tag{6.6}$$

where $\vec{w}(n)$ is the state of the weight vector at the discrete time n when $\vec{x}$ is presented for classification. α is zero if $\vec{x}$ is correctly classified and 1 (-1) if

$\vec{x}$ is misclassified in the low (high) state. If training vectors from a set $\{\vec{x}\}$ are presented in sequence, Eq. (6.6) is known to converge on a weight vector implementing an arbitrary prescribed dichotomy if such a weight vector exists.

One means of implementing a perceptron in a photorefractive system would be to update each interconnection in series as prescribed by Eq. (6.6). This approach has two disadvantages. The first is that the saturable nature of the photorefractive response limits the range of w_i to $(-w_{sat}, w_{sat})$ for coherent systems and $(0, w_{sat})$ for incoherent systems. The weights determined by Eq. (6.6) may not be guaranteed to lie within these bounds. The second disadvantage to this approach is that each weight must be updated independently. In order to update the weights in this way we would need to detect the value of each weight and generate optical beams specifically to change that weight by the prescribed amount. In a volume holographic implementation with spatially multiplexed weights it is not possible to change the weights independently.

These problems can be avoided by modifying the learning procedure to conform more closely to the dynamics of hologram formation. This can be done particularly simply in an incoherent system. An architecture for an incoherent photorefractive perceptron is shown schematically in Fig. 6.7. Much of the complexity of the systems described in the previous section is avoided in this system so that we can concentrate on the use of photorefractive dynamics in learning. The input to the system, $\vec{x}$, corresponds to a two-dimensional pattern recorded from a video monitor onto a liquid crystal light valve. The light valve transfers this pattern onto a laser beam. This beam is split into two paths which cross in a photorefractive crystal. The light propagating along each path is focused such that an image of the input pattern is formed on the crystal. The images along

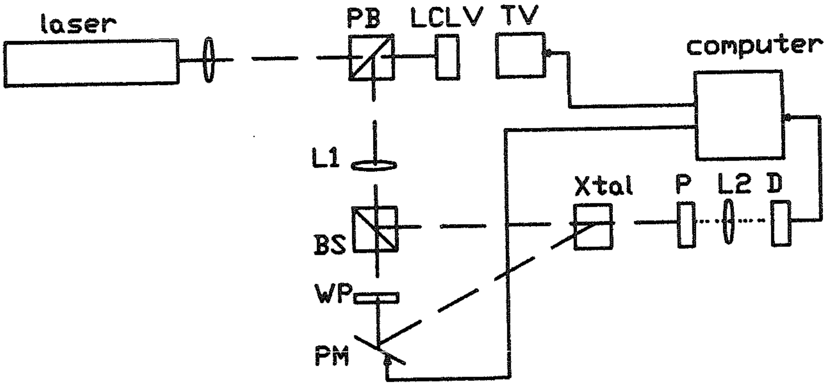


Figure 6.7. Photorefractive perceptron. PB is a polarizing beam splitter. L1 and L2 are imaging lenses. WP is a quarter waveplate. PM is a piezoelectric mirror. P is a polarizer. D is a detector. Solid lines show electronic control. Dashed lines show the optical path.

both paths are of the same size and are superimposed on the crystal. The intensity diffracted from one of the two paths onto the other by a hologram stored in the crystal is isolated by a polarizer and spatially integrated by a single output detector. The thresholded output of this detector corresponds to the output of a perceptron. The fact that the connections in this system are stored locally in the image plane of the input allows us to very simply control each connection independently. This is at a cost, however, of the loss of the high connection densities achieved by using the entire volume of the storage medium in the Fourier domain.

The i^{th} component of the input to this system corresponds to the intensity in the i^{th} pixel of the input pattern. The interconnection strength, w_i, between the i^{th} input and the output neuron corresponds to the diffraction efficiency of the hologram taking one path into the other at the i^{th} pixel of the image plane. In analogy with Eq. (6.6), w_i may be updated by exposing the crystal with the input along both paths. If the modulation depth between the light in the two paths is high then where x_i is high w_i is increased. If the modulation depth is low between the two paths then where x_i is high w_i is reduced. The modulation depth between two optical beams can be adjusted by a variety of simple mechanisms. In Fig. 6.7 we choose to control $m(t)$ using a mirror mounted on a piezoelectric crystal. By varying the frequency and the amplitude of oscillations in the piezoelectric crystal we can electronically set both $m(t)$ and $\phi(t)$ over a continuous range without changing the intensity in the optical beams or interrupting readout of the system.

We have implemented the architecture of Fig. 6.7 using a SBN60:Ce crystal. We used the 488 nm line of an argon ion laser to record holograms in this crystal.

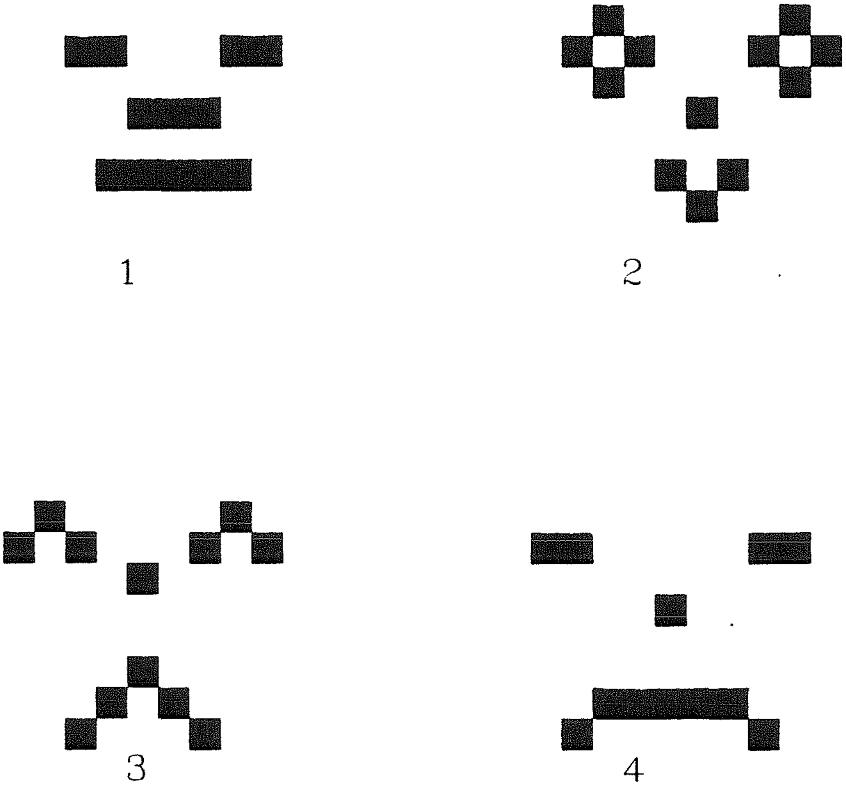
Most of the patterns considered were laid out on 10 × 10 grids of pixels, thus allowing 100 input channels. Ultimately, the number of channels which may be achieved using this architecture is limited by the number of pixels which may be imaged onto the crystal with a depth of focus sufficient to isolate each pixel along the length of the crystal.

Using the variation on the perceptron learning algorithm described below with fixed exposure times Δt_r and Δt_e for recording and erasing, we were able to correctly classify various sets of input patterns. An example of such a set is shown in Fig. 6.8. In one training sequence, we grouped patterns 1 and 2 together with a high output and patterns 3 and 4 together with a low output. After all four patterns had been presented four times, the system gave the correct output for all patterns. The weights stored in the crystal were corrected seven times, four times by recording and three by erasing. Fig. 6.9(a) shows the output of the detector as pattern 1 is recorded in the second learning cycle. The dashed line in this figure corresponds to the threshold level. Fig. 6.9(b) shows the output of the detector as pattern 3 is erased in the second learning cycle.

Applying the results of chapter 4 to this system we find that

$$w_i(n) = w_{sat} \left| \sum_{s=1}^{n} m_i(s) e^{-\sum_{s'>s}^{n} \frac{t(s')}{\tau}} \left(1 - e^{-\frac{t(s)}{\tau}}\right) \right|^2 \tag{6.7}$$

Two problems prevent the use of the exposure schedule in this system. The first is that the assumption of approximately constant intensity in each exposure is violated in an incoherent image plane system. The second is that, while the perceptron algorithm is known to converge, the number of training steps needed to reach convergence can be very large. This second problem could be corrected by using the periodic copying technique described in chapter 4.



1

2

3

4

Figure 6.8. Training patterns.

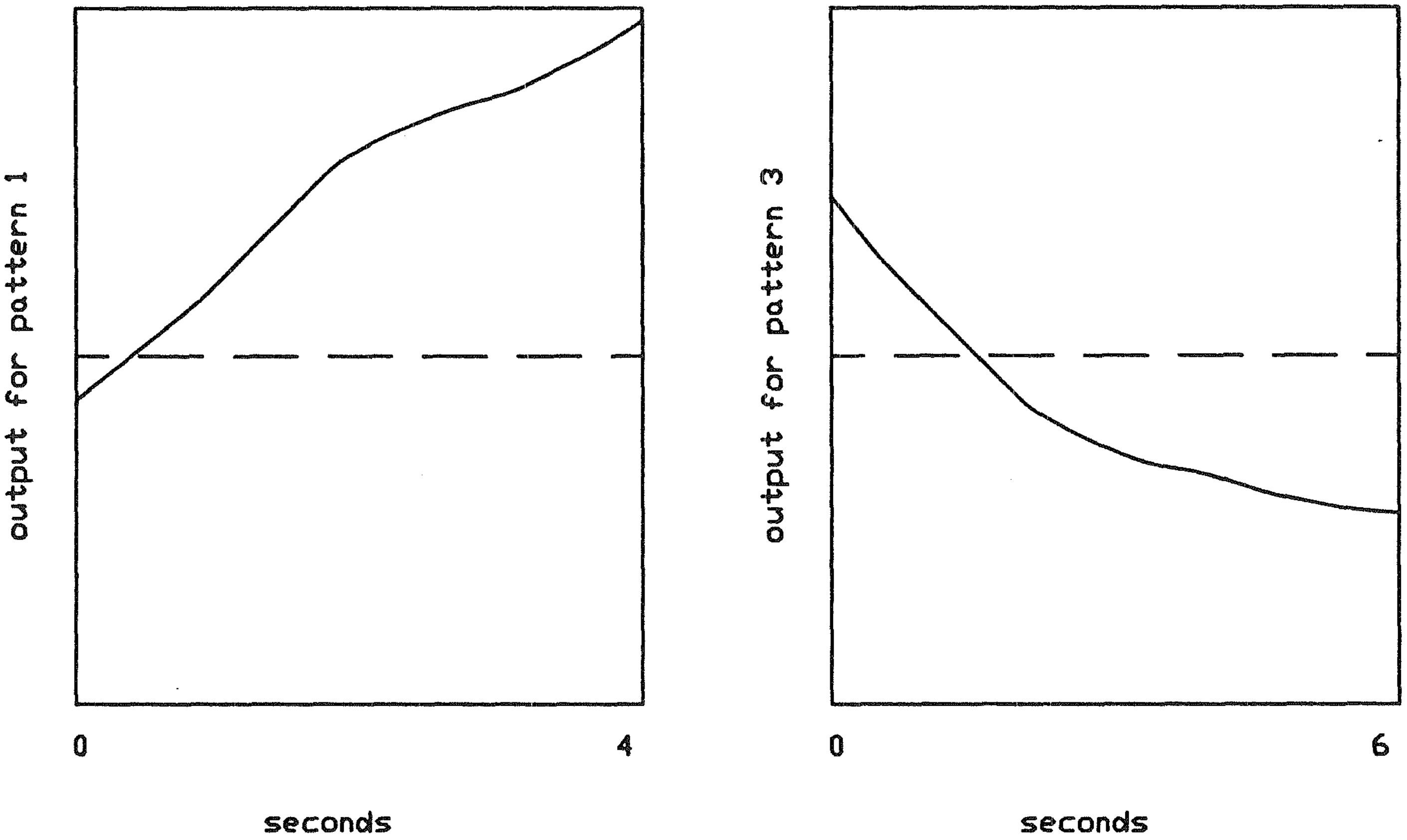


Figure 6.9. Detector output during training.

In lieu of using an exposure schedule in our simple perceptron we assume that $t(s) = \Delta t_e$ in cycles in which erasure occurs and $t(s) = \Delta t_r$ in cycles in which writing occurs. Since τ is inversely proportional to the optical intensity, we can express $\frac{1}{\tau}$ at each pixel of the input as αx_i. Letting $m_i(t) = 0$ when the n^{th} training vector yields too high an output, we find from Eq. (6.7) that the i^{th} component of the weight vector is updated according to

$$w_i(n+1) = e^{-2\Delta t_e \alpha x_i} w_i(n) \tag{6.8}$$

Let $m_i(t) = m$ when the n^{th} training vector yields too low an output, we find that the the i^{th} component of $n+1^{th}$ weight vector is

$$\begin{aligned} w_i(n+1) =& e^{-2\Delta t_r \alpha x_i} w_i(n) + m^2 w_{sat}(1 - e^{-\Delta t_r \alpha x_i})^2 \\ &+ 2m\sqrt{w_i(n) w_{sat}} e^{-\Delta t_r \alpha x_i}(1 - e^{-\Delta t_r \alpha x_i}) \end{aligned} \tag{6.9}$$

Note that, as one might intuitively wish, w_i decreases monotonically with x_i if the output for $\vec{x}$ is too high and increases monotonically with x_i if the output for $\vec{x}$ is too low. Eqs. (6.8) and (6.9) may be greatly simplified if we assume that x_i takes on only the values 0 and 1. Since the output power incident on the detector may be arbitrarily renormalized, we may also assume without loss of generality that $m^2 w_{sat} = 1$. Under these assumptions Eq. (6.8) may be recast in the form

$$\Delta w_i(n) = w_i(n+1) - w_i(n) = -(1 - \gamma_e^2) x_i w_i \tag{6.10}$$

where $\gamma_e = e^{-\Delta t_e \alpha}$. Eq. (6.9) becomes

$$\Delta w_i(n) = -(1 - \gamma_r^2) x_i w_i + 2\gamma_r(1 - \gamma_r) x_i \sqrt{w_i} + (1 - \gamma_r)^2 x_i \tag{6.11}$$

where $\gamma_r = e^{-\Delta t_r \alpha}$.

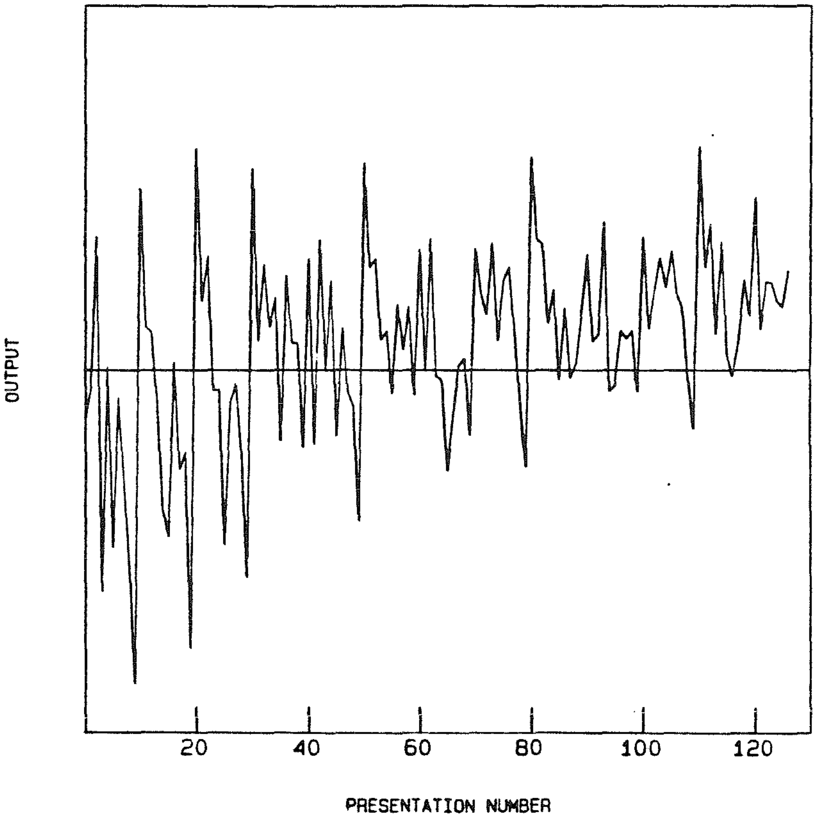


Figure 6.10. Learning curve from simulations.

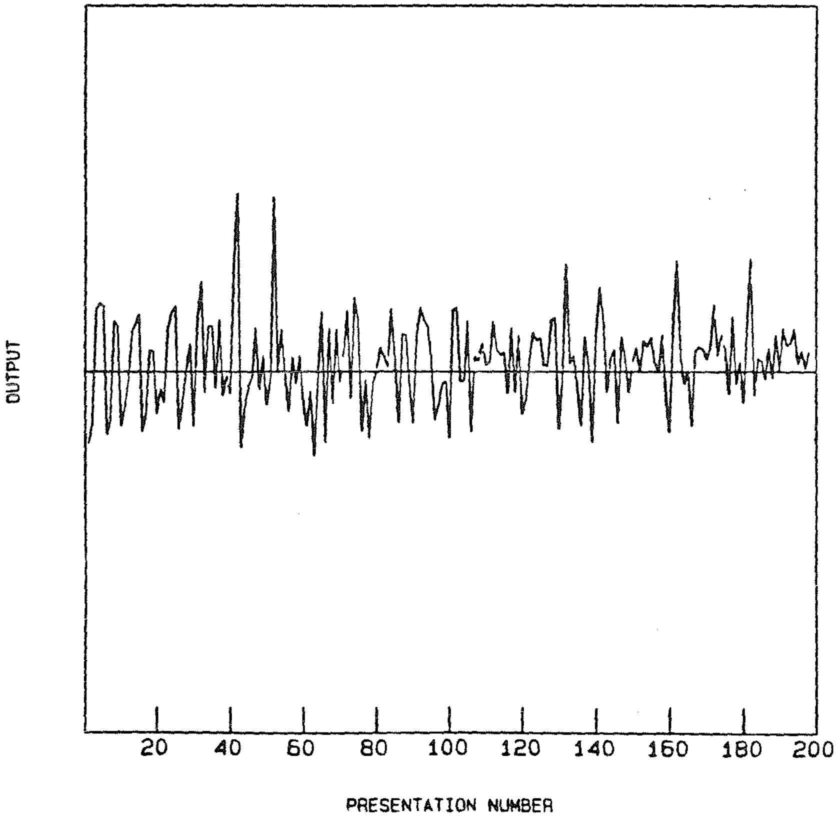


Figure 6.11. Experimental learning curve.

We assumed both in our simulations and in our experiments that an acceptable value for ω_o could be guessed *a priori*. This assumption is unnecessary if one of the stored weights is used as a threshold. An example of convergence in simulations run under this training algorithm is shown in Fig. 6.10. In this figure we plot $\delta \times (\vec{w}(n) \cdot \vec{x}^{(n)} - \omega_o)$ as a function of n. δ is 1 (-1) if $\vec{x}^{(n)}$ is in Ω_+ (Ω_-). The system has converged when a cycle through all stored vectors yields all positive outputs. This example involves ten randomly selected and classified training vectors drawn from a hundred-dimensional space. Ω_+ and Ω_- each contain five vectors. The training vectors are presented twelve times in sequence before a solution vector was found. In the thirteenth sequence through the vectors $\delta(\vec{w}(n) \cdot \vec{x}^{(n)} - \omega_o) > 0$ for all n, indicating that a solution has been obtained. γ was 0.9 and the weights were randomly initialized. A learning curve under similar circumstances in the experimental system is shown in Fig. 6.11. In this case nineteen cycles were necessary to converge on a solution vector. Most of the discrepancy between the experimental system and the simulations arises from the difficulty involved setting the learning parameters and initial conditions to be identical. The key trend to notice in both figures is that the amplitude and frequency of negative outputs decreases with each cycle until convergence is reached. We found that it was relatively easy to achieve convergence both in simulations and in the experimental system.

A final point to notice about Fig. 6.10 and Fig. 6.11 is that convergence is achieved even though the exposure involved in each training cycle is large. We know from Eq. (6.7) that the effect of a given exposure on the final state of w_i is decreased by at least a factor of γ^q, where q is the number of exposures made after the exposure of interest. Since $\gamma^q \ll 1$ for $q > 10$, information recorded early in the training sequences is erased from the crystal before convergence is reached. The early exposures drive the system toward a region of weight space from which convergence is achieved, rather than store information.

6.4 OPTICAL VERSUS ELECTRONIC NETWORKS

In "electronic" implementations of neural networks, neurons and the channels which interconnect them are integrated on the same 2-D surface. In large scale networks the area devoted to the interconnections is much greater than the area devoted to the processing nodes [152, 153]. In "optical" implementations neurons are again arranged on a plane, but the addition of light sources, modulators and detectors allows internodal communication to be done optically through the third dimension. The advantages of electronic neural processors are that they can be constructed using conventional processing techniques, typical device sizes are small and device speeds are respectable. The advantages of the optical implementations are that no expensive lithographic area need be devoted to linear devices and wiring and adaptation can be implemented in a direct and simple manner [102]. In this section we compare optical and electronic implementations as a means of putting the techniques described in this thesis in perspective and outlining situations in which holographic optics may be of use in massively parallel processing.

A number of previous studies have considered the relative advantages of optical and electronical systems. Barakat and Reif derive a constraint on the scale of problems which an optical system in a volume V can resolve in time T [154]. Their analysis is based on an analogy with similar constraints which have been derived relating the integrated area of VLSI circuits and T to the scale of the problem the circuit can resolve. In this thesis we are interested in a much more limited class of problems than was considered by Barakat and Reif. The optical systems we consider are also considerably less powerful, since we allow only changes to the linear properties of each voxel. We have assumed that each layer of the optical system performs the largest possible linear transformation in each time step.

Studies closer to the systems of interest in this thesis were completed by Goodman *et al.* [155], Feldman *et al.* [156], Wu *et al.* [157], and Miller [158]. These studies consider the relative energy efficiencies of optical and electronic interconnections. For our purposes, the key result of these analyses is that the power in communicating a bit of information is lower bounded by

$$P = \frac{CV^2}{2\tau}, \tag{6.12}$$

where C is the capacitance charged to bring the detecting device to its threshold voltage V. τ is the time used to communicate the bit.

Consider the following question: *What are the maximum size and speed for a network with* c *connections per neuron implemented using a chip of area* A *dissipating power* P *if (a) the neural interconnections are purely electronic or (b) the neural interconnections are optical.* We answer this question for three different types of connectivity:

(1) The input to each neuron is a weighted sum of the activities of its c nearest neighbors, as is the case in the volume holographic interconnection systems described in this thesis.

(2a) Each neuron receives unweighted the activities of its c nearest neighbors. The electronic implementation uses an independent wire for each connection.

(2b) Each neuron receives unweighted the activities of its c nearest neighbors. The electronic implementation uses active "axons" to broadcast the activity of each neuron.

Let R_s be the ratio of the number of neurons in the optical network and the number of neurons in the electronic network, R_τ be the ratio of the response time of the optical net and the response time of the electronic net, and $R_I = \frac{R_S}{R_\tau}$ be the ratio of the rates at which interconnections are made in the two networks. ρ_A is the ratio between the area of a single processing element in the optical network, δ_o^2, and the area, δ_e^2, of a single processing element in the electronic network. ρ_E is the ratio of the energy required to charge a single element of the optical network and the energy required to charge a single element of the electronic network. η_o describes the maximum diffraction efficiency of a single holographic grating in the optical system.

In cases (1) and (2b), the electronic layout can be done using a crossbar network. In this case each connection takes up one unit of area on the chip. Sketches of appropriate layouts for the electronic circuits are shown in Fig. 6.12 and Fig. 6.13. In the optical case only active devices are integrated on the chip. Thus, R_S in these two cases is $\rho_A c$. In case (2a), the area of the connections between neurons grows as c^2 as the scale of the chip grows, as is seen in Fig. 6.14. In this case, R_s is $\rho_A c^2$. In every case, the capacitance of the system is proportional to the area of the chip, which is fixed. As we saw in this chapter and in chapters

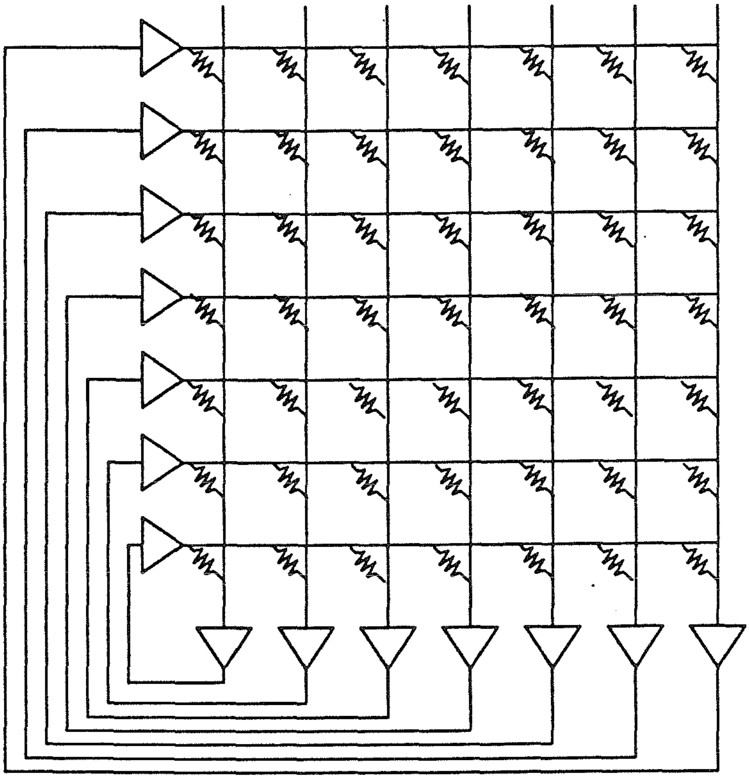


Figure 6.12. Electronic layout for case (1).

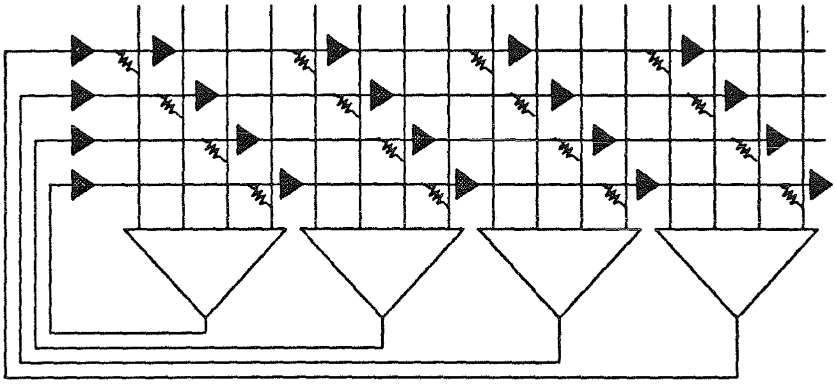


Figure 6.13. Electronic layout for case (2b).

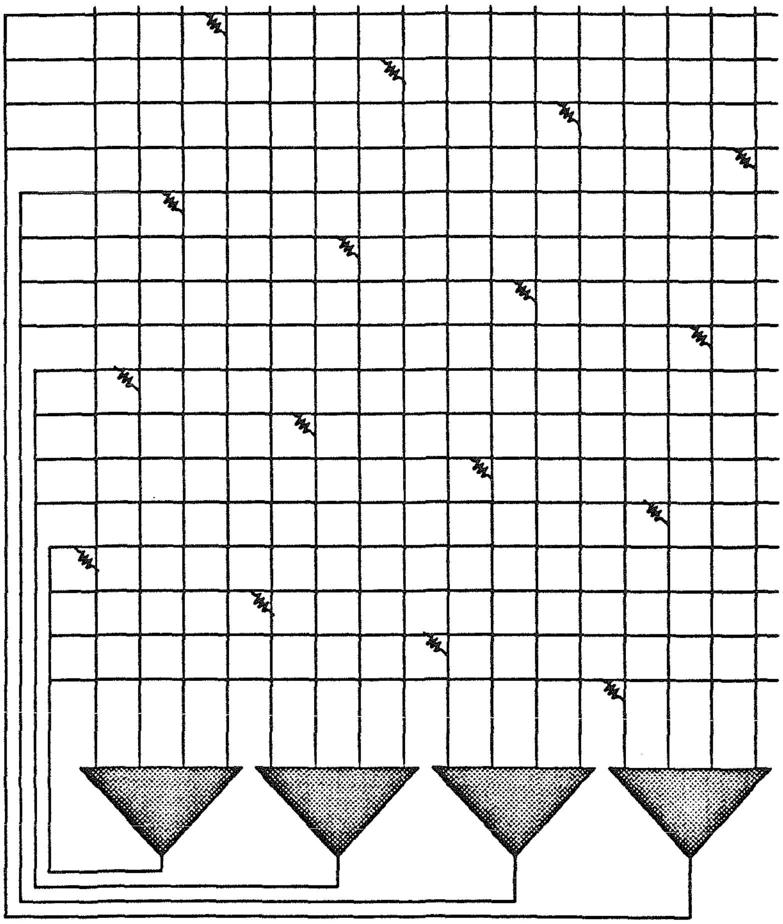


Figure 6.14. Electronic layout for case (2a).

3 and 4, however, the diffraction efficiency of weighted volume holographic interconnections scales inversely with c^2. This implies that $R_\tau = \rho_E c^2/\eta_o$ in case 1, where η_o is the ratio of the maximum perturbation in a hologram to the perturbation which causes diffraction efficiency 1. Our results are summarized in the following table.

OPTICAL VS. ELECTRONIC NETWORKS			
Case	1	2a	2b
R_S	$\rho_A c$	$\rho_A c^2$	$\rho_A c$
R_τ	$\frac{\rho_E c^2}{\eta_o}$	$\frac{\rho_E}{\eta_o}$	$\frac{\rho_E}{\eta_o}$
R_I	$\frac{\rho_A}{\rho_E}\frac{\eta_o}{c}$	$\frac{\rho_A}{\rho_E}\eta_o c^2$	$\frac{\rho_A}{\rho_E}\eta_o c$

The significance of these results depends on the relative costs of integrated area and power and on the size of the scaling parameters. Obviously, if integrated area is the key parameter, the advantage goes to optics when large scale interconnections are required. For networks of of sufficient size and interconnection density, optical connections yield advantages in both speed and size in cases (2a) and (2b). These systems forgo what may be one of the key advantages of optics, however, which is that holography allows us to implement Hebbian style adaptation directly in hardware. This advantage is achieved in case (1) at a cost of loss in dynamic range, or diffraction efficiency.

This problem brings us back to where we began this thesis, with the the hope volume holography offers for implementing very large linear transformations, if a means can be found to control the hologram without wasting all its dynamic range. In this thesis, we have demonstrated this problem by experimentally documenting the decay in diffraction efficiency of multiply exposed holograms. We have shown that η_o can be made larger than 1 in photorefractive crystals by increasing the imaginary component of the time constant. We should note that in certain crystals η_o is much greater than one in any event [125]. We have considered copying techniques which allow us to keep adaptively recorded holograms at the c^{-2} decay rate over arbitrarily long exposure cycles. We have considered the possibility that the scaling may be reduced to c^{-1} using polychromatic control. And we have shown how this problem can be overcome in integrated architectures.

With these results in mind, it is my belief that R_T for case (1) can be made to remain within one or two orders of magnitude of 1 out to the maximum connectivity which can be supported in a single aperture. If this goal is achieved the long postulated advantages of optical holography in signal processing will be accessible.

REFERENCES

1. P. J. van Heerden, "Theory of optical information storage in solids," Appl. Opt., **2**, 393(1963).
2. W. L. Bragg, "A new type of x-ray microscope," Nature, **143**, 678(1939).
3. D. Gabor, "A new microscopic principle," Nature, **161**, 777(1948).
4. H. Hauptman, "The phase problem of x-ray crystallography," Physics Today, **42**, no. 11, 24(1989).
5. E. N. Leith and J. Upatnieks, "Reconstructed wavefronts and communication theory," JOSA, **52**, 1123(1962).
6. E. N. Leith and J. Upatnieks, "Wavefront reconstruction with diffused illumination and three-dimensional objects," JOSA, **54**, 1295(1964).
7. Y. N. Denisyuk, "Photographic reconstruction of the optical properties of an object in its own scattered radiation field," Sov. Phys. Dokl., **7**, 543(1962).
8. G. Lippman, J. de Phys., **3**, 97(1894).
9. H. Kogelnik, Bell Syst. Tech. J., **48**, 2909(1969).
10. R. J. Collier, C. B. Burckhardt, and L. H. Lin, *Optical holography* , Academic, (New York)1971.
11. P. St. J. Russel, "Optical volume holography," Phys. Repts., **71**, 209(1981).
12. L. Solymar and D. J. Cooke, *Volume holography and volume gratings*, Academic, (London)1981.

13. A. Ashkin, G. D. Boyd, J. M. Dziedzic, R. G. Smith, A. A. Ballman, and K. Nassau, "Optically-induced refractive index inhomogeneities in $LiNbO_3$ and $LiTaO_3$," Appl. Phys. Lett., **9**, 72(1966).

14. R. L. Townsend and J. T. La Macchia, "Optically induced refractive index changes in $BaTiO_3$, J. Appl. Phys., **41**, 5188(1970).

15. J. B. Thaxter, "Electrical control of holographic storage in strontium barium niobate," Appl. Phys. Lett., **15**, 210(1969).

16. J. P. Huignard and F. Micheron, "High sensitivity read-write volume holographic storage in BSO and BGO crystals," Appl. Phys. Lett., **29**, 591(1976).

17. M. B. Klein, "Beam coupling in undoped GaAs at 1.06 μm using the photorefractive effect," Opt. Lett., **9**, 350(1984).

18. A. M. Glass, A. M. Johnson, D. H. Olson, W. Simpson, and A. A. Ballman, "Four-wave mixing in semi-insulating InP and GaAs using the photorefractive effect," Appl. Phys. Lett., **44**, 948(1984).

19. N. V. Kuktarev, V. B. Markov, S. G. Odulov, M. S. Soskin, and V. L. Vinetskii, "Holographic storage in electro-optic crystals. I. Steady State.," Ferroelectrics, **22**,949(1979).

20. P. Gunter and J.-P. Huignard, eds., *Photorefractive materials and their applications I and II*, Springer-Verlag, (Berlin)1989.

21. T. J. Hall, R. Jaura, L. M. Connors, P. D. Foote, Prog. Quan. Electr. **10**,77(1985).

22. D. Psaltis, M. A. Neifeld, A. A. Yamamura, and S. Kobayashi, "Optical

memory disks in optical information processing systems," to be published in Appl. Opt., January 1990.

23. E. N. Leith, A. Kozma, J. Upatnieks, J. Marks, and N. Massey, "Holographic data storage in three-dimensional media," Appl. Opt., **5**, 1303(1966).
24. L. D'Auria, J. P. Huignard, and E. Spitz, "Holographic read-write memory and capacity enhancement by 3-D storage," IEEE Trans. Mag., **MAG-9**, 83(1973).
25. Special issue on optical storage of digital data, Appl. Opt., **13**, no. 4, (1974).
26. D. von der Linde and A. M. Glass, "Photorefractive effects for reversible holographic storage of information," Appl. Phys., **8** 85(1975).
27. K. Bløtekjaer, "Limitations on holographic storage capacity of photochromic and photorefractive media," Appl. Opt., **18**, 57(1979).
28. D. L. Staebler, W. J. Burke, W. Phillips, and J. J. Amodei, "Multiple storage and erasure of fixed holograms in Fe-doped $LiNbO_3$," Appl. Phys. Lett., **26**, 182(1975).
29. F. Mok, M. Tackitt, H. M. Stoll, OSA Annual Meeting, *1989 Technical Digest Series*, **12**, Optical Society of America, Washington, DC, 74(1989).
30. D. Gabor, "Associative holographic memories," IBM J. Res. Develop., **13**, 156(1969).
31. M. Sakaguchi, N. Nishida, and T.Nemoto, "A new associative memory system utilizing holography," IEEE Trans. Comp., **C-19**, 1174(1970).

32. An overview of past and present research in artificial neural networks can be found in *Neurocomputing*, J. A. Anderson and E. Rosenfeld, eds., MIT Press, (Cambridge)1988.

33. D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel Distributed Processing, Volume 1*, MIT Press, Cambridge, (1986).

34. T. Kohonen, *Self-Organization and Associative Memory*, Springer-Verlag, (Berlin)1984.

35. J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," Proc. Natl. Acad. Sci. USA, **79**,2554(1982).

36. M. L. Minsky and S. A. Papert, *Perceptrons*, MIT, (Cambridge) 1988.

37. C. A. Mead, *Analog VLSI and Neural Systems* , Addison-Wesley, (Reading, Massachusetts)1989.

38. D. Psaltis and N. Farhat, "Optical information processing based on an associative-memory model of neural nets with thresholding and feedback," Opt. Lett., **10**, 98(1985).

39. N. H. Farhat, D. Psaltis, A. Prata, and E. Paek,"Optical implementation of the Hopfield model," Appl. Opt, **24**, 1469(1985).

40. J. Goodman, A. R. Dias, and I. M. Woody, Opt. Lett., **2**, 1(1978).

41. D. Z. Anderson, "Coherent optical eigenstate memory," Opt. Lett., **11**, 56(1986).

42. B. H. Soffer, G. J. Dunning, Y. Owechko, and E. Marom, "Associative holographic memory with feedback using phase-conjugate mirrors," Opt.

Lett., **11**, 118(1986).

43. R. A. Athale, H. H. Szu, and C. B. Friedlander, "Optical implementation of associative memory with controlled nonlinearity in the correlation domain," Opt. Lett. ,**11**, 482(1986).

44. A. Yariv and S. K. Kwong, "Associative memories based on message-bearing optical modes in phase-conjugate resonators," Opt. Lett., **11**, 186(1986).

45. A. Yariv, S. K. Kwong, and K. Kyuma, "Optical associative memories based on photorefractive oscillations," Proc. SPIE, **613**, 1(1986).

46. A. D. Fisher, R. C. Fukuda, and J. N. Lee, "Implementations of adaptive associative optical computing elements," Proc. SPIE **625**, 196(1986).

47. M. S. Cohen, "Design of a new medium for volume holographic information processing," Appl. Opt., **25**, 2288(1986).

48. T. Jannson, H. M. Stoll, and C. Karaguleff, "The interconnectability of neuro-optic processors," Proc. SPIE, **698**, 157(1986).

49. B. Kosko and C. Guest, "Optical bidirectional associative memories," SPIE Proc. **758**, Jan. 1987.

50. D. Psaltis, J. Yu, X. G. Gu, and H. Lee, Second Topical Meeting on Optical Computing, Incline Village, Nevada, March 16-18, 1987.

51. Y. S. Abu-Mostafa and D. Psaltis,"Optical Neural Computers," Scientific American, **256**(3), 88(1987).

52. D. Z. Anderson and M. C. Erie, "Resonator memories and optical novelty filters," Opt. Eng., **26**, 434(1987).

53. Y. Owechko, G. J. Dunning, E. Marom, and B. H. Soffer, “Holographic associative memory with nonlinearities in the correlation domain,” Appl. Opt. **26**,(10),1900(1987).

54. E. G. Paek and D. Psaltis, “Optical associative memory using Fourier transform holograms,” **26**, 428(1987).

55. Special issue on neural networks, Appl. Opt., **26**, no. 23, (1987).

56. K. Wagner and D. Psaltis, “Multilayer optical learning networks,” Appl. Opt., **26**, 5061(1987).

57. D. Psaltis, K. Wagner, and D. J. Brady, “Learning in optical neural computers,” Proc. IEEE International Conf. on Neural Networks, San Diego, June, 1987.

58. K. Hsu, D. J. Brady, and D. Psaltis, “Experimental demonstrations of optical neural computers,” **Neural Information Processing Systems**, D. Z. Anderson, ed., American Institute of Physics, (New York)1988.

59. D. Psaltis, D. J. Brady, and K. Wagner, “Adaptive optical networks using photorefractive crystals,” Applied Optics, **27**, 1752(1988).

60. D. J. Brady and D. Psaltis, “Perceptron learning in optical neural computers,” Proceedings of the 34th Scottish Universities Summer School in Physics: Optical Computing, Edinburgh, 1988.

61. J. S. Jang, S. Y. Shin, and S. Y. Lee, “Optical implementation of quadratic associative memory with outer-product storage,” Opt. Lett., **13**, 693(1988).

62. D. Psaltis, C. H. Park, and J. Hong, “Higher order associative memories

and their optical implementations," Neural Networks, **1**, 149(1988).

63. D. Psaltis, D. J. Brady, X.-G. Gu, and K. Hsu, "Optical implementation of neural computers," *Optical Processing and Computing*, H. Arsenault, ed., Academic Press, (New York)1989.

64. M. Ishikawa, N. Mukohzaka, H. Toyoda, and Y. Suzuki, "Optical associatron: a simple model for optical associative memory," Appl. Opt., **28**, 291(1989).

65. L.-S. Lee, H. M. Stoll, and M. C. Tackitt, "Continuous-time optical neural network associative memory," Opt. Lett., **14**, 162(1989).

66. H. Yoshinaga, K. Kitayama, and T. Hara, "All-optical error-signal generation for backpropagation learning in optical multilayer neural networks," Opt. Lett., **14**, 202(1989).

67. E. G. Paek and A. Von Lehmen, "Holographic associative memory for word-break recognition," Opt. Lett., **14**, 205(1989).

68. H. Yoshinaga, K. Kitayama, and T. Hara, "Experimental learning in an optical perceptronlike neural network," Opt. Lett., **14**, 716(1989).

69. J. Ohta, M. Takahasi, Y. Nitta, S. TAi, K. Mitsunaga, and K. Kyuma, "GaAs/AlGaAs optical synaptic interconnection device for neural networks," Opt. Lett., **14**, 844(1989).

70. D. Psaltis, X.-G. Gu, and D. J. Brady, "Holographic implementations of neural networks," *An Introduction to Neural and Electronic Networks*, S. F. Zornetzer, J. L. Davis, and Clifford Lau, eds. Academic Press, (New York)1990.

71. K. Y. Hsu, "Optical neural computing for associative memories," Ph.D. Thesis, California Institute of Technology, 1990.

72. W. E. Ross, D. Psaltis, and R. H. Anderson, "Two-dimensional magneto-optic spatial light modulator for signal processing," Opt. Eng., **22**, 485(1983).

73. H. M. Gibbs, *Optical Bistability: Controlling Light with Light*, Academic, (New York)1985.

74. S. D. Smith, A. C. Walker, F. A. P. Tooley, and B. S. Wherrett, "The demonstration of restoring optical logic," Nature, **325**, 6099(1987).

75. D. A. B. Miller, "Electric field dependence of optical properties of quantum well structures," in *Electro-optic and Photorefractive materials*, P. Günter, ed., Springer-Verlag, (Berlin)1987, p. 35.

76. W. P. Bleha, L. T. Lipton, E. Wiener-Avenear, J. Grinberg, P. G. Reif, D. Casasent, H. B. Brown, and B. V. Markevitch, "Application of the liquid crystal light valve to real time optical data processing," Opt. Eng., **17**, 371(1978).

77. C. Warde, A. M. Weiss, A. D. Fisher, and J. I. Thachara, "Optical information processing characteristics of the microchannel spatial light modulator," Appl. Opt., **20**, 2066(1981).

78. J. H. Kim, S. H. Lin, J. Katz, and D. Psaltis, "Monolithically integrated two-dimensional arrays of optoelectronic threshold devices for neural network applications," SPIE Proc., **1043-7**, 1989.

79. X. G. Gu, *Optical neural networks using volume holograms*, Ph.D. thesis,

California Institute of Technology, 1990.

80. D. Pohl, "Stacked optical memories," Appl. Opt., **13**, 341(1974).

81. J. B. Thaxter and M. Kestigian, "Unique properties of SBN and their use in a layer optical memory," Appl. Opt., **13**, 913(1974).

82. R. V. Johnson and A. R. Tanguay, Jr., Opt. Eng., **25**, 235(1986).

83. J. A. Fleck, Jr., J. R. Morris, and M. D. Feit, Appl. Phys., **10**, 129(1976).

84. R. V. Johnson and A. R. Tanguay, Jr., Opt. Lett., **13**, 189(1988).

85. H. Lee, X. G. Gu, D. Psaltis, "Volume holographic interconnections with maximal capacity and minimal crosstalk," J. Appl. Phys., **65**, 2191(1989).

86. D. Psaltis, X.-G. Gu, and D. J. Brady, "Fractal sampling grids for holographic interconnections," ICO Topical Meeting on Optical Computing, SPIE proceedings, **963 70**, 1988.

87. S. Hudson, D. J. Brady, and D. Psaltis, "Properties of 3-D imaging systems," OSA Annual Meeting, *1988 Technical Digest Series*, **11**, Optical Society of America, Washington, DC, 74(1988).

88. E. Wolf, "Three-dimensional structure determination of semi-transparent objects from holographic data," Opt. Comm., **1**, 153(1969).

89. W. J. Burke and P. Sheng, "Crosstalk noise from multiple thick phase holograms," J. Appl. Phys., **48**, 681(1977).

90. A. C. Strasser, E. S. Maniloff, K. M. Johnson, and S. D. D. Goggin, "Procedure for recording multiple-exposure holograms with equal diffraction efficiency in photorefractive media," Opt. Lett., **14**, 6(1989).

91. R. McRuer, J. Wilde, L. Hesselink, and J. Goodman, "Two wavelength photorefractive dynamic optical interconnect," Proc. SPIE, **881**, 192(1988).

92. D. J. Brady, K. Hsu, and D. Psaltis, "Multiply exposed photorefractive holograms with maximal diffraction efficiency," Optical Computing, 1989 Technical Digest Series, **9** (Optical Society of America, Washington, D.C. 1989).

93. D. J. Brady, K. Hsu, and D. Psaltis, "Periodically refreshed multiply exposed photorefractive holograms," submitted to Optics Letters.

94. G. C. Valley, "Erase rates in photorefractive materials with two photorefractive species," Appl. Opt., **22**, 3160(1983).

95. V. E. Wood, P. J. Cressman, R. L. Holman, and C. M. Verber, in *Photorefractive Materials and Their Applications II*, P. Gunter and J. P. Huignard, eds., Springer-Verlag, (New York)1989.

96. A. Katzir, A. C. Livanos, J. B. Shellan, and A. Yariv, "Chirped gratings in integrated optics," IEEE J. Quantum Elec., **QE-13**, 296(1977).

97. D. Psaltis and D. J. Brady, "A photorefractive integrated optical vector matrix multipler," SPIE proceedings, **825-19**, August 1987.

98. D. Brady and D. Psaltis, "Dynamic holographic interconnections for optical neural computers," Neural Networks, **1**, Supplement 1, Proc. of the INNS First Annual Meeting, 374(1988).

99. D. J. Brady, X.-G. Gu, and D. Psaltis, "Photorefractive crystals in optical neural computers," SPIE proceedings, **882-20**, January 1988.

100. J. Hong, S. Campbell, P. Yeh, "Optical learning machine for pattern classification," OSA Annual Meeting, *1989 Technical Digest Series*, **12**, Optical Society of America, Washington, DC, 74(1989).

101. E. G. Paek, J. R. Wullert II, and J. S. Patel, "Holographic implementation of a learning machine based on a multicategory perceptron algorithm", Opt. Lett., **14**, 1303(1989).

102. J. Yu, A. Johnston, D. Psaltis, and D. J. Brady, "Limitations of opto-electronic neural networks," SPIE Proceedings, **1053-04**, January 1989.

103. U. Langbein and F. Lederer, "Modal theory for thick holographic gratings with sharp boundaries," Optica Acta, **27**, 171(1980).

104. T. Tamir, H. C. Wang, and A. A. Oliner, "Wave propagation in sinusoidally stratified dielectric media," IEEE Trans. Microwave Theory Tech., **MTT-12**, 323(1964).

105. T. Tamir and H. C. Wang, "Scattering of electromagnetic waves by a sinusoidally stratified dielectric half-space, I," Can. J. Phys., **44**, 2073(1966).

106. T. Tamir, "Scattering of electromagnetic waves by a sinusoidally stratified dielectric half-space, II," Can. J. Phys., **44**, 2461(1966).

107. T. K. Gaylord and M. G. Moharam, "Analysis and applications of optical diffraction by gratings," Proc. IEEE, **73**, 894(1985).

108. See, for example, J. W. Goodman, *Introduction to Fourier optics*, McGraw-Hill, (New York)1968.

109. I. S. Gradshteym and I. M. Ryzhik, *Table of integrals, series, and products*,

Academic, (New York)1980, section 8.511.

110. F. Bowman, *Introduction to Bessel Functions*, Dover, (New York)1958, p. 87.

111. See, for example, D. M. Burton, *Elementary Number Theory*, Allyn and Bacon, (Boston)1980, page 40.

112. D. von der Linde, A. M. Glass, and K. F. Rodgers, Appl. Phys. Lett., **25**, 155(1974).

113. D. von der Linde, A. M. Glass, and K. F. Rodgers, J. Appl. Phys. **47**, 217(1976).

114. P. M. Rentzeipis and D. A. Parthenopoulos, "Three-dimensional optical memory," OSA Annual Meeting, *1989 Technical Digest Series*, **12**, Optical Society of America, Washington, DC, 74(1989).

115. T. Jannson, "Shannon number of an image and structural information capacity in volume holography," Optica Acta, **27**, 1335(1980).

116. See, for example, K. Hoffman and R. Kunze, *Linear Algebra*, Prentice-Hall, (Englewood Cliffs)1971, p. 125.

117. See, for example, J. W. Goodman, *Statistical Optics*, Wiley, (New York)1985.

118. See, for example, C.-T. Chen, *Linear system theory and design*, Holt, Rinehart and Winston, (New York)1984, p. 569.

119. M. G. Moharam, T. K. Gaylord, R. Magnusson, and L. Young, " Holographic grating formation in photorefractive crystals with arbitrary electron transport lengths," J. Appl. Phys., **50**, 5642(1979).

120. J. Feinberg, D. Heiman, A. R. Tanguay, Jr. and R. W. Hellwarth, "Photorefractive effects and light induced charge migration in barium titanate," J. Appl. Phys., **51**, 1297(1980).

121. A. M. Glass, D. von der Linde, and T. J. Negran, "High-voltage bulk photovoltaic effect and the photorefractive process in $\mathrm{LiNbO_3}$," Appl. Phys. Lett., 233(1974).

122. N. V. Kuktarev, "Kinetics of hologram recording and erasure in electrooptic crystals," Sov. Tech Phys. Lett., **2**, 438(1976).

123. A. Yariv and P. Yeh, *Optical waves in crystals*, Wiley, (New York)1984.

124. R. Magnusson and T. K. Gaylord, "Use of dynamic theory to describe experimental results from volume holography," J. Appl. Phys., **47**, 190(1976).

125. J. Hong, P. Yeh, D. Psalits, and D. J. Brady, "Diffraction efficiency of strong volume holograms," to appear in Optics Letters.

126. R. A. Mullen, *Time-Resolved Holographic Measurements of Bulk Space-Charge Gratings in Photorefractive* $Bi_{12}SiO_{20}$, Ph.D. thesis, University of Southern California, 1984.

127. J. K. Tyminski and R. C. Powell, "Analysis of the decay dynamics of laser-induced gratings in $\mathrm{LiNbO_3}$," J. Opt. Soc. Am., **2**, 440(1985).

128. Y. Tomita, *Modal dispersal of information and wave mixing in photorefractive crystals for information retrieval, processing and sensing*, Ph.D. thesis, California Institute of Technology, 1989.

129. G. C. Valley and J. F. Lam, "Theory of photorefractive effects in electro-

optic crystals," in *Photorefractive materials and their applications I* , P. Gunter and J.-P. Huignard, eds., Springer-Verlag, (Berlin)1989.

130. See, for example, H. Haken, *Synergetics*, Springer-Verlag, (Berlin)1983, p. 195.

131. J. C. Palais and J. A. Wise, Appl. Opt., **10**, 667(1971).

132. K. M. Johnson, M. Armstrong, L. Hesselink, J. W. Goodman, Appl. Opt, **24**, 4467(1985).

133. D. L. Staebler and J. J. Amodei, Ferroelectrics, **3**, 107(1972).

134. H. Vormann, G. Weber, S. Kapphan, and E. Krätzig, "Hydrogen as origin of thermal fixing in $LiNbO_3$:Fe," Solid State Comm., **40**, 543(1981).

135. F. Micheron and G. Bismuth, "Electrical control of fixation and erasure of holographic patterns in ferroelectric materials," Appl. Phys. Lett., **20**, 79(1972).

136. F. Micheron and G. Bismuth, "Field and time thresholds for the electrical fixation of holograms recorded in $(Sr_{0.75}Ba_{0.25})Nb_2O_6$ crystals," Appl. Phys. Lett., **23**, 71(1973).

137. See, for example, R. G. Hunsperger, *Integrated Optics: Theory and Technology*, Springer-Verlag, (Berlin)1985.

138. T. Jannson, "Information capacity of Bragg holograms in planar optics," JOSA, **71**, 342(1981).

139. See, for example, D. Marcuse, *Theory of Dielectric Optical Waveguides* Academic Press, (New York)1979.

140. D. Hebb, *The Organization of Behavior*, Wiley, (New York)1949.

141. S. S. Venkatesh and D. Psaltis, "Information storage and retrieval in two associative nets," Conf. on neural network models for computing, Santa Barbara, California, April 1985.

142. L. Personnaz, I. Guyon, and G. Dreyfus, "Information storage and retrieval in spin-glass like neural networks'," J. Physique Lett., **46**, L359(1985).

143. D. Psaltis and C. Park, "Nonlinear discriminant functions and associative memories", *Neural Networks for Computing*, APS conference proc. No. 151 (1986).

144. T. Maxwell, C. L. Giles, Y. C. Lee and H. H. Chen, "Nonlinear dynamics of artificial neural systems", *Neural Networks for Computing*, APS conference proc. No. 151 (1986).

145. E. B. Baum, On the capabilities of multilayer perceptrons, to be published.

146. F. Rosenblatt, *Principles of Neurodynamics: Perceptron and the Theory of Brain Mechanisms*, Spartan Books, (Washington)1961.

147. B. Widrow and M. E. Hoff, "Adaptive switching circuits," IRE Wescon Conv. Rec., pt. 4, 96(1960).

148. D. B. Parker, "Learning logic," Invention report, S81-64, File 1, office of Technology Licensing, Stanford University, October 1982.

149. J. S. Denker (Ed.), *Neural Networks for Computing*, APS conference proc. No. 151 (1986).

150. K. Fukushima, "A hierarchical neural network model for associative memory," Biol. Cybern. **50**,105(1984).

151. S. Grossberg, *Studies of Mind and Brain*, D. Reidel Pub. Co., (Boston)1982.

152. I. E. Sutherland and C. A. Mead, "Microelectronics and computer science," Scientific American, **237**, 210(1977).

153. D. Hammerstrom, *A Connectivity Analysis of a Class of Simple Associative Neural Networks*, Tech. Report CS/E-86-009, Department of Computer Science/Engineering, Oregon Graduate Center, Beaverton, Oregon, January 1988.

154. R. Barakat and J. Reif, "Lower bounds on the computational efficiency of optical computing systems," Appl. Opt., **26**, 1015(1987).

155. J. W. Goodman, F. Leonberger, S. Kung, and R. A. Athale, "Optical Interconnections for VLSI Systems," *Proc. IEEE*, **72**, 850(1984).

156. M. R. Feldman, S. C. Esener, C. C. Guest, and S. H. Lee, "Comparison between optical and electrical interconnects based on power and speed considerations," Appl. Opt., **9**, 1742(1988).

157. W. Wu, L. Bergamn, A. Johnston, C. Guest, S. Esener, P. Yu, M. Feldman, S. Lee, "Implementation of Optical Interconnections for VLSI," *IEEE trans. Electron Devices*, **ED-34**, 706(1987).

158. D. A. B. Miller, "Optics for low-energy communication inside digital processors: quantum detectors, sources, and modulators as efficient impedance converters," Opt. Lett., **14**, 146(1989).